



HAL
open science

Catégorisation des comportements de conduite en termes de consommation en carburant : une méthode de découverte de connaissances contextuelles à partir des traces d'interactions

Assitan Traoré

► **To cite this version:**

Assitan Traoré. Catégorisation des comportements de conduite en termes de consommation en carburant : une méthode de découverte de connaissances contextuelles à partir des traces d'interactions. Modélisation et simulation. Université de Lyon, 2017. Français. NNT : 2017LYSE1012 . tel-01489950

HAL Id: tel-01489950

<https://theses.hal.science/tel-01489950>

Submitted on 14 Mar 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



N°d'ordre NNT : 2017LYSE1012

THESE de DOCTORAT DE L'UNIVERSITE DE LYON
opérée au sein de
l'Université Claude Bernard Lyon 1

Ecole Doctorale N° accréditation
Informatique et Mathématique

Spécialité de doctorat :
Discipline : Informatique

Soutenue publiquement le 19 janvier 2017, par :
Assitan TRAORE

**Catégorisation des comportements de
conduite en termes de consommation
en carburant :**
**Une méthode de découverte de connaissances
contextuelles à partir des traces d'interactions**

Devant le jury composé de :

Despres, Sylvie, Professeur, Université Paris 13, LIMICS-INSERM	Présidente
Brézillon, Patrick, Professeur, LIP6 CNRS, Paris 6	Rapporteur
Kolski, Christophe, Professeur, LAMIH CNRS Université de Valenciennes	Rapporteur
Midenet, Sophie, Docteur, Chercheur au GRETTIA/ Ifsttar	Examinatrice
Mille, Alain, Professeur Université Lyon1, LIRIS UMR 5205 CNRS	Directeur de thèse
Tattegrain, Hélène Docteur, Directrice du LESCOT/Ifsttar	Co-directrice de thèse

Remerciements

Je tiens à exprimer ici ma gratitude envers tout le monde, tous ceux qui ont contribué à la réussite de cette thèse. Cette thèse a été effectuée grâce aux soutiens de plusieurs personnes, aussi bien sur le plan professionnel que personnel.

Mes remerciements vont tout d'abord aux membres du jury, qui ont accepté d'évaluer mon travail de thèse. Je remercie mes rapporteurs, Patrick Brézillon et Christophe Kolski, pour leur travail de relecture. Je remercie également Sylvie Despres et Sophie Midenet d'avoir accepté de participer à mon jury de thèse.

Je remercie mes encadrants. Hélène Tattegrain, d'avoir toujours cru en moi et en ma capacité de faire cette thèse. Alain Mille, mon Directeur de thèse, pour son soutien infaillible, sa disponibilité et sa patience. Vous avez toujours su me guider durant cette thèse grâce à vos interventions et vos remarques pertinentes, je vous en remercie infiniment.

Je remercie l'ensemble de mes collègues et amis du Lescot et du LIRIS. J'ai passé de très bons moments avec vous et je vous remercie de vos soutiens. Je remercie l'équipe de badminton du Lescot constitué de : Christophe, Fabien, Franck, Marion, Caroline et Fanny pour les moments sportifs que nous avons passés ensemble. Je tiens à remercier Jean et Jonathan pour nos discussions pendant les pauses café.

Je remercie ma famille et mes amis les plus proches. Merci à toi, ma sœur Korotoumou Traoré, qui a été toujours là quand j'en avais besoin et qui m'a soutenue tout le long de cette thèse. Un grand merci à mes parents et à toute la famille Traoré. Je remercie également Yvan Doyeux ingénieur en informatique de sa présence au quotidien et d'avoir été là aux moments difficiles de la thèse.

Je remercie mes relecteurs qui ont contribué à la qualité de ce mémoire grâce à leurs retours pertinents. Il s'agit de mes encadrants bien sûr, les rapporteurs de la thèse, Aline Alauzet, Marie-Pierre Bruyas et Yvan Doyeux. Je remercie Rochdi Trigui pour sa coopération lors de cette thèse pour la mise en place du modèle d'optimisation.

Et pour terminer, je remercie le CEESAR pour la mise en disposition des données sans lesquelles cette thèse n'aurait pas été possible.

Résumé

Cette thèse propose une méthode d'ingénierie des connaissances contextuelles qui permet la modélisation et l'identification du contexte explicatif d'un critère observé. Le contexte est constitué de connaissances explicatives situées permettant une représentation élicitée valide d'un objet dans la situation visée. Ces connaissances sont découvertes généralement découvertes lors de l'observation de la réalisation de l'activité dans laquelle cet objet est impliqué. Elles sont donc difficiles à décrire en début d'analyse d'une activité. Toutefois, elles restent nécessaires pour la définition, l'explication et la compréhension efficace d'une activité selon un critère observé caractérisant cette dernière. Cette thèse propose la définition progressive du *contexte* satisfaisant pour expliquer un critère observé lors de l'observation d'une activité. Cette recherche mobilise les traces d'interaction de l'activité analysée, précise la notion de contexte et exploite les méthodes de fouille de données pour réaliser la catégorisation et la classification d'un critère observé en distinguant les paramètres contextuels et non contextuels. L'environnement développé sur les principes des traces d'interaction, permet d'assister la découverte du contexte explicatif par une approche interactive, à l'aide des connaissances de l'analyste, de distinguer ce qui est contexte de ce qui ne l'est pas. Nous montrons qu'il est possible de construire un contexte valide, en le « découvrant » et en le formulant sous une forme générique, telle que proposée dans la littérature. Une application de la méthode a été effectuée en situation de conduite automobile pour modéliser et identifier le contexte explicatif de la consommation en carburant. En s'appuyant sur les connaissances existantes du domaine, la validation de la méthode est effectuée en étudiant qualitativement les connaissances produites sur la consommation réelle en carburant. La méthode est validée quantitativement en appliquant les règles de classifications établies sur des données collectées de l'activité de conduite. Cette illustration de l'analyse de l'activité de conduite automobile avec la méthode de découverte de connaissances contextuelles, pour déterminer le contexte explicatif de la consommation en carburant, a été effectuée à l'Ifsttar sur des données réelles collectées lors de l'activité de conduite en situation naturelle. Les expérimentations menées montrent des résultats encourageants et permettent d'envisager l'intégration de la méthode de découverte de connaissances contextuelles dans les pratiques des analystes de l'Ifsttar.

Mots-clés : Découverte interactive de connaissances, contexte, traces modélisées, transport, consommation en carburant.

Abstract

This thesis proposes an engineering method of contextual knowledge that allows modelling and identification of explanatory context of observed criteria. The context consists of located explanatory knowledge allowing valid representation of an object in the covered situation. This knowledge is generally discovered when observing the activity performance in which the object is involved. They are therefore difficult to describe in the beginning of activity analysis but are necessary for the definition, explanation and effective understanding of an activity according to an observed criterion characterizing this activity. This thesis proposes a progressive definition of adequate context to explain an observed criterion during activity observation. The research mobilizes interaction traces of the analysed activity, clarify context notion and uses data mining methods for classification or categorization of an observed criterion by distinguishing contextual parameters and no contextual parameters. The developed environment, based on interaction traces principles, allows to assist explanatory context discovery by interactive approach, using context analyst knowledge. We demonstrate that it's possible to build a valid context, by discovering it and by formulating it in a generic form as proposed in literature. An application of the method was performed in driving situation to identify and model the explanatory context of fuel consumption. The method validation is performed by studying produced knowledge on fuel consumption, qualitatively by relying on existing domain knowledge and quantitatively by applying classification rules established trough data collected from driving activity. This illustration of driving activity analysis with the contextual knowledge discovery method to determine explanatory context of fuel consumption was conducted at Ifsttar on real data, collected during driving activity in natural driving situation. The led experiments show encouraging results and allows considering the integration of contextual knowledge discovery method in Ifsttar analyst practices.

Keywords: Interactive knowledge discovery, context, modelled traces, transport, fuel consumption.

Table des matières

Remerciements	2
Résumé	3
Abstract	4
Table des matières	5
Table des figures	9
Table des tableaux	13
Introduction	15
Chapitre I. État de l'art	21
1 Analyse de l'activité de conduite automobile pour la réduction de la consommation en carburant.....	22
2 Le Contexte : sa définition et sa modélisation	29
2.1 Définitions du contexte	30
2.2 Modélisations du contexte pour des systèmes sensibles au contexte.....	34
2.2.1 Modèles Attribut/Valeur.....	34
2.2.2 Modèles de représentation par balise (Mark-up Scheme)	35
2.2.3 Modèles graphiques et les modèles orientés objets	36
2.2.4 Modèles basés sur la logique	40
2.2.5 Modèles basés sur l'ontologie	41
2.2.6 Graphes conceptuels	44
2.2.7 Modèles avec le Réseau de Pétri	46
2.2.8 Modèles spatiaux et temporels	48
2.2.9 Modèles hybrides.....	49
2.3 Approche de découverte de connaissances contextuelles	50
3 La notion de Trace Modélisée (M-Trace)	53
4 Discussion de l'état de l'art.....	63
Chapitre II. Contribution : Méthode et Application informatique.....	67
1 Méthode de découverte de connaissances contextuelles.....	67

1.1	Etape 1 : Définition des composants génériques du contexte.....	69
1.1.1	Description du contexte de l'activité.....	69
1.1.2	Modélisation du contexte de l'activité.....	70
1.2	Etape 2 : Collecte et préparation des données.....	72
1.2.1	Collecte et exploration des données.....	73
1.2.2	Opération de préparation des données.....	73
1.2.3	Opération de contextualisation des données.....	75
1.3	Etape 3 : Construction de la base de connaissances (les modèles des M-Traces) .	78
1.3.1	Elaboration des règles de segmentation de l'activité en séquences.....	79
1.3.2	Elaboration des règles de catégorisation et de classification des séquences en fonction du critère observé.....	79
1.3.3	Formalisation des règles dans la base de connaissances.....	80
1.4	Etape 4 : Processus d'assistance à la découverte de connaissances contextuelles .	81
1.4.1	Représentation de l'activité par les M-Traces.....	82
1.4.2	Découverte des règles de classification pour identifier les paramètres contextuels explicatifs du critère observé.....	83
1.5	Etape 5 : Evaluation et validation des connaissances découvertes.....	84
1.6	Itérations entre les différentes phases.....	85
2	Application informatique : Outils de mise œuvre de la méthode.....	86
2.1	Outils utilisés pour la mise en œuvre de la méthode de découverte de connaissance contextuelles.....	86
2.2	Outil informatique pour l'Analyse des Traces modélisées (M-Traces) pour la Découverte de Connaissance Contextuelles (ATraDeCC).....	89
2.2.1	Structure de la base de trace.....	90
2.2.2	Les fonctionnalités implémentées.....	91
3	Synthèse du chapitre contribution.....	100
Chapitre III. Application de la méthode de découverte de connaissances contextuelles ..		101
1	Description des données (expérimentation).....	101
1.1	Instrumentation des véhicules.....	102

1.2	Protocole expérimental	105
2	Application de la méthode de découverte de connaissances contextuelles.....	106
2.1	Définition des composants contexte de l'activité de conduite.....	107
2.2	Collecte et préparation des données.....	109
2.2.1	Collecte des données	109
2.2.2	Exploration des données.....	110
2.2.3	Opération de préparation des données de l'activité de conduite	113
2.2.4	Opération de contextualisation des données de l'activité de conduite	117
2.3	Construction de la base de connaissances.....	124
2.3.1	Segmentation de l'activité de conduite en sous-parties (situations de conduite) :.....	124
2.3.2	Catégorisation du critère observé des situations de l'activité de conduite :.	126
2.3.3	Classification des situations de conduite en fonction du critère observé « consommation en carburant ».....	130
2.3.4	Stockage des connaissances issues de la fouille de données dans la base de connaissances	132
2.4	Processus d'assistance à la découverte de connaissance contextuelles	134
2.4.1	Représentation de l'activité de conduite par les M-Traces.....	134
2.4.2	Découverte des règles de classification pour identifier les paramètres contextuels explicatifs de la consommation en carburant.....	137
2.5	Évaluation et validation des connaissances contextuelles	146
2.6	Itérations entre les différentes phases	150
2.7	Application de la méthode sur le critère d'écart de consommation.....	151
3	Synthèse du chapitre application de la méthode de découverte de connaissances contextuelles dans le domaine du transport.....	152
Chapitre IV. Discussion.....		155
Conclusion et Perspectives.....		161
Bibliographie.....		165
Annexes.....		173

1	Données Collectées	173
2	Répartition des consommations en carburant.....	177
3	Compléments pour la classification de la consommation en carburant avec les paramètres non contextuels	178
4	Application de la méthode de découverte de connaissances contextuelles sur l'écart entre la consommation en carburant réelle et optimisée.	182
	Liste des Acronymes	192

Table des figures

Figure I.1 : Méthode interactive de découverte de connaissance à partir de M-Traces de Mathern et al., dans [6].....	22
Figure I.2 : Répartitions des déplacements des français par modes de transport en 2013 (voyageurs-km), extrait de [2].....	23
Figure I.3 : Répartitions des marchandises par modes de transport en 2013, extrait de [2]	23
Figure I.4 : Evolution de la température de l'hémisphère Nord depuis 1860, extrait [11]	25
Figure I.5 : Répartitions de la consommation d'énergie par modes de transport en 2013, extrait de [2]	26
Figure I.6 : Modèle des composants contexte de Bazire et Brazillon [41].	33
Figure I.7 : Exemple du modèle attribut/valeur de Schilit et al. dans [26]	35
Figure I.8 : Modélisation du contexte avec le modèle graphique d'Henricksen et al	37
Figure I.9 : Exemple de modélisation du contexte avec le langage CML dans [41].....	38
Figure I.10 : Modèle UML de l'approche Hydrogen d' Hofer et al., 2003.....	40
Figure I.11 : Exemple d'un modèle basé sur la logique de Ye et al., dans [55].....	41
Figure I.12 : Les classes de l'ontologie de service de Miraoui [59]	42
Figure I.13 : Modèle ASC, extrait de [60]	44
Figure I.14 : Représentation des trois types de contexte de Brézillon et Pomerol [62]	45
Figure I.15 : Exemple de graphe contextuel, extrait de [63].....	45
Figure I.16 : Approche utilisant le réseau de Pétri pour générer des interfaces sensible au contexte	47
Figure I.17 : Processus de découverte de connaissances contextuelles de Ngoc et al.	51
Figure I.18 : Processus d'identification du contexte dans le processus de conception d'un système de Mena	51
Figure I.19 : Contexte des entités pour la résolution de problèmes	52
Figure I.20 : Approche de modélisation <i>MUSETTE</i> de Champin et al.[93].	57
Figure I.21 : Principe d'un Système à Base de Traces (<i>SBT</i>) [96].....	58
Figure I.22 : Démarche de modélisation de trace, extrait de [96]......	59
Figure I.23 : Illustration de visualisation de l'activité de conduite avec ABSTRACT, extrait de [6]	62
Figure II.1 : Schéma de la méthode de découverte de connaissances contextuelles.....	68
Figure II.2 : Modèle des composants contexte de Bazire et al., [41].....	70
Figure II.3 : Diagramme de classe de la modélisation du contexte d'une activité.....	71
Figure II.4 : Collecte et prétraitement des données.....	73

Figure II.5 : Construction de la base de connaissances	78
Figure II.6 : Processus d'assistance à la découverte de connaissances contextuelles.....	82
Figure II.7 : Evaluation et validation des connaissances découvertes	84
Figure II.8 : Utilisation de la méthode de découverte de connaissances contextuelles.....	85
Figure II.9 : Outils utilisés pour la mise en œuvre de la méthode de découverte de connaissances contextuelles	88
Figure II.10 : Fonctionnalités de l'outil ATraDeCC	89
Figure II.11 : Modèle relationnel des données utilisé pour le stockage des M-Traces	90
Figure II.12 : Page de création ou de chargement d'une M-Trace.....	91
Figure II.13 : Exemple de fichier contenant la M-Trace primaire	92
Figure II.14 : Page de résultat de la création d'une M-trace	92
Figure II.15 : Page de visualisation et de modification des attributs d'une M-Trace	93
Figure II.16 : Saisie d'informations complémentaire d'un attribut.....	93
Figure II.17 : Page résultat de la modification de la M-Trace	94
Figure II.18 : Page de transformation d'une M-Trace	94
Figure II.19 : Page résultat récapitulatif d'une transformation	95
Figure II.20 : Page de création de séquence dans une M-Trace.....	97
Figure II.21 : Page de contextualisation des M-Traces	98
Figure II.22 : Page de description d'une la M-Trace	98
Figure II.23 : Page de visualisation des obsels d'une M-Trace transformée	99
Figure II.24 : Visualisation graphiques des obsels.....	99
Figure III.1 : Illustration du premier niveau d'instrumentation du véhicule.....	103
Figure III.2 : Illustration du second niveau d'instrumentation du véhicule.....	104
Figure III.3 : Instrumentation du véhicule d'expérimentation de CEESAR	105
Figure III.4 : Répartition des périodes d'enregistrement de l'expérimentation	106
Figure III.5 : Répartition des véhicules de second niveau d'instrumentation	106
Figure III.6 : Illustration des composants <i>contexte</i> de Bazire et al., [41] appliqué à l'activité de conduite	108
Figure III.7 : Collecte et préparation des données par l'analyste	110
Figure III.8 : Interface de visualisation des données de BIND	111
Figure III.9 : Affichage des paramètres collectés avec les différentes fonctionnalités de BIND	112
Figure III.10 : Illustration de la visualisation des vidéos de l'activité de conduite avec BIND	113
Figure III.11 : Illustration du calcul des rapports de boite de vitesse d'un véhicule Clio.....	114

Figure III.12 : Répartition de la moyenne de la vitesse sur l'ensemble des situations de conduite	117
Figure III.13 : Illustration du calcul de la consommation en carburant à partir d'un profil de vitesse optimisé de Mensing et al., extrait de la thèse de [101]	118
Figure III.14 : Figures des profils de vitesse et de rapport de boîte de vitesse calculé avec la méthode d'optimisation de Mensing et al., extrait de [100].....	119
Figure III.15 : Résultats de la méthode d'optimisation du profil vitesse de Mensing et al. extrait de [102]	120
Figure III.16 : Illustration des questionnaires sur les attitudes pendant la conduite.	121
Figure III.17 : Illustration de l'arbre de décision pour la segmentation de l'activité de conduite	125
Figure III.18 : Répartitions des situations de conduite après la segmentation	126
Figure III.19 : Répartition de la consommation en carburant réelle en l/100km des situations de conduite	127
Figure III.20 : Répartition de la consommation en carburant réelle calculé par le modèle d'optimisation de Mensing et al., [45] en l/100km	127
Figure III.21 : Répartition de la consommation en carburant optimisée calculée par le modèle d'optimisation de Mensing et al., [45] en l/100km	128
Figure III.22 : Répartition de l'écart entre les consommations en carburant réelle et optimisée du modèle d'optimisation de Mensing et al., [45].....	128
Figure III.23 : Arbre de décision de la classification de la consommation en carburant avec les paramètres non contextuels.	131
Figure III.24 : Construction de la base de connaissances à partir des arbres de décision	133
Figure III.25 : Page de création d'une transformation d'une M-Trace	135
Figure III.26 : Page de résultat de la transformation d'une M-Trace.....	135
Figure III.27 : Liste des obsels issues d'une transformation.....	136
Figure III.28 : Visualisation des obsels mal classés pour identifier les paramètres contextuels	136
Figure III.29 : Arbre de décision avec les paramètres contextuels du composant contexte « <i>Environnement de conduite</i> ».....	138
Figure III.30 : Arbre de décision avec les paramètres contextuels du composant contexte « <i>Conducteur</i> ».....	140
Figure III.31 : Arbre de décision avec les paramètres contextuels du composant contexte « <i>Environnement de conduite- Infrastructure</i> »	143

Figure III.32 : Arbre de décision à partir du noeud1 avec les paramètres relatif à l'infrastructure.....	144
Figure III.33 : Extrait arbre de décision par le paramètre contextuel <i>type de route</i>	145
Figure III.34 : Arbre de décision par le paramètre contextuel <i>infrastructure avant</i>	145
Figure IV.1 : Le contexte au cœur de l'analyse de l'activité de conduite	155
Figure A.1 : Arbre de décision de la classification de la consommation en carburant avec les paramètres non contextuels	178
Figure A.2 : Arbre de décision de la classification à partir du nœud 4	179
Figure A.3 : Arbre de décision de la classification à partir du nœud 5	180
Figure A.4 : Arbre de décision de la classification à partir du nœud 15	181
Figure A.5 : Arbre de décision de la classification avec les paramètres non contextuels.....	182
Figure A.6 : Arbre de classification décrivant l'infrastructure de la route.....	185
Figure A.7 : Suite arbre de décision à partir du nœud 6 de la classification décrivant l'infrastructure de la route	187
Figure A.8 : Arbre de décision de la classification avec les paramètres contextuels décrivant le conducteur	189

Table des tableaux

Tableau I.1 : Récapitulatif des limites et atouts de quelques travaux de recherche étudiés dans le chapitre III	65
Tableau II.1 : Liste des fonctions d'agrégation standard sur les séquences de conduite	75
Tableau II.2 : Fonctions d'agrégation mis en place pour logiciel	96
Tableau III.1 : Extrait de la liste des paramètres calculés et leurs descriptions	115
Tableau III.2 : Discrétisation du régime moteur avec la méthode d'amplitude.....	116
Tableau III.3 : Tableau des indicateurs calculés sur les situations de conduite	117
Tableau III.4 : Liste de paramètres contextuels de l'entité « <i>Analyste</i> »	121
Tableau III.5 : Liste de paramètres contextuels de l'entité « <i>Conducteur</i> »	122
Tableau III.6 : Liste de paramètres contextuels de l'entité « <i>Véhicule</i> »	122
Tableau III.7 : Liste des paramètres contextuels de l'entité « <i>Environnement de Conduite</i> ».....	123
Tableau III.8 : Catégorisation de la consommation en carburant des situations de conduite	129
Tableau III.9 : Catégorisation du gain de consommation en carburant des situations de conduite	130
Tableau III.10 : Classification de la consommation en carburant avec les deux échantillons sans les paramètres contextuels.....	147
Tableau III.11 : Classification de la consommation en carburant avec les deux échantillons avec les paramètres contextuels relatifs au trafic	148
Tableau III.12 : Classification de la consommation en carburant avec les deux échantillons avec les paramètres contextuels relatifs au conducteur.....	148
Tableau III.13 : Classification de la consommation en carburant avec les deux échantillons avec les paramètres contextuels relatifs à l'infrastructure.....	149
Tableau IV.1 : Liste des questions de recherche avec la méthode proposée.....	159
Tableau A.1 : Liste des paramètres issus du bus CAN	173
Tableau A.2 : Liste des paramètres issus de la caméra multi fonction	173
Tableau A.3 : Liste des paramètres issus du Radar	174
Tableau A.4 : Liste des paramètres issus du GPS	174
Tableau A.5 : Liste des paramètres issus de la phase de mise en correspondance de la position GPS et la base de données cartographique (Map Matching)	175
Tableau A.6 : Liste des paramètres calculés et leurs descriptions	176
Tableau A.7 : Tableau des percentiles de l'indicateur de la consommation en carburant <i>IND_ConsoReelle</i>	177
Tableau A.8 : Tableau des percentiles de l'indicateur <i>IND_Mod_EcartConsoOpt</i>	177

Tableau A.9 : Tableau de classification du critère écart entre la consommation réelle et optimisée du modèle d'optimisation avec les paramètres non contextuels.....	184
Tableau A.10 : Tableau de classification du critère écart de consommation avec les paramètres contextuels décrivant l'infrastructure de la route	188
Tableau A.11 : Tableau de classification du critère écart de consommation avec les paramètres contextuels décrivant le conducteur.....	191

Introduction

Le travail de recherche présenté dans cette thèse s'est déroulé au Lescot (Ifsttar) et au LIRIS (Université Lyon 1 et CNRS), à la frontière de la recherche dans le domaine du transport et de la recherche en ingénierie de la dynamique de la connaissance.

Le questionnement de l'Ifsttar sur la relation entre comportements de conduite et impact sur le réchauffement climatique

L'Ifsttar, et en particulier le Lescot, est questionné sur l'évolution des comportements de conduite et par suite sur l'évolution de l'environnement de conduite automobile qui permettraient de réduire l'impact de cette activité sur le réchauffement climatique. Le changement climatique se manifeste aujourd'hui par plusieurs caractéristiques observables comme l'augmentation générale de la température terrestre, la diminution de la fréquence des froids extrêmes. Les vagues de chaleur sont de plus en plus fréquentes, la fonte des glaciers de montagnes s'accélère et le niveau de la mer s'élève [1]. Ces dégradations climatiques ne sont pas seulement un problème environnemental, elles ont également un impact économique et politique qui nous concerne tous.

En effet, en moins d'une trentaine d'années le réchauffement climatique est devenu un problème mondial qui nécessite une réponse adéquate d'où la tenue de la dernière conférence mondiale annuelle sur le climat, la Conférence des Parties (COP21) en décembre 2015 à Paris. Cette COP21 a permis de prendre des décisions devant permettre d'atteindre les objectifs de lutte contre les changements climatiques qui avaient été annoncés lors de la Convention Climat de Rio en 1992. Les décisions ne peuvent être prises qu'à l'unanimité des parties ou par consensus. L'enjeu de la dernière COP21 était donc de trouver un accord mondial pour limiter, d'ici à la fin de nouveau siècle, le dérèglement climatique à 2°C par rapport à la période préindustrielle.

Le rôle de l'homme dans le processus du réchauffement climatique à travers ses émissions de gaz à effet de serre (GES) est bien connu maintenant du grand public. En effet, les activités humaines comme l'industrie, le transport, l'agriculture et le résidentiel-tertiaire, nécessitent de l'énergie (gaz, pétrole, gasoil, essence, etc.) et, par conséquent, produisent des GES (par exemple, CO₂, H₂O, NO_x, CH₄, O₃).

En conséquence, le domaine d'application de cette thèse est particulièrement intéressé à la réduction des émissions liées au secteur des transports, l'un des principaux émetteurs de GES en France et en Europe. En effet, si l'on observe une baisse des GES (-11 %) depuis 1990, cette dernière est due à la réduction des émissions de l'industrie (-40 %), les émissions liées

aux transports ayant augmenté de 12 % sur la même période dans le monde. En 2013, plus de 80 % des transports de personnes et de marchandises ont été effectués en France par la route, ce qui fait du secteur routier le plus grand consommateur d'énergies fossiles, avec un taux de 70 % devant les autres modes de transports (aérien, fluvial et ferroviaire) [2]. De nombreux efforts ont été faits pour diminuer la consommation de carburant en optimisant les véhicules (aérodynamique, motorisation, ...). Elle a ainsi été réduite de 1,7 litre au 100 kms depuis 1990. Des solutions pour favoriser l'éco-conduite ont également été mises en place. Ces solutions ont un potentiel de gain de consommation qui peut atteindre jusqu'à 20 % en incitant les conducteurs à modifier leurs comportements au volant [3].

Le questionnement de l'équipe TWEAK du LIRIS sur l'ingénierie de la dynamique des connaissances

Le CNRS et en particulier l'équipe TWEAK du LIRIS est questionné sur la manière dont la connaissance se construit dynamiquement au cours d'une activité humaine dans un environnement de plus en plus technologique et numérique. La notion de *contexte* est souvent utilisée pour construire des environnements sensibles à la dynamique des situations et recommandant des comportements *adaptés au contexte*. Le cadre théorique de la thèse précisément s'inscrit dans la découverte de connaissances contextuelles lors de l'analyse d'une activité. Cette découverte se fera lors de l'analyse de l'activité de conduite de conducteurs afin d'identifier les contextes explicatifs permettant d'expliquer la consommation en carburant, dans le but de la réduire et, par conséquent, de réduire les émissions de GES du secteur routier.

L'identification des contextes explicatifs d'un critère observé pendant la réalisation d'une activité est une tâche très ardue lorsque plusieurs éléments sont impliqués dans la réalisation de cette activité. Chacun d'eux possède son propre contexte. La multiplicité des contextes et la complexité de l'activité observée font qu'il est nécessaire de se concentrer sur certains comportements pour pouvoir identifier les contextes explicatifs d'un critère observé de l'activité.

L'analyste doit repérer, comprendre et prendre en compte les contextes explicatifs de certains comportements pour mettre en évidence l'influence qu'ils peuvent avoir sur le critère observé. Pour cela, il a besoin d'un cadre méthodologique et des outils pour l'assister dans ce processus. Cette thèse aborde cette question précise de découverte de **contextes explicatifs appliqués au domaine des transports**, au travers de l'analyse de **l'activité de conduite**, avec pour finalité de réduire **la consommation en carburant**. Elle s'inscrit dans le domaine de l'ingénierie des connaissances contextuelles.

La facette transport de la thèse

Les travaux d'analyse et l'expérimentation des résultats de la recherche ont été effectués dans le département TS2 (Transport, Santé, Sécurité) de l'Ifsttar au sein du laboratoire de recherche Lescot (Laboratoire Ergonomie et Sciences Cognitives pour les Transports). Les missions de recherche de ce laboratoire ont pour objectif principal l'étude des mécanismes cognitifs et comportementaux mis en œuvre par une personne pour adapter et optimiser son comportement dans des situations de déplacement, afin de lui permettre de maintenir une performance optimale en termes de sécurité et de confort. Cette thèse s'inscrit dans le premier thème de recherche du Lescot (Connaissance des usagers et de leurs pratiques des transports) et s'appuie sur les connaissances des chercheurs du laboratoire sur l'activité de conduite.

L'objectif applicatif de cette thèse est d'identifier des situations de conduite précises pour lesquelles il serait possible de conseiller au conducteur des actions de conduite qui réduisent la consommation en carburant, conseils prenant en compte le contexte de conduite. Pour cela, cette thèse vise à produire des connaissances contextuelles expliquant la consommation en carburant en fonction des situations de conduite, répondant ainsi à l'un des objectifs de l'Ifsttar, qui est de réduire les émissions de GES du secteur routier.

Cette thèse profite de l'accès à un grand volume de données de l'activité de conduite en situation naturelle collecté grâce au projet européen EuroFot (European Field Operational Test), des méthodes d'analyse de données (fouille de données) et de la collaboration avec le LIRIS (Laboratoire de l'InfoRmatique en Image et Système d'informatique).

La facette recherche ingénierie de la dynamique des connaissances de cette thèse

L'équipe TWEAK (Traces, Web, Education, Adaptation, Knowledge) du pôle interactions et cognition est l'équipe d'accueil de cette thèse au sein du LIRIS. Ses travaux s'inscrivent dans les disciplines de l'intelligence artificielle et de l'Ingénierie des Connaissances (IC). Elle s'intéresse en particulier à la coévolution des systèmes utilisateurs-machines au sein des environnements informatisés, en adoptant un point de vue orienté connaissances.

Les travaux en IC concernent une ingénierie de la dynamique des connaissances dans laquelle des traces modélisées constituent une représentation de l'expérience et une importante source de connaissances. Cette équipe propose la notion de Système de Gestion de Base de Traces (SGBT), spécifiée et mise en œuvre au sein d'une plate-forme ouverte. Ces SGBT, couplés à une instrumentation des systèmes informatiques tracés, placent l'utilisateur dans un rôle

central pour la co-construction des connaissances. Pour la prise en compte de cette dynamique, l'équipe développe le paradigme du raisonnement à partir de l'expérience tracée, et s'appuie sur les sciences cognitives pour soutenir l'utilisation réflexive des traces d'activité.

Les contributions principales de la thèse reprennent et enrichissent les travaux d'Ingénierie de Connaissance du LIRIS. Par exemple, les concepts et méthodes d'ingénierie de la dynamique des connaissances des traces modélisées (M-Traces) sont utilisés pour proposer une solution à la question de recherche traitée dans cette thèse. Les M-Traces sont déjà utilisées dans le transport pour la modélisation de l'analyse de l'activité de conduite [4]–[6]. Les atouts des M-traces tels que la représentation de l'expérience tracée, la construction dynamique de connaissances à partir de l'expérience tracée sont exploités dans notre travail.

Cette thèse s'intéresse à la méthode de découverte de connaissances basée sur l'observation de l'activité. L'activité est décrite comme une interaction entre un utilisateur et un système, dans un environnement donné et avec un but précis, le système pouvant être un objet ou une application informatique. L'activité de conduite est alors décrite comme une interaction entre le conducteur et son véhicule dans l'environnement routier, interaction ayant pour but de se déplacer d'un endroit à un autre.

Pour mieux appréhender la variabilité de l'activité de conduite, cette thèse est focalisée sur la découverte du **contexte**, c'est-à-dire sur **la modélisation et l'identification du contexte explicatif** lors de l'analyse d'une activité. La notion de contexte est très vaste et sa définition varie d'un domaine à l'autre ou d'une situation à une autre. Nous présenterons donc plusieurs de ces définitions et celle sur laquelle nous nous appuyerons pour la découverte de connaissances contextuelles.

Cette thèse va coupler les travaux existants sur le contexte et l'analyse de l'activité à partir de l'expérience tracée, pour proposer une **méthode et un outil d'assistance à la découverte de connaissances contextuelles à partir de l'observation tracée de l'activité**. Cette méthode, basée sur l'observation de l'activité, permet d'identifier et modéliser les *contextes explicatifs* d'un critère observé d'une sous-partie de l'activité ou de l'activité en général. Une application de cette méthode dans le domaine du transport est effectuée pour l'identification des contextes explicatifs de la consommation de carburant lors de l'activité de conduite. Cette thèse est composée de quatre chapitres, qui sont définis de la manière suivante :

Chapitre I : État de l'art

Dans cette thèse pluridisciplinaire, l'objectif de ce chapitre est de faire un état des lieux selon le point de vue des différentes disciplines impliquées. L'état de l'art étudie les travaux revendiqués dans chaque discipline et précise les choix effectués dans cette thèse. Ce chapitre récapitule dans un tableau les apports et les limitations des contributions étudiées dans l'état de l'art. Ce chapitre pose et situe ainsi les questions de recherche de la thèse.

Chapitre II: Les contributions de la thèse

Ce chapitre détaille les deux contributions principales de la thèse : la méthode de découverte de connaissances contextuelles et l'environnement informatique permettant l'analyse de l'activité de conduite automobile selon la méthode des traces modélisées M-Traces.

Chapitre III : Application de la méthode

Ce chapitre décrit la mise en application de la méthode proposée pour la découverte de connaissances contextuelles dans les traces de l'activité de conduite automobile. Cette méthode est utilisée pour découvrir les paramètres contextuels explicatifs de la consommation en carburant lors de l'activité de conduite. Deux critères de consommation en carburant sont analysés pour expliquer les différences de consommation en carburant d'une même situation de conduite.

Chapitre IV : Discussion de la thèse

Ce chapitre discute les résultats obtenus lors de l'application de la méthode au domaine du transport. Les contributions sont ensuite discutées au regard des questions de recherche posées en fin de chapitre I. Nous discutons alors de l'approche proposée, de la méthode et de l'outil expérimenté pour la découverte de connaissances contextuelles lors d'une activité complexe et dynamique.

Conclusion et Perspective

Une conclusion et les perspectives de cette thèse sont fournies.

Chapitre I. État de l'art

Ce travail de recherche répond à une question posée au sein de du Lescot : « **Comment construire dynamiquement des connaissances sur les contextes explicatifs d'un comportement de conduite à partir de son observation** » et valide l'approche choisie par son application à l'identification des contextes explicatifs de la consommation de carburant. C'est avec cette question en tête que nous réalisons l'état de l'art, afin de mettre en évidence la question de recherche sous-jacente et de définir précisément les contributions attendues dans ce travail de thèse.

Nous commençons par clarifier l'état de l'art concernant les connaissances sur le cas d'application : la réduction des émissions de gaz à effet de serre (GES) du secteur routier. En effet, pour atteindre cet objectif, il faut comprendre comment il est possible de réduire la consommation en carburant, dont la combustion produit une part significative des GES. Pour pouvoir élaborer de nouvelles connaissances relatives à la consommation en carburant, nous analysons les travaux existants sur l'analyse de l'activité de conduite pour la réduction de la consommation en carburant. Nous étudierons ensuite la place du *Contexte* tel qu'il est défini par la communauté scientifique « **CONTEXT**¹ ». Enfin, nous nous focaliserons sur la place de la *découverte de connaissances* dans l'analyse de l'activité de conduite.

La coopération Ifsttar-LIRIS a déjà abordé cette question dans des travaux de recherche qui ont débouché sur la définition d'une méthode d'ingénierie de la dynamique des connaissances (M-Traces). Notre état de l'art analyse ces travaux en cherchant cette fois à répondre à la question de recherche concernant la prise en compte du contexte lors de l'analyse de l'activité de conduite en considérant, non seulement les comportements de conduite, mais aussi le contexte associé à cette activité.

La prise en compte du contexte nécessite une observation par des analystes de l'activité qui évolue dynamiquement dans le temps. La thèse s'intéresse donc spécifiquement à la partie interactive de la construction de connaissances contextuelles. Les observations, les interprétations humaines et les méthodes de fouilles de données sont mobilisées comme des outils à articuler dans une méthode de découverte interactive. Cette approche a été validée par les travaux antérieurs qui en ont montré la faisabilité [4]–[6]. Par exemple dans [5], Mathern et al., modifient le cycle de découverte de connaissance proposé par Fayyad et al., [7] pour

¹ <http://cyprusconferences.org/context2015/>

analyser des séquences de signatures d'une activité selon une méthode d'analyse des traces modélisées (M-Traces) proposée par Georgeon et al., dans [4]. A partir des séquences de signatures (M-Traces), des automates sont implémentés sous la forme d'un Réseau de Pétri pour faire une synthèse de l'ensemble des séquences et ainsi représenter l'activité analysée.

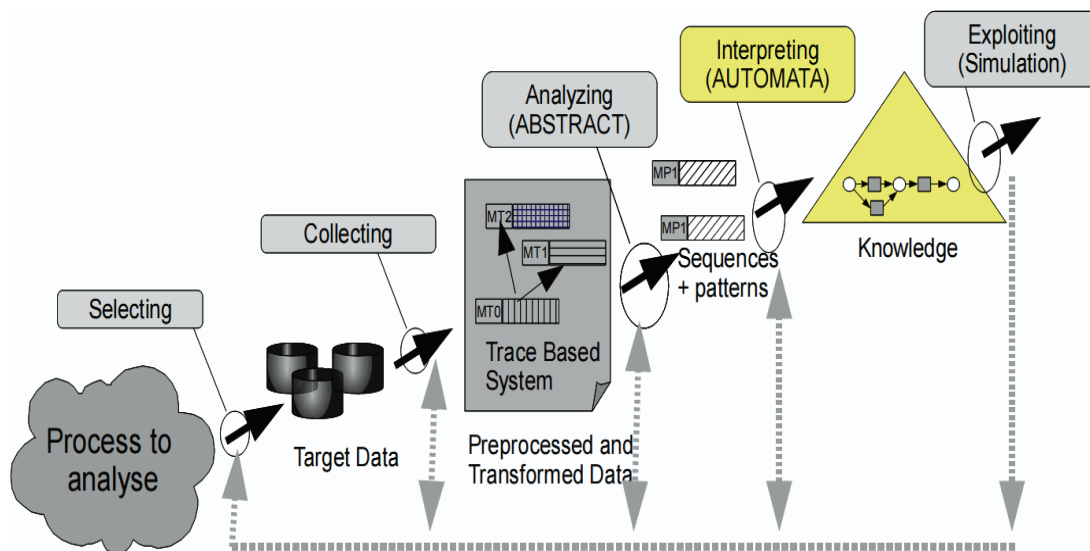


Figure I.1 : Méthode interactive de découverte de connaissances à partir de M-Traces de Mathern et al., dans [6]

Dans cette thèse une forme de découverte interactive de connaissances à partir des M-Traces est proposée et un prototype réalisé.

La première partie de l'état de l'art étudie l'analyse de l'activité de conduite automobile pour la réduction de la consommation en carburant, la seconde interroge la notion de contexte, la troisième s'intéresse à la découverte de connaissances à partir des traces d'observation. La quatrième partie portera sur la discussion de l'état de l'art et définira précisément la question de recherche de la thèse et les contributions qui sont visées pour faire évoluer les connaissances en la matière.

1 Analyse de l'activité de conduite automobile pour la réduction de la consommation en carburant

Les véhicules sont indispensables pour le déplacement des personnes (véhicule personnel, véhicule de transport en commun) et des marchandises. Malgré les autres moyens de transport disponibles (ferroviaire, fluvial et aérien), la route reste le moyen le plus utilisé en France. Cela explique la croissance constante du parc automobile français d'une moyenne annuelle de 0,6% sur les dix dernières années. En 2013, le parc automobile français a connu une baisse de

0,4% par rapport aux dix dernières années. Il a eu donc une croissance de seulement 0,2% en 2013 [2].

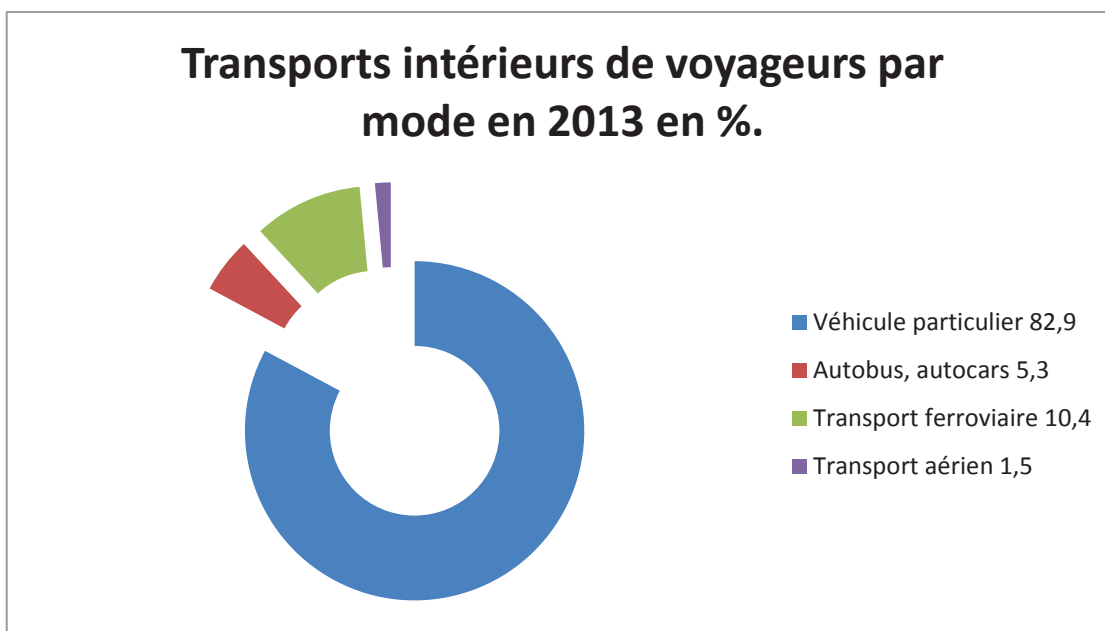


Figure I.2 : Répartitions des déplacements des français par modes de transport en 2013 (voyageurs-km), extrait de [2]

La Figure I.2 de la répartition des déplacements des français par modes de transport montre bien leur préférence pour la route qui représente à elle seule presque 83% des déplacements, laissant 17% pour l'ensemble des autres modes de transport.

On peut faire le même constat pour le transport des marchandises en France qui se fait essentiellement par la route avec un pourcentage de 85 % de transport de marchandises par la route et 15% pour les autres moyens de transport.

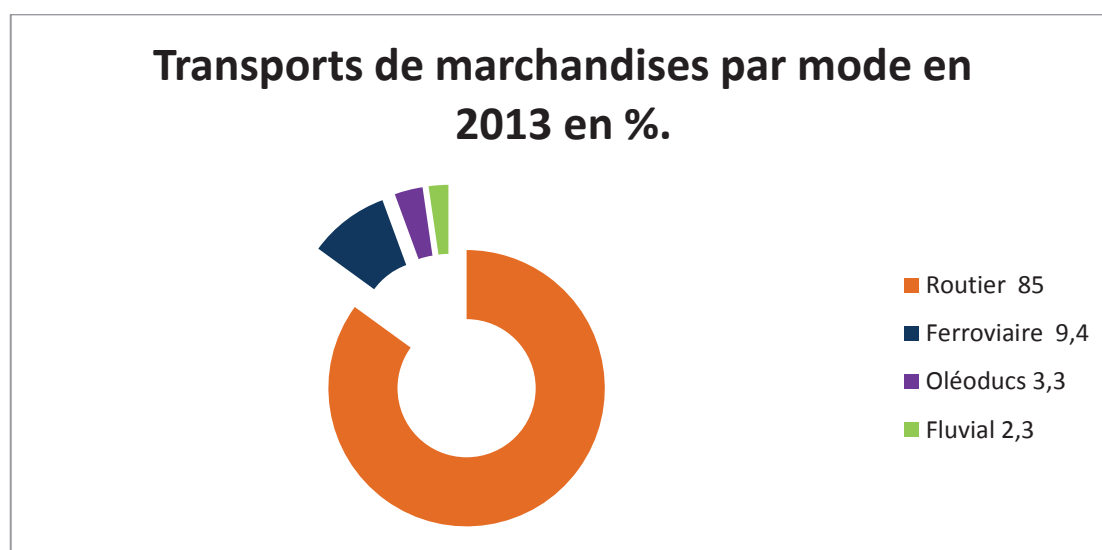


Figure I.3 : Répartitions des marchandises par modes de transport en 2013, extrait de [2]

Et pourtant, les impacts des émissions de gaz à effet de serre (GES) produites par la combustion du carburant nécessaire au déplacement des véhicules sur l'environnement sont bien connus depuis 1895 [8]. En effet, en 1895, le mécanisme de l'effet de serre est expliqué pour la première fois par *Arrhénius*, un chercheur suédois. Il explique que le mécanisme de l'effet de serre est bénéfique pour la terre car il lui permet d'avoir une température moyenne compatible à la vie. Il explique que ce mécanisme existe grâce aux GES (par exemple, CO₂, HO, NOX, CO) présents dans l'atmosphère en faible quantité. Ces gaz ayant la propriété d'absorber le rayonnement infrarouge réémis par le sol permettent à la terre d'avoir une température moyenne de 15°C au lieu d'une température estimée à -18°C [9] en leur absence. Mais il met en garde contre le développement de l'usage des combustibles fossiles, qui pour lui favoriseraient l'augmentation de la température moyenne de la terre. L'hypothèse des responsabilités de l'homme dans le processus du réchauffement climatique n'est pas prise au sérieux à l'époque, ni d'ailleurs les autres études qui défendront cette hypothèse plusieurs années après.

En 1979, la première Conférence Mondiale sur le Climat permet aux scientifiques de faire adopter l'hypothèse d'*Arrhénius* par les gouvernements membres des Nations-Unies sur l'environnement. La lutte contre le réchauffement climatique est alors lancée, des études scientifiques sont effectuées dans le monde entier pour mieux comprendre et expliquer les causes et les conséquences du changement climatique. En 1988, lors de la conférence de Toronto, le Groupe Intergouvernemental d'Étude du Climat (GIEC) est créé pour suivre le problème du réchauffement climatique. Le GIEC n'est pas un organisme de recherche, mais un groupe international d'expertise visant à synthétiser les travaux de recherche du monde entier sur les changements climatiques, leurs impacts et les mesures de prévention et d'adaptation envisageables. Le premier rapport du GIEC en 1990 met en évidence le rôle des activités humaines dans le processus du réchauffement climatique et constate son influence sur l'environnement comme l'augmentation de la température moyenne de la terre et du niveau de la mer d'environ respectivement 0.5°C et 15 cm [10].

Evolution de la température

de l'hémisphère Nord

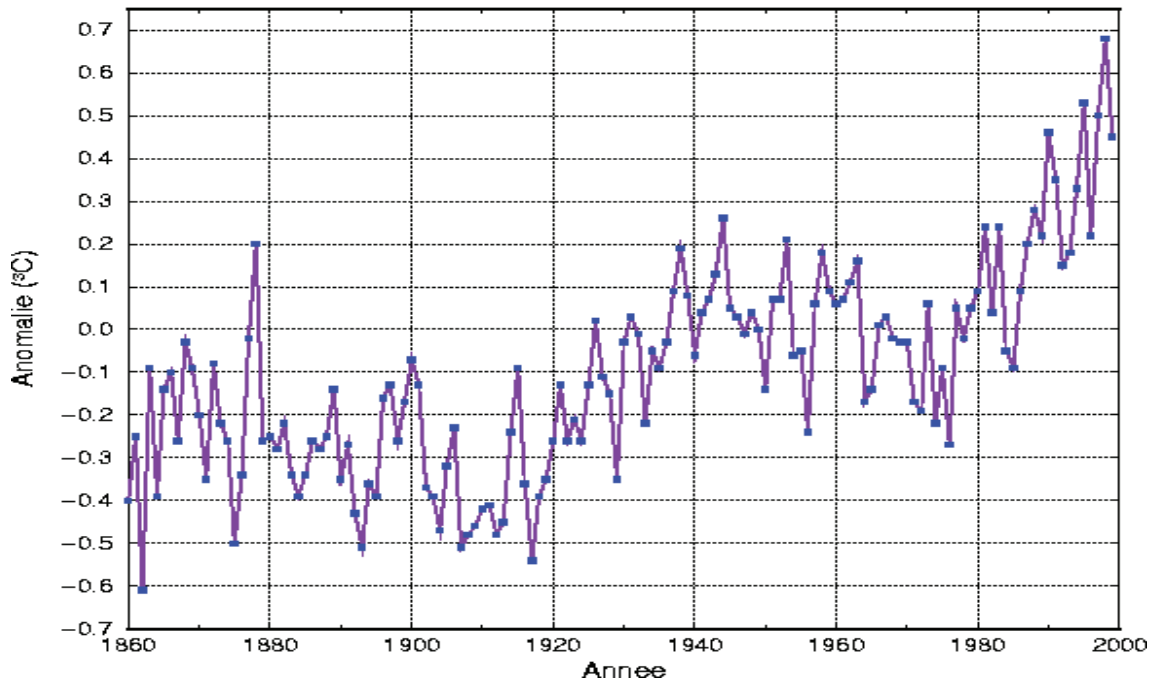


Figure I.4 : Evolution de la température de l'hémisphère Nord depuis 1860, extrait [11]

Les activités humaines qui favorisent le réchauffement climatique sont : les productions industrielles, le transport, le résidentiel-tertiaire, l'agriculture. En effet, depuis le début de l'ère industrielle (entre 1750 à nos jours), l'augmentation de la température moyenne de la terre est liée aux émissions de GES, conséquence des activités humaines. Sur cette période, l'augmentation de la quantité dans l'atmosphère de ces gaz est fulgurante, une augmentation de 64 % de gaz carbonique, de 19 % de méthane, de 6 % de protoxyde d'azote et les 10 % restants pour les autres GES [9]. Ces augmentations de la quantité des GES dans l'air sont dues à l'intensification de l'agriculture, de l'élevage et de l'utilisation des combustibles fossiles dans l'industrie [9].

Malgré la mobilisation mondiale contre le réchauffement climatique et les efforts mis en place, ce problème qui était seulement environnemental au début est devenu au fil du temps un problème économique et politique. Le réchauffement climatique est toujours actuel et devient une préoccupation mondiale. En effet, la dernière conférence sur le climat qui a eu lieu récemment était la 21^{ème} Conférence des Parties (COP21) à Paris en décembre 2015. L'objectif de la COP21 était de trouver un accord mondial pour maintenir la température de la terre à 2°C, ce qui a été une réussite d'ailleurs puisqu'un accord a été signé et les mesures nécessaires pour atteindre cet objectif devraient être mise en œuvre dans les années à venir.

Parmi les études pour la réduction de l'émission des GES, nous nous intéressons à celles liées aux transports. Le secteur des transports en France et en Europe est classé comme le premier

producteur de GES [12], [13]. En effet, d'après le Réseau Action Climat France, le transport est le principal émetteur de GES principalement le CO₂, produit par la combustion du carburant devant l'industrie, le résidentiel-tertiaire et l'agriculture selon une étude menée en 2015 [12]. Ceci a été aussi relevé en 2011, par le Ministère de l'Écologie, du Développement Durable et de l'Énergie (MEDDE) que le transport reste le premier émetteur de GES en France avec 32 % de l'énergie finale consommée et 27 % des émissions de GES [13]. Les émissions du secteur des transports n'ont pas cessé d'augmenter depuis 1990 : en 2011, une augmentation de 13% avait été enregistrée par rapport à celles de 1990. Mais cette croissance n'a pas été homogène, elle a atteint son niveau maximum en 2004 pour ensuite avoir une diminution annuelle de 0.82 % entre 2005 et 2011, grâce à l'amélioration des infrastructures et des modes de transports alternatifs à la voiture en solo comme les transports en commun (trains, bus, tram, métro, etc.) et les co-voiturages [13].

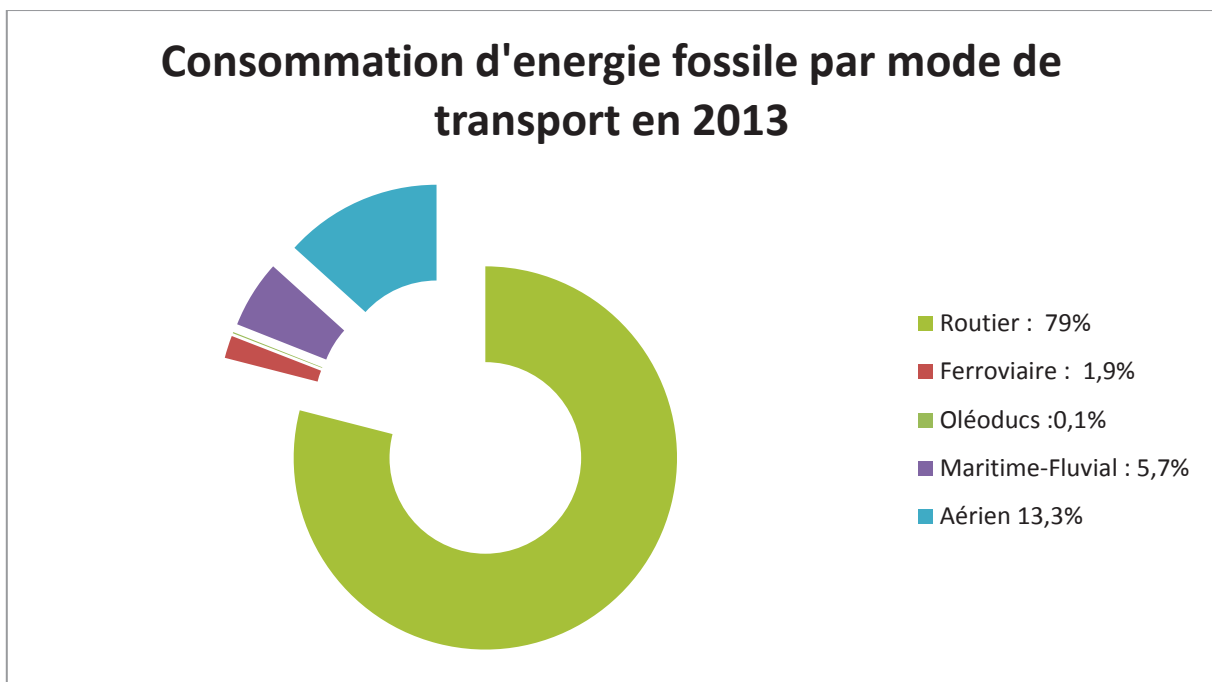


Figure I.5 : Répartitions de la consommation d'énergie par modes de transport de la France en 2013, extrait de [2]

Pour réduire les émissions liées au secteur routier des transports, des études ont été menées pour comprendre la consommation en carburant selon l'activité de conduite. Ces études montrent que la consommation en carburant lors de l'activité de conduite et ses émissions de GES dépendent des comportements au volant du conducteur comme le style de conduite (calme, normal ou agressif), le mode de démarrage du véhicule (démarrage à froid, démarrage à chaud), le type de véhicule, le type de la route (urbain, rural, autoroute) et les caractéristiques de la route et du trafic [14]–[16]. En effet, Karin Brundell-Freij et al.,

montrent en 2005, que les caractéristiques de la route, du trafic et de la performance du véhicule affectent les comportements de conduite [15] et en 2006, Ericsson et al., montrent que la consommation en carburant dépend du type de véhicule mais aussi du style de conduite [16]. Avec ces études, on constate que la consommation en carburant est fortement dépendante des comportements de conduite qui à leur tour dépendent des caractéristiques de la route, du trafic, de la performance du véhicule. Une forte consommation en carburant résulte d'un style de conduite agressif. La consommation en carburant sur autoroute est différente de la consommation en milieu urbain, elle varie en fonction du type de véhicule, de l'état du véhicule, de la météo et de l'environnement de conduite.

En 1999, la notion d'Eco-Conduite définie comme conduite écologique est proposée par Smith [17]. L'Eco-Conduite consiste à adopter les actions de conduite efficaces pour avoir une consommation en carburant optimale lors de l'utilisation d'un véhicule. À lui seul, un changement de comportement lors de la conduite automobile peut permettre une réduction jusqu'à 20 % de la consommation en carburant [18]. Les conseils d'Eco-Conduite à adopter sont par exemple : utiliser la voiture dans une plage de fonctionnement optimal du moteur (changer la boîte de vitesse le plus tôt possible après le démarrage, rouler avec un rapport de vitesse le plus haut possible), rouler à une vitesse stable, anticiper le changement de trafic, éviter les freinages brusques et les accélérations. Plusieurs pays ont fait la promotion de l'Eco-conduite en la considérant comme un moyen immédiat et peu cher pour réduire les émissions de CO₂ sans vérifier son impact sur la sécurité routière. L'union européenne recommande même l'apprentissage de l'Eco-Conduite aux jeunes conducteurs et le Japon est arrivé à atteindre ses objectifs environnementaux de réduction d'émissions de CO₂ en 2010 en encourageant l'Eco-Conduite [18]. Mais des résultats récents des recherches de l'impact de l'Eco-Conduite sur la consommation en carburant démontrent que la réduction de la consommation ne perdure pas dans le temps. En effet, seulement quatre mois après les cours d'Eco-Conduite, on ne constate en moyenne que 5,8 % de gain de consommation en carburant au lieu de 20 % [3]. Cela s'explique par le fait que dans la majorité des cas les conducteurs reprennent leurs anciens comportements de conduite quelques mois après les cours d'éco-conduite. D'une part, ceci montre la difficulté pour les conducteurs d'adopter les actions d'Eco-Conduite lors de l'activité de conduite et d'autre part, une incompatibilité apparaît entre les actions de conduite recommandées par l'Eco-Conduite et les situations de conduite réelles. Par exemple : l'Eco-Conduite recommande de rouler avec un rapport de vitesse le plus haut possible alors qu'en réalité, il est très difficile de rouler en urbain par exemple en 5^{ème} au risque de caler et de provoquer un accident ou bien d'essayer de rouler à une vitesse constante alors qu'il y a un véhicule devant. Le maintien de la conduite à une vitesse constante est

difficile dans cette situation de conduite, ce qui oblige le conducteur à baisser sa vitesse ou à freiner à la dernière minute pour éviter un accident. Le conducteur adapte donc sa conduite à l'environnement de conduite pour assurer sa sécurité. En général, l'adaptation des actions de conduite au contexte de conduite implique un surplus de consommation en carburant. Les actions d'Eco-Conduite sont des règles générales sur la dynamique du véhicule telles que la gestion de la vitesse et de l'accélération. Elles ne tiennent pas compte des caractéristiques de l'environnement de conduite telles que le trafic, l'infrastructure, les contraintes d'arrêt lors de l'activité de conduite, dues par exemple au passage d'un piéton dans un carrefour. Les études suivantes d'évaluation de l'Eco-conduite montrent que les actions de conduite conseillées par l'Eco-Conduite ne sont pas toujours réalisables en situation de conduite réelle. En effet, la baisse de consommation en carburant attendue n'est pas toujours atteignable [3] et des questions de sécurité routière sont posées dans [18], telle que la conduite à une vitesse constante en situation de trafic dense ou le passage d'un piéton. De nouveaux systèmes d'assistance à la conduite sont en train de voir le jour, qui permettent aux conducteurs de choisir un bon compromis entre la sécurité routière et la réduction de la consommation en carburant [18].

Les éléments qui ont une influence sur la consommation en carburant comme les caractéristiques de la route, le trafic, la performance du véhicule, le style de conduite sont déterminés d'une manière générale lors de l'activité de conduite par ces études [14]–[16]. Elles ne permettent pas de caractériser de manière précise l'influence de ces éléments sur la consommation en carburant dans une situation de conduite précise, une situation de conduite étant définie par l'état de l'ensemble des éléments impliqués dans l'activité de conduite. Par exemple, quelle est la consommation en carburant dans une situation de conduite définie par « un jeune conducteur sur autoroute avec un fort trafic ». *Quelle est l'influence du trafic sur la consommation en carburant dans cette situation de conduite ? Les comportements du jeune conducteur dans la situation de conduite auront-ils une influence sur la consommation en carburant ? Si oui, comment identifier cette influence de manière précise sur la consommation en carburant ?*

Pour répondre à ces questions, il est nécessaire d'introduire la notion de **Contexte** dans l'analyse de l'activité de conduite pour déterminer de manière précise l'influence des éléments impliqués dans l'activité de conduite sur la consommation en carburant. La partie suivante s'intéresse aux études scientifiques du **Contexte** en général, sa définition et sa modélisation lors de l'analyse d'une activité pour déterminer le contexte explicatif d'un comportement observé lors de l'activité.

2 Le Contexte : sa définition et sa modélisation

La notion de contexte a été étudiée très tôt en intelligence artificielle par un des pionniers, John Mc Carthy [19]: la relation de base est $ist(c,p)$ qui établit que la proposition p est vraie dans le *contexte* c . Ceci permet d'étendre ou de restreindre un raisonnement en fonction d'un contexte, qu'il faut alors établir. Le contexte est défini en extension par les variables qui sont contenues dans les propositions valides dans ce contexte, c'est-à-dire des variables qui permettent de produire une inférence valide dans ce contexte.

Comment identifier ces variables reste une question non résolue. Ces travaux préliminaires ont été rapidement approfondis et comparés et le lecteur intéressé trouvera une étude comparative très bien documentée dans [20]. Dans cet article, Akman et Surav font une étude comparative des travaux traitant la formalisation du contexte pour la résolution des problèmes en général en intelligence artificielle. Dans cette étude, il est ressorti rapidement que la définition d'un contexte était problématique et soulevait aussi bien des questions de complexité de calcul que de difficultés à en définir les composants : la thèse de Guha [21] encadrée par Mc Carthy précisait bien ces questions difficiles.

La complexité de l'ingénierie de cette modélisation pour la représentation des connaissances dans le système Cyc de Lenat [22] impose une méthode de réduction de complexité. En effet, une grande variété de combinaisons de valorisations de variables permet de définir un contexte, et chaque contexte doit être « affirmé » en examinant la validité des hypothèses de validité des règles que ses variables peuvent (ou non) déclencher.

Dans [23], Patrick Brézillon faisait le point des connaissances sur la notion de contexte en y associant les notions de révision, de raisonnement hypothétique, de raisonnement analogique également très étudiées en Intelligence Artificielle. C'est avec Fausto Giunchiglia ([24], [25]) étudiant formellement le raisonnement en contexte, que Patrick Brézillon anime la série des conférences CONTEXT¹ depuis 2007. Cette conférence a permis d'élargir considérablement l'étude de la notion de contexte à d'autres champs de recherche que l'intelligence artificielle. La section suivante se fonde principalement sur les travaux sur la définition, la modélisation du contexte et la découverte de connaissances contextuelles avec le prisme de l'ingénierie des connaissances pour l'appliquer au cas particulier de l'observation d'une activité.

2.1 Définitions du contexte

De nombreuses définitions de contexte sont proposées dans la bibliographie, nous en citons quelques-unes parmi les plus référencées et par ordre chronologique de proposition par leurs auteurs :

En 1994, Schilit et al., dans [26] ont considéré que le contexte possède trois aspects importants qui répondent aux questions suivantes : Où es-tu ? Avec qui es-tu ? De quelles ressources disposes-tu à proximité ? Dans [27], Schilit et al., précisent le contexte comme constitué de la localisation, la description de personnes et d'objets dans l'entourage ainsi que les changements de ces objets.

En 1995, Brown dans [28] donne une définition très générale mais adaptée au contexte informatique : « les éléments de l'environnement d'un utilisateur dont l'ordinateur a connaissance ». En 1997, Ward et al., dans [29], voient le contexte comme les états des environnements possibles de l'application. Tandis que Brown et al., dans [30] introduisent une dimension spatiale en définissant le contexte comme un ensemble d'éléments extensibles avec les éléments suivants : la localisation, l'ensemble d'objets dont l'utilisateur a besoin, le temps et l'orientation spatiale (direction).

Pour réduire la complexité du contexte, Lenat dans [22], dès 1998, propose de décrire « presque » n'importe quel contexte selon « seulement » 12 dimensions qui permettraient de décrire 99 % des contextes qu'il a imaginés. Ces dimensions sont (nous reprenons la liste en anglais telle qu'elle a été publiée):

- Absolute Time: a particular time interval in which events occur
- Type Of Time: a non-absolute type of time period, such as “just after eating”
- Absolute Place: a particular location where events occur, such as “Paris”
- Type Of Place: a non-absolute type of place, such as “in bed”
- Culture: linguistic, religious, ethnic, age-group, wealth, etc., of typical actors
- Sophistication/Security: who already knows this, who could learn it, etc.
- Topic/Usage: drilling down into aspects and applications – not subsets
- Granularity: phenomena and details which are (and are not) ignored
- Modality/Disposition/Epistemology: who wants/believes this content to be true?
- Argument-Preference: local rules for how to resolve pro-con argument disputes
- Justification: are things in this context generally proven, observed, on faith...
- Let's: local bindings of variables etc. that hold true in that context.

Ces dimensions sont surtout descriptives et chaque définition possède des raffinements importants, en particulier dans la gestion du temps. Notons que le marquage de chaque affirmation par son contexte de validité, exprimé selon les différentes dimensions, revient à organiser des collections d'affirmations dans un espace à 12 dimensions dont certaines sont potentiellement infinies (le temps, la localisation spatiale) et d'autres finies (type de culture par exemple). En pratique, avant de lancer une inférence, il faut définir un « motif contextuel » (un intervalle, une zone géographique, une contrainte sur telle ou telle autre dimension) et sélectionner les règles d'inférence qui sont valides pour ce motif de contexte (ce qui laisse supposer que toute règle est annotée par le motif contextuel pour lequel elle est valide).

En 1999, Ryan et al., dans [31] assimilent le contexte à l'environnement, l'identité et la localisation de l'utilisateur ainsi que le temps. Brézillon et Pomerol dans [32] proposent une définition plus générique en définissant le contexte comme « ce qui contraint la résolution d'un problème sans y intervenir explicitement ».

Dans leur étude sur la plasticité des interfaces homme machine (IHM) en 1999, Thevenin et al. [33], aboutissent à une définition assez proche avec la notion de contexte d'interaction. Le contexte d'interaction est un triplet « plate-forme – environnement – utilisateur » où l'environnement est l'ensemble des entités (objets, personnes et événements) périphériques à la tâche courante et pouvant avoir un impact sur le comportement du système ou de l'utilisateur.

En 2001, Geenberg dans [34] introduit une dimension sociale avec la définition suivante du contexte : le contexte est considéré comme une construction dynamique avec cinq dimensions : (1) temps, (2) épisodes d'utilisation, (3) interactions sociales, (4) objectifs internes et (5) des influences locales. La même année, Dey dans [35], propose la définition du contexte comme « toute information qui peut être utilisée pour caractériser la situation d'une entité. Une entité étant une personne ou un objet qui est considéré pertinent à l'interaction entre l'utilisateur et l'application, incluant l'utilisateur et l'application eux-mêmes. »

En 2002, Henricksen et al., dans [36] proposent une définition spécifique aux systèmes informatiques, le contexte étant « la circonstance ou la situation dans laquelle une tâche informatique se déroule ». La même année, G. Rey et al., dans [37] définissent le contexte comme un espace infini d'informations évolutives, qui ne sont pas connues en avance.

En 2007, Mena et al., dans [38] proposent une définition plus formelle qui dépend de trois choses qui sont : le domaine d'application, l'entité et le problème à résoudre. Ils définissent le contexte de la manière suivante : « le contexte c_i d'une entité e_i pour résoudre un problème p_k est définie comme un ensemble d'entités e_j (i différent de j) relié à p_k , en interaction avec e_i et un ensemble de relation $R_{i,j}$ entre les entités e_i et e_j .

En 2007, Miraoui et al., définissent le contexte comme « toute information dont le changement de sa valeur déclenche un service ou change la qualité (forme) d'un service » [39]. C'est une définition abstraite du contexte spécifique au système informatique diffus qui limite l'ensemble des informations contextuelles à celles qui sont reliées aux services.

Dans la communauté scientifique du contexte la définition la plus connue du contexte est celle de Dey [35]. Dey définit le contexte de la manière suivante : « *toute information qui peut être utilisée pour caractériser la situation d'une entité. Une entité étant une personne ou un objet qui est considéré pertinent à l'interaction entre l'utilisateur et l'application, incluant l'utilisateur et l'application eux-mêmes* ».

Par exemple dans un accident de la route le fait que le conducteur ait consommé récemment de l'alcool ou pas est une information relevant du contexte. En effet, elle permet de savoir si le conducteur était dans un état normal pour conduire ou non. Cette information caractérisée la situation² de l'entité « conducteur » engagé dans l'activité de conduite au moment de l'accident. Cette définition est générale car, quel que soit le domaine, elle permet de déterminer les informations relevant du contexte ou pas. Si cette définition est retenue, la question de la modélisation reste posée. C'est la définition du contexte qui est adoptée dans cette thèse.

A partir des définitions du contexte Bazire et Brézillon [41], ont proposé un modèle de contexte en s'appuyant sur une décomposition générique de l'activité. Ce modèle représente les entités ou composants impliquées dans la réalisation d'une activité et les relations qui peuvent exister entre eux pour définir le contexte de l'activité.

² Travaux argumentant sur la nécessité de formaliser les contextes revendiquent une approche « située » de la cognition, telle que Clancey l'a théorisée par exemple [40].

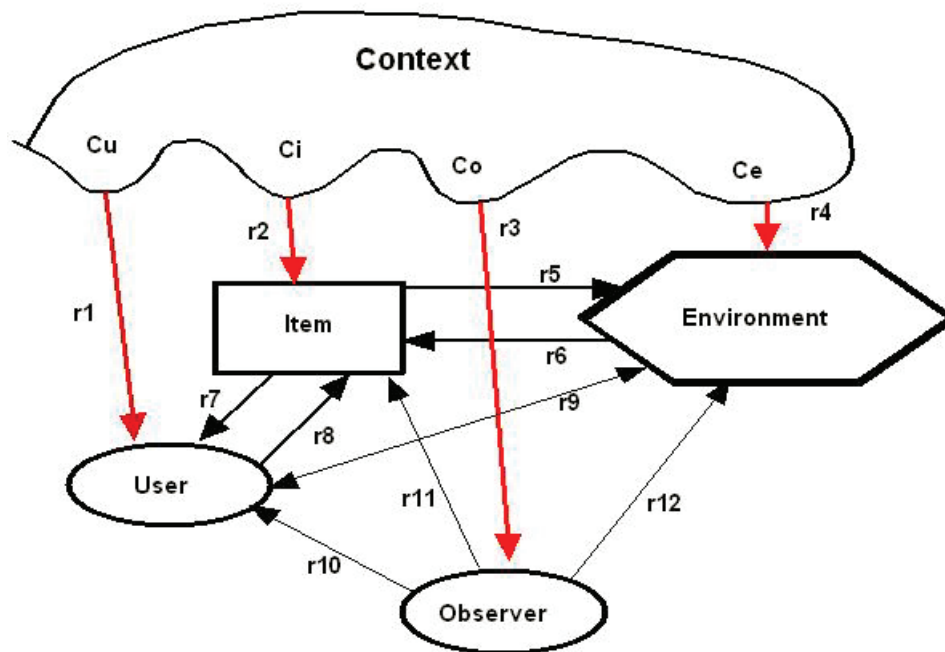


Figure I.6 : Modèle des composants contexte de Bazire et Brazillon [41].

Ce modèle de composant contexte a été mis en place grâce à l'analyse d'un corpus de 150 définitions du contexte tout domaine confondu. Il représente le contexte comme l'ensemble des contextes des différents composants (User, Item, Observer, Environment) impliqués dans une situation, chaque composant possède son propre contexte (Cu, Ci, Co et Ce). La situation pouvant être décrite par un utilisateur, un système ou une application dans un environnement particulier et éventuellement par un observateur. Le modèle décrit aussi les relations entre les différents composants par exemple les relations r8 et r9 : r8 représentant les interactions entre le composant utilisateur et le composant système et r9 celle entre le composant utilisateur et le composant environnement. Ce modèle a permis aux auteurs de faire l'hypothèse que les définitions du contexte sont différentes par ce qu'elles ne se focalisent pas sur les mêmes composants contextes. Le modèle des composants contexte de Bazire et al., dans [41] permet de catégoriser les informations candidates au contexte des entités d'une activité mais elle ne dit rien sur l'identification des paramètres contextuels pertinents pour une activité. ce problème relève d'une question d'ingénierie de la connaissance qui associe d'une part, les méthodes issues de l'acquisition des connaissances auprès des experts, et d'autre part les méthodes cherchant à découvrir des connaissances dans les données disponibles, considérées comme caractéristiques de l'activité mobilisant ces connaissances. Cette approche de modélisation du contexte est utilisée par Traoré et al., dans [42] pour modéliser le contexte de l'activité de conduite automobile dans le but d'expliquer la consommation en carburant.

A partir de ce modèle du composant contexte, le contexte peut être défini comme « l'ensemble des contextes des entités impliquées dans une activité. À savoir le contexte de

l'entité utilisateur, contexte de l'entité système ou application, le contexte de l'entité environnement et le contexte de l'entité observateur » pour notre besoin de découverte de connaissances contextuelles lors de l'analyse d'une activité.

2.2 Modélisations du contexte pour des systèmes sensibles au contexte

Après la phase de définition du contexte, nous allons maintenant nous intéresser au problème de la modélisation et de l'utilisation du contexte dans des systèmes sensibles à ce dernier. Pour construire des systèmes sensibles au contexte il est nécessaire de collecter, stocker et analyser des informations contextuelles dans le but de les exploiter pour adapter les services fournis par le système sensible au contexte à l'utilisateur. Un système conscient du contexte est un système capable de s'adapter à un changement de situation en fournissant des services adéquats à cette nouvelle situation. Dans la littérature sur les systèmes sensibles au contexte plusieurs travaux récapitulent les structures de modélisation de contexte les plus pertinents. En effet, les travaux [43], [43]–[45] sur les approches de modélisation de contexte nous permettent de citer certaines de ces structures de modélisation de contexte et de donner quelques exemples d'utilisation.

2.2.1 Modèles Attribut/Valeur

La représentation attribut/valeur est la structure des données la plus simple pour la modélisation des informations contextuelles. Elle a été proposée au départ par Schilit et al., dans [26] pour la gestion des informations contextuelles d'un environnement. L'attribut est une métadonnée qui décrit la nature (type) de la valeur à laquelle il est associé. Par exemple, pour une application de localisation, une information sur l'occupation d'une pièce ou un bâtiment est représentée de la manière suivante :

OCCUPANTS = adams:schilit:theimer:weiser:Welch

Cette approche est très utilisée dans les services distribués. Les services sont décrits en général avec une liste d'attributs sous forme d'attribut/valeurs et la découverte de service est ensuite appliquée par un algorithme qui utilise ces paires attributs/valeur. Le modèle attribut valeur est facile à gérer mais ne convient pas pour les structures complexes et ne permet pas de faire un raisonnement efficace sur le contexte.

Category	Example
Date and time	after April 15 between 10 and 12noon
Location	in room 35-2200
Co-location	with {User Adams} with {Type Display} having {Features Color}

Figure I.7 : Exemple du modèle attribut/valeur de Schilit et al., dans [26]

2.2.2 Modèles de représentation par balise (Mark-up Scheme)

Ce modèle utilise une structure hiérarchique de données constituée de balises avec des attributs et des contenus. Ces balises peuvent être définies récursivement par d'autres balises. Pour la gestion des données sous forme de balise ce modèle utilise le langage XML (eXtended Markup Language). Ce langage dérive du langage SGML (Standard Generic Markup Language) et les informations contextuelles liées aux entités sont contenues dans les pages web qui peuvent être extraites par adresse web (URL). Comme exemple de langage utilisant le modèle de représentation par balise nous avons le CC/PP (Composite Capabilities/ Preferences Profiles) et le CSCP (Comprehensive Structure Context Profile).

Le CC/PP (Composite Capabilities/ Preferences Profiles) permet de supporter la négociation de contenu entre un navigateur web et un client. Il utilise RDF (Resources Description Framework) pour créer des profils décrivant les capacités des terminaux et les préférences de l'utilisateur. Il est utilisé pour personnaliser le contenu d'une page en définissant ces capacités et ces préférences.

Le langage CSCP (Comprehensive Structure Context Profile) est une extension du langage CC/PP. Contrairement au CC/PP, il ne définit aucune hiérarchie pour résoudre les problèmes de capacités et de préférences (CC/PP). Il résout les limites du langage CC/PP concernant la structuration et étend le mécanisme pour exprimer les préférences de l'utilisateur en exploitant les avantages hérités de l'utilisation de RDF, c'est-à-dire l'interchangeabilité, la décomposabilité et l'extensibilité. Le langage CSCP est un métalangage basé sur RDF et défini par Held et al., dans [46].

Il existe d'autres approches de modélisation du contexte par le langage des balises mais elles sont pour la plupart spécifiques à un domaine et limitées à un ensemble d'aspects du contexte (localisation, environnement, etc.) telles que le PDDL (Pervasive Profile Description

Language) de Pascoe dans [47] ou le CCML (Centaurus Capability Markup Language) proposé par Kagal et al., dans [48].

Le modèle de représentation par balise est limité pour tenir compte des différents types de contexte, des relations et des dépendances pouvant exister entre les données contextuelles. Il ne permet pas de faire du raisonnement sur les informations contextuelles ce qui n'assure pas une bonne utilisabilité des informations contextuelles.

2.2.3 Modèles graphiques et les modèles orientés objets

Cette approche consiste à modéliser les informations contextuelles selon un graphe conceptuel tel que le langage UML (Unified Modelling Language). Grâce à sa structure générique très forte, le langage UML est approprié pour modéliser le contexte. En effet, dans [49] Bauer et al., utilisent le diagramme UML pour la modélisation des informations contextuelles pour le système de contrôle du trafic aérien. D'autres auteurs ont utilisé le diagramme UML pour la modélisation du contexte tel que Kubicki et al., dans [50], qui exploitent le diagramme UML pour modéliser le contexte des interactions sur les tables interactives avec les objets tangibles. Dans cet article ils proposent un modèle de contexte pour la gestion de l'interface utilisateur dans un système collaboratif. Ce modèle de contexte prend en compte les interactions individuelles sur une table interactive mais aussi les interactions distribuées et collaboratives. Ce modèle de contexte définit le contexte comme un triplé « *Utilisateur, Plate-forme et Environnement* » et il tient compte des caractéristiques de la table interactive. Chacun des triplés est caractérisé par des informations contextuelles permettant de décrire l'interaction entre l'utilisateur et la table. Le modèle de contexte pour l'utilisation collaborative de système interactif est utilisé dans cet article, pour permettre à des utilisateurs parents travaillant sur une autre table connectée d'interagir avec la table de leur enfant effectuant un exercice d'apprentissage de couleur. Par rapport au type de contexte que nous souhaitons analyser dans cette thèse, l'observateur doit pouvoir être représenté explicitement. Ceci impliquerait de modifier ce modèle pour l'appliquer à notre recherche. Baas et al., dans [51] utilisent la même technique pour la modélisation du contexte dans des situations de saisie de données médicales.

Dans la littérature d'autres travaux utilisent la représentation graphique basée sur le langage UML pour la modélisation du contexte comme le travail de Henricksen et al., dans [36] qui utilisent le modèle « entité /association » et le diagramme de classe du langage UML pour modéliser le contexte d'un système sensible au contexte. Une entité pouvant être une personne, un canal de communication ou un dispositif de communication. Les propriétés

d'une entité comme le nom d'une personne ou le numéro d'identification d'un canal de communication sont représentées par des attributs. Une entité est liée avec ces attributs et d'autres entités grâce à des relations unidirectionnelles appelées associations. Dans ce modèle graphique orienté de modélisation de contexte les entités et leurs attributs représentent les nœuds et les associations sont représentées par des arcs permettant de lier les nœuds entre eux.

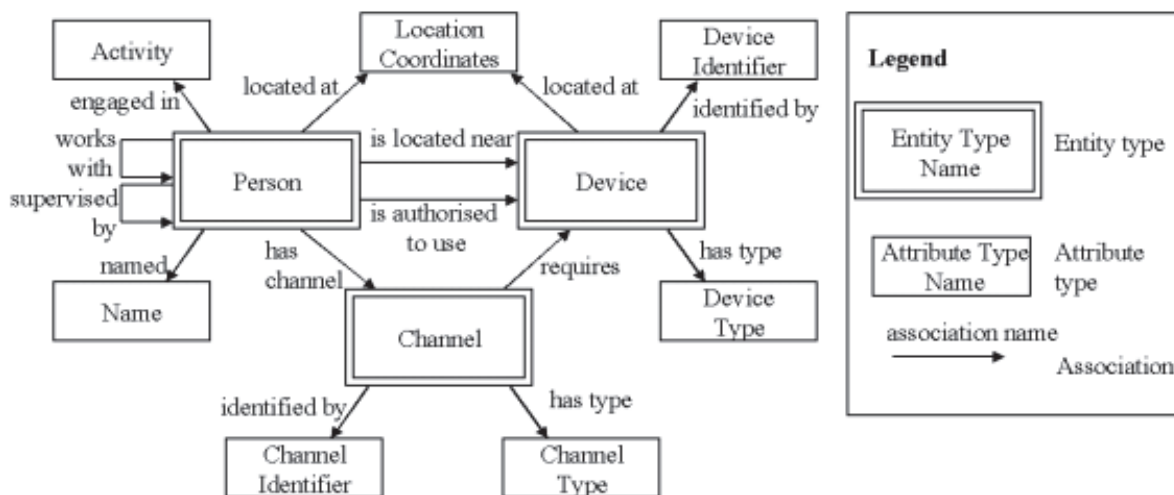


Figure I.8 : Modélisation du contexte avec le modèle graphique d'Henricksen et al., dans [36]

Un autre modèle graphique de modélisation de contexte est le langage CML (Context Modeling Language). Le langage CML est une amélioration de l'approche de modélisation graphique utilisant la méthode ORM (Object Role Modeling). La méthode ORM est basée sur les faits, elle est utilisée pour la modélisation des informations contextuelles imparfaites dans [52]. Dans ce travail, Henricksen et Indulska définissent quatre types d'imperfections sur les informations contextuelles : les informations inconnues, les informations ambiguës, les informations imprécises et les fausses informations. Une information contextuelle est inconnue quand elle est indéfinie donc impossible à déterminer. Il y a ambiguïté d'information contextuelle quand plusieurs valeurs différentes de la même propriété sont disponibles. Par exemple quand deux dispositifs de positionnement fournissent des emplacements différents pour la même personne. Les informations sont imprécises quand l'état rapporté est une approximation incorrecte du vrai état d'une propriété. Par exemple quand la position d'une personne ne peut être définie exactement. Les informations contextuelles sont fausses quand il y a une disparité entre les états réels et l'état rapporté d'une propriété. La modélisation consiste à identifier les types des faits et les rôles des types d'entités. Selon les auteurs, cette extension est plus formelle et expressive pour capter

différents types d'informations contextuelles, elle permet le raisonnement sur le contexte et décrit bien les informations imparfaites et résout l'ambiguïté dans les informations contextuelles.

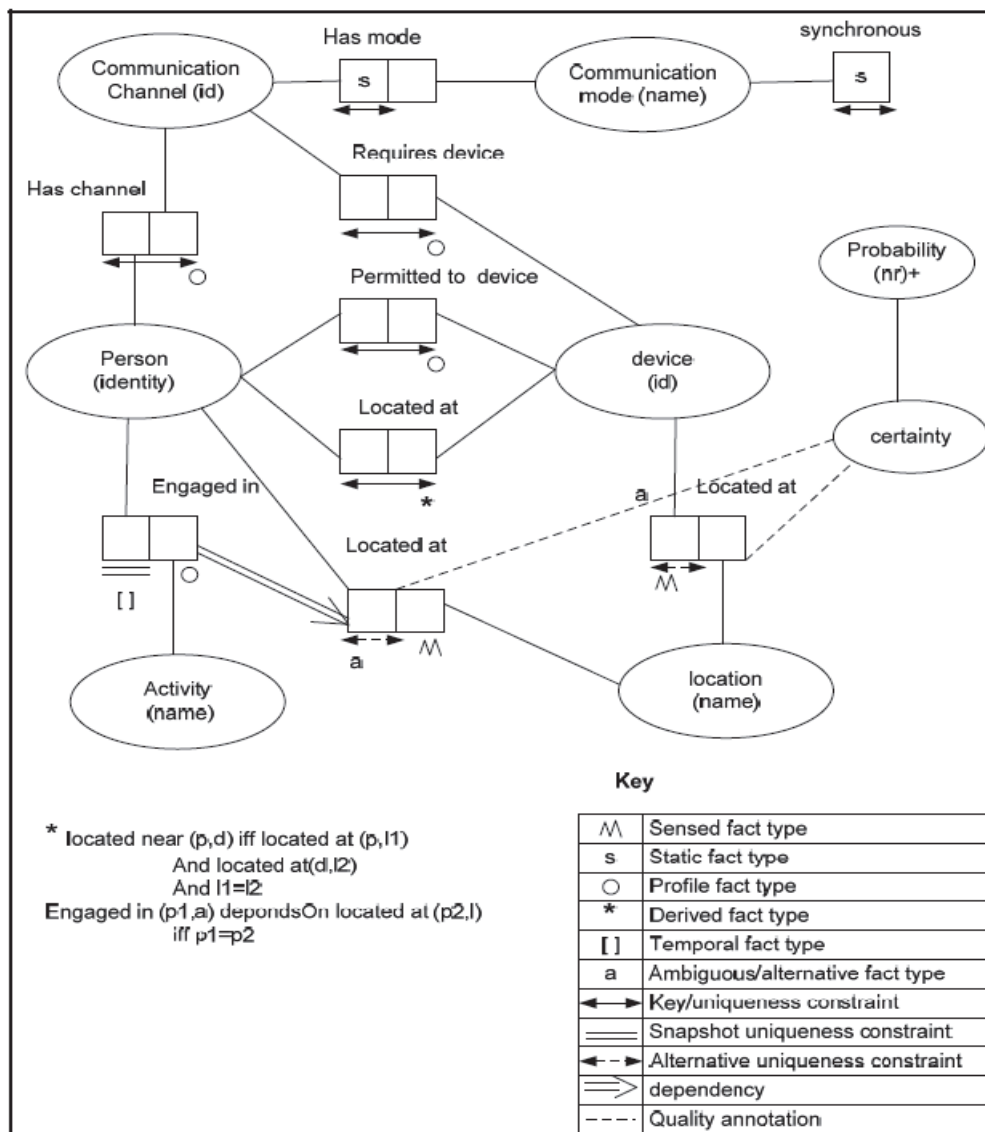


Figure I.9 : Exemple de modélisation du contexte avec le langage CML dans [41]

Le langage CML est un outil d'aide dans l'exploration et la spécification des exigences d'un système sensible au contexte pour les concepteurs. Il fournit un modèle de construction pour décrire les différents types d'informations (en termes de types de faits), leurs classifications (capteur, statique, profile), une métadonnée de qualité et les dépendances entre les différents types d'informations. Le CML permet aux types de faits d'être annotés pour indiquer si les informations sont ambiguës (des lectures alternatives d'emplacements) et si des informations historiques sont conservées. Il permet de mettre des contraintes sur les types de faits telles que

la contrainte de cardinalité entre les types de faits et de durée de vie. Un exemple de modélisation avec le langage CML est donné par la Figure I.9.

Le modèle orienté objet a été utilisé pour la modélisation du contexte pour ses principaux avantages d'encapsulation, d'héritage, de réutilisation, etc. Ces avantages permettent de tenir compte de l'aspect dynamique du contexte dans les systèmes sensibles au contexte dans un environnement ubiquitaire. Généralement, les entités sont représentées par les classes et les informations par les attributs de la classe. Le détail du contexte est invisible aux autres objets, grâce à la technique d'encapsulation. L'accès et la mise à disposition des informations sont contrôlés grâce aux interfaces des classes. L'approche objet est utilisée pour pouvoir intégrer facilement la représentation du contexte au sein de l'application qui en dépend. Cette représentation du contexte s'appuie sur les propriétés de nommage, d'encapsulation, de réutilisation et d'héritage. Ce modèle orienté objet est utilisé par plusieurs systèmes sensibles au contexte tels que : le projet Guide qui propose un guide touristique sensible au contexte fondé sur l'intégration d'un modèle orienté objet et un modèle d'information hypertexte. Le modèle contexte utilisé par le projet Guide ne contient que deux objets distincts : l'objet point de navigation et l'objet localisation. Ce modèle ne tient pas compte de l'aspect général du contexte, il n'est focalisé que sur la localisation. Le projet guide est proposé par Cheverest et al., dans [53].

Hofer et al., dans [54], ont introduit l'approche « HYDROGEN ». Ils ont proposé une architecture à 3 niveaux pour la modélisation du contexte dans le domaine de l'informatique mobile. Le contexte est modélisé sous forme des diagrammes de classes UML. Ils ont décomposé le contexte en contexte local et en contexte distant (les machines distantes en communication avec la machine locale). Chaque type de contexte est composé de plusieurs objets contexte qui constituent la superclasse de plusieurs éléments du contexte, notamment : le temps, le réseau, la localisation, l'utilisateur, la machine et autres qui peuvent être ajoutés par héritage de la super classe (Figure I.10). Cette modélisation offre la possibilité de représenter le contexte comme une hiérarchie et de décrire chaque élément indépendamment des autres en utilisant la technique d'encapsulation. Cela permet de favoriser la réutilisation de ces éléments dans d'autres applications mais ne donne pas la possibilité de représenter les relations entre les objets. Cette approche de modélisation est bien efficace en termes de distribution et d'abstraction. Cependant, elle reste propre à une application spécifique et ne permet pas de partager le contexte entre applications.

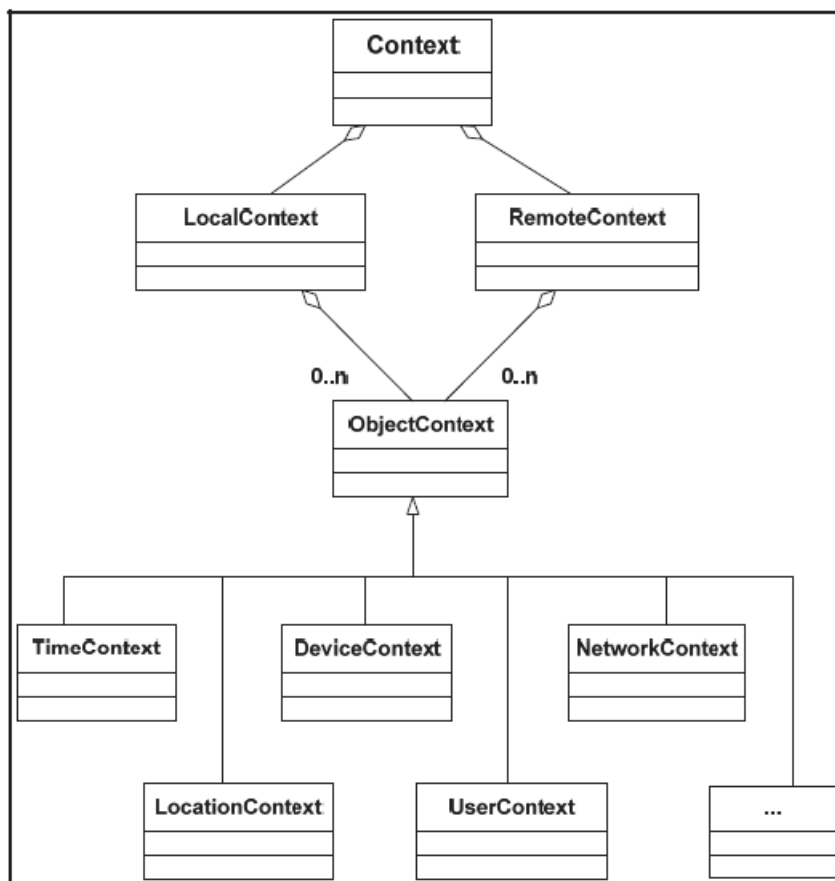


Figure I.10 : Modèle UML de l'approche Hydrogen d' Hofer et al., 2003

2.2.4 Modèles basés sur la logique

Le modèle basé sur la logique consiste à formaliser les informations contextuelles comme des faits à partir desquels il est possible d'inférer de nouveaux faits. Cette approche permet une représentation formelle du contexte et offre un mécanisme d'inférence ou de raisonnement sur les informations contextuelles. Ainsi les informations contextuelles peuvent être ajoutées, mises à jour ou supprimées de la base des faits. Le premier à utiliser le modèle logique pour modéliser le contexte est McCarty dans [19], il introduit le contexte comme un objet formel et définit la relation $ist(p,c)$. C'est la relation de base qui formalise le contexte, elle décrit le contexte c dans lequel la proposition p est vraie. Le modèle de contexte basé sur la logique McCarty supporte la notion d'héritage. Par exemple la Figure I.11 traduit une situation dans laquelle une personne ne peut pas être interrompue (en réunion ou au téléphone).

$$\begin{aligned}
Occupied(person) = & \exists t_1, t_2, event | engagedIn(person, event, t_1, t_2) \\
& \cdot ((t_1 \leq timenow() \wedge (timenow() \leq t_2 \vee isnull(t_2))) \\
& \vee ((t_1 \leq timenow() \vee isnull(t_1)) \wedge timenow() \leq t_2)) \\
& \wedge (event = 'inmeeting' \vee event = 'takingcall')
\end{aligned}$$

Figure I.11 : Exemple d'un modèle basé sur la logique de Ye et al., dans [55]

2.2.5 Modèles basés sur l'ontologie

Une ontologie est un ensemble structuré de concepts utilisée pour la représentation des connaissances. Les concepts sont organisés dans un graphe dont les relations peuvent être des relations sémantiques ou des relations de composition et d'héritage. Elle permet de modéliser l'ensemble des connaissances d'un domaine donné à travers une représentation utilisable par une machine. Une ontologie fournit un vocabulaire représentatif pour un domaine donné, un ensemble de définition et d'axiomes qui contraignent le sens des termes de ce vocabulaire de manière suffisante pour permettre une interprétation consistante des données représentées au moyen de ce vocabulaire. Une ontologie permet d'organiser des informations ou concepts pour construire de la connaissance. Les ontologies décrivent généralement les :

- Individus : objets de base
- Classes : ensembles, collection ou types d'objets
- Attributs : propriétés, fonctionnalités, caractéristiques ou paramètres que les objets peuvent posséder ou partager
- Relations : les liens que les objets peuvent avoir entre eux
- Événements : changement subis pas des attributs ou des relations.
- Métaclasses : des collections de classe qui partagent certaines caractéristiques.

Les ontologies possèdent quatre propriétés majeures intéressantes pour modéliser le contexte :

- Le partage de la connaissance, grâce à un langage flexible et extensible
- Le raisonnement logique, pour la déduction de faits implicites au contexte
- La description structurée de la connaissance
- La sémantique qui caractérise les relations existantes entre concepts.

Les ontologies sont utilisées pour la modélisation du contexte car elles fournissent un moyen formel pour représenter les données du capteur, du contexte et des situations dans une terminologie bien structurée, ce qui les rend compréhensibles, partageables et réutilisables par les humains et les machines (Chen et al., dans [56]). Plusieurs modèles basés sur les ontologies ont été proposés pour les systèmes sensibles au contexte. La plupart de ces modèles utilisent le langage OWL-DL car ce langage permet de modéliser un domaine particulier en définissant les classes, les individus, les caractéristiques des individus et les relations entre les individus. Les

descriptions complexes des classes et des propriétés peuvent être construites par la composition des descriptions élémentaires via des opérateurs spécifiques fournis par le langage. Ainsi des données contextuelles complexes de haut niveau peuvent être inférées à partir des données brutes issues des capteurs et représentées par des expressions structurées OWL-DL. C'est grâce à la définition de la sémantique entre concepts qu'il est possible d'effectuer les opérations de raisonnement pour extraire les données contextuelles complexe de haut niveau. Parmi les propositions de modèles basés sur les ontologies nous peut citer : SOUPA (Standard Ontology for Ubiquitous and Pervasive Applications) dans [57] par Chen et al., SOUPA est une ontologie dédiée à la modélisation du contexte dans un environnement ubiquitaire et CONON (CONtext ONtology) définie par Zhang et al., dans [58], est une ontologie dédiée aux environnements domestiques intelligents tels que le téléphone intelligent. Ce dernier se met au mode silencieux quand le propriétaire est couché et transfère tous les appels automatiquement vers le répondeur ou un système de surveillance intelligent. Ce système envoie un message quand il y a quelque chose d'anormal dans une maison et envoie un message au propriétaire pour qu'il puisse voir ce qui se passe chez lui.

Pour rendre employable et extensible les ontologies pour les SID (Système Informatique Diffus), Miraoui dans sa thèse propose une ontologie de service générique pour modéliser des informations contextuelles dans un SID [59]. Pour cela il définit une ontologie constituée de cinq classes, qui sont : équipement informatique, capteur, service, forme et contexte.

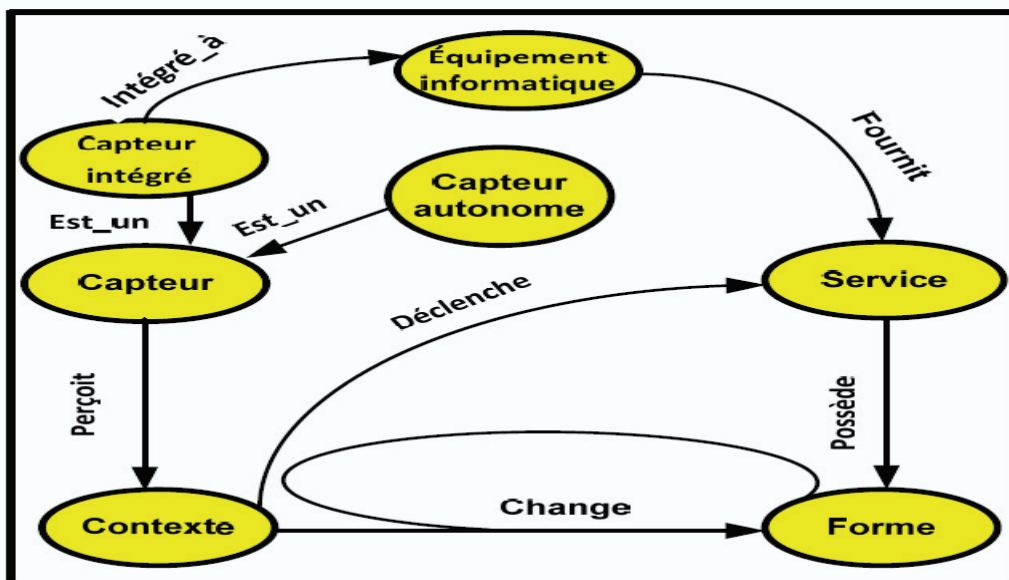


Figure I.12 : Les classes de l'ontologie de service de Miraoui [59]

La Figure I.12 représente les différentes classes de l'ontologie de service de Miraoui et les relations qui peuvent exister entre ces classes. La classe équipement informatique peut être un ordinateur ou un téléphone portable qui fournit des services à l'utilisateur et aux applications. Il peut être équipé d'un ou plusieurs capteur(s). Il doit communiquer avec les autres pour échanger des informations contextuelles. La classe capteur définit un capteur qui peut être intégré à un équipement informatique ou autonome. Cette source importante d'informations permet de recueillir beaucoup de types d'informations contextuelles (lumière, température, localisation, temps etc.). La classe service décrit le service fourni à l'utilisateur et aux applications. Chaque service possède une certaine fonctionnalité et peut être fourni sous plusieurs formes. Un service possède un ensemble d'informations de déclenchement qui permettent de le démarrer quand cet ensemble prend certaines valeurs spécifiques. Un service a également certaines caractéristique comme la durée, les entrées, les sorties, etc. La classe forme décrit la façon de fournir un service ou la qualité d'un service. Chaque forme possède ses propres caractéristiques (médias utilisés, modalité utilisée, etc.). Pour chaque forme, il y a un ensemble d'informations de changement qui permettent de changer d'une forme de service à une autre quand une information prend certaines valeurs spécifiques. La classe contexte décrit un ensemble d'informations recueillies par les différents capteurs du SID. Celui-ci spécifie la situation actuelle qui regroupe des informations et des événements qui se sont produits dans le système diffus. Cette classe contient deux catégories d'informations contextuelles : a) informations de déclenchement de services dont le changement de valeurs démarre les services et b) informations de changement de formes dont le changement de valeurs change les formes des services.

Un autre langage de formalisation du contexte, le langage CoOL dérive du modèle de contexte ASC (Aspect-Scale Context) qui peut être utilisé pour permettre la sensibilité au contexte, l'interopérabilité du contexte durant la découverte des services et l'exécution dans une architecture de système distribué. Les ontologies sont utilisées par CoOL dans le but d'avoir les mêmes définitions et les mêmes interprétations des choses par les différents éléments impliqués dans un système conscient du contexte. Le raisonneur de l'architecture distribuée infère des conclusions sur le contexte à partir de l'ontologie construite avec CoOL. Pour la mise en œuvre du langage CoOL plusieurs langages de spécification et d'interrogation d'ontologies ont été analysés en se concentrant sur deux points : la représentation des connaissances avec l'ontologie (la capacité à décrire les concepts, les attributs et les liens qui existent entre eux) et la connaissance encapsulée dans la requête (la capacité à interroger

l'ontologie pour inférer des conclusions). Les auteurs n'ont pas défini CoOL comme un langage monolithique, mais ils se basent sur deux sous-ensembles de représentation :

- OWL et DAML+OIL sont deux langages basés sur XML et RDF/S. Ils font partie des ontologies du web sémantique.
- F-Logic est un langage logique combinant les caractéristiques de l'orienté objet et de la logique de prédicats.

Le modèle ASC (Aspect-Scale Context) porte le nom des principaux concepts qui sont : aspect, échelle et information de contexte. Chaque aspect agrège une ou plusieurs échelles, chaque échelle agrège une ou plusieurs informations contextuelles. Ces concepts étant reliés entre eux par des relations. Et un aspect regroupe un ensemble cohérent d'échelles.

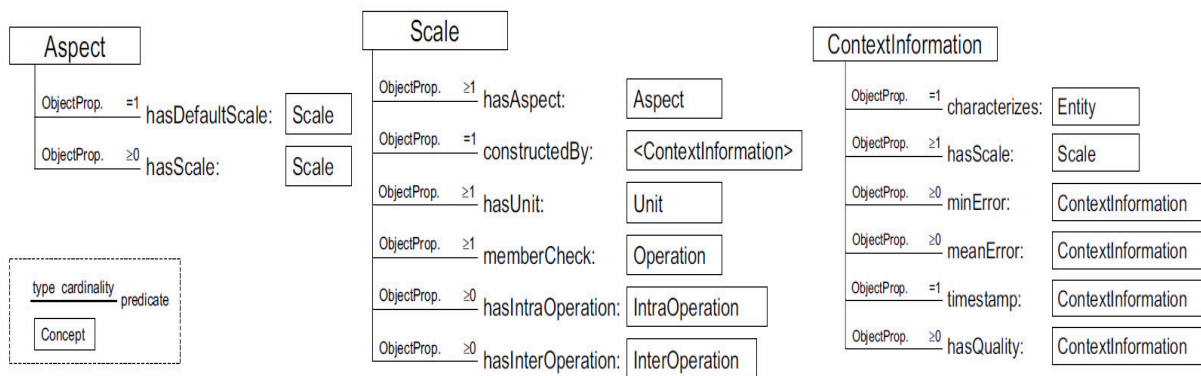


Figure I.13 : Modèle ASC, extrait de [60]

2.2.6 Graphes conceptuels

Les graphes contextuels sont utilisés pour représenter des connaissances et des raisonnements liés au contexte. Pour cela, ils exploitent trois types de contexte : le contexte externe, la connaissance contextuelle et le contexte procéduralisé. Le contexte externe est le contexte qui est sans intérêt pour un opérateur dans une situation donnée. La connaissance contextuelle est la partie du contexte qui présente un intérêt pour l'opérateur dans une situation donnée. Et le contexte procéduralisé est une structure de connaissances contextuelles mise en place par l'opérateur pour construire un raisonnement, une interprétation ou une explication d'une situation. Cette structure de connaissances contextuelles est relative à chaque opérateur, à la situation et au moment auquel l'opérateur travaille. C'est une structure non permanente qui évolue en fonction du focus d'attention de l'opérateur. Les graphes contextuels peuvent être utilisés dans plusieurs domaines pour la gestion d'une situation en tenant compte du contexte et des pratiques non formalisés des opérateurs. Par exemple,

ils ont été utilisés par la RATP pour la gestion des incidents sur les lignes de trains [61] et dans la santé pour le suivi du traitement du cancer du sein dans la thèse d'Aroua dans [62].

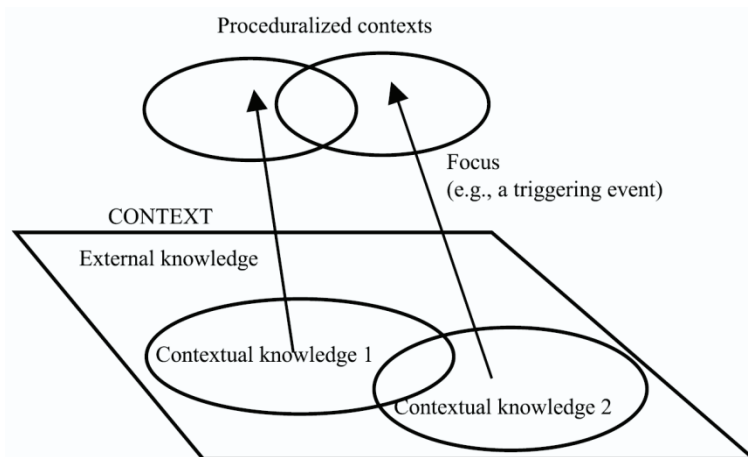


Figure I.14 : Représentation des trois types de contexte de Brézillon et Pomerol [61]

Un graphe contextuel est un graphe orienté acyclique avec une entrée et d'une sortie. Il est constitué de trois types de nœuds A_i , C_j et R_i . Les nœuds A_i correspondent aux actions et sont représentés par des boîtes carrées. Les nœuds C_j correspondent aux éléments contextuels, ils sont représentés par des cercles. Et les nœuds R_i sont des nœuds de recombinaison permettant de regrouper les différents chemins de graphe pour arriver à la sortie. Un chemin de graphe est constitué d'une séquence d'actions, des éléments contextuels et des nœuds de recombinaisons. Par exemple dans la figure I.15, le chemin $\{A1, A2, A7 \text{ et } A9\}$ passe par les éléments contextuels $C1$, défini par le contexte $C1.2$ et $C4$ défini par le contexte $C4.1$. À partir d' $A7$, le chemin passe par les nœuds de recombinaison $R4$ et $R1$ pour atteindre la dernière action $A9$.

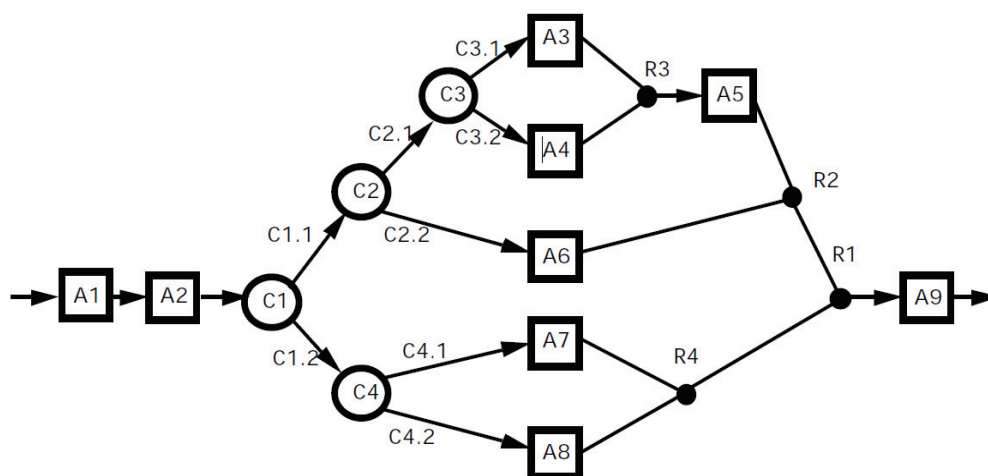


Figure I.15 : Exemple de graphe contextuel, extrait [63]

Par rapport à notre objectif de recherche qui est l'identification des connaissances contextuelles d'un critère donné, l'approche des graphes contextuels n'est pas très adaptée car elle fait la représentation graphique des solutions alternatives de résolution d'un problème par un opérateur en tenant compte du contexte. Les différents contextes sont représentés en parallèle pour atteindre la sortie du graphe. L'approche des graphes contextuels est faite pour décrire les pratiques de résolution d'un problème par un opérateur en tenant compte du contexte.

2.2.7 Modèles avec le Réseau de Pétri

Un réseau de Pétri est une méthode de modélisation graphique utilisée pour la représentation des systèmes dynamiques, discrets ayant des entrées, des sorties et des états. Ils sont utilisés pour la modélisation des systèmes sensibles au contexte en raison de la sémantique formelle, de la description du système basée sur des états et ce formalisme permet de nombreuses des techniques d'analyses. Par exemple, Sun et al., dans [64], utilisent le réseau de Pétri pour représenter les données du contexte d'une voiture intelligente. Les données du contexte sont représentées en trois couches selon le degré d'abstraction et de la sémantique en trois couches : capteur, atomes du contexte et la couche situation de contexte. La couche capteur est la source des données contextuelles. La couche atomes du contexte joue le rôle de médiateur entre le monde physique et le monde sémantique en récupérant les données de la couche capteur. Elle fournit des données conceptuelles élémentaires et utilise l'ontologie pour fournir la définition des classes et des attributs. Elle n'est pas capable de représenter les connaissances contextuelles complexes. Pour représenter ces connaissances, une couche de situation est construite sur la couche atomes pour ajouter les atomes individuels du contexte dans la situation de contexte significative. La définition du modèle de représentation des données contextuelles est donnée par le tuple suivant : $SN = (S, T, F, W, M_0)$

- $S = \{s_1, s_2, \dots, s_n\}$: est l'ensemble fini des situations de contexte
- $T = \{t_1, t_2, \dots, t_n\}$: est l'ensemble fini de transition entre les situations de contexte
- $F \subseteq (S \times T) \cup (T \times S)$: est une fonction qui définit la corrélation par des arcs de transition liés aux situations
- $W \rightarrow \{1, 2, 3, \dots\}$: le poids de chaque transition
- $M_0 : S \rightarrow \{1, 2, 3, \dots\}$: l'inscription initiale du système, c'est-à-dire la mission initiale de jeton.

Riahi Dans sa thèse [65], décrit le réseau de Pétri comme un graphe orienté, bipartite, défini par un tuple $\langle C, P, T, F, B \rangle$

- C : représente la structure du réseau
- P : l'ensemble de place
- T : l'ensemble de transition
- F : $P * T \rightarrow N$, la fonction d'incidence ascendante
- B : $P * T \rightarrow N$, la fonction d'incidence descendante.

Riahi et al., dans [59], proposent une approche pour la spécification et la génération d'interfaces sensibles au contexte c'est-à-dire des interfaces qui s'adaptent au contexte courant de l'utilisateur. Pour cela, les informations contextuelles sont captées et modélisées avec le modèle « utilisateur, plateforme et environnement » et le réseau de Pétri est utilisé pour modéliser le contexte et la tâche de l'utilisateur.

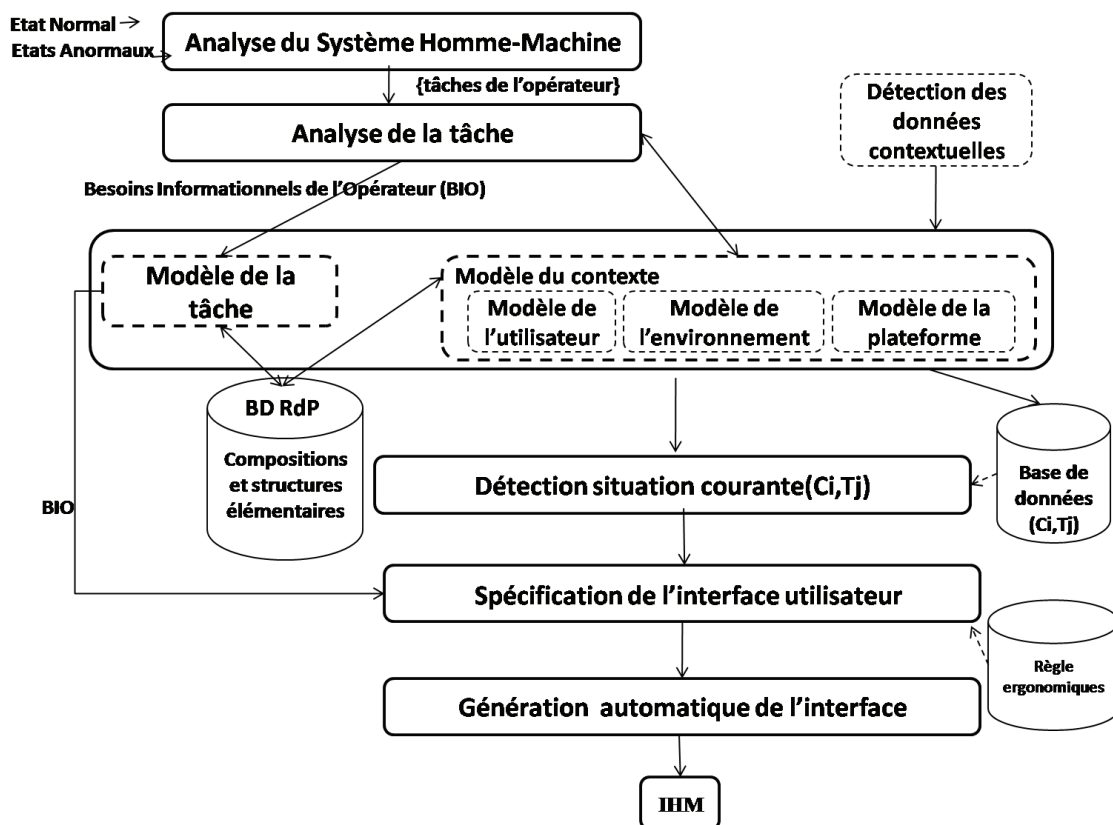


Figure I.16 : Approche utilisant le réseau de Pétri pour générer des interfaces sensible au contexte

Pour cela, les places sont utilisées pour représenter l'état du contexte/comportement de l'utilisateur vis-à-vis de l'évolution de l'environnement et les transitions pour décrire l'évolution entre ces différents états. Il est à noter que chaque changement au niveau du contexte ou au niveau de la tâche doit affecter l'interface utilisateur afin de l'informer de la situation courante. Comme chaque tâche de l'utilisateur est spécifique à un contexte donné,

une base de couples contexte C_i et la tâche T_j sont construites pour spécifier et générer l'interface de l'utilisateur. Pour cela l'approche dispose d'un détecteur de situation courante utilisant la base de couples (C_i, T_j) et le BIO (Besoins Informationnels de l'Opérateur) pour spécifier l'interface utilisateur. La spécification de l'interface utilisateur est alors utilisée pour générer automatiquement l'interface utilisateur en fonction de la tâche T_j effectuée par utilisateur et du contexte C_i d'utilisation.

2.2.8 Modèles spatiaux et temporels

Les modèles spatiaux sont des modèles qui sont focalisés sur la localisation ou l'emplacement physique. Cet emplacement peut être un emplacement réel des entités qui est décrit dans les informations du contexte ou l'emplacement du capteur qui mesure les informations du contexte. Cette information de localisation est soit prédéfinie (si les entités sont statiques), ou obtenue par les systèmes de positionnement qui permettent de suivre les objets mobiles par rapport à leur position. Ces modèles gèrent deux types de coordonnées spatiales, les coordonnées géométriques et les coordonnées symboliques. Les coordonnées géométriques représentent des points ou des zones dans un espace métrique, telles que les coordonnées du GPS (latitude, longitude et altitude). Les coordonnées symboliques sont représentées par un identifiant, tel qu'un numéro de chambre ou l'ID d'un point d'accès dans un téléphone mobile.

Les modèles spatiaux et temporels sont utilisés en général pour des applications de maison intelligente. Ces applications sont capables de raisonner sur les caractéristiques spatiales et temporelles du contexte. Pour spécifier des connaissances temporelles sur les activités humaines Augusto et al., dans [66], ont introduit des opérateurs temporels dans les règles « Evènement-Condition-Action ». Gottfried et al., dans [67], utilisent la logique temporelle d'Allen [68], pour décrire, contraindre et raisonner sur l'ordre temporel des événements dans une maison intelligente. Par exemple pour la suite d'évènement d'une personne qui rentre dans une pièce et déclenche un capteur, il y a trois événements qui sont : ouverture de la porte (I1) accès de la personne à la pièce (I2) et déclenchement du capteur pendant la présence de la personne dans la pièce (I3). Ces trois événements sont temporellement liés et la composition de ces événements indique que l'évènement I1 doit se réaliser avant les événements I2 et I3. La logique d'Allen permet de contraindre le déclenchement du capteur après l'ouverture des portes.

2.2.9 Modèles hybrides

Les modèles hybrides consistent à combiner plusieurs approches de modélisation du contexte dans le but de satisfaire les exigences des applications sensibles. Plusieurs travaux proposent des modèles hybrides pour la modélisation du contexte comme celui de Henricksen et al. dans [69], combinent les approches basées sur les ontologies avec les modèles basés sur les faits CML (Context Modélisation Language). L'objectif est de combiner les avantages du CML (le traitement des informations du contexte ambigu et imparfait) avec le soutien de l'interopérabilité et des divers types de raisonnement fournis par les modèles ontologiques. Cette approche est basée sur une cartographie de modélisation CML pour OWL-DL. Il est intéressant de noter qu'en raison de certaines limites d'expressivité d'OWL-DL, une cartographie complète entre CML et OWL-DL ne peut être obtenue. En ce qui concerne les questions d'interopérabilité, les avantages procurés par une représentation ontologique du modèle de contexte sont clairement reconnaissables.

Un autre exemple de modèle hybride est celui d'Agostini et al., dans [70] proposent une combinaison des approches basées sur les langages de balises avec les modèles basés sur les ontologies. En effet, les auteurs intègrent la représentation des données du contexte (profils CC/PP) avec les classes et les relations OWL-DL. Un autre modèle hybride combine les modèles de balises et les modèles basés sur l'ontologie.

En plus de tous ces types de modélisation du contexte, d'autres approches de modélisation de contexte pour les systèmes sensibles au contexte peuvent être trouvées dans la littérature comme les modèles probabilistes et les modèles issus de l'intelligence artificielle. Les modèles issus de l'intelligence artificielle utilisent les approches basées sur la logique floue, les arbres de décisions, les réseaux de neurones et les arbres de suffixes pour traiter les données contextuelles. Les différentes approches de modélisation du contexte pour les systèmes sensibles au contexte sont comparées suivant des critères à satisfaire dans un environnement ubiquitaire. Ces critères peuvent être trouver dans les synthèses réalisées par Bettini et al., dans [44], Strang et Linnhoff-Popien dans [43], Baldauf et al., dans [71]. Une étude comparative de ces approches est faite dans la thèse de Riahi dans [65].

On constate dans cette partie, que l'on sait maintenant gérer les connaissances de contexte, y compris pour les exploiter par des applications tierces. Toutefois, si la question de la capture puis de l'exploitation d'informations contextuelles est bien gérée, la découverte du contexte pertinent lors de l'analyse d'une activité n'est pas traitée dans cette littérature.

2.3 Approche de découverte de connaissances contextuelles

Dans un travail se référant à la question de l'ingénierie des connaissances contextuelles dans [72] Castelli et al., considèrent la manière de « produire » les informations contextuelles à partir de capteurs RFID, sans aborder la question de la construction du sens qui devra ensuite avoir lieu pour exploiter cette information estampillée « *contexte* ». Un travail plus orienté sur la découverte de connaissances contextuelles de Ngoc et al., [73] dans le cadre d'un middleware robotique, repose sur l'association de plusieurs méthodes d'apprentissage automatique à partir d'un jeu de capteurs et actionneurs avec une mise en correspondance avec une ontologie contextuelle pour l'interprétation. Les auteurs ne s'intéressent pas à la construction de l'ontologie mais à la découverte de connaissances contextuelles exploitables par un système sensible au contexte. Un travail effectué par Miléa et al., dans [74] de même nature mais plus général encore, propose une manière d'intégrer la médiation contextuelle dans un processus de découverte de connaissance. La notion de contexte y est bien définie et les connaissances contextuelles permettent de multiplier les points de vue sur les patterns découverts. La connaissance découverte est en relation avec le contexte qui lui est attribué, mais cette connaissance contextuelle doit venir de l'expert médiateur de la découverte de connaissance. Plusieurs travaux s'intéressent à la représentation ontologique des contextes avec par exemple le langage CoOL (Context Ontology Language) [60], tandis que des travaux européens [75] se sont efforcés d'exploiter ces possibilités dans le cadre de l'interopérabilité ubiquitaire.

Les travaux qui traitent la question de la découverte de connaissances contextuelles sont très rares, nous allons détailler deux d'entre eux, ceux de Ngoc et al., dans [73] et Mena et al., dans [38].

Ngoc et al., dans [73] proposent un processus de découverte de connaissances contextuelles modélisant les comportements de l'utilisateur pour identifier les besoins de l'utilisateur d'un système informatique qui devraient être satisfaits par le système sensible au contexte. Ces besoins sont alors gardés sous forme de connaissances (règles de classification, clusters) qui seront exploitées par le module de raisonnement et d'apprentissage pour la mise en place du système sensible au contexte.

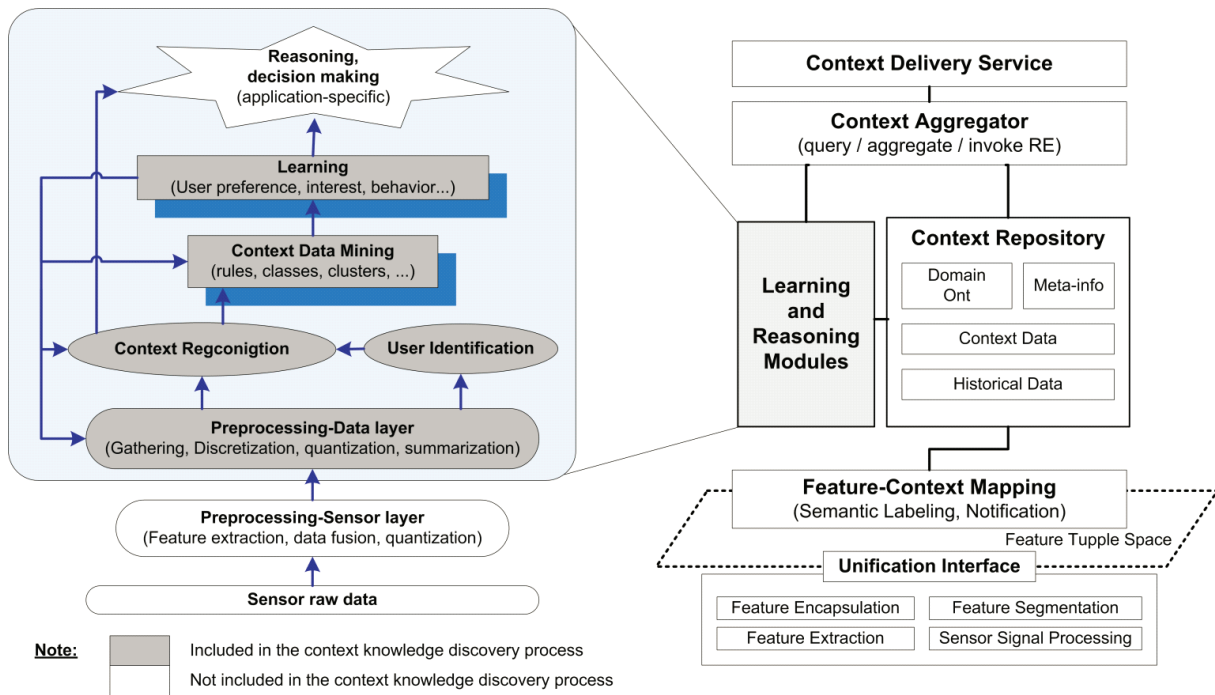


Figure I.17 : Processus de découverte de connaissances contextuelles de Ngoc et al.

Mena et al., dans [38] présentent un modèle générique de contexte et une méthode d'identification du contexte pour le développement d'un système sensible au contexte. Cette méthode permet d'identifier et de construire le contexte en fonction des exigences du système à développer. Le processus de modélisation et d'identification du contexte est fait en parallèle avec le processus classique de conception du système. La Figure I.18 montre comment le processus d'identification du contexte intervient dans le processus classique d'analyse du système.

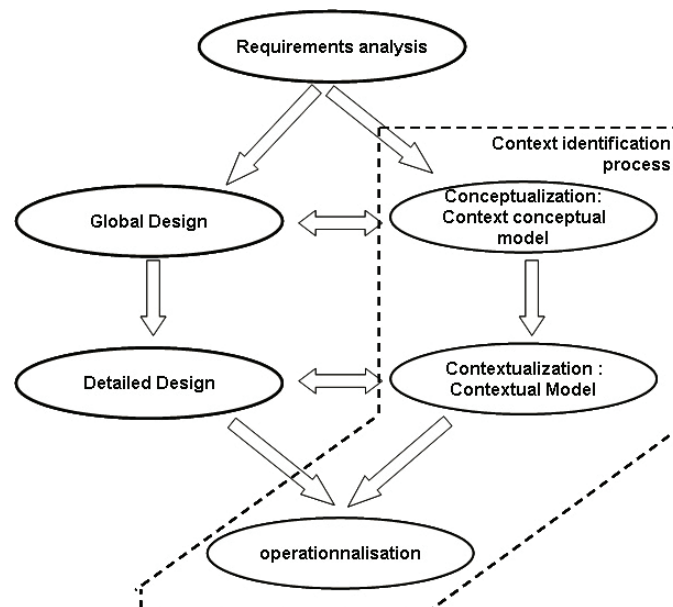


Figure I.18 : Processus d'identification du contexte dans le processus de conception d'un système de Mena

Le processus d'identification du contexte est composé de trois étapes : la conceptualisation du contexte, la contextualisation et l'opérationnalisation du contexte.

L'étape de conceptualisation (conception du modèle de contexte) permet au concepteur de déterminer le domaine du système à développer, par exemple un simulateur, un système sensible au contexte, un système d'interaction homme-machine. Une fois identifié, le domaine est divisé en sous domaines dont l'analyse identifiera les entités impliquées dans le développement du système. Durant cette étape d'identification du domaine, le concepteur détermine les entités impliquées et les relations qui existent entre elles pour ensuite les classer par groupe. Parmi les entités sélectionnées celles qui doivent être sensibles au contexte doivent être définies. Cette étape permet également de faire une cartographie des entités et des problèmes qu'elles doivent résoudre mais aussi des relations qui peuvent exister entre les différentes entités.

Pour l'étape de contextualisation les auteurs utilisent la définition du contexte suivante : « le contexte c_i d'une entité e_i pour résoudre un problème p_k est définie comme un ensemble d'entités e_j (i différent de j) reliées à p_k en interaction avec e_i et un ensemble de relation $R_{i,j}$ entre les entités e_i et e_j ».

A partir de cette définition du contexte et de la cartographie des entités et des problèmes de l'étape de conceptualisation, le contexte de chaque entité est identifié. Cette étape permet de structurer le contexte en construisant les relations et les hiérarchies entre les contextes en relations avec les différentes entités. Pour déterminer le modèle de contexte, deux types de raisonnement sont mis en place et des opérations de fusion, de mise à jour, etc, sont utilisées sur le contexte. Le premier type de raisonnement est relié à l'extraction de contexte sur le contexte des entités et le second sur le contexte de résolution des problèmes.

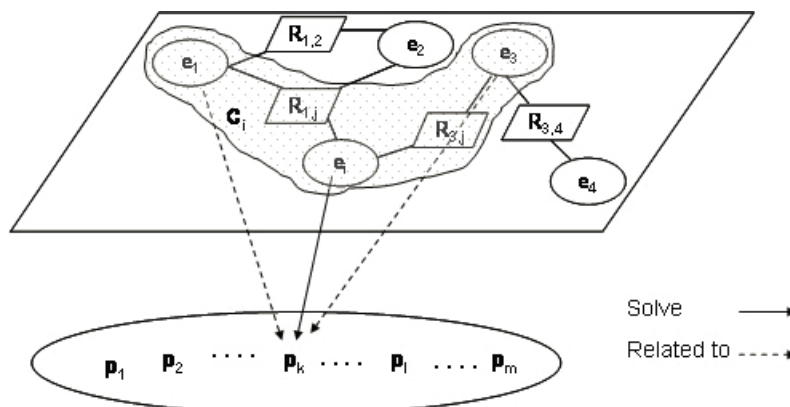


Figure I.19 : Contexte des entités pour la résolution de problèmes

L'étape d'opérationnalisation permet la transcription du modèle de contexte identifié dans l'étape de contextualisation dans un langage formel et opérationnel de représentation de connaissances. Cette étape est itérative car plusieurs boucles sont nécessaires pour construire des contextes opérationnels adaptés au système et aux exigences de ce dernier.

Cette méthode de développement d'un système sensible au contexte en considérant le contexte comme un triplé « domaine, entité et problème » a été utilisé par Mena et al., dans [38] pour le développement d'un simulateur multi-agent de sauvetage de cas d'urgence d'accident.

Nous n'avons pas repéré d'autres travaux spécifiquement dédiés à la question de la découverte de connaissances contextuelles que ce soit à partir d'un dispositif d'observation d'activité (notre cadre) ou non. Il est pourtant difficile pour un expert de nommer et de qualifier les contextes explicatifs pour chaque situation observable dans une activité donnée et c'est parfois même une tâche très complexe, comme le montre le contexte de l'Ifsttar où l'on cherche à élucider les contextes explicatifs de la consommation en carburant de l'activité de conduite. Par contre, il y a beaucoup d'autres travaux sur le contexte qui n'ont pas pour objectif de faire de la découverte de connaissances contextuelles mais qui proposent une modélisation du contexte pour permettre aux systèmes d'être sensibles au contexte comme décrits dans la partie 2.2.

Toute phase d'identification et de modélisation d'un contexte d'une activité nécessite de pouvoir observer et caractériser l'activité même. Dans la suite de l'état de l'art, nous choisissons de détailler la manière de décrire comment formaliser les traces issues de l'observation d'une activité pour les considérer comme une source de connaissance à la représentation de l'activité et par suite aux contextes de cette activité. Nous considérons dans la section suivante la notion de trace « modélisée », c'est-à-dire formalisée pour permettre la représentation et certains calculs sur l'activité observée.

3 La notion de Trace Modélisée (M-Trace)

Tracer une activité consiste à la décrire à partir des empreintes qu'elle laisse dans l'environnement lors de sa réalisation. La trace résultante constitue un point de départ pour l'analyse et l'exploration de l'activité ainsi observée. La construction de traces à des fins d'analyse et de partage d'informations constitue une activité très ancienne. Elle est à la base de nombreuses découvertes de connaissance pour une meilleure compréhension d'un comportement observé ou d'une activité. Elle était utilisée par nos lointains ancêtres et est utilisée de manière formelle par les scientifiques. Elle est par exemple utilisée en archéologie

pour comprendre et expliquer le mode de vie des humains et leur évolution par l'observation des traces observables bien après leur disparition [76].

Il n'est pas étonnant que la construction de traces d'activité à partir de l'observation des empreintes laissées pendant la réalisation d'une activité dans un environnement donné, puisse être utilisée pour la découverte de nouvelles connaissances exploitables pour la compréhension et l'amélioration de cette activité. En informatique, les fichiers d'erreurs (logs) sont souvent analysés pour corriger les erreurs d'un programme ou pour améliorer les interactions d'un utilisateur avec une application informatique. Observer les interactions homme-machine en informatique, permet de fabriquer des traces très riches sur les comportements observables pendant une activité médiée par un environnement informatique. Ces informations peuvent être exploitées pour comprendre et expliquer les comportements des agents humains ou artificiels interagissant lors de l'activité. Elles peuvent également être utilisées pour améliorer les interactions homme-machine avec des services personnalisés en fonction des habitudes et besoins des utilisateurs. Une synthèse de l'état de l'art sur les travaux réalisés dans différentes disciplines sur la notion de trace est disponible dans [77] pour une approche anthropocentrée. Cette état de l'art pluridisciplinaire est particulièrement détaillé dans [78] pour une approche en sciences cognitives, intégrant la facette de l'intelligence artificielle. Le numéro spécial d'Intellectica « De la trace à la connaissance à l'ère du Web » rassemble des auteurs des sciences de l'information et de la communication qui montrent la nature profonde d'une trace construite de manière volontaire ou non comme documentarisation d'une expérience située d'un observateur avec son corpus de connaissances interprétatives. La réutilisation informée des traces nécessite donc que la genèse de construction et d'interprétation soit associée à ces traces. Plusieurs contributions sont ensuite proposées par des auteurs des sciences de l'informatique pour décrire des formalisations, des modèles de traces, des opérations possibles sur les traces et des usages démonstratifs de ces approches pour accompagner l'analyse de l'activité d'apprentissage humain. Une approche empirique de l'usage des traces de manière réflexive est proposée par des auteurs en sciences de la psychologie en montrant les difficultés d'appropriation des traces lorsqu'elles ne sont pas explicitées et manipulables par leurs auteurs.

Des travaux plus spécifiquement liées à l'élaboration d'indicateurs à partir de traces se sont multipliés depuis quelques années. Ces travaux se sont particulièrement développés dans le domaine de l'apprentissage humain [3], [79] avec des efforts de modélisation pour permettre

un usage partagé des traces entre chercheurs [80], ou pour concevoir des environnements pour l'apprentissage humains intégrant le calcul d'indicateur à partir de traces [81].

En ingénierie des connaissances, plusieurs travaux se sont intéressés à l'utilisation des traces pour modéliser l'expérience [82], [83]. Plusieurs travaux se sont intéressés à l'utilisation des traces pour l'assistance à l'activité comme [84], [85]. La modélisation de traces est considérée en général uniquement comme un problème technique de représentation, en particulier dans le domaine du web et l'on peut citer par exemple le format xAPI, se concentrant sur la collecte d'événements observés à des fins d'interopérabilité [86], [87]. Plus récemment dans le cadre du projet ANR Hubble³, plusieurs systèmes de représentation de traces à des fins d'analyse pour des chercheurs confrontent leurs approches [88] : UnderTracks⁴ pour l'équipe Metah du LIG qui est un outil de préparation, de gestion de corpus de traces pour l'analyse de données et la fouille de données issues d'apprentissage humain, SMOOPLE⁵ [89] qui est surtout une plateforme de déploiement pour des Moocs (Massively Open Online Courses) sémantiques, UTL⁶ qui est un langage permettant de représenter et réaliser des traitements sur des traces d'utilisation d'un environnement pour l'apprentissage humain [90], et enfin le système de gestion de base de traces *modélisées* (kTBS⁷) qui est la seule approche revendiquant une orientation *connaissance et raisonnement* et considère la *trace modélisée* comme objet informatique à part entière. Comme nous allons le voir dans les sections suivantes, le choix de l'approche des traces modélisées était d'autant plus pertinent pour soutenir le travail de recherche que cette approche trouve ses fondements dans des travaux réalisés à l'INRETS d'abord avec Olivier Georgeon[91], puis Benoît Mathern [6]. Nous allons détailler dans la section suivante les principales caractéristiques de cette approche des traces *modélisées* avec une approche *ingénierie des connaissances*.

Depuis plusieurs années l'équipe Silex maintenant TWEAK (Traces, Web, Education, Adaptation, Knowledge) du LIRIS étudie la notion des traces d'interactions pour exploiter les traces numériques construites à l'issue de l'observation des interactions entre un utilisateur et son environnement informatique. L'objectif de l'équipe Silex est d'utiliser l'observation des interactions d'un agent humain ou artificiel avec son environnement (numérique lorsqu'il s'agit d'un agent humain) [4], [82], [84], [92]–[94] pour construire des traces d'interactions considérées comme sources de connaissance pour des tâches de découverte ou d'apprentissage

³<http://hubblelearn.imag.fr/>

⁴<https://www.liglab.fr/fr/presentation/pole-ingenierie/plate-forme-undertracks>

⁵<http://3s.enstb.org/drupal/?q=node/7> pour un accès à Smoople

⁶<http://eiah.univ-lemans.fr/UTL/UTL.xml>

⁷<http://tbs-platform.org/tbs/doku.php/tools:ktbs>

interactif. Pour y parvenir, l'équipe Silex a développé la notion de *Traces Modélisées* appelés *M-Traces* [95], elles permettent de définir une structure de données propre à l'activité observée. Chaque *M-Trace* est constituée d'un *Modèle de Trace* qui correspond aux structures des données ou des observables de l'activité et d'une *Trace* qui contient les instances de ces observables conformément au *Modèle de Trace*. Les *M-Traces* proposent plusieurs avantages lors de l'analyse tracée d'une activité, comme la définition de la structure de données des observables en amont en fonction des besoins de l'analyse, l'accès et la manipulation des *M-Traces* en vue de les enrichir pour la découverte de nouvelles connaissances. Ces nouvelles connaissances peuvent alors être utilisées pour comprendre, expliquer et améliorer l'activité tracée. L'analyse de ces *Traces* permet d'assister un utilisateur en temps réel en exploitant les interactions antérieures de ce dernier en lui proposant des options personnalisées.

L'utilisation des *Traces* antérieures d'un utilisateur pour l'assister ou faciliter ses interactions avec un système observé relèvent d'une approche *réflexive* de l'observation. La notion de *réflexivité* dans les *M-Traces* est mise en évidence par Mille et al., dans [84]. Cette inscription de l'expérience tracée nécessite une modélisation de l'expérience tracée, selon l'approche des *Traces Modélisées (M-Traces)* [82]. Cette approche qui permet la création, la manipulation et l'enrichissement des *M-Traces* à partir des interactions homme-machine d'un système observé a été mise en place d'abord pour faciliter la réutilisation d'épisodes d'expériences comme dans *MUSETTE* [93] puis systématisée dans l'approche des Systèmes à Base de Traces (*SBT*)[92].

L'approche *MUSETTE* [93] concerne la modélisation de l'expérience tracée. Elle correspond au premier type d'utilisation des traces d'interactions à partir d'observations suivant un modèle d'observation de l'activité défini par l'analyste. Cette approche suppose donc que ce modèle d'observation ait été conçu par un processus d'ingénierie de connaissance en amont de l'analyse des traces d'interactions. L'activité est observée selon le modèle mis en place en amont, dans le but d'extraire des épisodes de l'activité. Ces épisodes sont alors réutilisés pour assister l'utilisateur lors de son interaction avec le système observé.

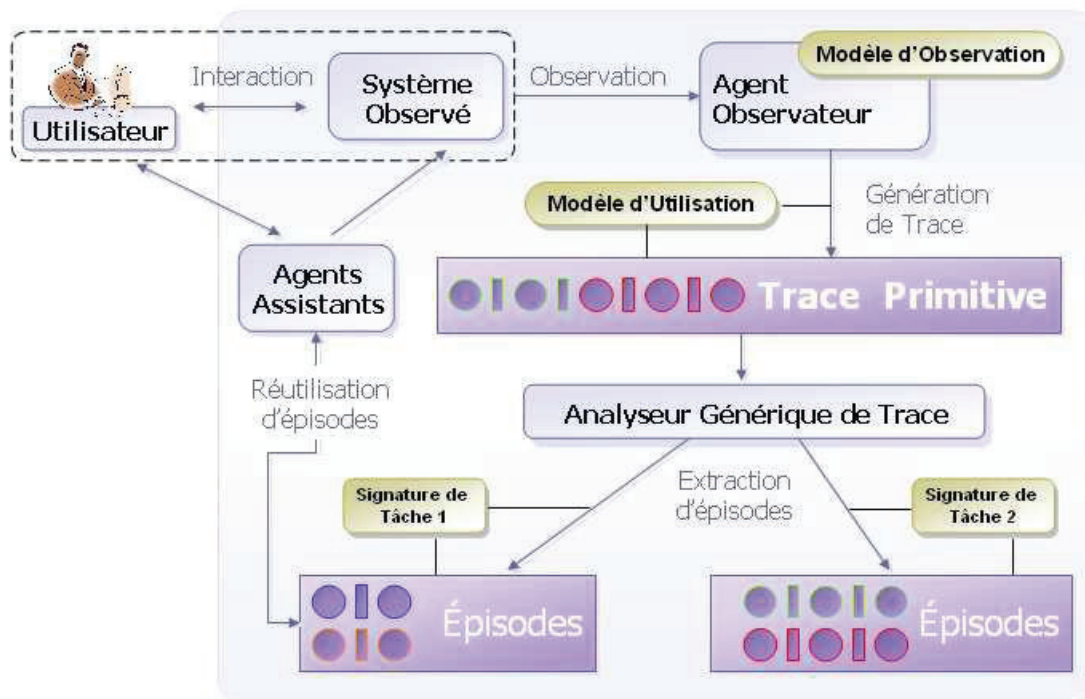


Figure I.20 : Approche de modélisation *MUSETTE* de Champin et al.[93].

La figure (Figure I.20) représente la modélisation *MUSETTE*, elle est constituée d'un *Utilisateur (U)*, d'un *Système Observé (SO)*, d'un *Agent Observateur(AO)* et d'un *Agent Assistant(AA)*. Les interactions entre l'*U* et le *SO* sont observés par l'*AO* (*analyste*) pour définir le *Modèle d'Observation*, qui est le *Modèle de Trace* décrivant la structure des données du *Système Observé*. L'*AO* utilise ce *Modèle de Trace* sur les traces collectées ou primaires d'interaction pour extraire des épisodes qui seront réutilisés par l'*AA* pour assister l'*U* lors de son interaction avec le *SO*. L'approche *MUSETTE* permet d'utiliser les traces d'interactions comme support d'assistance basé sur la réutilisation des épisodes de l'activité identifiés à partir d'un *Modèle de Trace* de l'interaction entre un utilisateur et son environnement. Le *Modèle de Trace* étant créé lors d'une phase d'ingénierie de connaissance basée sur l'observation des traces d'interactions d'une activité.

Le Système à Base de Traces (*SBT*) [92] a été mis en place pour généraliser l'approche *MUSETTE* à toutes sortes de découvertes de connaissances à partir de traces d'interaction. Il s'agit d'une plateforme permettant à l'analyste de faire la construction du *Modèle de Trace* et la création des *traces* qui sont associées au *Modèle de Trace*. Le *SBT* est un système de gestion de base de *M-traces*, un objet informatique original qui dispose de différentes opérations sur les *M-traces* comme la collecte et la transformation. Le *SBT* offre des fonctionnalités de gestion des *M-traces*, de persistance des bases de traces, de sécurité d'accès, et de la manipulation de *M-Traces*. Les fonctions d'un *SBT* sont analogues à celles d'un

Système de Gestion de Base de Données (*SGBD*) mais l'objet informatique géré est de type *M-Trace*. La démarche de l'approche d'un *SBT* formalise les traitements pouvant être appliqués sur une trace depuis sa collecte jusqu'à son utilisation comme support d'assistance dans un cadre très générique. Un *SBT* peut être utilisé pour analyser tous les types d'environnement observable qui peut donner lieu à une construction de *M-traces* comme par exemple, les EIAH (Environnement Informatique pour l'Apprentissage Humain) et l'analyse de l'activité de conduite.

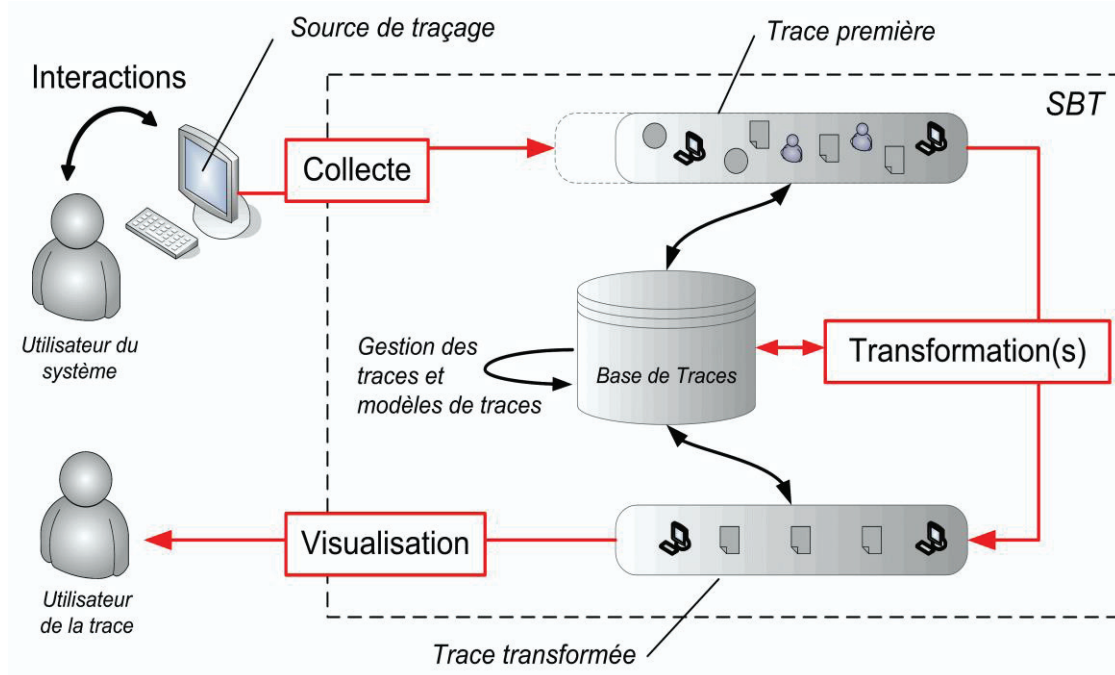


Figure I.21 : Principe d'un Système à Base de Traces (SBT) [96].

Un *SBT* offre trois fonctionnalités principales qui sont : la *Collecte*, la *Transformation* et la *Visualisation*. L'ensemble des traces manipulées dans le *SBT* sont stockées dans la *Base de Traces* et elles sont accessibles par l'intermédiaire des requêtes d'interrogations adressées au *SBT*.

La fonctionnalité *Collecte* permet de capturer les interactions entre l'utilisateur et son environnement numérique par l'intermédiaire d'une *source de traçage* et crée une première *M-Trace* qui est nommée la *trace première*. La fonctionnalité *Transformation* permet de créer une nouvelle *M-Trace* construite par transformation d'une ou plusieurs *M-Traces* (premières ou non). Les *M-traces* issues d'une transformation sont appelées *traces transformées* ou *calculées* qui sont également stockées dans la *Base de Traces*. La fonctionnalité *Visualisation* permet à un *utilisateur* d'une part, de visualiser les *traces collectées* pour l'exploration et la compréhension de la trace en vue de préparer leurs

enrichissements via les transformations, et de d'autre part, permet la documentation des *traces transformées* pour expliquer l'activité observée.

La notion d'Ingénierie des traces d'interaction comme inscription de connaissances est basée sur un processus de modélisation dynamique. Une Trace Modélisée est le couple *Modèle de Trace* et *Trace*, elle est notée *M-Trace* et représente une reformulation d'une trace première conforme à l'interprétation d'un analyste des interactions entre un utilisateur avec un système observé.

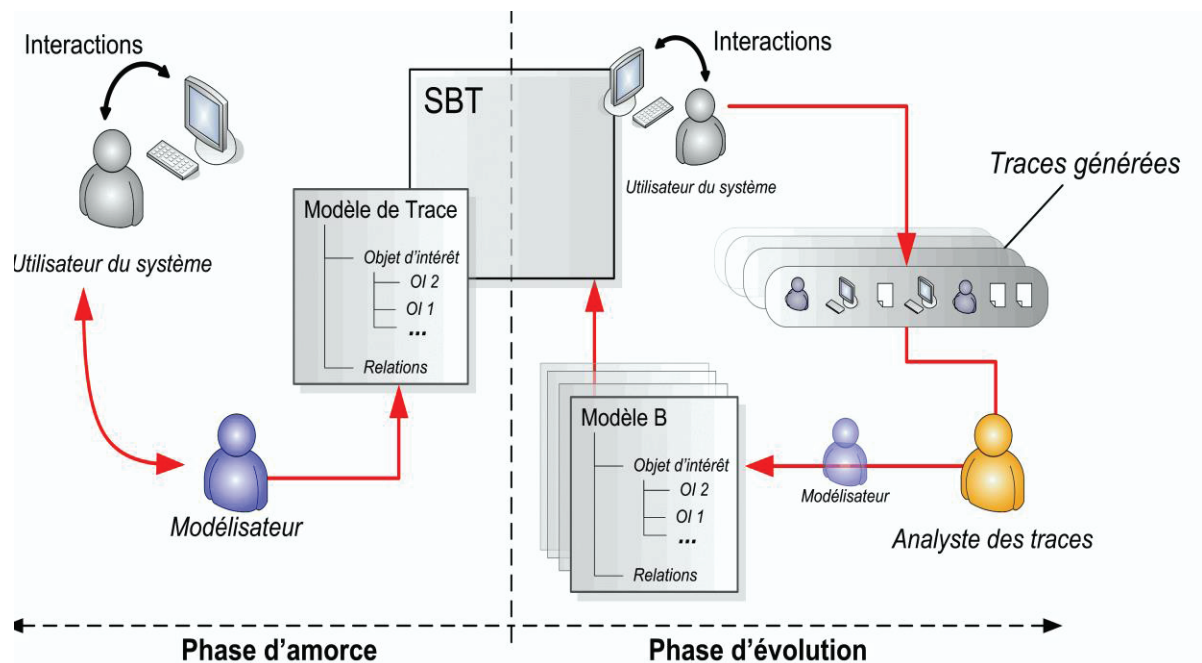


Figure I.22 : Démarche de modélisation de trace, extrait de [96].

L'approche des *M-Traces* est illustrée ci-dessus, elle est constituée de deux phases : la *phase d'amorçage* et la *phase d'évolution*.

La *phase d'amorçage* permet à l'analyste de définir la structure du *Modèle de trace première*, c'est-à-dire les objets d'intérêt et les relations qui existent entre eux. Les traces collectées seront des instances de ce *Modèle de trace première*. La phase d'amorçage a pour objectif de définir la structure des traces premières d'interactions qui seront créées dans le *SBT* à partir des interactions observées. Dans la seconde phase qui est celle de l'*évolution*, l'objectif est de permettre des ajustements ou des transformations successives pertinentes des traces pour inférer des nouvelles connaissances à partir de la première trace ou des traces transformées. Chaque transformation correspond à un niveau d'abstraction dans lequel il est possible d'avoir les informations nécessaires pour situer ce dernier par rapport à la *trace collectée*. Il est à noter qu'il y a une évolution conjointe de l'activité et du *Modèle de Trace*, c'est-à-dire

que chaque trace générée possède son modèle de trace qui définit ses objets d'intérêts et les relations qui peuvent exister entre ces objets. Dans cette approche de modélisation des traces, la dimension d'ingénierie de l'expérience tracée se manifeste par le dynamisme de l'évolution des *M-Traces* dans le temps, dans le sens où la connaissance se construit dynamiquement. Cette dynamique est explicitement gardée comme explicative de telle ou telle interprétation. En pratique, l'évolution des *M-Traces* se fait via des *transformations* à partir d'une trace collectée ou transformée, dans le but d'ajouter des nouvelles informations ou connaissances pour reformuler *la M-trace* en fonction de ce que l'on sait déjà ou de ce que l'on a appris dans les étapes de transformation précédentes. Chaque trace transformée est reliée au graphe de transformations qui y a mené depuis les traces premières, fournissant un outil puissant d'exploitation de la genèse de l'analyste. Les connaissances manipulées réfèrent à des objets d'intérêt situés dans le temps de l'activité analysée, ce sont des signatures ou des motifs remarquables de l'activité elle-même ou des sous-tâches de cette dernière.

L'approche des Traces Modélisées (*M-Trace*) développée par l'équipe Silex/TWEAK du LIRIS est utilisée aujourd'hui dans des domaines très variés comme l'enseignement, le web, le transport et d'une manière générale pour toute activité observable par une instrumentation adéquate. Dans l'enseignement, les traces sont construites pour analyser les interactions des apprenants avec leur environnement éducatif en ligne dans le but d'améliorer ses conditions d'apprentissage et son efficacité. Cela permet de tracer les interactions des apprenants pendant l'utilisation du système éducatif observé. Ces traces d'interactions sont très utiles aussi bien pour l'apprenant que pour l'enseignant. Par exemple, elles permettent à l'apprenant d'accéder à ses actions antérieures pour identifier les parties qui lui sont difficiles et de les travailler pour s'améliorer. Mais aussi, les traces d'interactions des apprenants permettent à l'enseignant de personnaliser les sessions de révision des examens en fonction des difficultés des apprenants. Grâce à ces traces, il est possible de fournir une assistance personnalisée aux apprenants.

Dans le transport, l'approche des M-Traces est utilisée pour analyser l'activité de conduite [4]–[6]. La notion de traces d'interactions s'adapte parfaitement à l'analyse de l'activité de conduite, le but de l'analyse étant d'analyser les interactions entre le conducteur et son environnement de conduite. Ces interactions sont enregistrées à l'aide de capteurs lors de l'activité de conduite. Les données sont collectées grâce à des capteurs installés dans le véhicule (Bus CAN, GPS, Radar). Les paramètres enregistrés sont par exemple la vitesse, l'accélération, les coordonnées GPS, les enfoncements des pédales de frein ou d'embrayage. Les véhicules sont également équipés de caméras pour filmer les actions du conducteur comme les enfoncements de pédales, les regards sur les rétroviseurs, l'environnement de

conduite (vue devant, sur les côtés et arrière). Ces véhicules équipés permettent de collecter une grande quantité de données de l'activité de conduite en situation réelle avec plusieurs types de données différentes (numériques, images). Ces données collectées peuvent être exploitées pour analyser l'activité de conduite dans le but de la décrire, la comprendre et l'expliquer avec l'approche des *M-Traces*. C'est ainsi que Georgeon et al., dans [4] se sont appuyés sur la notion de trace d'interaction numérique pour proposer un outil de modélisation d'activité à partir de l'analyse de données qualitatives séquentielles. **ABSTRACT** est un outil de gestion et d'analyse de traces créé par Georgeon et al., [4]. Cet outil est basé sur l'approche des *M-Traces* pour développer une ingénierie des connaissances tracées [96]. L'objectif de cet outil est d'assister l'analyste dans la recherche et la découverte d'une « *Signature d'une tâche* » correspondant au *modèle de l'activité* selon son point de vue. Ce modèle tient compte des besoins d'analyse et des observations des traces d'interactions de l'activité. **ABSTRACT** est un outil qui supporte la représentation symbolique de l'activité grâce à l'approche des *M-Traces*, facilitant la visualisation et l'interprétation pour une meilleure compréhension de l'activité. Cette représentation symbolique de l'activité de conduite permet de réduire sa complexité en représentant l'activité avec les aspects pertinents (objets d'intérêts) qui la caractérisent. Chaque aspect de l'activité est associé à un symbole ou icône pour faciliter la visualisation.

Cet outil peut être utilisé par des ergonomes qui cherchent à concevoir de meilleures interfaces utilisateur, des analystes cherchant à prédire le comportement des participants dans des conditions spécifiques, des formateurs qui cherchent à améliorer les techniques de formation ou même des utilisateurs cherchant à améliorer leur compréhension de leurs actions. Dans chaque cas, les modèles d'activité constituent des micro-hypothèses proposées par les analystes pour décrire, expliquer et essayer de prédire comment un sujet effectue l'activité. Dans [4], cet outil est utilisé dans l'analyse de l'activité de conduite pour modéliser la situation de conduite « *changement de voie sur autoroute* » à partir de l'analyse des traces d'interaction de l'activité de conduite en situation de conduite réelle. Les modèles de deux types de « *changement de voie sur autoroute* » ont été identifiés en fonction de la présence ou non d'un véhicule sur la voie de gauche du conducteur. La présence d'un véhicule sur la voie de gauche induit par exemple une forte accélération de la part du conducteur. Les deux types de « *changement de voie sur autoroute* » ont été décrits de façon symbolique par les aspects pertinents de cette situation de conduite tels que le regard sur le rétroviseur gauche du conducteur avant d'engager le changement de voie, le coup d'accélération, la position du véhicule sur la voie. Le travail de Georgeon et al., [4] a montré la possibilité d'utiliser les *M-Traces* dans l'analyse de l'activité de conduite comme support d'assistance pour la

construction d'un modèle générique de situations de conduite. La thèse de Benoît Mathern [6], a permis d'aller plus loin dans le processus d'analyse de l'activité de conduite en offrant un support pour synthétiser les modèles des connaissances découvertes. Ce travail permet à l'analyste de synthétiser un ensemble de modèles d'une situation de l'activité de conduite identifié avec l'outil **ABSTRACT** en un modèle unique grâce aux automates [6].

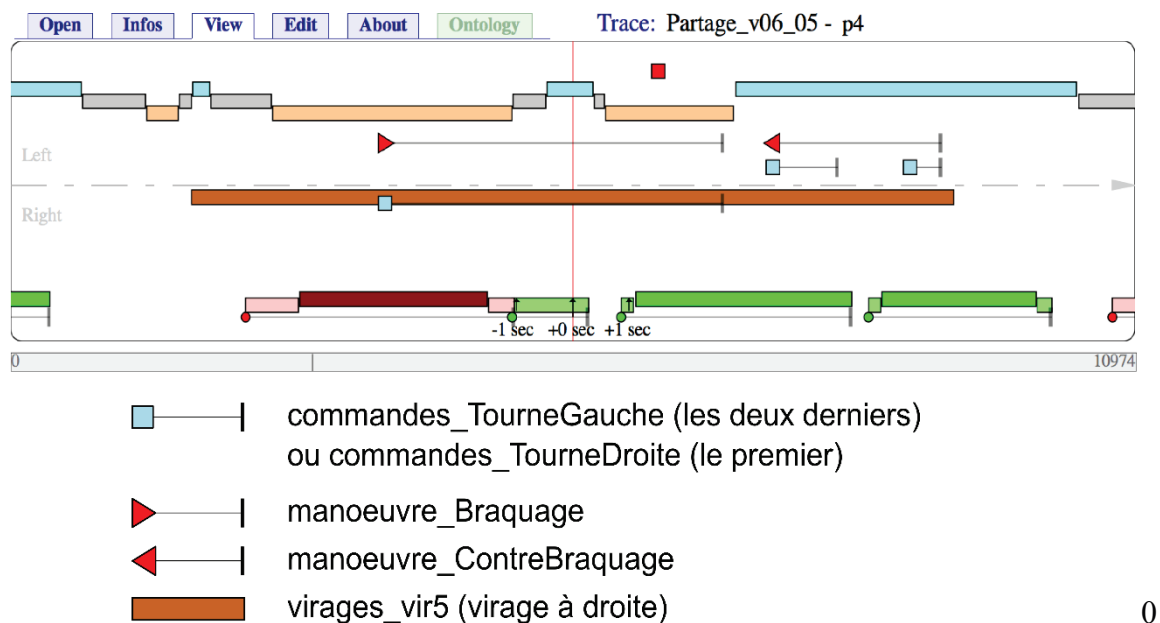


Figure I.23 : Illustration de visualisation de l'activité de conduite avec ABSTRACT, extrait de [6]

Dans cette thèse, nous utilisons la notion de Traces Modélisées (**M-Traces**) pour analyser l'activité de conduite dans le but *d'expliquer les comportements de conduite responsable d'une forte consommation en carburant en tenant compte du contexte de conduite.*

L'approche des **M-Traces** a été choisie car elle propose plusieurs avantages pour l'analyse de l'activité de conduite. Le premier avantage de l'utilisation des **M-Traces** est la place (rôle) attribuée à l'analyste dans le processus d'analyse. L'analyste est au centre du processus, ce qui lui permet de définir et de construire le *Modèle de Trace* selon son point de vue. En effet, le *Modèle de Trace* associé à l'activité par l'analyste est construit en fonction des besoins d'analyse, des connaissances de l'analyste sur l'activité et des contraintes liées à l'activité. Ceci permet à l'analyste d'être le guide de son analyse, de tenir compte des spécificités et des contraintes liées à l'activité telles que la **complexité de l'activité**, la prise en compte du **Contexte** pour une étude plus fine et ciblée. L'analyste a aussi la possibilité au cours du processus d'adapter son analyse en fonction des premiers résultats obtenus, des observations de l'activité et de l'évolution de ses connaissances au cours de l'analyse. Le deuxième avantage de l'approche M-Trace est la découverte de nouvelles connaissances, c'est-à-dire

l'Ingénierie des connaissances à partir des traces d'interactions. Elle fournit une assistance à la découverte des aspects inconnus ou des détails pertinents concernant l'activité grâce à la prise en compte de la dynamique de l'évolution des *M-Traces* dans le temps. Cette dynamique des *M-Traces* est *naturellement* représentée avec le Système à Base de Trace (SBT) qui permet la manipulation des *M-Traces* avec les fonctionnalités de *Collecte*, de *Transformation* et de *Visualisation*. Le troisième avantage des *M-Traces* est la possibilité de réduire la complexité de l'activité grâce à la représentation symbolique. La représentation symbolique de l'activité de conduite permet de réduire la complexité de l'activité en la modélisant en plusieurs sous-parties (ou sous tâches) pour une meilleure compréhension, exploration et visualisation de l'activité.

Par exemple, dans cette thèse, l'analyse de l'activité de conduite sera répartie en plusieurs situations de conduite comme des situations d'aller tout droit, de virage, ou de changement de voie qui seront ensuite analysées. Une fois ces situations de conduite modélisées sous forme de *M-Traces*, elles seront catégorisées en termes de consommation en carburant. Ces *M-Traces* représentant les situations de conduite pourront être exploitées pour déterminer les paramètres expliquant la consommation en carburant. Et grâce à la notion de *réflexivité* des *M-Traces* (utilisation des *Traces* antérieures d'un utilisateur pour l'assister ou faciliter ses interactions avec un système observé), il serait possible de faire une prédiction de la consommation en carburant en temps réel dans une situation de conduite précise. En effet, les *M-Traces* antérieures de conduite d'un conducteur d'une situation de conduite pourraient être utilisées pour prédire sa consommation en carburant dans des situations de conduite futures similaires. Enfin, les analystes de l'Ifsttar sont familiers avec les *M-Traces* qui ont déjà fait leurs preuves pour assister l'analyse de l'activité de conduite [4]–[6].

4 Discussion de l'état de l'art

La question de recherche abordée dans cette thèse est *la découverte de connaissances contextuelles* lors de l'analyse d'une activité dans le but d'identifier le contexte explicatif d'un critère observé comme la consommation en carburant de l'activité de conduite. D'après l'état de l'art, cette tâche est très complexe et demande la prise en compte de plusieurs aspects tels que la notion de contexte, la temporalité et le dynamisme de l'activité lors de son analyse. La notion de contexte permet de tenir compte des caractéristiques des différents éléments impliqués dans la réalisation de l'activité et de leur modélisation lors de l'analyse d'une activité. La modélisation du contexte nécessite la prise en compte de l'aspect dynamique et temporel de l'activité car le contexte ne se définit que pendant la réalisation de l'activité et est dynamique dans le temps. Ces aspects dynamiques et temporels du contexte représentent une

limite majeure pour la modélisation du contexte lors de l'analyse d'une activité. Pour proposer une solution à la question de recherche posée dans cette thèse, les traces modélisées M-Traces sont mobilisées car elles fournissent la possibilité de gérer le contexte de l'activité ainsi que les aspects temporels liés à la dynamique de cette activité. Les avantages des M-Traces tels que la réflexivité, la place de l'analyste et la prise en compte de l'évolution de l'activité dans le temps sont exploités dans cette thèse pour la mise en place de la méthode de découverte de connaissances contextuelles. Grâce aux M-Traces, les éléments impliqués dans la réalisation d'une activité peuvent être caractérisés de manière temporelle en tenant compte de leur évolution dans le temps. Ces éléments sont définis grâce au modèle des composants contexte de Bazire et al., [41], pour identifier et définir les paramètres caractérisant les composants contexte d'une activité. Les M-Traces permettent l'observation de l'activité et son évolution dans le temps, à travers les transformations permettant l'enrichissement de l'activité sur la base des connaissances de l'analyste et de son observation de l'activité. La méthode proposée dans cette thèse mobilise les techniques d'ingénierie de découverte de connaissances (apprentissage automatique) pour l'identification initiale des connaissances contextuelles candidates. Cette méthode de découverte de connaissances contextuelles repose sur une modélisation du contexte en reprenant les résultats des modèles génériques du *contexte* et sur l'accompagnement de l'analyste à partir de *traces modélisées (M-Traces)*.

Nous présentons dans Tableau I.1 un récapitulatif des atouts et limites de certains travaux cités dans l'état de l'art de cette thèse :

Travaux	Contexte			M-Traces		Consommation en carburant	
	Dey et al., [35]	Bazire et al., [41]	Ngo et al., [73]	Gorgeon et al., [4]	Mathern et al., [5], [6]	Ericsson [16]	Brundell-Freij et al., [97]
Questions de recherches							
L'activité de conduite est-elle analysée dans ce travail ?	Non	Non	Non	Oui	Oui	Oui	Oui
Est-ce que les méthodes d'analyse de l'activité proposent un cycle itératif de construction des connaissances ?	Non	Non	Non	Oui	Oui	Non	Non
Une prise en compte de la temporalité de l'activité analysée ?	Non	Non	Non	Oui	Oui	Non	Non
Y a-t-il une prise en compte du contexte lors de l'analyse ?	Oui	Oui	Oui	Non	Non	Non	Non
Est-ce qu'il y a la modélisation du contexte lors de l'analyse de l'activité ?	Non	Non	Oui	Non	Non	Non	Non
Est-ce que le contexte est explicitement considéré comme un type de connaissance spécifique ?	Non	Non	Oui	Non	Non	Non	Non
Est-ce que la M-Trace est utilisée lors de l'analyse de l'activité ?	Non	Non	Non	Oui	Oui	Non	Non
Est-ce que l'approche SBT est utilisée pour l'exploration de variables contextuelles dans les travaux cités ?	Non	Non	Non	Non	Non	Non	Non

Tableau I.1 : Récapitulatif des limites et atouts de quelques travaux de recherche étudiés dans le chapitre I.

Ce tableau donne les atouts et limites des travaux actuels étudiés dans cette thèse pour répondre à la question de recherche abordée par l'Ifsttar et le LIRIS. La question de recherche est la *découverte de connaissances contextuelles*, la réponse proposée dans cette thèse est une méthode qui profite des atouts fournis par les méthodes existantes pour proposer de dépasser les limites constatées en définissant une méthode de découverte de connaissances contextuelles. Cette méthode s'appuie sur le modèle des composants contexte de Bazire et al., dans [41] et adopte la définition du contexte de Dey [35]. Elle utilise les travaux sur les M-Traces de l'équipe du LIRIS [4], [6], [93] en y incluant la prise en compte du contexte et de sa modélisation lors de l'analyse de l'activité de conduite pour identifier les facteurs influençant la consommation dont certains proviennent de la littérature (Ericsson [16], Brundell-Freij et al., [97]). Cette méthode est itérative et interactive, elle mobilise le processus de découverte et d'extraction de connaissances tel que décrit par Fayyad dans [7], [98], [99]. Elle permet les interactions avec l'analyste tout au long du processus de découverte des connaissances.

Chapitre II. Contribution : Méthode et Application informatique

L'objectif de recherche de l'Ifsttar est : « *De comprendre et expliquer la consommation en carburant de l'activité de conduite en fonction des comportements des conducteurs et de leur environnement* ». A partir de cet objectif de recherche de l'Ifsttar et de l'éclairage des travaux étudiés dans le chapitre état de l'art, notre question de recherche a été formalisée pour répondre à ce type de besoin très présent lors de l'analyse d'une activité en général. Ce besoin de recherche est celui de la modélisation et de l'identification du contexte lors de l'analyse d'une activité. La question de recherche de cette thèse est formulée de la manière suivante : « *Comment construire dynamiquement des connaissances sur les contextes explicatifs d'un comportement de type conduite automobile à partir de son observation* » puis de valider l'approche par son application à l'identification des contextes explicatifs de la consommation en carburant. Pour répondre à cette question de recherche en permettant d'atteindre l'objectif de l'Ifsttar, nous proposons une méthode de découverte de connaissances contextuelles à partir de l'observation de l'activité de conduite. Cette méthode permet de tenir compte du contexte lors de l'analyse d'une activité et de modéliser le contexte explicatif d'un phénomène ou d'un critère observé de cette activité. Elle est définie et présentée dans la partie suivante.

1 Méthode de découverte de connaissances contextuelles

L'objectif de cette méthode est d'identifier les facteurs contextuels pouvant expliquer les variations d'un critère observé sur des données d'une activité temporelle. Nous cherchons à construire une méthode de découverte de connaissance interactive et itérative basée sur l'observation et les connaissances du domaine de l'analyste. Elle doit être *interactive* car c'est l'analyste qui pilote chaque étape et doit être *itérative* car l'analyste peut revenir en arrière à n'importe quelle étape. Dans cette approche, l'analyste a un rôle très important car c'est lui qui guide l'analyse en fonction de ses connaissances du domaine et de l'activité. C'est lui qui propose les transformations à mettre en œuvre sur une M-Trace (source) pour « voir » l'activité selon « son » interprétation (M-Trace transformée). Les transformations sont conservées dans la base de traces modélisées. Les règles de transformations sont soit initiées à l'aide de méthodes d'analyse de données (fouille de données) soit mises en place à partir des observations et des connaissances de l'analyste de l'activité. La trace transformée obtenue peut alors être explicitée facilement et confrontée à d'autres interprétations également

obtenues par des transformations. L'analyste pourra donc créer toutes les transformations nécessaires pour modéliser le contexte et celles permettant de différencier correctement les séquences d'activités ayant des valeurs de critères différentes. Quand l'analyste sera satisfait, il pourra comparer les transformations pour savoir quels éléments de contexte ont été utilisés pour différencier les séquences.

La méthode de découverte de connaissances contextuelles permet de déterminer le *Contexte explicatif* d'un critère observé à partir de l'analyse tracée. Cette méthode, *interactive* et *itérative*, est composée de cinq étapes décrites dans la figure ci-dessous (Figure II.1).

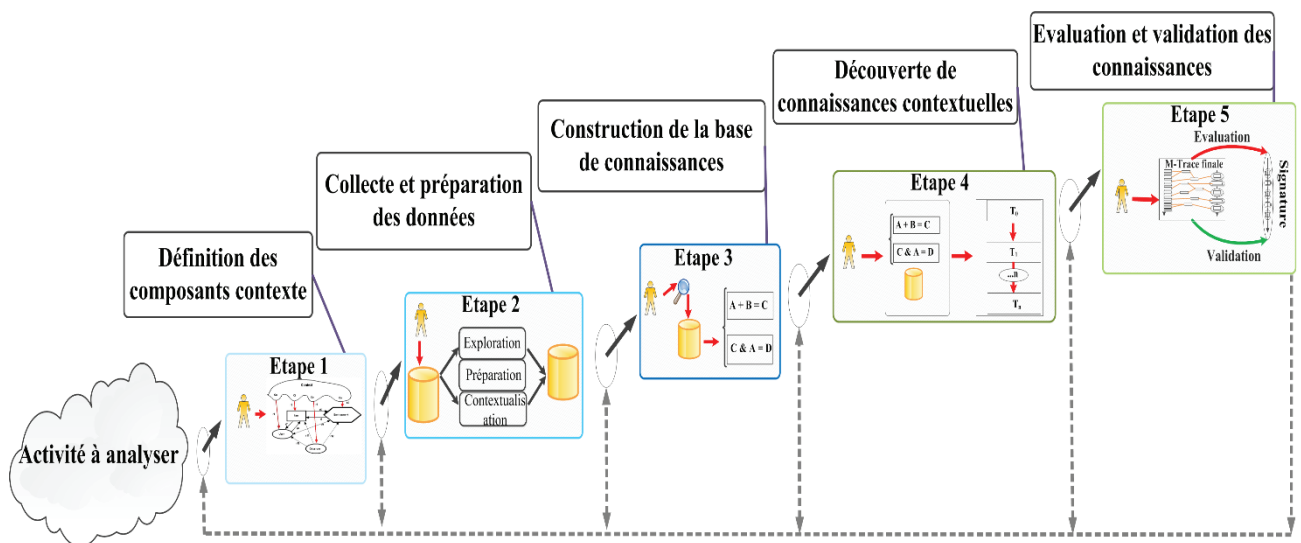


Figure II.1 : Schéma de la méthode de découverte de connaissances contextuelles

L'objectif de la méthode de découverte de connaissances contextuelles est de déterminer les éléments contextuels d'un critère observé d'une activité.

La première étape permet de définir les paramètres potentiels candidats au contexte explicatif du critère observé grâce au modèle des composants génériques du contexte de Bazire et al., [41].

La deuxième étape permet de faire la collecte et la préparation des données dans le but de préparer et de contextualiser la base de données qui sera utilisée pour la découverte de connaissances contextuelles.

La troisième étape a pour objectif de construire une base de connaissances qui sera utilisée dans l'étape suivante. Pour cela, l'activité est segmentée en séquences homogènes, chacune d'entre elles étiquetée en fonction des valeurs prises par le critère observé et classifiée par des paramètres connus non contextuels. La base de connaissance contient les règles de segmentation de l'activité générale et celles de classification du critère observé par les

paramètres non contextuels. Ces règles sont nécessaires à la mise en place des transformations préliminaires à la découverte de connaissances contextuelles de l'étape suivante.

La quatrième étape qui est la phase d'identification des paramètres contextuels du critère observé est basée sur l'observation et la classification de l'activité en fonction du critère observé. Les éléments mal classés avec les paramètres non contextuels sont ré-analysés avec des paramètres contextuels candidats pour identifier lesquels permettent de les reclasser correctement. Les paramètres contextuels candidats qui sont impliqués dans un meilleur reclassement prennent le statut de paramètres contextuels et construisent le *contexte explicatif* du critère observé.

La dernière étape qui consiste à évaluer et valider les paramètres contextuels explicatifs du critère observé a pour but d'utiliser les connaissances contextuelles découvertes sur de nouvelles données pour juger la robustesse des connaissances découvertes.

La méthode est totalement guidée par l'analyste du début à la fin de manière interactive et itérative. Cela lui permet d'itérer autant de fois le processus jusqu'à ce qu'il obtienne des résultats satisfaisants et d'interagir pendant l'analyse en introduisant ses connaissances dans le processus. En effet, il est très fréquent, au cours du processus de découverte, de revenir dans la phase de préparation des données pour créer de nouveaux indicateurs. De même, les phases d'analyse du contexte peuvent faire apparaître un certain manque d'informations qui peut nécessiter la collecte de nouvelles données. Ces différentes étapes de la méthode sont détaillées dans les sections suivantes.

1.1 Etape 1 : Définition des composants génériques du contexte

Cette étape nous permet à partir du modèle des composants contexte de Bazire et al., [41], d'identifier « toutes » les variables possibles « candidates » au contexte lors de l'analyse d'une activité en fonction des objectifs ou des besoins d'analyse.

1.1.1 Description du contexte de l'activité

Le modèle des composants contexte est constitué des contextes des entités qui interviennent lors de la réalisation d'une activité.

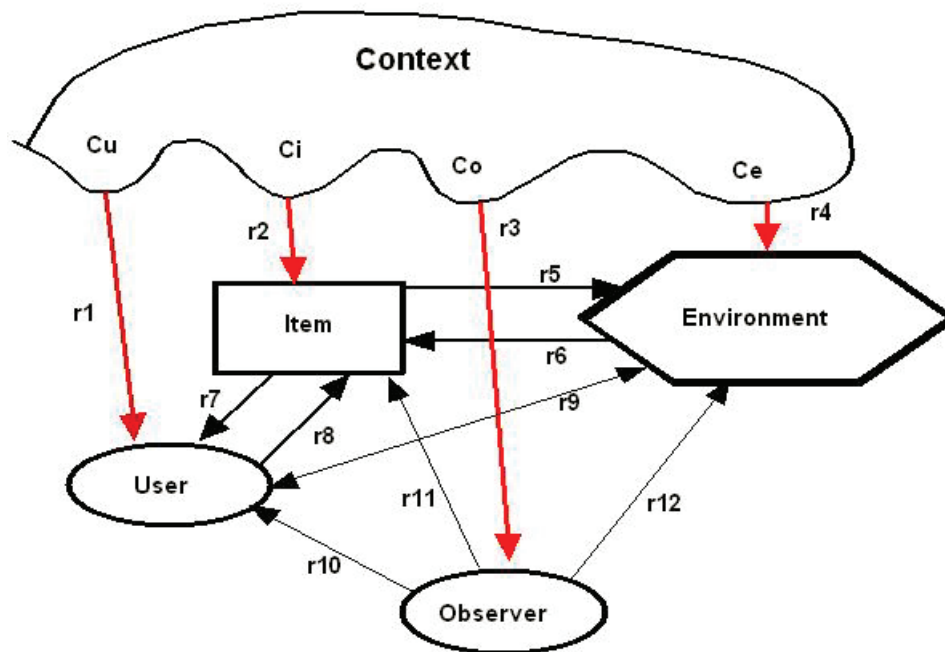


Figure II.2 : Modèle des composants contexte de Bazire et al., [41]

D'après ce modèle, le contexte d'une activité est défini par l'ensemble des contextes de chaque entité de cette activité qui sont : le contexte de l'utilisateur « *Cu* », le contexte de l'item (système, objet, application) « *Ci* », et le contexte environnemental « *Ce* ». Toutes les activités possèdent ces entités car une activité se traduit par l'interaction entre un utilisateur et un objet ou une application dans un environnement dans le but de réaliser une tâche. Le contexte de l'observateur « *Co* » permet de définir le contexte de l'analyse comme les objectifs et le périmètre de l'analyse d'une activité. Pour modéliser l'ensemble de ces contextes, il est nécessaire de définir les informations relatives aux entités impliquées dans cette activité. Elles peuvent être utilisées pour caractériser un segment d'activité particulier. Ceci nous renvoie à la définition du contexte de Dey [35], qui définit le contexte comme étant toute information pouvant être utilisée pour caractériser une entité. L'entité pouvant être une personne, un objet, une place. A ce niveau de l'analyse il s'agit d'informations *potentielles* de contexte de l'activité analysée. Pour identifier celles qui sont réellement explicatives, nous utiliserons donc les techniques de l'ingénierie des connaissances tracées dans l'étape 4.

1.1.2 Modélisation du contexte de l'activité

Le modèle de contexte utilisé dans cette thèse est le modèle des composants contexte de Bazire et al., dans [41] qui définissent le contexte comme l'ensemble des contextes des entités utilisateur, plate-forme, environnement et observateur. Le contexte d'une activité ou d'une situation de cette activité est défini par les informations contextuelles décrivant les différentes

entités impliquées dans l'activité mais aussi par les informations contextuelles qui caractérisent l'observation qui est faite de cette activité. Le diagramme de classe de la modélisation du contexte d'une activité avec le modèle des composants contexte est donné dans la Figure II.3.

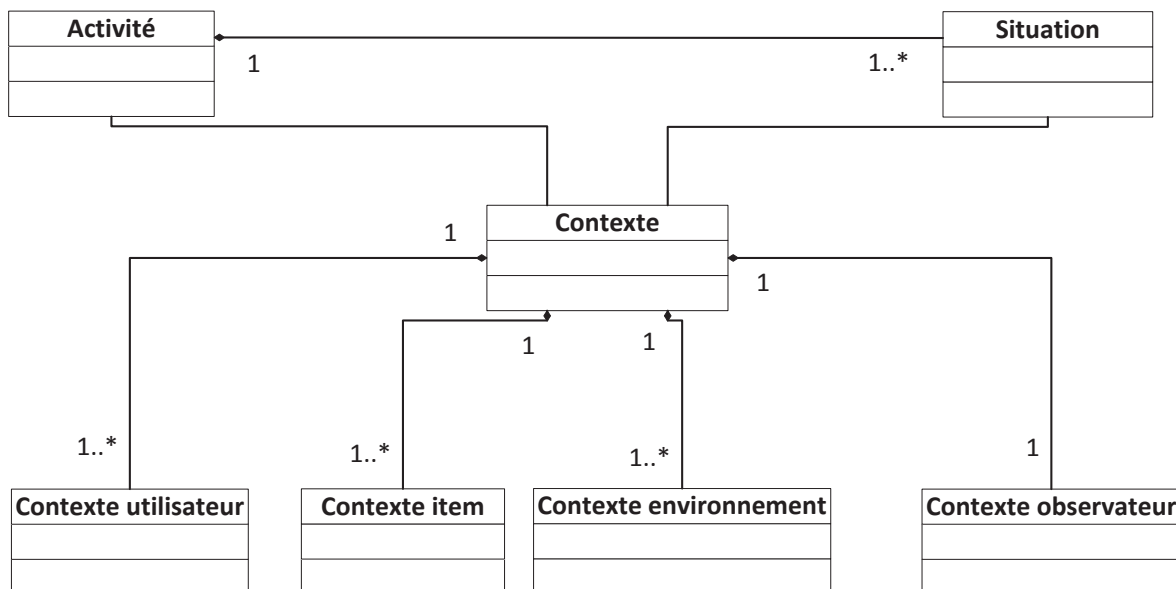


Figure II.3 : Diagramme de classe de la modélisation du contexte d'une activité

Dans ce diagramme l'activité est composée de plusieurs situations. Pour décrire le contexte d'une activité ou d'une situation il est nécessaire de définir le contexte des entités de l'activité.

Le contexte de l'utilisateur est décrit par les informations contextuelles qui permettent de décrire ou de caractériser l'utilisateur en fonction de l'activité. Par exemple les informations contextuelles caractérisant un utilisateur qui effectue une tâche de conduite automobile sont différentes de celles d'une personne qui effectue une recherche d'information sur le web. Ces informations peuvent être personnelles comme l'âge, profession, etc., ou en lien avec l'activité effectuée comme l'expérience de conduite, style de conduite pour l'activité de conduite et pour l'application web la connaissance de l'application ou de l'outil informatique etc.

Le contexte de l'item ou de la plate-forme permet de décrire la plate-forme c'est-à-dire l'élément avec lequel l'utilisateur interagit pour réaliser l'activité en question. Pour l'activité de conduite la plate-forme est la voiture qui est décrit par la marque, le type de carburant,

l'état, le nombre d'année, etc., et pour l'application web cela peut être un ordinateur ou un téléphone portable.

Le contexte de l'environnement est décrit par les informations contextuelles permettant de caractériser l'environnement dans lequel l'utilisateur interagit avec la plate-forme. Il décrit la localisation, le temps, la météo, le type d'infrastructure, etc. pour l'activité de conduite et pour l'application web il peut être défini par le système d'exploitation, la version de l'application utilisée, etc.

Le contexte de l'observateur permet de définir les informations contextuelles relatives à l'analyse de l'activité, aux objectifs de l'analyse et des conditions d'analyse. Le contexte de l'observateur permet de décrire le paramètre à expliquer ou le critère observé de l'activité et les contraintes qui lui sont liés. Ces informations décrivant les besoins d'analyse ont une influence sur les autres informations contextuelles car elles sont déterminées en fonction des besoins d'analyse. Par exemple, les informations contextuelles décrivant les entités utilisateur, plate-forme et environnement ne sont pas les mêmes pour une étude qui a pour objectif de déterminer l'influence de la vieillesse sur la conduite automobile qu'une étude dont l'objectif est d'expliquer la consommation en carburant.

1.2 Etape 2 : Collecte et préparation des données

Cette étape a pour objectif d'utiliser les techniques classiques de prétraitements des données pour préparer la phase d'analyse des données. A partir des données brutes collectées lors de l'activité, deux opérations vont être nécessaires pour la préparation des données.

La première opération est la préparation des données qui consiste à nettoyer les données, à créer de nouveaux paramètres, à transformer les paramètres via des opérations de discrétisation, de catégorisation ou de reformulation. La deuxième opération est la contextualisation des données analysées. Cette opération consiste à créer et à renseigner l'ensemble des paramètres candidats au contexte définis dans l'étape 1. Pour mener à bien cette phase, l'analyste doit se familiariser avec les données en les explorant et en les visualisant avant les opérations de préparation.

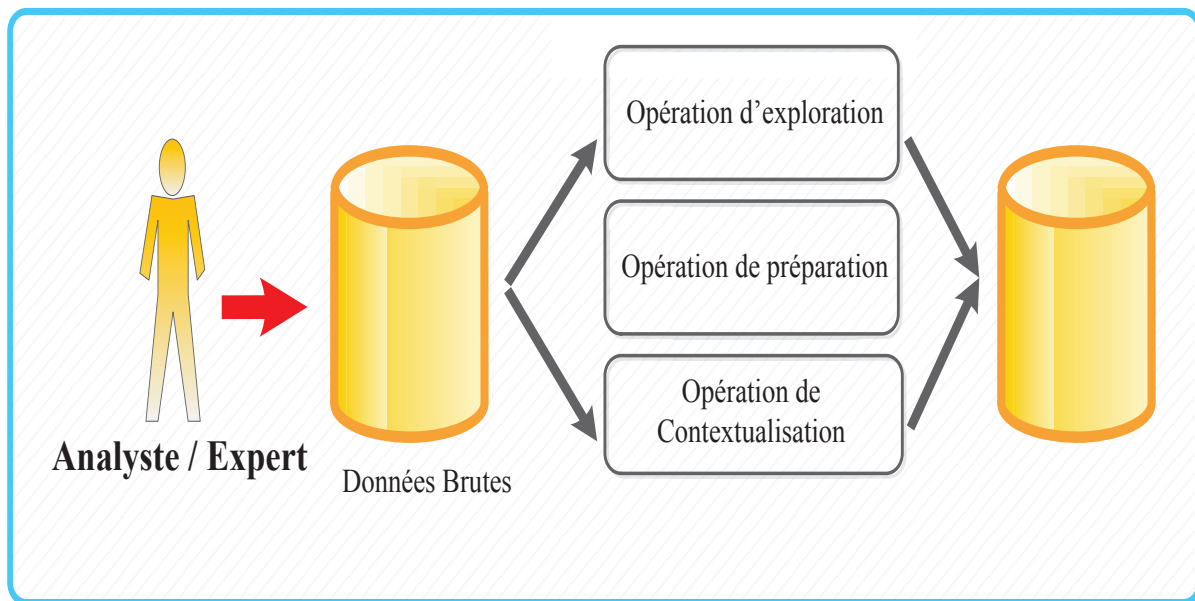


Figure II.4 : Collecte et prétraitement des données

1.2.1 Collecte et exploration des données

La collecte et l'exploration des données consistent à capturer et à étudier les données dans le but de les préparer pour leur analyse. La collecte se fait avec des capteurs pendant la réalisation d'une tâche ou par des caméras vidéo qui permettent de filmer la tâche effectuée. L'exploration des données consiste à se familiariser avec les données en les observant soit avec des techniques de la statistique descriptive (moyenne, écart type, maximum, minimum, etc.) ou par visualisation simultanée des données des capteurs enregistrées et des enregistrements vidéos.

1.2.2 Opération de préparation des données

L'opération de préparation des données est très importante dans l'analyse d'une activité, et elle nécessite une bonne connaissance des données pour les opérations de préparation des données suivantes : le **nettoyage des données**, la **création de nouveaux paramètres** et la **reformulation des paramètres** (discrétisation, renommage).

1.2.2.a Nettoyage des données

L'opération de nettoyage dans l'analyse d'une activité permet, de faire la gestion des valeurs manquantes des paramètres et, de vérifier la cohérence des données. Les données de l'activité étant collectées par des capteurs, il y a souvent des problèmes de mesure. Cette opération permet de gérer ces paramètres non renseignés pendant l'activité en les remplaçant, soit par la moyenne des valeurs sur l'ensemble de la séquence d'activité pour les paramètres numériques

quantitatifs, soit par la dernière valeur prise par le paramètre avant l'arrêt des enregistrements pour les paramètres qualitatifs. Elle permet également de gérer les valeurs extrêmes ou improbables enregistrées par les capteurs. Ces valeurs sont corrigées, soit par suppression, soit par remplacement par la moyenne, le maximum, le minimum sur le trajet ou par la dernière valeur prise par le paramètre. Grâce à cette étape, la cohérence des paramètres collectés est vérifiée. Ces opérations de nettoyage sont effectuées sur l'ensemble des paramètres enregistrés pendant l'activité.

1.2.2.b Création de nouveaux paramètres (paramètres calculés)

Cette étape permet de créer de nouveaux paramètres calculés à partir des paramètres collectés lors de l'expérimentation. Ce sont des paramètres déduits ou calculés à partir des paramètres collectés car ils ne sont pas directement disponibles après l'acquisition des données. Ce sont généralement des paramètres jugés pertinents par l'analyste pour l'explication du critère observé, d'où la nécessité de les créer si possible à partir des données disponibles.

1.2.2.c Reformulation des paramètres collectés

L'opération de reformulation permet de transformer un paramètre numérique (collecté ou calculé) en le catégorisant ou en le discrétisant. La discrétisation d'un paramètre nécessite un découpage en classes qui pose deux problèmes qui sont : le choix du nombre de classes et celui des bornes de classes. Pour cela, il existe plusieurs méthodes comme la méthode des quantiles, la méthode des amplitudes, la méthode des moyennes emboîtées et la méthode des grandes différences relatives. La méthode des quantiles est basée sur l'équipartition dans les différentes classes. La méthode d'amplitude garantit le critère d'égalité d'amplitude entre les classes, l'amplitude étant la différence entre la plus grande et la plus petite valeur du paramètre dans une classe. La méthode des moyennes emboîtées consiste à séparer la partition globale des valeurs du paramètre en N intervalles égaux (N doit être un nombre pair et correspond au nombre de classes). La valeur de séparation de chaque classe d'intervalle est la moyenne globale des valeurs sur cet intervalle de classe. La méthode des grandes différences relatives consiste à calculer les différences relatives successives entre une valeur et sa suivante, les valeurs ayant été préalablement triées par ordre croissant. On change de classe lorsque la différence relative est supérieure à un seuil arbitraire, classiquement 50 %.

1.2.2.d Création de paramètres agrégés ou indicateurs sur les séquences de conduite

Les paramètres enregistrés étant continus, la description d'une activité qui peut durer plusieurs heures correspond à un grand nombre de lignes de données, d'où la nécessité de

découper ces trajets en séquences temporelles qu'il faut caractériser. Pour cela, il faut mettre en place des *indicateurs* d'agrégation qui permettent de caractériser une séquence d'activité grâce aux fonctions d'agrégation sur les différents paramètres durant cette séquence de conduite. Pour calculer ces *indicateurs*, plusieurs fonctions d'agrégation ont été mises en place en fonction des besoins d'analyse de l'activité (moyenne, amplitude, maximum, ...). Elles sont définies dans le tableau suivant :

Descriptions fonctions d'agrégation	Notation
Calcule la durée de la séquence	IND_Durée
Calcule la somme des valeurs d'un paramètre numérique durant une séquence	IND_Sum
Calcule la moyenne d'un paramètre numérique durant une séquence	IND_Mean
Calcule la dérivée d'un paramètre numérique durant une séquence	IND_Derivée
Calcule l'écart type d'un paramètre numérique durant une séquence	IND_EcartType
Calcule l'amplitude du paramètre numérique, différence entre les valeurs maximale et minimale de ce paramètre durant une séquence	IND_Amplitude
Calcule la différence entre deux valeurs du même paramètre	IND_Diff
Calcule la valeur minimum prise par un paramètre numérique durant une séquence	IND_Min
Calcule la valeur maximum prise par un paramètre numérique durant une séquence	IND_Max
Calcule le pourcentage de temps d'utilisation des modalités d'un paramètre durant une séquence	IND_%
Détermine la modalité la plus utilisée d'un paramètre nominal durant une séquence	IND_PlusUse

Tableau II.1 : Liste des fonctions d'agrégation standard sur les séquences de conduite

1.2.3 Opération de contextualisation des données

L'opération de contextualisation des données utilise les opérations de préparation telles que la transformation, la création et le nettoyage des données mais uniquement sur des paramètres relatifs au contexte de l'activité de conduite. L'opération de contextualisation des données consiste à créer et à renseigner les paramètres candidats au **Contexte** de l'activité de conduite défini dans l'étape 1. Par exemple, les entités impliquées dans l'activité de conduite sont : le *conducteur*, le *véhicule*, et l'*environnement de conduite*. Dans cette étape, elles seront caractérisées par des paramètres qui sont considérés par l'analyste comme pertinents pour caractériser une situation de l'activité de conduite. Dans l'exemple d'un changement de voie effectué par un jeune conducteur sur autoroute avec une Clio, chacune des entités impliquées

dans la conduite doit être caractérisée, le conducteur par son âge (jeune) effectuant un changement de voie, le véhicule par la marque (Renault Clio) et l'environnement de conduite par le type de route (autoroute). En plus de ces entités, il faut aussi caractériser l'entité « *Observateur/Analyste* » qui consiste à définir les paramètres relatifs aux objectifs d'analyse et les contraintes qui lui sont liées par exemple, dans notre analyse le choix de s'intéresser qu'aux situations « *AllerToutDroit* ». Toujours avec l'exemple du jeune conducteur sur autoroute, il peut y avoir plusieurs objectifs d'analyse pour cette situation comme par exemple, une analyse qui répond au besoin de définir les comportements des jeunes conducteurs sur autoroute ou bien de déterminer la consommation en carburant de conduite des jeunes. Les paramètres définis pour cette entité correspondent aux critères observés répondant aux besoins d'analyse de l'activité, il est donc très important de bien définir ce critère car la définition des paramètres décrivant les autres entités est dépendante de ce dernier. L'objectif de cette étape est de définir, l'ensemble des paramètres qui, du point de vue de l'analyste, permettent de décrire ou de caractériser chaque entité impliquée dans l'activité et qui pourraient influencer les objectifs d'analyse donc le critère observé décrit dans l'entité « *Observateur / Analyste* ». Ces paramètres peuvent être statiques (constant dans le temps) ou dynamiques (évoluant au cours de l'activité).

1.2.3.a Les paramètres de l'entité « *Observateur /Analyste* »

Dans cette thèse l'entité « *Observateur /Analyste* » du modèle des composants contexte de Bazire et al. [41], permet de définir le critère observé qui correspond au paramètre à comprendre ou à expliquer par l'analyste. En général, une activité est analysée soit pour la comprendre dans sa globalité ou soit pour expliquer un phénomène ou un critère observé pendant l'activité. Une bonne définition des objectifs de l'analyse d'une activité est très importante car elle détermine l'intégralité du processus d'analyse. Cela reste vrai aussi pour notre méthode de découverte de connaissances contextuelles, c'est pourquoi l'entité « *Observateur /Analyste* » est décrite en premier lieu par le critère observé et les contraintes liées à l'analyse avant la définition des paramètres des autres entités de l'activité. Cela permet de bien identifier les paramètres candidats au contexte de l'activité analysée en fonction du critère observé. Par exemple pour déterminer l'influence du vieillissement sur la conduite automobile, les paramètres contextuels explicatifs de ce critère ne seront pas les mêmes que les paramètres contextuels explicatifs de la consommation en carburant. C'est pourquoi la définition des paramètres caractérisant les entités expliquées dans l'activité est dépendante du critère observé pendant la contextualisation des données. Il est donc très important de bien

définir les hypothèses de l'analyse avant de définir les paramètres contextuels des autres entités de l'activité.

1.2.3.b Les paramètres de l'entité « *Utilisateur* »

L'entité utilisateur est caractérisée par les informations qui du point de vue de l'analyste pourrait être utilisées pour décrire l'utilisateur lors de la réalisation de l'activité. L'entité utilisateur peut être caractérisée par des informations personnelles comme son âge, sa profession, sa santé, etc., et des informations en relations avec l'activité comme son expérience, sa manière utiliser la plate-forme ou outil utilisé pour réaliser l'activité, etc. Par exemple, pour un utilisateur de jeux vidéo, le joueur peut être décrit par son âge, son niveau, sa performance, sa connaissance de l'outil informatique (téléphone ou ordinateur sur lequel il joue), etc. Ces paramètres décrivant l'utilisateur de manière précise doivent être créés et renseignés pendant l'opération de contextualisation. Ils peuvent être utilisés pour expliquer une situation donnée de l'activité.

1.2.3.c Les paramètres de l'entité « *Item ou plate-forme* »

Les paramètres décrivant l'item ou la plate-forme utilisée doivent être également créés et renseignés car ces paramètres représentent le contexte de l'entité item. Ils représentent les caractéristiques de la plate-forme impliquée dans l'activité comme le type de matériel, la marque, sa performance ou puissance, etc. Par exemple pour le cas du jeu vidéo il s'agit de créer de renseigner les informations permettant de décrire le jeu vidéo comme le type de jeu, sa puissance, etc.

1.2.3.d Les paramètres de l'entité « *Environnement* »

L'entité environnement est caractérisée par les paramètres décrivant la location, les interactions avec l'extérieur, le temps, etc., ces paramètres permettent de décrire le contexte de l'entité environnement peuvent avoir une grande influence sur la réalisation d'une l'activité. Par exemple pour l'activité de conduite l'entité environnement est décrite par la météo, le trafic, le type d'infrastructure, la présence de feu tricolore. La contextualisation permet de créer et renseigner ces paramètres de l'entité environnement qui peuvent être utilisés pour caractériser la situation d'une activité.

1.3 Etape 3 : Construction de la base de connaissances (les modèles des M-Traces)

Cette étape a pour objectif de construire une base de connaissances qui sera utilisée dans l'étape suivante.

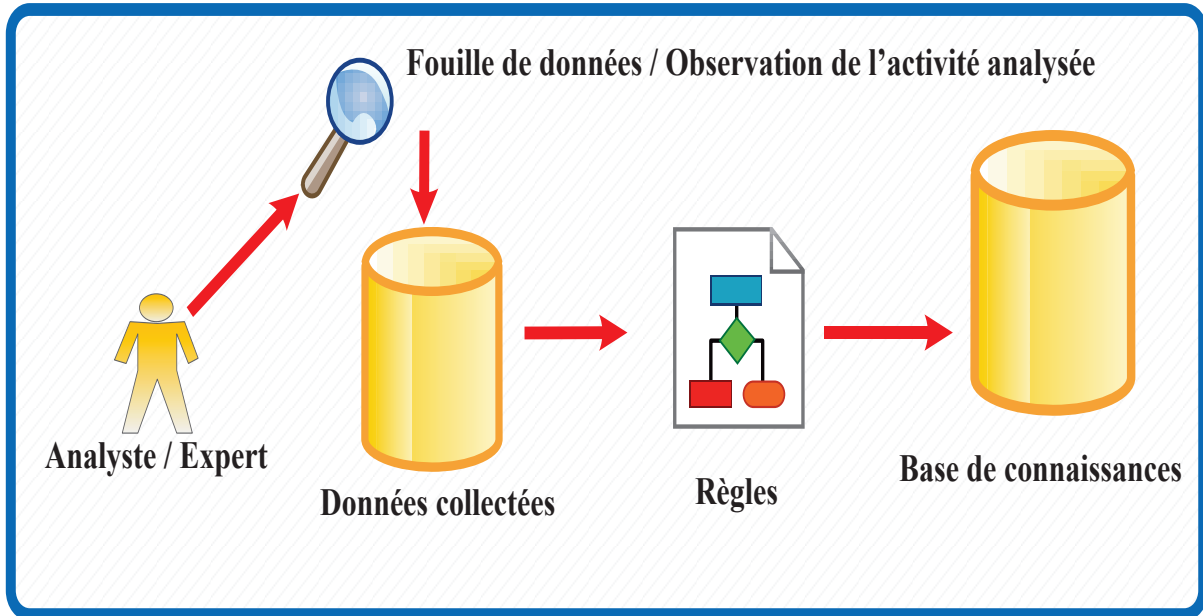


Figure II.5 : Construction de la base de connaissances

Afin de préparer l'étape d'identification des éléments contextuels explicatifs du critère observé, une base de connaissance est créée et sera exploitée pour la mise en place des transformations nécessaires à la découverte de connaissances contextuelles. Dans cette étape, les méthodes classiques de fouilles de données sont mobilisées comme outils d'analyse dans le processus de découverte interactive de connaissances pour faciliter l'observation de l'activité. Dans notre cas, nous avons utilisé la méthode de classification supervisée « Arbre de décision » avec l'outil d'analyse de données SPSS pour déterminer les règles de segmentation et de catégorisation des paramètres observés de l'activité.

Dans le cas de cette analyse, deux opérations ont été effectuées pour préparer la phase d'assistance à la découverte de connaissances à partir des M-Traces. La première opération sur les trajets de la base de données prétraitée est la segmentation de l'activité en séquences, et la deuxième opération est de catégoriser et de classifier ces situations en fonction du critère observé. La base de connaissance sera donc constituée de l'ensemble de règles permettant la segmentation et la catégorisation de ces séquences. Ces opérations sont détaillées dans les sections suivantes :

1.3.1 Elaboration des règles de segmentation de l'activité en séquences

Le contexte étant spécifique à la tâche en cours (sous-partie de l'activité globale), il est important de travailler sur des séquences homogènes de l'activité pour identifier quelles sont les entités impliquées dans l'activité dont les variations peuvent expliquer celle du critère étudié.

La segmentation consiste à découper l'activité analysée en sous-parties homogènes qui ont les mêmes caractéristiques. La segmentation peut être faite manuellement ou automatiquement en utilisation des méthodes de classification automatique. Ces méthodes ont pour but de créer en fonction d'une variable de classe des groupes d'individus de manière à ce que les individus dans un même groupe se ressemblent, et les individus dans des groupes différents soient dissemblables. Pour cela, des paramètres indépendants seront utilisés pour identifier les individus qui se ressemblent en fonction de la variable de classe qui est la variable dépendante. Par exemple, pour segmenter l'activité de conduite en situation de conduite, la variable dépendante est *situation_de_conduite* (aller tout droit, virage à gauche, virage à droite ou changement de voie) et les paramètres indépendants sont par exemple, *angle_du_volant*, *vitesse* ou *cap_avec_le_nord*. Ces paramètres indépendants sont utilisés pour prédire la situation de conduite. La segmentation de l'activité de conduite permet de simplifier sa complexité lors de l'analyse car elle est représentée comme une succession de séquences qui seront ensuite catégorisées en fonction du critère étudié.

Pour pouvoir identifier les paramètres contextuels explicatifs d'une situation il est important de travailler sur des séquences qui ont les même caractéristique c'est-à-dire les séquences qui ont le même type ou les séquences homogènes. Nous allons donc filtrer les séquences pour ne garder que celles d'un seul type.

1.3.2 Elaboration des règles de catégorisation et de classification des séquences en fonction du critère observé

Après avoir segmenté l'activité en plusieurs sous-parties (séquences), cette phase a pour objectif de faire une catégorisation des séquences en fonction du critère observé puis une classification de ces séquences pour supprimer l'influence des paramètres non contextuels. Chaque segment de l'activité est décrit par un ensemble de paramètres tels que les paramètres contextuels, les paramètres non contextuels et le critère observé. Les paramètres contextuels sont des paramètres qui peuvent être utilisés pour caractériser une entité contexte. Les

paramètres non contextuels permettent de décrire l'activité elle-même comme par exemple, pour l'activité de conduite, la vitesse de conduite, l'accélération, l'angle du volant. Et le critère observé est le paramètre que l'analyse doit expliquer. Les séquences possèdent une valeur du critère observé.

La catégorisation du critère observé consiste à créer des groupes suivant la valeur du critère observé. Par exemple, la catégorisation de la consommation en carburant consiste à construire 3 groupes (faible, moyenne et forte) en fonction de la consommation en carburant des séquences de conduite.

Les variations du critère observé ne sont pas dues uniquement aux variations des paramètres contextuels. En effet, certains phénomènes physiques peuvent les expliquer. Par exemple, la consommation de carburant est très dépendante du rapport de boîte de vitesse sélectionné. Pour mieux identifier les paramètres contextuels influant le critère, il est important de supprimer l'influence de ces paramètres physiques. Dans un premier temps, notre objectif est donc à partir des séquences homogènes de l'activité de les classer en fonction du critère observé en n'utilisant que les paramètres non contextuels. Pour cela, des règles de classifications sont créées pour que les séquences qui ont la même catégorie de critère observé soit dans le même groupe. Ces règles de classification seront stockées dans la base de connaissance pour être utilisées dans l'étape suivante.

1.3.3 Formalisation des règles dans la base de connaissances

La base de connaissances créée est composée de règles de reformulations pour décrire une activité ou une situation de l'activité comme une succession d'événements pertinents. Cette étape peut être très rapide si l'analyste maîtrise suffisamment le domaine pour créer directement ces règles lors de l'analyse tracée de l'activité. L'élaboration même de ces reformulations est très utile à l'analyste pour préciser et formaliser ses connaissances. La base de connaissance est constituée de l'ensemble des règles de segmentation et de classification de l'activité. Les règles de segmentation ont pour but de segmenter l'activité en plusieurs segments. Ces segments sont catégorisés en fonction du critère observé. Les règles de classification permettent de classer les segments homogènes en fonction du critère observé en créant des groupes segments qui ont le même critère observé. Les règles de segmentation et de classification sont représentées sous de la forme : Si *Condition* Alors *Action*.

Par exemple pour la segmentation la forme de la règle pourra être :

R= Si *Condition1* = 'vrai' ET *condition2* = 'Vrai' ET Alors *Segment* = 'typeN'

Et pour la classification la règle de classification pourra être :

$R = \text{Si } \textit{Condition1} = \textit{'vrai'} \textit{ ET } \textit{condition2} = \textit{'Vrai'} \textit{ ET } \dots \textit{ Alors Critère} = \textit{'classeN'}$

Par exemple, dans cette thèse, cette étape sera utilisée (dans la partie Chapitre III.2.3) pour identifier les règles de segmentation de l'activité de conduite en fonction des situations homogènes de conduite comme aller droit tout (AllerToutDroit), virage à droite (TournerADroite) ou changement de voie. En fonction des paramètres dynamiques du véhicule, une modélisation de ces situations de conduite est alors effectuée. En pratique, des paramètres comme l'angle du volant ou la vitesse sont utilisés pour prédire les situations de conduite et des paramètres non contextuels comme la vitesse moyenne, le rapport de boîte de vitesse sont utilisés pour prédire la consommation. Les règles de prédiction de ces situations de conduite sont validées par observation de l'analyste avant d'être stockées dans la base de connaissances. Ces règles seront utilisées dans l'étape suivante pour créer les situations de conduite homogènes.

A la fin de cette étape, une base de connaissances initiale exploitable pour l'étape de la découverte de connaissances contextuelles a été créée. Elle a permis de mettre en place :

- Les règles de segmentation automatique de l'activité en plusieurs segments pour lesquelles on a affecté la valeur du critère à observer.
- Les règles de classification des séquences en fonction du critère observé avec des paramètres non contextuels.

1.4 Etape 4 : Processus d'assistance à la découverte de connaissances contextuelles

Le processus d'assistance à la découverte du contexte d'une activité est effectué à partir de l'analyse des traces construites à l'occasion de l'observation de l'activité. *Tracer une activité* consiste à décrire cette activité avec les aspects pertinents (*obsels*) à partir d'observations et des connaissances de l'analyse. Les observations collectées constituent les M-Traces collectées (M-Traces premières) et les M-Traces transformées constituent les différentes façons d'interpréter les choses selon telle ou telle expertise. La découverte est supervisée et les séquences de comportement à caractériser sont étiquetées par l'analyste qui y « reconnaît » quelque chose qui fait sens pour lui. Le processus d'assistance à la découverte de connaissances contextuelles est simple dans son principe. L'analyste propose une interprétation contextuelle qui d'après ses connaissances permettrait d'expliquer tel ou tel comportement correspondant à une étiquette qu'il connaît. Les transformations sont alors propagées sur l'ensemble des séquences étiquetées, et les séquences qui ne seraient pas

étiquetées correctement par la transformation précédente sont identifiées et utilisées pour raffiner la transformation. Ce processus est itéré jusqu'à obtenir un taux d'étiquetage correct satisfaisant pour les experts. L'ensemble des transformations est créé manuellement par l'analyste et appliqué automatiquement par l'outil ATraDeCC (voir partie II.2.2) sur l'ensemble des traces d'activité sélectionnées. La comparaison entre les éléments mobilisés dans la requête décrivant la première transformation imaginée par l'analyste et la dernière requête de transformation considérée comme satisfaisante, met en évidence les paramètres de contexte (les éléments non pris en compte initialement par l'analyste) candidats pour expliquer le comportement étiqueté. Cette méthode a donc pour but de définir interactivement le contexte explicatif d'un comportement lors de l'analyse d'une activité.

Cette figure décrit les différentes étapes du processus d'assistance à la découverte de connaissance contextuelles, Une fois l'activité décrite par les M-Traces, l'objectif est alors de découvrir les règles contextuelles afin d'explicitier le contexte de cette situation comportementale.

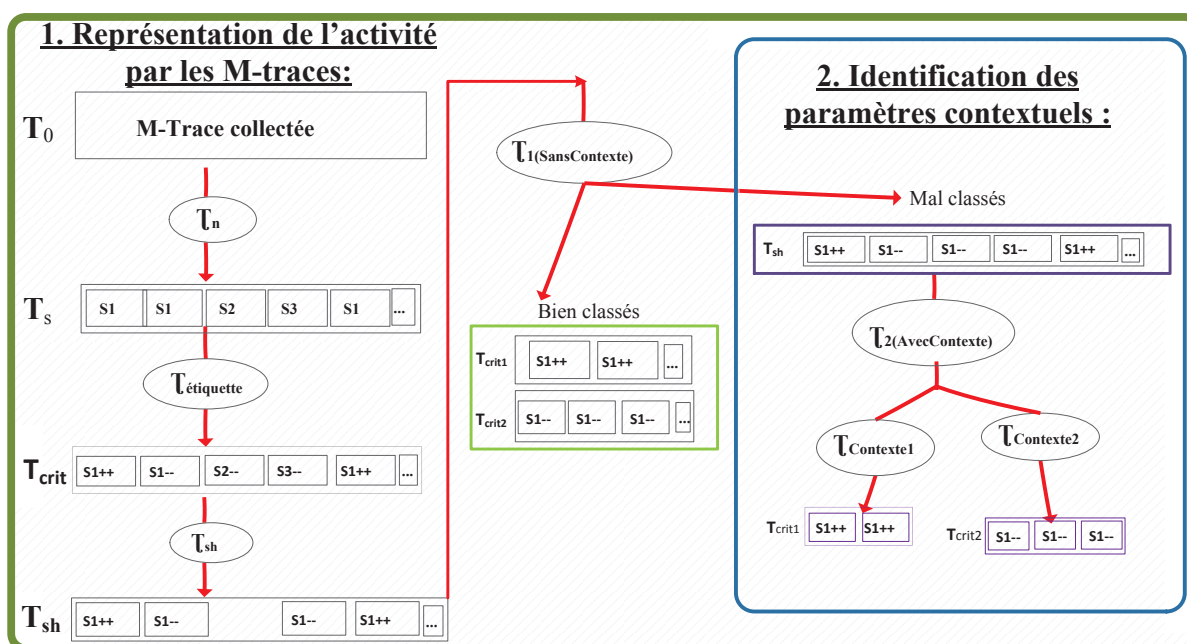


Figure II.6 : Processus d'assistance à la découverte de connaissances contextuelles

1.4.1 Représentation de l'activité par les M-Traces

La représentation de l'activité par les M-Trace a pour but de la formaliser avec les aspects pertinents qui la caractérise. Dans un premier temps, la M-Trace première collectée T_0 est créée à partir de la base de données prétraitées et contextualisées de l'étape 1.2. Puis, les

premières transformations \mathbb{T}_n vont permettre de créer les différentes séquences temporelles de l'activité correspondantes aux différentes tâches qui constituent l'activité analysée.

Puis la transformation $\mathbb{T}_{\text{étiquette}}$, permet d'étiqueter les séquences en fonction des valeurs du critère à expliquer (par exemple, les signes -- et ++ sur le schéma indiquent les différentes classes du critère observé). Ces transformations sont faites grâce aux règles de catégorisation élaborées lors de l'étape 1.3 et complétées par l'expertise de l'analyste.

Ensuite pour réaliser l'opération de filtrage de séquences homogènes, la transformation \mathbb{T}_{sh} permet de sélectionner dans ces séquences, celles qui correspondent à la tâche de l'activité que l'on souhaite analyser (ex. les séquences S1 dans la Figure II.6).

Plusieurs transformations de type ($\mathbb{T}_{1(\text{sansContexte})}$) de la Figure II.6 implémentent les règles de classification des séquences en fonction du critère observé avec des paramètres non contextuels puis différencient les séquences qui ont été bien classées par ces paramètres de celles qui ont été mal classées. Ces dernières sont celles qui seront utilisées pour la découverte des connaissances contextuelles.

1.4.2 Découverte des règles de classification pour identifier les paramètres contextuels explicatifs du critère observé

L'objectif de cette étape est de découvrir quels sont les paramètres contextuels qui pourraient expliquer des valeurs différentes du critère observé que celui qui a été prédit avec les paramètres non contextuels. En effet, les paramètres physiques expliquent une partie de cette variation mais pas la totalité. Nous faisons donc l'hypothèse que les paramètres contextuels peuvent expliquer le reste de ces variations. Le processus de découverte consiste maintenant de découvrir lesquels sont les plus influents. La visualisation des M-Traces permet à l'analyste d'identifier des paramètres supplémentaires qu'il faudrait créer (par exemple, pour spécifier les enchaînements temporels).

La découverte des connaissances contextuelles s'effectue alors en cherchant à identifier les règles de classification responsables des séquences d'activité mal classées par les paramètres de l'état des connaissances en cours, c'est-à-dire sans prise en compte d'autres paramètres explicatifs qui seraient alors considérés comme contextuels. Cette classification est recommencée itérativement avec les paramètres contextuels candidats identifiés grâce à la visualisation des M-Traces associées aux séquences d'activité concernées. Plus précisément, la découverte nécessite plusieurs tentatives de classifications des séquences encore mal classées : une classification avec les paramètres représentant une entité impliquée dans l'activité. Une entité dont la prise en compte démontre un meilleur taux de bonne

classification est gardée et les paramètres correspondants sont intégrés dans les classifications suivantes. Chacune des entités (via les paramètres le représentant) est ainsi introduite dans de nouvelles classifications intégrant la valeur de leurs attributs et le résultat de la classification précédente, pour observer si elle présente un meilleur taux de classement des séquences observées. Si c'est le cas, l'entité est alors gardée à son tour. Ceci est réitéré jusqu'à ce que toutes les entités aient été considérées. Ces règles sont formalisées par les transformations $T_{Contexte1}$, $T_{Contexte2}$ donnant les M-Traces transformées T_{crit1} et T_{crit2} de la figure Figure II.6.

Le contexte explicatif sera représenté par l'ensemble des règles de classification définies par les différentes entités. Par exemple le contexte explicatif C d'une situation pourra être défini par : $C = règles_{ce} Et règles_{ci} Et règles_{ci}$, les règles_{ce}, règles_{ci} et règles_{ci} étant les règles issues de la classification du critère observé des différentes entités impliquées dans l'activité (environnement, item et user).

1.5 Etape 5 : Evaluation et validation des connaissances découvertes

L'étape d'évaluation consiste à appliquer les connaissances découvertes sur de nouveaux jeux de données afin de déterminer la validité de ces connaissances. Cela consiste à voir si les connaissances découvertes sont robustes. Pour cela, nous les appliquons sur de nouveaux jeux de données pour vérifier que les taux de bonne classification sont semblables entre les différents jeux de données. Cette validation est faite à l'aide des règles considérées comme satisfaisantes, en les appliquant sur de nouveaux jeux de données du corpus d'observation complet. Si le taux de réussite des règles est supérieur à 5%, les variables retenues dans ces règles sont intégrées dans la représentation du contexte à prendre en compte.

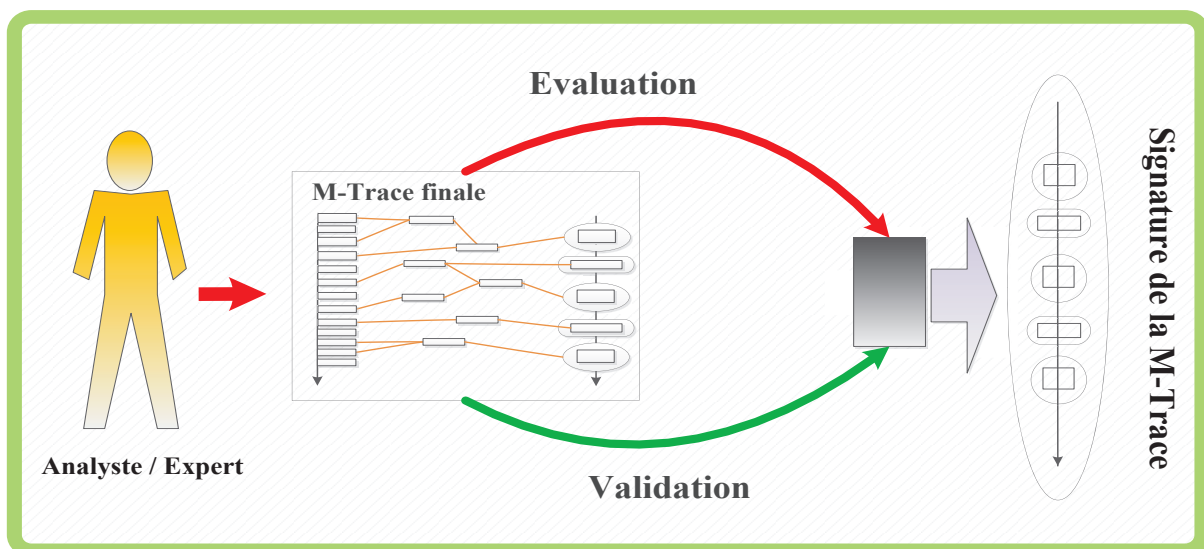


Figure II.7 : Evaluation et validation des connaissances découvertes

1.6 Itérations entre les différentes phases

Dans l'application de la méthode de découverte de connaissances contextuelles à l'analyse de l'activité, plusieurs types d'itération peuvent être nécessaires jusqu'à l'obtention des résultats satisfaisants pour l'analyste. Ces types d'itération permettent de passer d'une étape à une autre pour les besoins de l'analyse.

Le premier type d'itération est la « Révision d'observés bruts », il consiste à revenir de n'importe quelle étape à l'étape 1. Comme par exemple, passer de l'étape 2 à l'étape 1, ou des étapes 3, 4 ou 5 à l'étape 1. Elle permet d'ajouter de nouveaux paramètres contextuels, soit pour décrire d'autres caractéristiques de l'activité elle-même (par exemple, pour décrire son évolution temporelle) ou pour prendre en compte des observations de l'analyste pour introduire ses connaissances dans la boucle d'analyse.

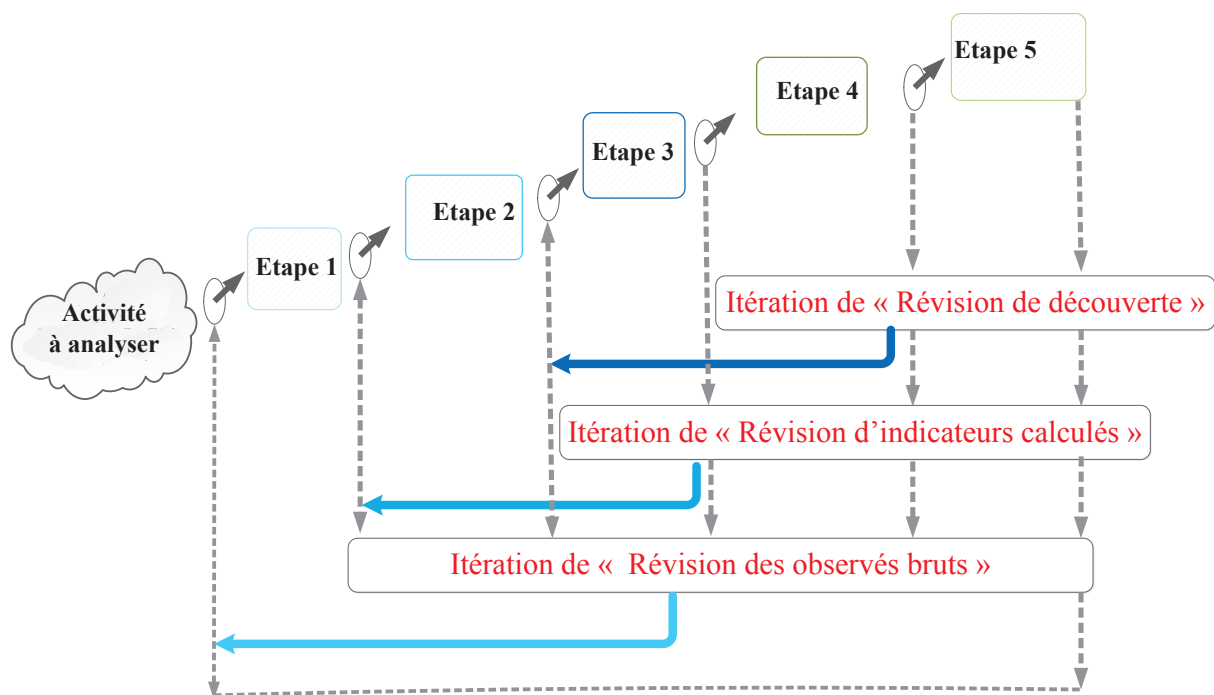


Figure II.8 : Utilisation de la méthode de découverte de connaissances contextuelles

Le deuxième type d'itération est la « Révisions d'indicateurs calculés », il consiste à passer de n'importe quelle étape à l'étape 2, dans le but de créer de nouveaux paramètres calculés ou des indicateurs pour l'enrichissement des connaissances produites sans forcément modifier les attributs des composants contexte. Comme exemple d'itération de ce type, nous avons la création des indicateurs décrivant les sous-parties de l'activité à l'aide des fonctions d'agrégation des paramètres continus telles que le pourcentage de temps d'utilisation d'un

rapport de boîte de vitesse ou la densité de feux de signalisation par km sur une séquence de conduite.

Enfin le dernier type d'itération est la « Révision de découverte », il consiste à passer des étapes 4 et 5 à l'étape 3. Les itérations de ce type ont permis de découvrir et valider les règles de classification avec les paramètres contextuels et non contextuels pour les besoins de l'analyse.

2 Application informatique : Outils de mise œuvre de la méthode

Pour la mise en œuvre de la méthode de découverte de connaissances contextuelles à partir des M-Traces plusieurs outils peuvent être utilisés, dont un a été développé au cours de la thèse pour les besoins spécifiques au domaine d'application pour l'analyse de l'activité de conduite. L'outil informatique développé dans cette thèse respecte le modèle des traces modélisée (M-Traces) développé par l'équipe Tweak du LIRIS. Cette équipe propose déjà plusieurs outils pour l'analyse des traces comme le KTBS (Kernel for Trace-Based Systems), SamoTrace, et Abstract [4], [6], [93]. Le choix a été fait de développer un nouvel outil informatique d'analyse de trace pour répondre aux besoins d'analyse spécifiques à l'activité de conduite. Par exemple, les fonctionnalités de calcul d'indicateurs sur les situations de conduite ont été implémentées spécialement pour répondre aux besoins de la thèse. Dans cette partie, l'architecture mise en place pour la découverte des connaissances contextuelles utilisée dans cette thèse est détaillée et l'outil informatique d'analyse trace est présenté.

2.1 Outils utilisés pour la mise en œuvre de la méthode de découverte de connaissance contextuelles

La méthode de découverte de connaissances contextuelles est une méthode basée sur l'observation d'une activité pour identifier le contexte qui explique les changements de comportements lors de la réalisation de cette activité. Pour cela, elle définit dans un premier temps les paramètres candidats aux contextes des entités impliquées dans cette activité à partir du modèle des composants contexte défini par Bazire et al., dans [41]. Concernant les étapes collecte et préparation des données et de la construction de la base de connaissances (les étapes 2 et 3 Figure II.1), les outils et méthodes classiques d'analyse de données peuvent être utilisés pour effectuer les prétraitements nécessaires et la contextualisation des données. Pour l'étape de construction de la base de connaissances toutes les méthodes et outils de fouille de

données peuvent être utilisés en adéquation avec les besoins d'analyse et des connaissances de l'analyste. Comme exemple de méthode de fouille, il y a les méthodes d'apprentissage supervisé ou non supervisé. Les méthodes d'apprentissage non supervisé consistent à construire des groupes homogènes d'éléments de telle sorte que les éléments qui se ressemblent soient dans le même groupe et ceux qui ne se ressemblent pas dans des groupes différents et l'apprentissage supervisé fonctionne de la même manière sauf que les groupes doivent être constitués suivant une variable donnée appelée variable de classe. Dans les étapes 4 et 5, les outils d'analyse de trace sont nécessaires pour l'identification du contexte explicatif d'un critère observé.

Les différents outils utilisés dans cette thèse pour l'analyse de l'activité de conduite avec la méthode de découverte de connaissances contextuelles sont les suivantes. L'activité de conduite est analysée pour définir les composants contexte dans l'étape 1. Dans l'étape 2, SPSS et Matlab sont utilisés pour la préparation des données collectées lors de l'activité de conduite : SPSS pour la création de nouveaux paramètres, l'analyse et l'exploration des données, Matlab pour le prétraitement des données collectées avec les différents capteurs lors de l'activité de conduite. Dans l'étape 3, SPSS est utilisé pour la découverte des règles de segmentation et de classification de l'activité de conduite qui seront stockées dans une base de connaissances. A partir de cette base de connaissances, dans l'étape 4, notre outil d'analyse de trace ATraDeCC (voir la partie 2.2) a été utilisé pour l'identification des paramètres contextuels explicatifs de la consommation en carburant et SPSS pour l'identification des règles. Dans l'étape 5, l'outil ATraDeCC d'analyse des traces d'interactions et SPSS sont être mobilisés pour la validation des connaissances produites.

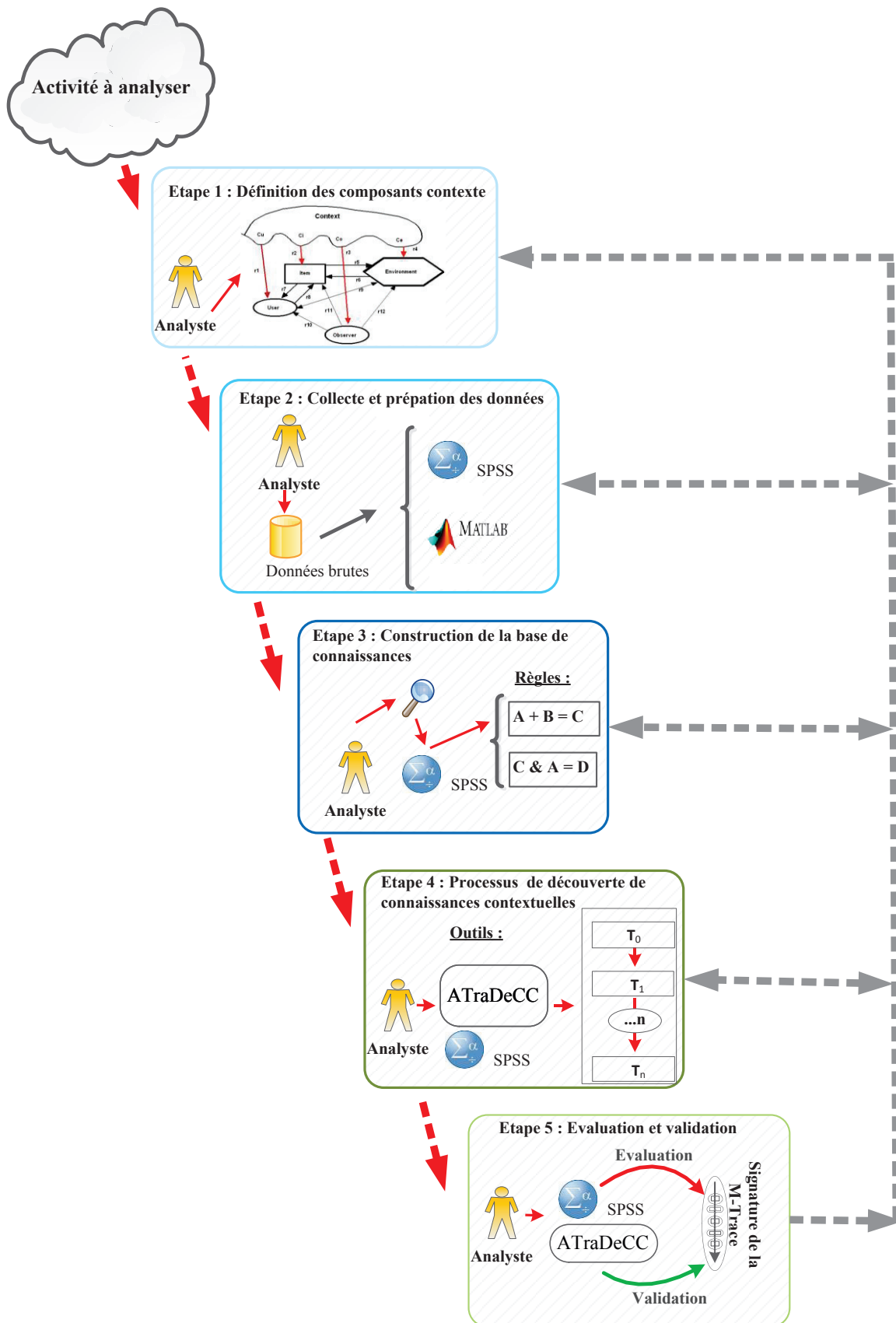


Figure II.9 : Outils utilisés pour la mise en œuvre de la méthode de découverte de connaissances contextuelles

2.2 Outil informatique pour l'Analyse des Traces modélisées (M-Traces) pour la Découverte de Connaissances Contextuelles (ATraDeCC)

Cet outil d'analyse de traces développé dans cette thèse respecte la notion de M-Trace modélisée définie dans [93]. En effet, les M-traces manipulées possèdent un modèle de trace dans lequel est définie la structure des observables « *Obsels* » avec leurs attributs. Chaque type d'observable est situé dans le temps avec une unité temporelle définie dans le modèle de trace. Ce logiciel peut être considéré comme un système à base de traces modélisées car il permet de créer une M-Trace collectée et d'effectuer des opérations de transformation pour enrichir la M-Trace avec de nouvelles connaissances issues de l'observation d'une activité.

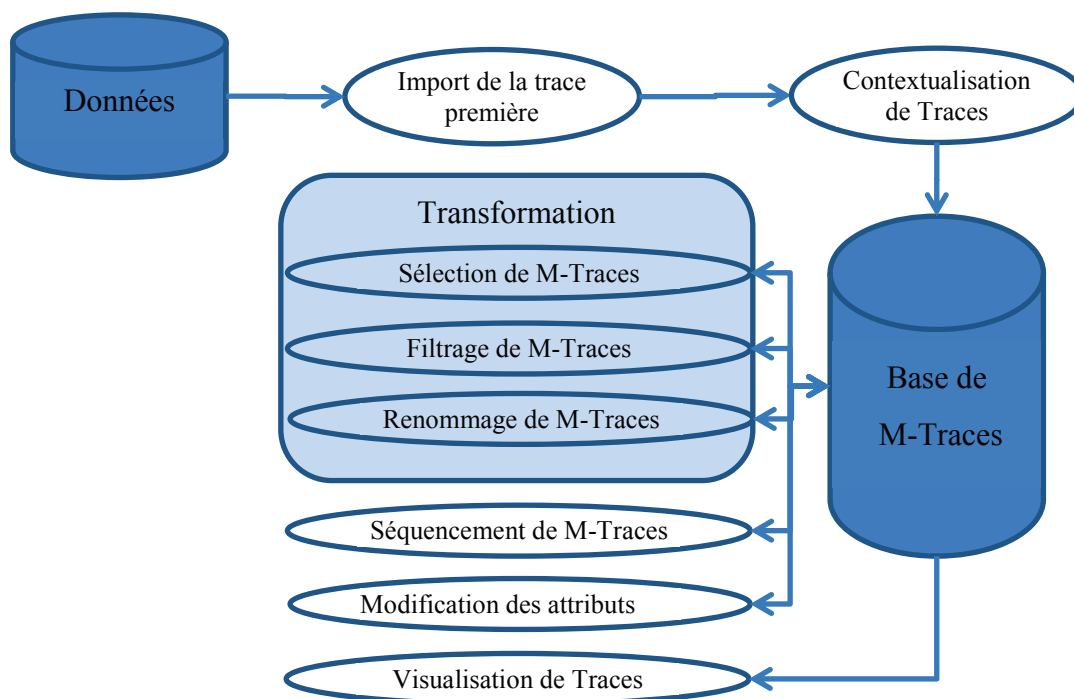


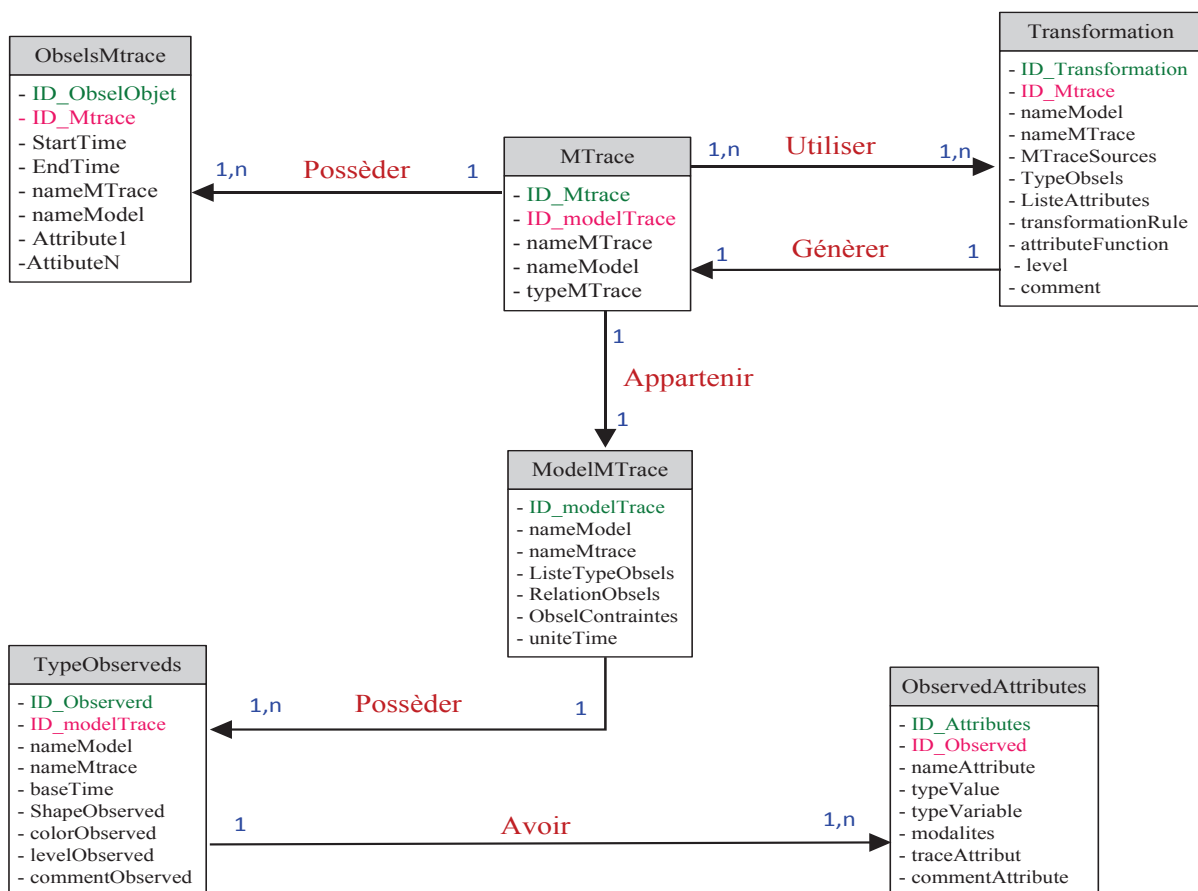
Figure II.10 : Fonctionnalités de l'outil ATraDeCC

Ce diagramme montre le fonctionnement de l'outil ATraDeCC. Une M-Trace première est créée à partir des données collectées. Puis elle peut être contextualisée en ajoutant les paramètres contextuels statiques pour les entités User et Item. A partir de cette M-trace première d'autres M-Trace sont obtenues par transformation. Une M-Trace calculée est une trace créée à partir d'une trace source sur laquelle est appliquée une opération de transformation (renommage, sélection, filtrage) ou de séquencement. Une transformation ou un séquencement donne toujours lieu à la création d'une nouvelle M-Trace qui contiendra de nouveaux obsels déduits des obsels de la M-Trace source qui répondent aux conditions des opérations de la transformation. Chaque transformation crée une M-Trace qui aura un modèle

de trace qui définira les types d'obsels et leurs attributs issus de la transformation. Ces M-Traces peuvent être visualisées.

2.2.1 Structure de la base de trace

Cet outil est développé sur un serveur WAMP et utilise donc du PHP et MySQL pour la gestion de la base de données. Contrairement au SBTm de [1] qui utilise RDF pour la gestion des M-Traces, ici on utilise la notion de gestion de base de données relationnelle pour gérer les données liées aux M-Traces. Voici le modèle relationnel que nous utilisons pour manipuler les M-Traces collectées et transformées :



Code couleurs:

Vert : Clé primaire de la table

Rose : Clé étrangère

Bleu : Contrainte d'unicité du champ

Figure II.11 : Modèle relationnel des données utilisé pour le stockage des M-Traces

Ce modèle relationnel permet la gestion des M-Traces collectées et des transformations qui leur sont associées. Dans ce modèle, toutes les M-Traces ont un modèle de trace qui est composé des types d'obsels « des observables de l'activité à analyser ». Chaque obsel possède des attributs qui le décrivent. Une M-trace est composée d'un ou de plusieurs obsels qui

peuvent être utilisés pour faire une transformation qui donne une nouvelle M-trace calculée ou transformée. Chaque transformation crée une M-Trace qui aura un modèle de trace qui définira les types d'obsels et leurs attributs issus de la transformation. Avec les relations de clés primaires et secondaires entre les différentes tables de notre modèle relationnel, il est possible d'associer une M-Trace à un modèle de trace, de définir la structure de ce dernier avec les types d'obsels et leurs attributs. Il est possible également de trouver les M-traces transformées d'une M-Trace source, d'avoir les types d'obsels liés à une transformation et les obsels de la M-Trace transformée. Ce modèle relationnel du SBTm et l'interface d'analyse de traces créé avec PHP permet (à partir des données brutes d'une activité) de stocker une M-Trace collectée et les M-traces transformées issue des opérations de transformations sur les traces collectées et transformées.

2.2.2 Les fonctionnalités implémentées

2.2.2.a Création de la M-Trace primaire ou collectée

L'objectif de cette fonctionnalité est de créer une M-Trace à partir d'un fichier contenant les données décrivant une activité. Pour cela, la M-Trace créée est enregistrée dans la base de données relationnelle décrite dans la figure II.10. La fonction de création de la M-Trace primaire prend en entrée le fichier contenant les données, les informations relatives à la base de données dans laquelle la M-Trace doit être créée et le modèle de trace qui doit être associé à la M-Trace. Ces informations sont recueillies grâce à l'interface de chargement de la M-trace.

Chargement M-trace collectée: Nouvelle M-Trace

Téléchargez la M-Trace collectée *:	<input type="text" value="D:/traore/Desktop/MTraces/TraceCollectee_02_02_16.csv"/>
Nom de la base de données *:	<input type="text" value="BDDrive"/>
Nom de la Trace collectée *:	<input type="text" value="Trajet1"/>
Nom du modèle de Trace *:	<input type="text" value="StructureTrajet1"/>
Unité temporelle du Modèle de Trace *:	<input type="text" value="Seconde"/>
Liste des types obsels du Modèle de la M-Trace *:	<input type="text" value="InfoDrive"/>
Relations binaire entre les types d'obsels:	<input type="text" value="Rien"/>
Contraintes du Modèle de la M-trace:	<input type="text" value="Temporele"/>

Figure II.12 : Page de création ou de chargement d'une M-Trace

A partir de l'en-tête du fichier, du nom du modèle de trace et de la liste d'obsel fourni par l'utilisateur, la fonction de création de M-Trace alimente la base de données relationnelle en créant le modèle de trace, les obsels, les attributs associés à la M-trace et la M-trace elle-même.

1	TimeStart	TypeObsel	TimeEnd	Speed	Bearing	SteeringWhe	SpeedLimit	TypeRoad	Cat_EginRP	FuelReal	FuelOpt	GearBoxReal
2	0	InfoDrive	1	0	0	-127.7	-1	3	0	0	0	0
3	1	InfoDrive	2	0	0	-127.7	-1	3	0	0	0	0
4	2	InfoDrive	3	0	0	-127.7	-1	3	0	0.01	0.01	0
5	3	InfoDrive	4	0	0	-127.7	-1	3	0	0.01	0.01	0
6	4	InfoDrive	5	0	0	-127.7	-1	3	1000	0.01	0.01	0
7	5	InfoDrive	6	0	0	-127.7	-1	3	0	0.01	0.01	0
8	6	InfoDrive	7	0	0	-136.38	-1	3	1000	0.01	0.01	0
9	7	InfoDrive	8	0.88	0	-241.33	-1	3	1000	0.01	0.01	1
10	8	InfoDrive	9	4.35	0	-302.21	-1	3	1000	0.01	0.01	1
11	9	InfoDrive	10	4.08	0	-328.07	-1	3	1000	0.02	0.02	1
12	10	InfoDrive	11	2.68	0	-341.9	-1	3	1000	0.02	0.02	1
13	11	InfoDrive	12	0.54	0	-287.88	-1	3	1000	0.02	0.02	1
14	12	InfoDrive	13	0	0	107.61	-1	3	1000	0.02	0.02	0
15	13	InfoDrive	14	1.25	0	324.82	-1	3	1000	0.03	0.03	1
16	14	InfoDrive	15	0.76	0	348.13	-1	3	1000	0.01	0.01	1
17	15	InfoDrive	16	0	0	345.7	-1	3	1000	0.02	0.02	0
18	16	InfoDrive	17	0	0	345.7	-1	3	1000	0.02	0.02	0
19	17	InfoDrive	18	0	0	351.61	-1	3	1000	0.01	0.01	0
20	18	InfoDrive	19	0.47	0	362.69	-1	3	1000	0.01	0.01	1
21	19	InfoDrive	20	2.68	0	337.89	-1	3	0	0.01	0.05	1
22	20	InfoDrive	21	5.04	0	129.21	-1	3	1000	0.04	0.09	1
23	21	InfoDrive	22	7.85	0	-161.39	-1	3	1000	0.04	0.04	1

Figure II.13 : Exemple de fichier contenant la M-Trace primaire

A la fin de traitement de cette fonction la base de M-Trace contient une M-trace primaire décrite pour les obsels ainsi que les attributs qui caractérisent les obsels ainsi qu'un modèle de trace qui décrit les types d'obsels et les attributs.

Les informations sur la M-Trace chargée

Nom M-Trace collectée : Trajet1

Nom du modèle de la M-Trace : StructureTrajet1

Unité temporelle : Seconde

Liste des types d'obsels dans le modèle : InfoDrive

Liste des attributs associés aux obsels : Speed, Bearing, SteeringWheelAngle, SpeedLimit, TypeRoad, Cat_EginRPM, FuelReal, FuelOpt, GearBoxReal

Relations binaire entre les types d'obsels : Rien

Les contraintes du Modèle de Trace : Temporelle

Voir les attributs d'un type observé:

Selectionnez un Observé: InfoDrive

Figure II.14 : Page de résultat de la création d'une M-trace

2.2.2.b Spécifications ou Modifications des caractéristiques des attributs des obsels des M-Trace

Chaque attribut peut être complété en renseignant les informations complémentaires le concernant.

Informations de l'obsel : InfoDrive

Nom de la M-Trace : **Trajet1**

Nom du modèle de la M-Trace : **StructureTrajet1**

Les attributs de l'obsel **InfoDrive** sont :

N°	Nom Attributs	Type Attributs	Type variable	Valeurs possibles	Opération de mise à jour
1	Speed	FLOAT	No define	No define	Compléter l'attribut
2	Bearing	FLOAT	No define	No define	Compléter l'attribut
3	SteeringWheelAngle	FLOAT	No define	No define	Compléter l'attribut
4	SpeedLimit	FLOAT	No define	No define	Compléter l'attribut
5	TypeRoad	FLOAT	No define	No define	Compléter l'attribut
6	Cat_EnginRPM	FLOAT	No define	No define	Compléter l'attribut
7	FuelReal	FLOAT	No define	No define	Compléter l'attribut
8	FuelOpt	FLOAT	No define	No define	Compléter l'attribut
9	GearBoxReal	FLOAT	No define	No define	Compléter l'attribut

Figure II.15 : Page de visualisation et de modification des attributs d'une M-Trace

A partir de cette page de visualisation des attributs de la M-trace collectée, l'utilisateur peut compléter les informations relatives à un attribut grâce à la page suivante :

Modification de l'attribut : Speed

Nom du type d'Obsel de l'attribut * : **Trajet1**

Nom Attribut * : **Speed**

Type Attribut * : **FLOAT**

Type variable (Numeric, Nominal, Ordinal,...)* : Numeric

Liste valeurs possible* : 0 à 150

Compéter l'attribut

Figure II.16 : Saisie d'informations complémentaire d'un attribut

Après avoir cliqué sur le bouton « Compléter l'attribut », l'attribut sélectionné est complété et la page résultat suivante est affichée pour confirmer la mise à jour de l'attribut en question.

Modification de l'attribut : Speed Opérateurs de Transformations

L'attribut Speed a été modifié avec succès.

Nom du type d'Obsel de l'attribut *	: Trajet1
Nom Attribut *	: Trajet1.Speed
Type Attribut *	: Speed
Type variable (Numeric, Nominal, Ordinal,...)*	: Numeric
Liste valeurs possible*	: 0 à 150

Figure II.17 : Page résultat de la modification de la M-Trace

2.2.2.c Transformation ou reformulation

La fonction de transformations a pour but d'enrichir la M-Trace collectée ou transformée en inférant de nouvelles connaissances à partir des observations de l'activité analysée. Cette fonction prend en entrée la M-Trace sur laquelle la transformation sera réalisée appelé la M-Trace source, le nom de la M-Trace transformée, le nom du modèle qui sera associé à la nouvelle M-Trace transformée, la liste des types d'obsels, la liste des attributs, les fonctions de calcul correspond aux attributs et la clause ou condition de transformation.

Transformation de la M-trace collectée: Trajet1 Visualiser la M-Trace

Nom du type observé :

Liste des attributs:

Liste type des attributs:

Fonction des attributs dans la clause :

Clause de Transformation(Where):

Commentaire observé:

Niveau de l'observé:

Voir les attributs d'un type d'observé de la M-Trace:

Selectionnez un Observé:

Figure II.18 : Page de transformation d'une M-Trace

A partir de la règle de transformation la fonction va sélectionner les obsels de la M-Trace source qui respectent cette condition. Chaque obsel est décrit par un temps de début, un temps de fin et les attributs qui le décrivent. Les obsels sont décrits par les attributs durant sa durée avec une liste de valeurs observables.

Après avoir renseigné les informations nécessaires à la transformation, en cliquant sur le bouton « Exécuter la transformation », la transformation est effectuée et la page de résultat de la transformation est affichée. Cette page contient la liste des obsels créés au cours de la transformation, la liste des attributs, les règles de transformation et un commentaire pour décrire la transformation.

Les Transformations de la M-Trace : Trajet1

Visualiser la M-Trace Effectuer une Transformation Créer une séquence de type d'obsels

Liste des transformations de la M-trace Trajet1

N°	Type Obsel	Liste des attributs	Règles de transformations	Commentaire
1	AngleVolantToutDroit	MeanAngleVolant, MaxAngleVolant, MinAngleVolant, MaxBearing, MinBearing, Distancetravel, MeanSpeed, FuelConsumption	InfoDrive.SteeringWheelAngle BETWEEN -10 AND 10	Angle volant virrage moyen.Sequence angle volant entre -10 et 10.

Voir les obsels des transformations de la M-Trace: Trajet1

Selectionnez un type d'obsel transformé: Liste type d'obsels Voir les observés de ce type d'obsel

Figure II.19 : Page résultat récapitulatif d’une transformation

Pour caractériser les attributs des obsels de la M-trace transformée, des fonctions d’agrégation doivent être utilisées sur la liste des valeurs des attributs des obsels de la M-Trace source qui sont compris entre le début et la fin de l’obsel de la trace transformée. Ces fonctions d’agrégations prédéfinies sont :

Descriptions fonctions d'agrégation	Notation
Calcule la durée de la séquence	dure(\$start, \$end)
Calcule la somme des valeurs d'un paramètre numérique durant une séquence	Somme (listeValeurs)
Calcule la moyenne d'un paramètre numérique durant une séquence	moyenneSequence(\$listeValue)
Calcule la dérivée d'un paramètre numérique durant une séquence	Derivée((\$listeValue)
Calcule l'écart type d'un paramètre numérique durant une séquence	EcartType ((\$listeValue)
Calcule l'amplitude du paramètre numérique, différence entre les valeurs maximale et minimale de ce paramètre durant une séquence	Amplitude ((\$listeValue)
Calcule la différence entre deux valeurs du même paramètre	Diff(\$listeValue)
Calcule la valeur minimum prise par un paramètre numérique durant une séquence	minListeValue(\$listeValue)
Calcule la valeur maximum prise par un paramètre numérique durant une séquence	maxListeValue(\$listeValue)
Calcule le pourcentage de temps d'utilisation des modalités d'un paramètre durant une séquence	Pourcentage((\$listeValue, \$listeModalite)
Détermine la modalité la plus utilisée d'un paramètre nominal durant une séquence	PlusUtilisé(listeValeurs, \$listeModalite)
Détermine la modalité qui a le moins d'occurrence dans la liste des valeurs.	minOccurence(\$listeValue, \$listeModalite)
Détermine la modalité qui a le plus d'occurrence dans la liste des valeurs.	maxOccurence(\$listeValue, \$listeModalite)

Tableau II.2 : Fonctions d'agrégation mis en place pour logiciel

A la fin de l'exécution de cette fonction de transformation, la M-Trace transformée est créée avec les obsels et les attributs qui les décrivent et la structure de la M-Trace transformée est décrite dans le modèle de trace.

2.2.2.d Séquencement d'une M-Trace

Cette fonction permet de créer une séquence temporelle à partir des obsels d'une M-Trace. Les séquences temporelles sont créées grâce à la logique temporelle d'Allen définie dans [68]. La logique d'Allen définit les relations temporelles qui peuvent exister entre les événements. Chaque relation est implémentée par une fonction qui prend en entrée les temps de début et fin des obsels et la fonction vérifie si ces obsels respectent ou non cette relation. La figure suivante illustre une transformation permettant de créer une situation de virage à gauche à partir des différentes valeurs des angles du volant.

Séquence des transformations de la M-trace : Trajte1

Nom de la séquence :

Liste des types d'obsels:

Clause de la séquence de Transformation:

Commentaire observé:

Niveau de transformation:

Les types d'obsels transformés de la M-Trace collectée: Trajte1

Le type d'obsel de la M-Trace source : InfoDrive

Les types d'obsels transformés : AngleVolantToutDroit AngleVolantGauche1 AngleVolantGauche2 AngleVolantGauche3 AngleVolantDroite1
AngleVolantDroite2 AngleVolantDroite3

Figure II.20 : Page de création de séquence dans une M-Trace

Cette transformation a pour but de créer une situation de virage à gauche, qui est décrite par la règle de transformation suivante : (((AngleVolantToutDroit Meet AngleVolantGauche1) Before AngleVolantGauche2) Before AngleVolantGauche3). Cette règle décrit le séquençage temporel des différents angles volant décrivant la situation virage à gauche.

2.2.2.e Contextualisation

La fonction de contextualisation a pour but de renseigner les informations contextuelles statiques relatives aux entités impliquées dans l'activité. Pour le cas d'analyse de l'activité de conduite il s'agit des entités « Conducteur » et « Véhicule ». Les informations contextuelles décrivant ces entités sont considérées statiques car elles sont constantes sur un trajet donné. Par exemple sur un trajet les informations telles que l'identifiant du conducteur, l'âge, le style de conduite, etc. ne changent pas pendant le trajet puisqu'on a le même conducteur et pareil pour le véhicule, les informations comme la marque du véhicule, l'état du véhicule, le type de carburant qu'il utilise ne changent pas non plus. Ces informations contextuelles décrivant les entités « Conducteur » et « Véhicule » sont connues et sont invariables sur un trajet. La fonction de contextualisation permet de renseigner ces informations contextuelles pour chaque M-Trace primaire.

Contextualisation de la M-Traces : Trajet1 Opérateurs de Transformations

Contexte de l'utilisateur: 'Conducteur'

Identifiant *: 4

Age *: 28

Experience de conduite *: 10

Port verre correcteur *: Oui

Type de conduite *: Souple

Autre: Fatigue au volant

Contexte de l'Item: 'Véhicule'

Identifiant *: Clo1

Marque *: Renault

Age *: 2

Etat *: Bon

Type de carburant *: Essence

Autre: RAS

Enregistrer ces Informations du Contexte

Figure II.21 : Page de contextualisation des M-Traces

2.2.2.f Description

Cette fonction permet de décrire une M-Trace, pour cela elle prend le nom de la M-Trace à décrire et elle renvoie les informations de description de la M-Trace telles que le nom du modèle de trace, les types d'obsels qui sont associés à la M-Trace et les attributs qui décrivent ces types d'obsels.

Description de la M-Trace Source : Trajet1 Opérateurs de Transformations

Modèle de la M-Trace	: StructureTrajet1
Liste des types d'obsels	: InfoDrive
Liste des attributs du modèle de la M-Trace	: Speed, Bearing, SteeringWheelAngle, SpeedLimit, TypeRoad, Cat_EnginRPM, FuelReal, FuelOpt, GearBoxReal,

Figure II.22 : Page de description d'une la M-Trace

2.2.2.g Visualisation

La fonction de visualisation a pour but de faire une visualisation d'une M-Trace (primaire ou transformée). Elle se fait sous deux formes :

- Affichage de la liste des obsels avec leurs attributs.

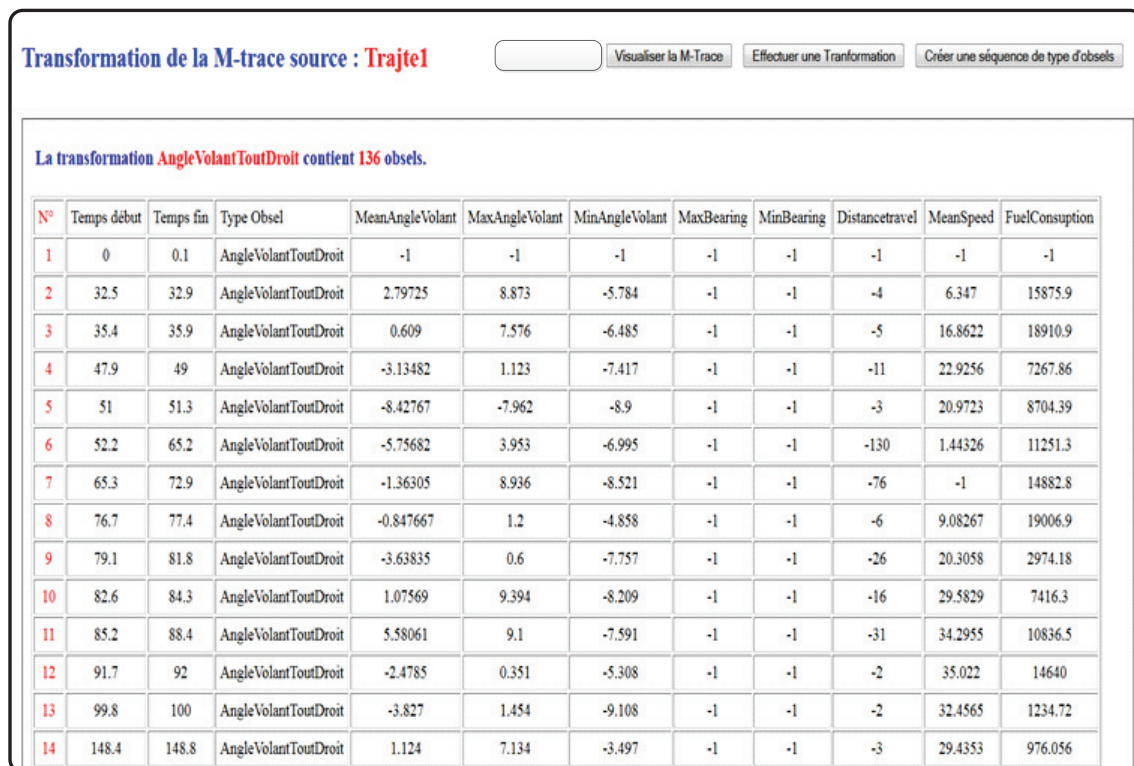


Figure II.23 : Page de visualisation des obsels d'une M-Trace transformée

Sur cette page de visualisation la liste des obsels de la M-Trace sélectionnée est affichée avec leurs attributs. Chaque couleur correspond à un type d'obsel et chaque type d'obsel décrit une situation de conduite. Par exemple, il y a quatre situations de conduite qui sont représentées dans la figure suivante : les obsels bleus correspondent aux situations de conduite aller tout droit, les verts aux virages à droite (virage1), les rouges aux virages à gauche (virage2) et les oranges aux arrêts.

- Affichage graphique

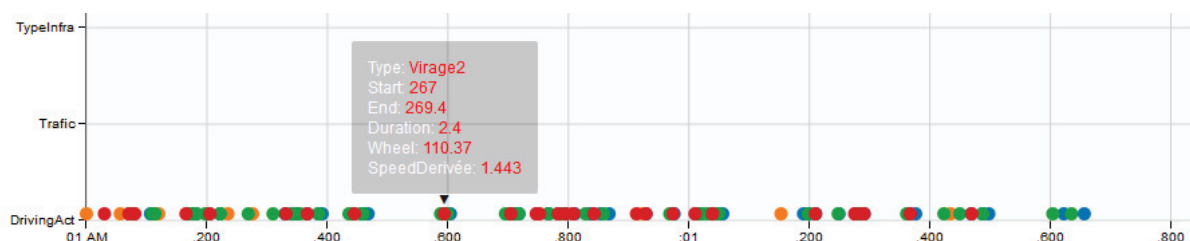


Figure II.24 : Visualisation graphiques des obsels

3 Synthèse du chapitre de contribution

Dans ce chapitre, les deux contributions principales de la thèse sont présentées. Il s'agit de la méthode de découverte de connaissances contextuelles et de l'outil informatique développé permettant d'analyser les traces d'interactions.

La méthode de découverte de connaissances contextuelles proposée est composée de cinq étapes et elle est interactive et itérative. Elle permet de découvrir le contexte explicatif d'un critère observé de l'activité en définissant au préalable le contexte de chacune des entités impliquées dans la réalisation de cette activité. Le contexte des entités est défini à travers les paramètres qui le caractérisent. Ces paramètres sont alors collectés et préparés pour l'identification du contexte explicatif en fonction du critère observé. Cette méthode utilise le modèle des composants contexte de Bazire et al., [41] pour définir les paramètres candidats au contexte de l'activité analysée. Les M-Traces (Traces Modélisée) et les méthodes de fouille de données sont mobilisées pour l'identification parmi les paramètres candidats de ceux qui ont une influence sur le critère observé. L'outil informatique développé permet la création des M-Traces et leur enrichissement selon le point de vue de l'analyste pour la découverte de connaissances contextuelles. Un modèle relationnel a été mise en place pour la gestion de la base de M-Traces contrairement au Système à Base de Traces (SBT) de Silex qui utilise la notion de graphe (RDF : Resource Description Framework) pour la gestion des M-Traces. Cet outil permet le chargement d'une M-Trace collectée, d'effectuer des transformations sur cette M-Trace et la visualisation des obsels issus des transformations pour l'identification du contexte explicatif d'un critère observé.

La méthode de découverte de connaissances contextuelles et l'outil ATraDecc seront utilisés dans le chapitre suivant pour identifier les contextes explicatifs de la consommation en carburant.

Chapitre III. Application de la méthode de découverte de connaissances contextuelles

L'objectif de ce chapitre est de présenter l'expérimentation de la méthode de découverte de connaissances contextuelles dans le domaine du transport pour identifier les éléments contextuels de l'activité de conduite modifiant le niveau de consommation en carburant. Il s'agit d'une application concrète menée avec une analyste de l'Ifsttar. Le critère observé pour cette étude est la consommation en carburant et l'objectif est donc d'identifier le *contexte* explicatif de ce critère. En d'autres termes, nous voulons déterminer les paramètres contextuels influençant fortement le niveau de consommation en carburant lors de la conduite automobile et le sens de variation de cette dernière en fonction des valeurs de ces paramètres. L'identification de ces éléments contextuels permettra d'une part, d'évaluer de manière précise et fiable l'influence du contexte de conduite sur la consommation en carburant et d'autre part, de définir les éléments contextuels explicatifs de la consommation en carburant qui pourraient être utilisés pour prédire la consommation en carburant en fonction de la situation de conduite. Cette prédiction pourrait, à terme, être implémentée dans un système d'assistance de conduite écologique en temps réel pour conseiller le conducteur dans des situations de conduite précises en lui fournissant les actions de conduite nécessaires pour une consommation de carburant optimisée.

Ce chapitre décrit les données expérimentales qui ont été utilisées dans cette thèse et l'application de la méthode de découverte de connaissances contextuelles sur ces dernières.

1 Description des données (expérimentation)

Les données ont été collectées dans le cadre du projet EuroFOT (European Field Operational Test), projet européen ayant pour objectif d'évaluer les impacts des systèmes d'assistance à la conduite sur la sécurité routière, l'environnement et l'efficacité de la conduite du conducteur. L'acquisition des données a été effectuée sur une période d'un an pour chacun des 35 participants. Pour l'acquisition des données, deux niveaux d'instrumentation des véhicules ont été mis en place, un niveau léger et un lourd. Le premier niveau d'instrumentation (léger) était constitué d'un dispositif minimum pour la collecte des données sur le réseau CAN du véhicule, d'un radar et d'une antenne GPS installée dans le véhicule. Le second niveau d'instrumentation (lourd) disposait en plus des équipements du premier niveau, des caméras vidéo et d'autres capteurs que nous verrons en détail dans la partie suivante. Le CEESAR (Centre Européen d'Etudes de Sécurité et d'Analyse des Risques) a mis en œuvre cette

collecte de données, en instrumentant les véhicules de tous les participants au premier niveau et en leur prêtant cinq véhicules de leur parc, instrumentés au second niveau. Les véhicules instrumentés au second niveau ont été prêtés aux sujets pendant 45 jours (3 fois 15 jours). Le reste du temps, ils ont utilisé leur propre véhicule pour leurs déplacements quotidiens, au cours desquels les données ont été collectées à l'aide d'une instrumentation spécifique. Les 19 participants qui conduisaient une Clio parmi les 35 ont été retenus pour notre étude et fournissent 1248 trajets de conduite observés en détail car le modèle d'optimisation utilisé n'avait pas de modèle de Laguna disponible lors de cette thèse.

1.1 Instrumentation des véhicules

L'acquisition de données en situation naturelle de conduite nécessite une instrumentation des véhicules adaptée avec un fonctionnement automatique et quasi invisible. Dans le projet EuroFOT, toutes les données ont été collectées en situation naturelle de conduite. Les véhicules ont été équipés de dispositifs nécessaires à la collecte des données lors de l'activité de conduite et leur instrumentation a été effectuée de manière très discrète pour ne pas influencer les attitudes des participants lors de la conduite. Les enregistrements des données ne nécessitaient aucune action du conducteur dans leur fonctionnement. Dans le but d'acquérir des données dans des situations de conduite naturelle, il n'y avait pas de parcours prédéfinis lors de cette expérimentation, les participants utilisaient au quotidien selon leurs besoins leur propre véhicule instrumenté légèrement ou des véhicules lourdement équipés prêtés par le CEESAR. Les véhicules instrumentés légèrement (au premier niveau) enregistraient des données sur le bus CAN (*Controller Area Network*) du véhicule et ont été équipés d'une antenne GPS, d'un Radar, et d'un enregistreur compact pour stocker les données enregistrées, transférées à distance grâce à une antenne GSM. Le bus CAN a permis d'enregistrer les paramètres dynamiques sélectionnés comme la vitesse, l'accélération, la consommation en carburant en se connectant sur le réseau du véhicule. La position du GPS a été collectée via une antenne sans fil et une étape de post-traitement a permis d'extraire d'une base de données cartographique des informations sur l'environnement de conduite. Le radar automobile d'une portée de 200m a permis la détection du véhicule de devant et de collecter des informations le concernant comme ses position, accélération et vitesse relatives (c'est-à-dire définies dans le référentiel du véhicule instrumenté).

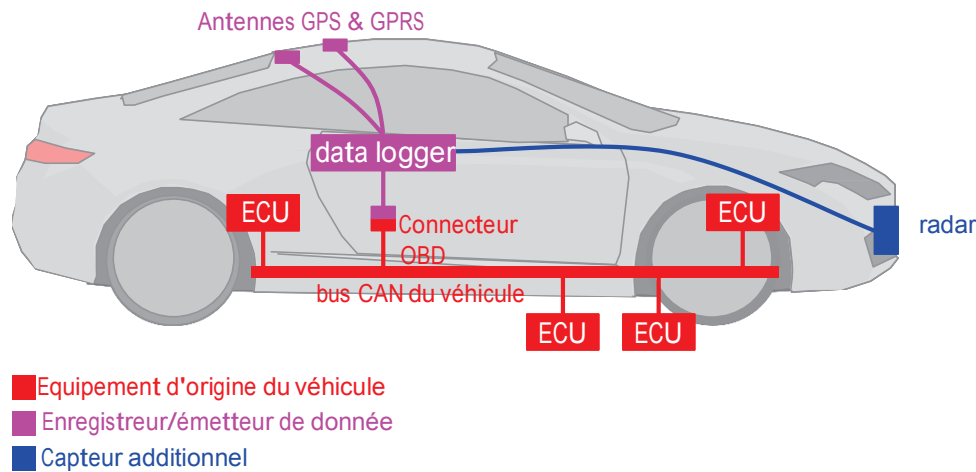


Figure III.1 : Illustration du premier niveau d'instrumentation du véhicule

En plus des équipements du premier niveau d'instrumentation, les véhicules équipés lourdement (second niveau) disposaient de quatre caméras vidéos, une caméra multifonction, un oculomètre et un ordinateur pour stocker les enregistrements vidéos. Les caméras ont permis de filmer l'environnement de conduite et les actions de conduite du conducteur. La caméra multifonction fournit la position latérale du véhicule (franchissement de ligne, déviation standard de la trajectoire). L'oculomètre permet de mesurer la position de la tête et l'orientation du regard.

Ces vidéos ont été enregistrées pendant la conduite automobile pour être exploitées lors de l'analyse de l'activité de conduite, afin d'interpréter certaines situations complexes de la conduite comme les incidents, ainsi que pour le codage de l'activité de conduite, la vérification et la validation des données collectées et de certains résultats obtenus lors de l'analyse de l'activité. Elles sont essentielles pour la prise en compte du contexte de l'activité de conduite car elles permettent de renseigner les paramètres contextuels comme par exemple le trafic, le type de route, la présence ou non de feux de signalisation. Et en plus, elles représentent une source d'informations pour décrire les comportements au volant des conducteurs tels que les enfoncements de pédales de frein, d'embrayage et d'accélérateur, les regards sur les rétroviseurs avant les virages, l'utilisation du téléphone portable pendant la conduite.

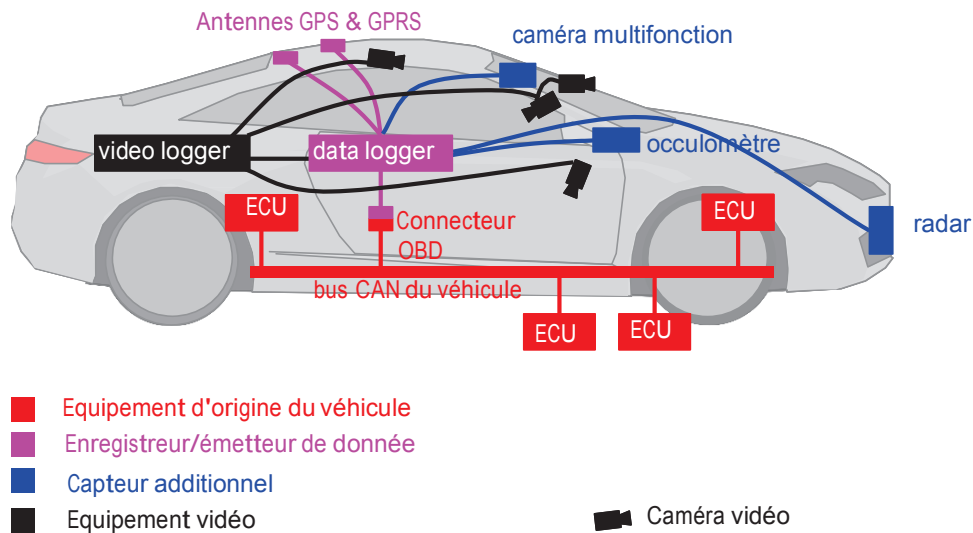


Figure III.2 : Illustration du second niveau d'instrumentation du véhicule

Les quatre canaux de caméras ont permis de filmer la scène avant, le visage du conducteur, le poste de conduite, la position des pieds. Ces caméras avaient une haute-sensibilité et un éclairage infrarouge pour les caméras internes à l'habitacle, permettant l'enregistrement de la conduite de nuit. L'enregistreur vidéo disposait d'une autonomie de 175 heures, largement suffisante pour les périodes de minimum deux semaines de prêt des véhicules du second niveau d'instrumentation



Figure III.3 : Instrumentation du véhicule d'expérimentation de CEESAR

Les enregistrements ont été réalisés de manière continue, afin de permettre tout type d'analyse ultérieure. Les différents sources des données étaient synchronisées (signaux capteurs, vidéos, GPS et Radar).

1.2 Protocole expérimental

L'expérimentation a été réalisée sur une période d'un an durant laquelle des données ont été collectées lors des déplacements quotidiens des 35 participants en situation naturelle de conduite automobile. L'objectif du projet EuroFOT étant d'évaluer les impacts des systèmes d'assistance à la conduite automobile, l'expérimentation a été effectuée en deux phases (*Baseline* et *Traitement*). La durée de la phase *Baseline* était de 3 mois et avait pour objectif de collecter des données sans utilisation de systèmes d'assistance. La phase *Traitement*, d'une

durée de 9 mois, avait pour objectif de collecter des données avec utilisation des systèmes d'assistances. Ceci a permis de connaître le type de conduite habituelle des sujets afin de mesurer les différences entre le comportement sans système et celui avec système. De plus, des passations de questionnaires ont été effectuées à différents moments (Time 1 2 3a 3b et 4) pour collecter des informations complémentaires sur les participants (caractéristiques telles que âge, genre, habitude de conduite, attitudes au volant) et sur l'usage des systèmes. Trois mois minimum ont été observés sans système et 9 mois minimum avec système.

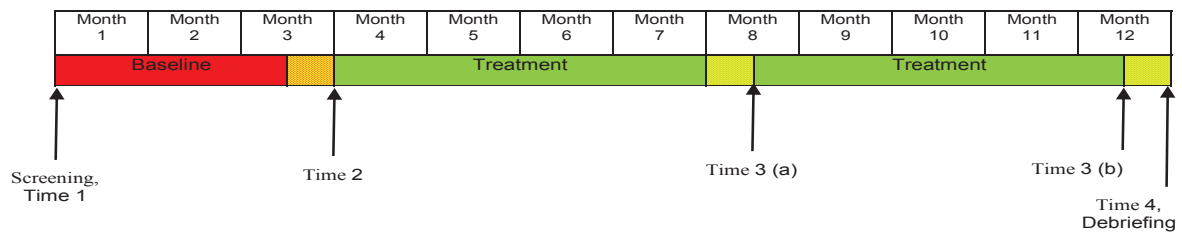


Figure III.4 : Répartition des périodes d'enregistrement de l'expérimentation

D'un point de vue pratique, un protocole glissant de l'expérimentation a été nécessaire pour pouvoir prêter les véhicules lourdement équipés (3 fois 15 jours) à chacun des participants pendant l'expérimentation.

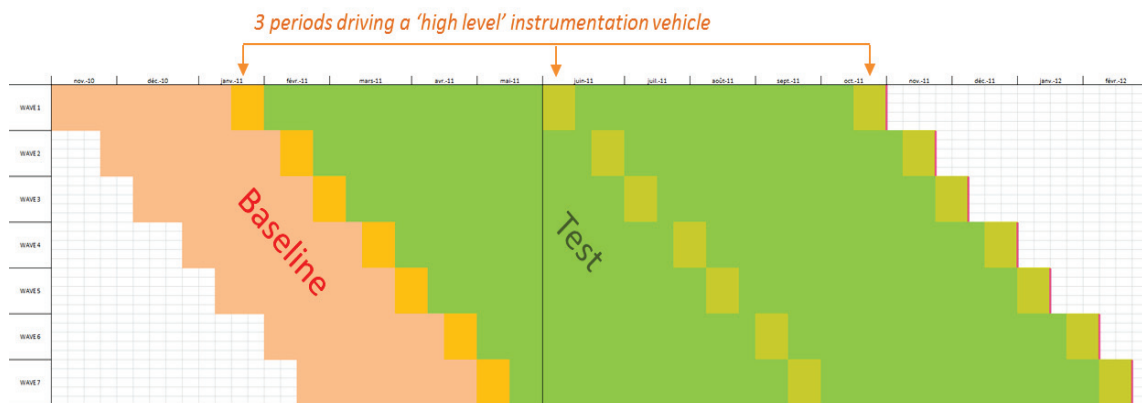


Figure III.5 : Répartition des véhicules de second niveau d'instrumentation

2 Application de la méthode de découverte de connaissances contextuelles

La méthode de modélisation des connaissances contextuelles a pour objectif d'identifier et de modéliser les éléments *contextuels* d'un critère observé d'une activité analysée. Ici, l'activité analysée est celle de conduite et le critère observé est la consommation en carburant. Le carburant est nécessaire à l'activité de conduite automobile car sa combustion lors de l'activité de conduite fournit l'énergie pour le déplacement du véhicule d'un endroit à un

autre. Cette combustion du carburant lors de la conduite automobile produit des GES qui contribuent fortement au phénomène du réchauffement climatique. Dans la littérature, il est montré que la consommation en carburant de l'activité de conduite est fortement liée aux caractéristiques routières (infrastructure, type de route, trafic), aux comportements du conducteur (style de conduite) et aussi au type de véhicule. Notre objectif dans cette analyse est de comprendre et expliquer l'influence de ces différents éléments sur la consommation en carburant. Pour cela, il faut au préalable identifier ces éléments pour déterminer leur influence sur la consommation en carburant. Il faut donc faire une modélisation de ces éléments. Dans le cadre de cette thèse, ces éléments sont considérés comme étant le *Contexte explicatif* de la consommation en carburant. Notre objectif est dans cette partie de décrire comment la méthode de découverte de connaissances contextuelles est utilisée sur les données de l'activité de conduite collectées en situation naturelle pour identifier l'influence du contexte sur la consommation.

2.1 Définition des composants contexte de l'activité de conduite

Notre objectif est dans cette étape de modéliser le *contexte* de l'activité de conduite expliquant la consommation en carburant. L'activité de conduite automobile consiste à se déplacer d'un point A à un point B au moyen d'un véhicule. Pour cela, les entités *conducteur* et *véhicule* interagissent ensemble dans un *environnement* donné qui est l'environnement de conduite (la route). Non seulement ces entités interagissent ensemble, mais elles ont une influence sur l'activité de conduite. Par exemple, en cas de trafic dense, la conduite peut être plus lente car l'environnement extérieur contraint le conducteur à tenir compte des autres véhicules. Lors de la conduite automobile, le conducteur adapte en permanence sa conduite en fonction de ce qui se passe autour de lui, donc de son environnement de conduite (trafic, météo, type de route, ...). Cette adaptation aux conditions extérieures influence les comportements et les actions de conduite. Ces changements de comportements de conduite influencent à leur tour la consommation en carburant [17]. La consommation en carburant étant dépendante des différentes entités de l'activité de conduite, les valeurs prises par les paramètres décrivant ces entités lors de l'activité de conduite pourraient potentiellement expliquer la consommation en carburant de manière très précise. Les paramètres décrivant ou caractérisant ces entités peuvent être considérés comme des paramètres *candidats au contexte* de l'activité de conduite d'après la définition du *contexte* de Dey [35]. Pour définir ces paramètres, le modèle des composants contexte (ou entités) de Bazire et al., [41] est utilisé dans cette thèse. Ils correspondent aux informations relatives aux entités impliquées dans l'activité de conduite telles que le *conducteur*, le *véhicule*, l'*environnement de conduite*.

Dans le cadre de cette thèse, l'entité Observateur fait référence à l'analyste avec ses objectifs et ses besoins d'analyse.

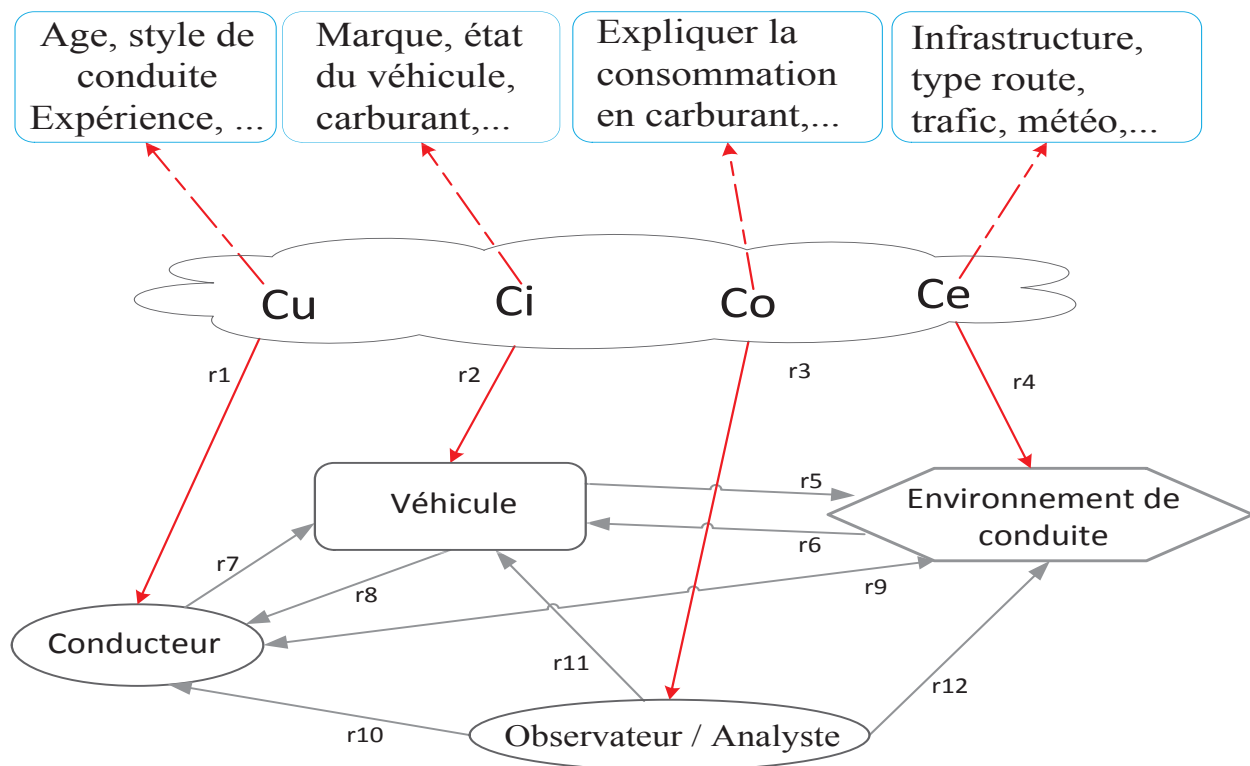


Figure III.6 : Illustration des composants *contexte* de Bazire et al., [41] appliqué à l'activité de conduite

Dans la figure ci-dessus les entités de l'activité de conduite sont caractérisées par les paramètres candidats potentiels au contexte de l'activité de conduite. Chaque entité possède son propre *contexte*, et ce dernier est défini par les paramètres qui caractérisent cette entité.

- Par exemple « **Cu** » qui est le contexte de l'entité Conducteur (User dans le modèle des composants contexte de Bazire et al.[41]) est défini comme par exemple les paramètres âge, style de conduite, expérience, nombre de kilomètres parcourus par année. Ces paramètres permettent de décrire un conducteur et de pouvoir le caractériser grâce à ces informations.
- Le contexte de l'Environnement « **Ce** » qui correspond aux caractéristiques de la route comme par exemple le trafic, l'infrastructure, le type de route, la présence de feux de signalisation. Ce composant contexte est celui qui a le plus évolué au cours de l'analyse. En effet, si des éléments comme le type de route ont été introduits dès le début de l'analyse, l'analyste a introduit d'autres éléments comme la présence des feux et le trafic au fur et à mesure du processus.
- Le contexte de l'entité « *Véhicule* » est caractérisé par les paramètres tels que la marque du véhicule, le type de carburant ou le système d'assistance actif.

- les paramètres décrivant le contexte de l'entité « *Observateur /Analyste* » est défini par les objectifs et les besoins d'analyse comme par exemple le critère à expliquer et les contraintes liées à l'analyse.

Ces paramètres décrivant les entités de l'activité de conduite peuvent être de différents types (quantitatif, qualitatif ou ordinal). Par exemple, le paramètre genre est de type qualitatif avec les modalités suivantes *homme, femme*. Le paramètre âge du conducteur qui est quantitatif varie entre *18 et 80 ans*. Le paramètre style de conduite peut être ordinal avec les modalités suivantes : *pas agressif, agressif et très agressif*.

La phase de définition des composants contexte permet de définir les paramètres candidats potentiels aux contextes des entités de l'activité analysée. Dans cette étape, il est préférable de définir le maximum de paramètres susceptibles d'avoir une influence sur le critère observé pour une meilleure identification de son **Contexte explicatif**. A chaque itération, les paramètres potentiels candidats au **Contexte** sont définis et les données nécessaires pour les renseigner sont collectées et préparées dans la phase suivante.

2.2 Collecte et préparation des données

2.2.1 Collecte des données

Grâce à l'expérimentation mise en place pour la collecte des données dans la partie (1.2), une grande base de données de l'activité de conduite a été collectée, en tout 1248 trajets ont été enregistrés dont 300 avec vidéos. Sur chaque trajet, plusieurs paramètres sont enregistrés en continu et de manière instantanée, grâce aux dispositifs installés à bord des véhicules d'expérimentation tels que le bus CAN, le GPS, le Radar et les caméras. Grâce à ces dispositifs, 47 paramètres et des vidéos ont été collectés, synchronisés et stockés lors de l'activité de conduite. Les paramètres, définis dans l'annexe 1 (Tableau A.1 : Liste des paramètres issus du bus CAN, Tableau A.2 : Liste des paramètres issus de la caméra multi fonction, Tableau A.3 : Liste des paramètres issus du Radar, Tableau A.4 : Liste des paramètres issus du GPS, Tableau A.5 : Liste des paramètres issus de la phase de mise en correspondance de la position GPS et la base de données cartographique (Map Matching)) sont enregistrés toutes les 100 millisecondes.

2.2.2 Exploration des données

Cette étape a pour objectif d'utiliser les techniques classiques de prétraitements des données pour préparer la phase d'analyse des données. A partir des données brutes collectées lors de l'activité de conduite, deux opérations vont être nécessaires pour la préparation des données.

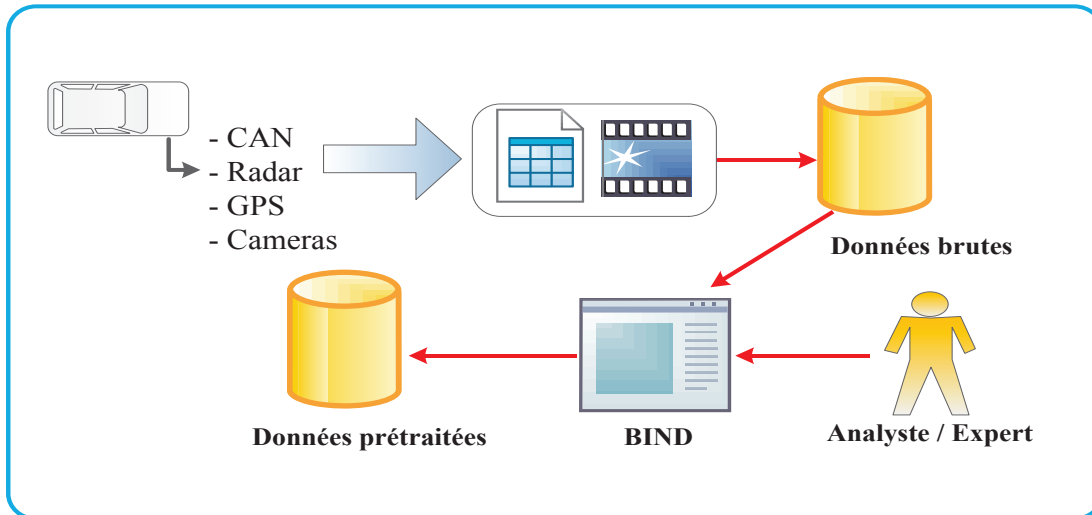


Figure III.7 : Collecte et préparation des données par l'analyste

L'exploration des données permet à l'analyste de bien connaître les données pour préparer le prétraitement des données. Dans cette thèse, les données collectées ont été explorées et visualisées avec l'outil BIND développé par le laboratoire Lescot de l'Ifsttar. Cet outil développé sous MATLAB est conçu pour l'analyse de l'activité de conduite dans le but de préparer les opérations de prétraitement des données collectées de l'activité de conduite. Il permet, à partir d'un trajet (trip) sélectionné, la visualisation synchronisée des paramètres (valeurs instantanées ou courbes d'évolution temporelle) et des vidéos enregistrées pendant ce trajet. Ces dernières permettent d'observer l'environnement de conduite et les actions du conducteur telles que la position des pieds sur les pédales ou celle des mains sur le volant.

Ajout et sélection des trajets à visualiser

Les fonctions de visualisation des paramètres

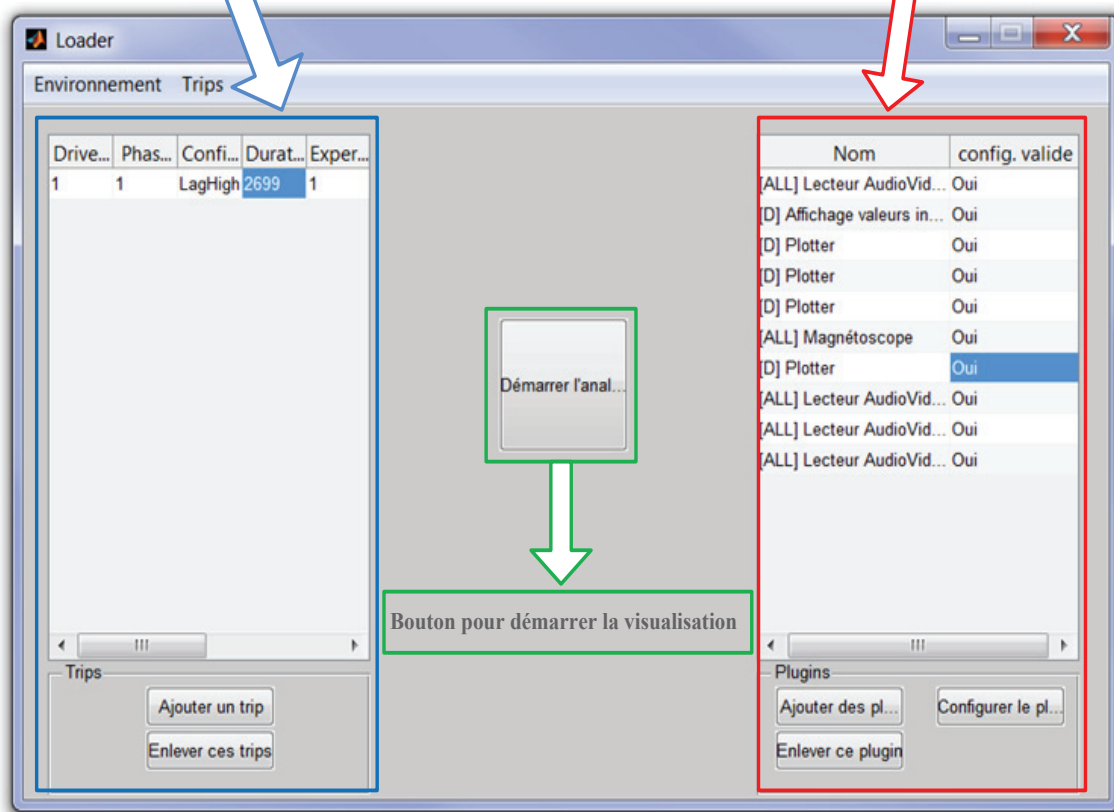


Figure III.8 : Interface de visualisation des données de BIND

L'interface de l'outil BIND est composée de trois parties. La première partie sur la gauche de la Figure III.8 permet d'ajouter les trajets à visualiser. Plusieurs trajets peuvent être ajoutés et la visualisation des trajets est faite un à un. L'analyste doit donc sélectionner un trajet à la fois. La deuxième partie, permet de choisir et de configurer les fonctionnalités de visualisation des paramètres. Les fonctionnalités de visualisation sont par exemple, l'affichage instantané d'un ou plusieurs paramètres, l'affichage en courbe des paramètres en fonction du temps, l'affichage des vidéos et du magnétoscope pour la gestion du temps (avancer ou reculer la lecture des données dans le temps). La configuration consiste à sélectionner les paramètres à visualiser et choisir l'emplacement des affichages sur l'écran de l'analyste. Pour la même visualisation, plusieurs fonctionnalités (plugins) peuvent être utilisées. Pour finir, la troisième partie contient le bouton pour lancer la visualisation avec BIND.

Visualisation sous forme de courbe

Vue d'avant et magnétoscope

Visualisation des paramètres instantanés

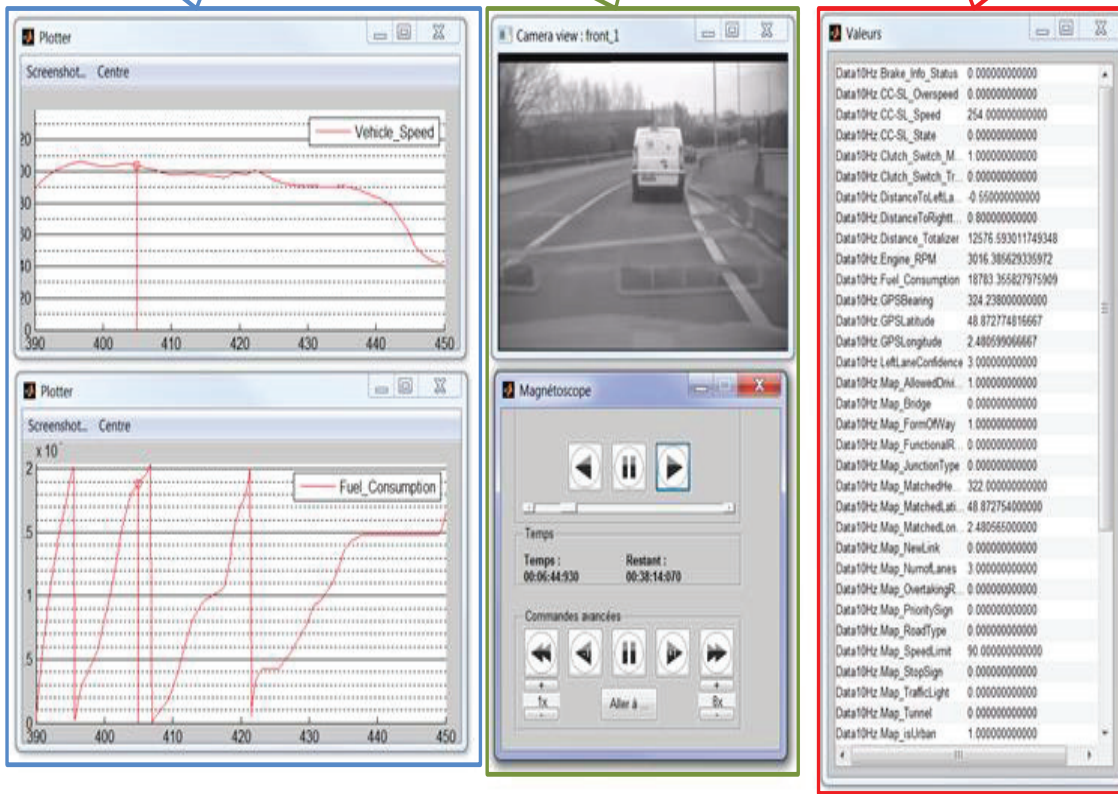
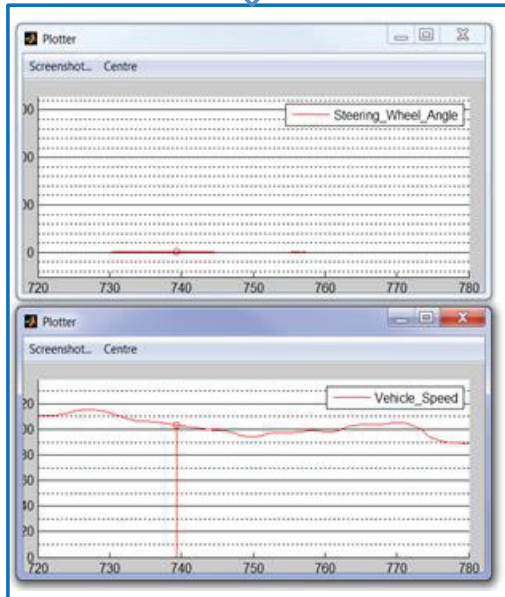


Figure III.9 : Affichage des paramètres collectés avec les différentes fonctionnalités de BIND

Dans la figure ci-dessus, un exemple de visualisation de l'activité de conduite avec l'outil BIND est illustré. Il est constitué des trois types de visualisation : (1) à gauche, la visualisation des paramètres (Vehicle_Speed et Fuel_Consumption) sous forme de courbes en fonction du temps à gauche, (2) au centre, la vidéo de la scène routière devant le véhicule pendant la conduite et en bas le magnétoscope, (3) à droite l'affichage de l'ensemble des valeurs instantanées des paramètres collectés pendant la conduite.

La figure ci-dessous (Figure III.10) illustre la visualisation des vidéos enregistrées pendant l'activité de conduite avec l'outil BIND. Les quatre vues affichées sont : la vue en avant, le visage du conducteur, les pieds du conducteur et la vue des actions de conduite.

Courbes des paramètres en fonction du temps



Les quatre vues enregistrées pendant la conduite



Figure III.10 : Illustration de la visualisation des vidéos de l'activité de conduite avec BIND

BIND permet l'exploration des données relatives à l'activité de conduite grâce aux fonctions de visualisation des trajets enregistrés par des dispositifs installés à bord du véhicule lors de la conduite. La visualisation des données collectées permet à l'analyste de se familiariser avec les données, de préparer la phase de prétraitement des données (création et transformations des paramètres, gestion des valeurs manquantes et suppression des données erronées). Une fois les données explorées avec BIND, les deux opérations nécessaires à la phase de préparation vont être mises en œuvre dans les étapes suivantes sur les données collectées de l'activité de conduite.

2.2.3 Opération de préparation des données de l'activité de conduite

2.2.3.a Nettoyage des données

L'opération de nettoyage dans l'analyse d'une activité permet, de faire la gestion des valeurs manquantes des paramètres et, de vérifier la cohérence des données. Les données de l'activité de conduite étant collectées par des capteurs (bus CAN, GPS, Radar) intégrés au véhicule, il y a souvent des problèmes de couverture de réseau comme dans les tunnels, où les données GPS ne sont plus disponibles. Ces valeurs manquantes peuvent être remplacées par la dernière valeur prise par le paramètre avant l'arrêt des enregistrements pour les paramètres qualitatifs.

Elle permet également de gérer les valeurs extrêmes ou improbables enregistrées par les capteurs. Par exemple, les valeurs des paramètres suivants : GPSBearing, Map_SpeedLimit, souvent respectivement à 900 et 400 ou bien la vitesse du véhicule qui dépasse 150km/h sont supprimées. Ces opérations de nettoyage sont effectuées sur l'ensemble des paramètres enregistrés pendant l'activité de conduite des différents participants.

2.2.3.b Création de nouveaux paramètres (paramètres calculés)

Cette étape permet de créer de nouveaux paramètres calculés à partir des paramètres collectés lors de l'expérimentation. Par exemple, les paramètres tels que le rapport de boîte vitesse, le trafic, le type de route, la consommation en carburant optimisée sont des paramètres calculés car ils sont utiles pour notre analyse. L'exemple de paramètre calculé qui est décrit ici est le rapport de boîte de vitesse. Les capteurs qui équipaient les véhicules ne permettaient pas la collecte de ce paramètre pendant la conduite, mais il est possible de le calculer à partir des paramètres enregistrés suivants : la vitesse, le régime moteur et l'enfoncement de la pédale d'embrayage. Pour cela, plusieurs trajets ont été analysés pour définir la droite d'équation de chacun des rapports de boîte vitesse pendant la conduite. La droite d'équation est décrite par :

$$Y = a * X + b, \text{ dans lequel } X \text{ correspond à la vitesse du véhicule et } Y \text{ à la valeur du régime moteur.}$$

Pour calculer les droites d'équation, pour chaque rapport de boîte de vitesse instantanée, en fonction de la vitesse pendant la conduite, il faut déterminer les valeurs prises par les constantes a_i et b_i pour les droites D_i (i allant de 1 à 6 pour chaque rapport de boîte). Les rapports de boîte de vitesse sont calculés toutes les 100 millisecondes grâce aux paramètres collectés pendant la conduite. Ci-dessous les droites d'équations de D1 à D6 correspondant aux rapports de boîtes de vitesse d'un véhicule Clio de la première vitesse à la sixième.

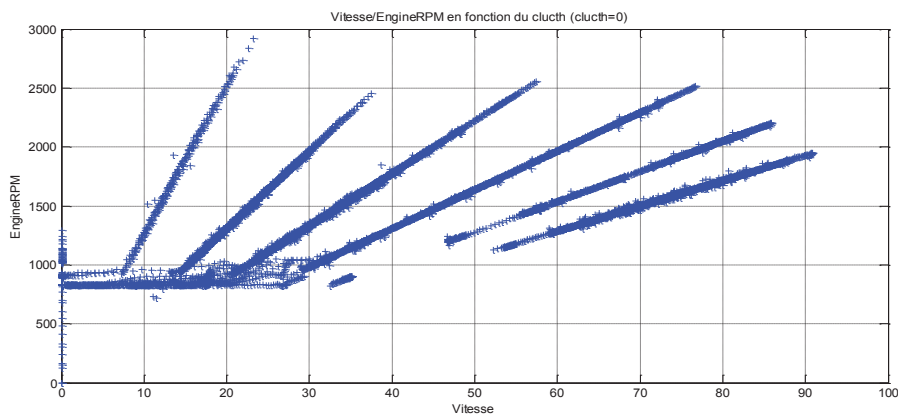


Figure III.11 : Illustration du calcul des rapports de boîte de vitesse d'un véhicule Clio

Grâce au calcul de ce paramètre, les différents rapports de boîte vitesse sont calculés toutes les 100 millisecondes sur l'ensemble des trajets enregistrés pendant l'expérimentation. Cette étape a permis de créer plusieurs paramètres calculés nécessaires à l'analyse de l'activité de conduite. L'ensemble des paramètres calculés dans cette étape sont décrits dans l'annexe 1 (Tableau A.6) et un extrait dans le tableau suivant :

Notation du paramètre	Description du paramètre calculé	Paramètres utilisés pour le calcul
Paramètres calculés initialement		
PC_RBV	les rapports de boîte vitesse instantanés du véhicule.	Vehicle_Speed, Engine_RPM et Clutch_Switch_Travel.
PC_ObjectifConduite	Indique les actions de conduite effectuées par le conducteur comme aller tout droit, tourner à droite, changer de voie.	Vehicle_Speed, Steering_Wheel_Angle, IndSteering_Wheel_Angle, Map_DerivedMatchedHeading
PC_TypeRoute	Type de route (Urbain, rural ou autoroute)	Map_RoadType
PC_TimeToCollision	Valeur du temps à la collision entre le véhicule équipé et le véhicule le précédant.	R_Range, R_RelativeVelocity
Paramètres calculés lors d'itération « Révision d'observés bruts »		
PC_TraficVitesseLimite	Ce paramètre calculé correspond à l'écart entre la vitesse limite et la vitesse du véhicule, plus cet écart est grand plus il y a du trafic. C'est un paramètre calculé avec les modalités [fluide = (écart <20) ; contraint (écart entre 20 et 40) ; bouchon = (écart >40)].	Vehicle_Speed, Map_SpeedLimit
PC_NbFeuTricolorParKm	Nombre de feux tricolores par km.	Map_TrafficLightSig, timeCode, Vehicle_Speed
PC_Mod_ConsoOpt	Consommation en carburant optimisée calculée par le modèle d'optimisation durant la situation de conduite.	Timecode, Vehicle_Speed, R_Range, R_RelativeVelocity, R_Acceleration, MapSpeeLimit, Engine_RPM

Tableau III.1 : Extrait de la liste des paramètres calculés et leurs descriptions

2.2.3.c Reformulation des paramètres collectés

Un exemple de la discrétisation du régime moteur avec la méthode d'amplitude (amplitude est fixée à 1000 pour toutes les classes) est donné dans le tableau suivant :

Nom paramètre	Classe de Discrétisation
Engine_RPM (entre 0 et 4500)	[0, 1500]→1500 ; [1500, 2500]→2500 ; [2500, 3500]→3500; [3500, +]→[→4500 ;
Age_conducteur	[18, 25]→1; [26, 55]→2; [56, 75]→3; [76, +]→[→4
Experience_conduite	[0, 5 ans [→ 1, [5, 10 ans [→ 2 et [10,+]→[→3

Tableau III.2 : Discrétisation du régime moteur avec la méthode d'amplitude

2.2.3.d Création de paramètres agrégés ou indicateurs sur les séquences de conduite

Comme exemple d'indicateurs sur les séquences de conduite, il y a la moyenne de la vitesse, la vitesse maximale ou minimum durant une situation de conduite. De même, avec le paramètre calculé PC_RBV (Rapport de boîte de vitesse instantané), des indicateurs sur des situations de conduite sont créés par les fonctions de d'agrégation définies dans le tableau (Tableau III.3). Par exemple, les fréquences d'utilisation des rapports de boîte de vitesse pendant une séquence de conduite ou le rapport de boîte de vitesse le plus utilisé durant une séquence de conduite.

Description des indicateurs décrivant les situations de conduite	Notation
Paramètres calculés lors d'itération « Révision d'indicateurs calculés »	
Objectif de conduite pendant la situation	IND_ObjectifConduite
Type d'infrastructure pendant la situation	IND_Infrastructure
Durée de la situation de conduite calculée en faisant la différence entre le temps de début et fin de la situation de conduite.	IND_Duree
Valeurs moyenne, maximale minimale, amplitude et écart-type de la vitesse pendant la situation de conduite	IND_VitesseMoy, IND_VitesseMax, IND_VitesseMin, IND_VitesseEcartType IND_VitesseAmplitude
Valeurs moyenne, maximale, minimale et écart-type de la dérivée de la vitesse pendant la situation de conduite	IND_DeriveVitesseMoy, IND_DeriveVitesseMax, IND_DeriveVitesseMin, IND_DeriveVitesseEcartType
Valeurs moyenne, maximale minimale et écart-type de l'angle volant pendant la situation de conduite	IND_AngleVolantMoy, IND_AngleVolantMax, IND_AngleVolantMin, IND_AngleVolantEcartType
Pourcentage d'utilisation de la pédale de frein durant la situation de conduite	IND_%Freinage
Valeur minimum du temps à la collision durant la situation de conduite	IND_TTCmin
Valeur minimum du temps à la collision dynamique durant	IND_TTCDynamique

la situation de conduite	
Pourcentage de temps durant lequel les 6 rapports de boîte vitesse sont utilisés durant la situation de conduite ou	IND_%RBV1, IND_%RBV2, IND_%RBV3, IND_%RBV4, IND_%RBV5, IND_%RBV6
Pourcentage de temps durant lequel les 6 rapports de boîte vitesse calculé par le modèle d'optimisation sont utilisés durant la situation de conduite	IND_%RBVMod1, IND_%RBVMod2, IND_%RBVMod3, IND_%RBVMod4, IND_%RBVMod5, IND_%RBVMod6
Rapport de boîte de vitesse le plus utilisé pendant la situation de conduite.	IND_MoreUseRBV

Tableau III.3 : Tableau des indicateurs calculés sur les situations de conduite

Un exemple de la distribution d'un indicateur calculé sur les situations de conduite est donné dans la Figure III.12. Cet indicateur correspond à la moyenne de la vitesse sur l'ensemble des situations de conduite.

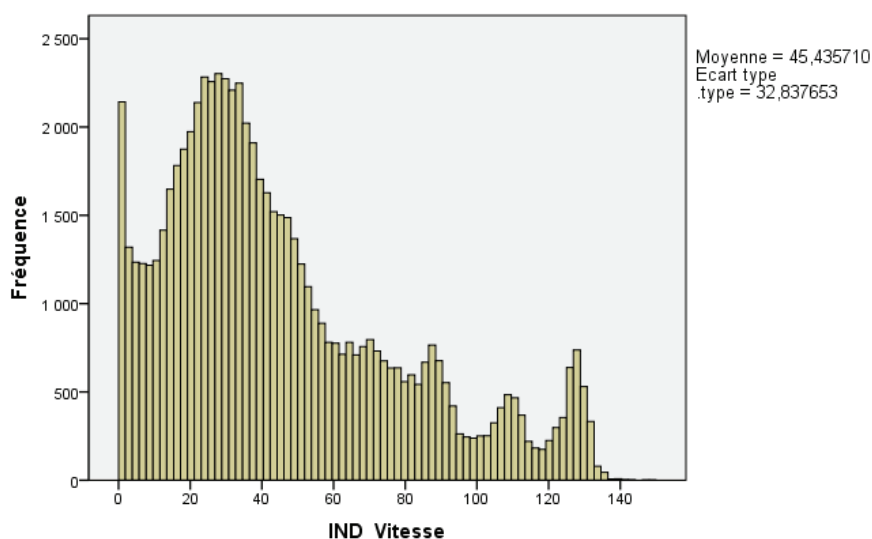


Figure III.12 : Répartition de la moyenne de la vitesse sur l'ensemble des situations de conduite

2.2.4 Opération de contextualisation des données de l'activité de conduite

Les entités impliquées dans l'activité de conduite sont : le *conducteur*, le *véhicule*, et l'*environnement de conduite*.

2.2.4.a Les paramètres de l'entité « *Observateur /Analyste* »

L'objectif de cette étude est d'expliquer la consommation en carburant de l'activité de conduite en tenant compte de l'environnement de conduite. Le critère observé étant la consommation en carburant, deux indicateurs de ce critère ont été mis en place : la

consommation réelle en carburant et l'écart de consommation avec une consommation optimisée. La consommation en carburant réelle est celle enregistrée pendant un trajet. L'écart de consommation en carburant est la différence des consommations calculées entre deux profils de vitesse. Le premier profil correspond à celui réalisé par le conducteur. Le deuxième profil est obtenu grâce à une méthode d'optimisation de la consommation en tenant compte de la sécurité et des contraintes liées aux trafics. Cette méthode a été définie par Mensing et al. dans [100], elle repose sur un algorithme de programmation dynamique pour calculer un profil de vitesse optimisé. Ce profil permet de parcourir la même distance avec la même durée, en tenant compte des contraintes liées à l'environnement de conduite telles que la vitesse limite et le trafic via la distance inter véhiculaire. La consommation optimisée du trajet est alors déduite du profil de vitesse optimisé calculé par l'algorithme de programmation dynamique.

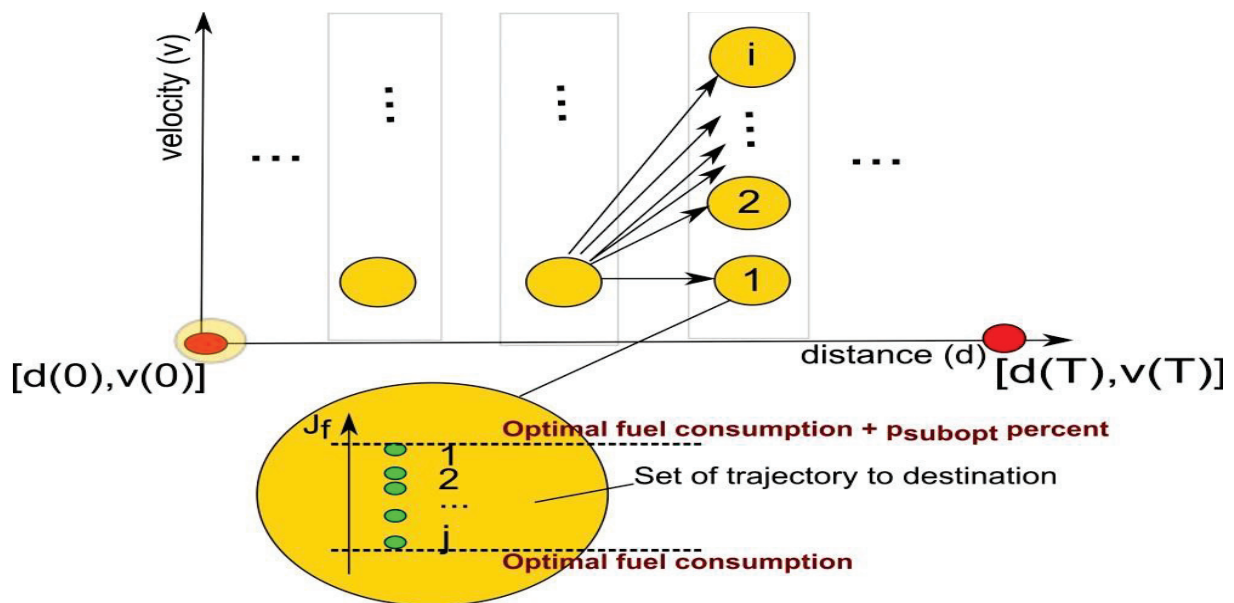


Figure III.13 : Illustration du calcul de la consommation en carburant à partir d'un profil de vitesse optimisé de Mensing et al., extrait de la thèse de [101]

L'algorithme de programmation dynamique calcule l'ensemble des profils de vitesse possibles sur le trajet. A chaque profil de vitesse optimisé est associé à une consommation en carburant optimisée. Parmi les consommations en carburant possibles, le choix de la consommation optimisée est fait en utilisant une méthode de coût calculée à partir des contraintes liées au rendement du moteur mais aussi à la sécurité et au trafic. Par exemple, plus la distance inter véhiculaire est faible, plus le coût affecté à la consommation en carburant calculé par l'algorithme de programmation dynamique est élevé pour éviter la collision avec le véhicule de devant [101]. De même, les contraintes induites par les vitesses

maximales autorisées induisent un coût très élevé si la vitesse calculée est supérieure à celle autorisée. La figure ci-dessous donne le profil de vitesse normal d'un trajet et celui du profil optimisé, ainsi que les rapports de boîte de vitesse normaux et optimisés.

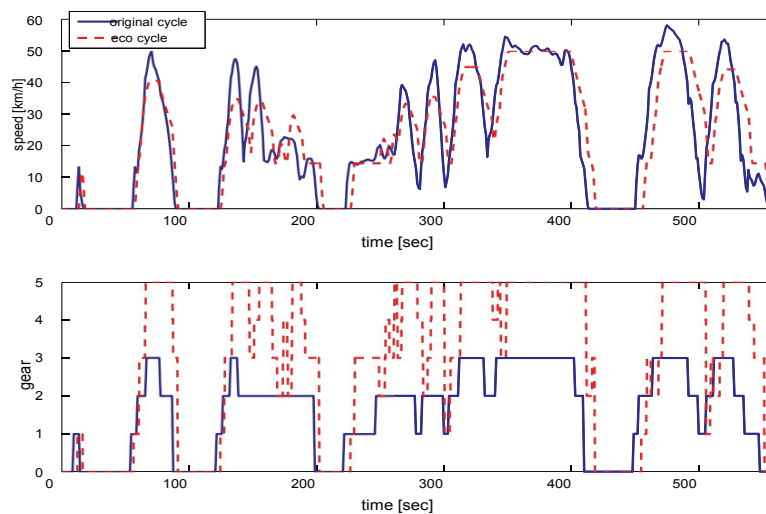


Figure III.14 : Figures des profils de vitesse et de rapport de boîte de vitesse calculé avec la méthode d'optimisation de Mensing et al., extrait de [100]

Une deuxième version de l'algorithme a été développée dans [102] pour introduire des contraintes liées à la pollution. En effet, la première version avait tendance à privilégier des profils de vitesse induisant une forte baisse de consommation mais augmentait les émissions de particules dangereuses. Pour estimer l'impact de ces méthodes d'optimisation du profil vitesse, Mensing et al., dans [102] ont comparé la consommation en carburant et les émissions de GES de trois cycles de conduite (cycle original, cycle économique de la version 1 et le cycle écologique de la version 2). Les résultats montrent que les cycles économique et écologique consomment moins de carburant que le cycle original (presque 1/3 de réduction) et les consommations en carburant des cycles économique et écologique sont sensiblement équivalents (respectivement 6.5 et 6.7 en l/100km) contre 9 l/100km pour le cycle original. Concernant les émissions, il y a une grande disparité entre les trois cycles. Le cycle original est celui qui produit le plus de CO_2 et de NO_x . Le cycle économique diminue de plus de moitié les émissions en CO_2 , mais double celle de CO et HC et diminue légèrement celle de NO_x par rapport au cycle original. Par contre, le cycle écologique avec presque la même consommation en carburant que le cycle économique, réduit considérablement l'émission des particules fines comme NO_x , CO et HC avec une légère augmentation d'émission de CO_2 par rapport au cycle économique. De plus, il permet de bien réduire la consommation et l'émission de NO_x , tout en n'augmentant pas ou très peu l'émission de CO et HC.

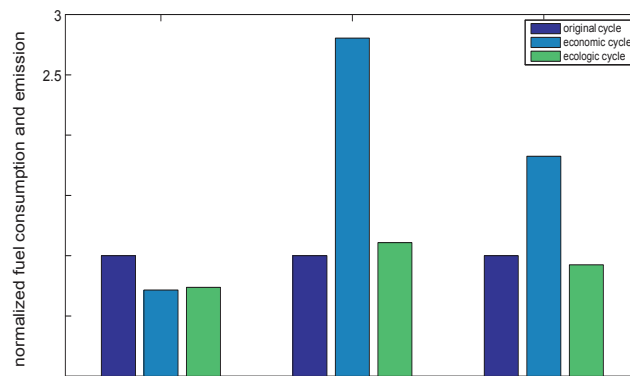


Fig. 10. Fuel consumption and emission results.

Table 3
Emission measurement final results.

Emission in g/km	CO ₂	CO	NO _x	HC	Fuel consumption [l/100 km]
Original cycle	206.96	2.06	0.0055	0.068	9.0
Economic cycle	140.96	5.78	0.0046	0.12	6.5
Ecologic cycle	151.51	2.18	0.0025	0.063	6.7
Euro IV norm (NEDC)	–	1	0.08	0.1	

Figure III.15 : Résultats de la méthode d'optimisation du profil vitesse de Mensing et al., extrait de [102]

Cette méthode qui permet de déduire la consommation en carburant optimisée à partir du profil de vitesse optimisé est choisie dans cette thèse pour plusieurs raisons. La première raison est qu'elle est économique car elle permet de réduire la consommation en carburant (1/3 de réduction de consommation en carburant en l/100km), mais aussi écologique car certaines émissions des gaz à effet de serre (particules fines) comme le NO_x [102] sont également réduites.

En résumé, l'entité « *Observateur / Analyste* » est caractérisée par les deux critères : la consommation en carburant réelle (*ConsoReelle*) et l'écart de consommation entre les profils de vitesse réel et optimisé (*EcartConsoOpt*). Des indicateurs associés à ces paramètres décrivant la consommation en carburant réelle (*CA_ConsoReellePour100km*) et optimisée (*CA_Mod_EcartConsoOpt*) sont alors calculés pour décrire le contexte « *Observateur* ». Dans cette analyse, nous n'avons qu'une seule contrainte : la méthode de découverte de connaissances contextuelles ne peut être appliquée que sur des sous-parties homogènes de l'activité car le contexte est évolutif et dynamique. Pour notre analyse, le choix a été fait de travailler sur la sous tâche « *AllerToutDroit* » de l'activité de conduite.

Description des paramètres collectés de l'entité « Analyste »	Notation (CA= Contexte Analyste)
Paramètre initial	
Consommation au 100 km enregistrée	CA_ConsoReellePour100km
Paramètres calculés lors d'itération « Révision d'observés bruts »	
La différence entre les consommations en carburant réelle et optimisée calculée par le modèle d'optimisation.	CA_Mod_EcartConsoOpt

Tableau III.4 : Liste de paramètres contextuels de l'entité « Analyste »

2.2.4.b Les paramètres de l'entité « Conducteur »

L'entité *Conducteur* est caractérisée par les paramètres contextuels suivants : l'âge, l'expérience de conduite, le nombre de kilomètre parcourus, le style de conduite, les attitudes au volant (respect des limitations de vitesse, respect de la distance inter-véhicule, habitude de conduite sur autoroute, en urbain et en campagne, conduite en endroit familier), et également les informations relatives à la santé du conducteur comme la prise de médicaments, le port des lunettes de vue. Ces informations ont été collectées grâce aux questionnaires mis en place lors de l'expérimentation pour collecter des informations supplémentaires sur les participants. Plusieurs paramètres ont été collectés pour décrire les attitudes des participants pendant la conduite. Une attitude correspond à une opinion du conducteur vis-à-vis d'un comportement. Par exemple, à la question « Quelle importance accordez-vous aux comportements « Dépasser la limitation de vitesse », le participant devra entourer une valeur dans une échelle de 1 à 10.

« très mineure 1 2 3 4 5 6 7 8 9 10 très sérieuse »

Figure III.16 : Illustration des questionnaires sur les attitudes pendant la conduite.

Les paramètres décrivant l'entité « *Conducteur* » sont décrits dans le tableau suivant :

Description des paramètres collectés de l'entité « Conducteur »	Notation (CC= Contexte Conducteur)
Paramètres initiaux	
Age du conducteur (numérique)	CC_Age
Genre (binaire : homme ou femme)	CC_Genre
Nombre d'années de permis (numérique)	CC_NbPermis
Nombre de kilomètres parcourus (numérique)	CC_NbKilometre
Port de verres de correction visuelle (binaire)	CC_CorrectionVisuelle
Problème de perception des couleurs (binaire)	CC_Daltonien
Problème auditif (binaire)	CC_ProblemeAudi
Déjà fait des accidents routiers (binaire)	CC_AccidentRoutier
Prise quotidienne de médicament (binaire)	CC_UtiliseMedicament
Habitué à rouler sur autoroute (échelle)	CC_UtiliseAutoroute
Habitué à rouler en rural (échelle)	CC_UtiliseRural
Habitué à rouler en urbain (échelle)	CC_UtiliseUrbain

Habitué à rouler sur des routes familières (échelle)	CC_RouteFamiliere
Habitué à rouler sur des routes non familières (échelle)	CC_RouteNonFamiliere
Attitude par rapport à la vitesse limite (échelle)	CC_AttVitesseLimite
Attitude par rapport à la distance de sécurité inter véhiculaire (échelle)	CC_AttDistanceSecurite
Conduite avec une vitesse souple (échelle)	CC_AttVitesseSouple
Attitude par rapport à regarder l'angle mort avant virage (échelle)	CC_AttRegardAngleMort
Attitude de sortie de voie (échelle)	CC_AttSortirVoie
Préparation d'itinéraire (échelle)	CC_AttPrepaItineraire
Fatigue au volant (échelle)	CC_AttFatigueAuVolant
Attitude par rapport au rapport boîte de vitesse (échelle)	CC_AttRapportdeBoite
Attitude par rapport à faire des accélérations rapides (échelle)	CC_AttAccelerationRapide
Attitude par rapport à prendre de la vitesse aux virages (échelle)	CC_AttVitesseVirage
Attitude par rapport à conduire sous l'effet de l'alcool (échelle)	CC_AttAlcoolVolant
Attitude par rapport à téléphoner au volant (échelle)	CC_AttTelephone
Paramètres calculés lors d'itérations « Révision d'observés bruts »	
Indique le style de conduite du conducteur (nominale souple, agressive, très agressive) :	CC_StyleConduite

Tableau III.5 : Liste de paramètres contextuels de l'entité « Conducteur »

Ces paramètres permettent de caractériser chaque participant de l'expérimentation, ils seront donc ajoutés aux autres paramètres collectés pendant la conduite des trajets de chaque participant.

2.2.4.c Les paramètres de l'entité « Véhicule »

L'entité véhicule est caractérisée par les paramètres contextuels suivant: le type du véhicule, le type de carburant, l'état du véhicule. Dans l'expérimentation, deux types de véhicule ont été utilisés, la Clio et la Laguna de la marque Renault. Dans cette thèse, seuls les trajets effectués avec la Clio Renault sont utilisés car la méthode d'optimisation du profil vitesse à partir duquel la consommation optimisée est calculée n'était pas disponible pour la Laguna. Ceci implique que l'entité véhicule est caractérisé par un seul type de véhicule avec les mêmes caractéristiques qui sont définies dans le tableau suivant (Tableau V.8). En plus de cette caractéristique, les paramètres d'utilisation des systèmes d'assistance à la conduite peuvent être pris en compte comme l'utilisation du régulateur et du limiteur de vitesse pendant la conduite.

Description des paramètres collectés de l'entité « Véhicule »	Notation (CV= Contexte Véhicule)
Marque du véhicule	CV_Marque = «Renault »
Type du véhicule	CV_Type = «Clio »
Type de carburant	CV_Carburant= « Diesel ou essence »
Utilisation du régulateur de vitesse (binaire)	CV_Utilise_CC
Utilisation du limiteur de vitesse (binaire)	CV_Utilise_SL

Tableau III.6 : Liste de paramètres contextuels de l'entité « Véhicule »

2.2.4.d Les paramètres de l'entité « Environnement de conduite »

L'entité environnement de conduite est caractérisée par toutes les informations pouvant décrire la route tels que le type de route (autoroute, urbain et campagne), le trafic, le type d'infrastructure, la météo et le nombre de voies. Ce sont des paramètres collectés ou calculés qui sont décrits dans le tableau suivant :

Description des paramètres contextuels de l'entité « Environnement de conduite »	Notation (CEC= Contexte Environnement de Conduite)
Paramètres calculés initiaux	
Paramètre indiquant le moment de la journée*	CEC_DayLight
La valeur de la vitesse limite de la route la plus fréquente sur la situation **	CEC_MapSpeedLimit
La moyenne du nombre de voies de la route **	CEC_MapNumofLanes
Le type de jonction fourni par le GPS*	CEC_MapJunctionType
Temps à la collision utilisant aussi l'accélération**	CEC_DynTimeToCollision
Distance de collision avec le véhicule précédent**	CEC_TimeToCollision
Le type de route sur la situation courante*	CEC_TypeRoute
Type d'infrastructure sur la situation courante*	CEC_TypeInfrastructure
Objectif de conduite sur la situation courante : Indique les actions de conduite effectuées par le conducteur*	CEC_ObjectifConduite
Paramètres calculés lors des itérations « Révision d'observés bruts »	
Le trafic routier avec la vitesse limite et la vitesse du véhicule (Fluide, Contraint, Bouchon) *	CEC_TraficVitesseLimite
Nombre de panneaux stop par km **	CEC_NbStopParKm
Nombre de panneaux priorité par km**	CEC_NbPrioriteParKm
Nombre de feux de signalisation par km sans arrêt**	CEC_NbFeuSigParKmSansArret
Nombre de feux de signalisation par km avec arrêt**	CEC_NbFeuSigParKmAvecArret
Indique le trafic autour du véhicule instrumenté dans une zone de 50m, 100m ou de 300m sur une période de 2 secondes, 1 minute ou 5 minutes. Ce sont des paramètres calculés renommés**	CEC_FD50m1s, CEC_FD50m30s, CEC_FD50m150s, CEC_FD50m300s, CEC_FD100m1s, CEC_FD100m30s, CEC_FD100m150s, CEC_FD100m300s CEC_FD300m1s, CEC_FD300m30s, CEC_FD300m150s, CEC_FD300m300s
Objectif de conduite effectué par le conducteur sur la situation précédente*	CEC_ObjectifConduiteAprès
Objectif de conduite effectué par le conducteur sur la situation suivante*	CEC_ObjectifConduiteAprès
Type d'infrastructure suivante* (par exemple, Ligne droite, virage, intersection)	CEC_infrastructureAprès
Type d'infrastructure précédente* (Ligne droite, virage, intersection)	CEC_infrastructureAvant

* donnée brute qui ne varie pas durant la situation de conduite

** paramètre agrégé comme indicateur sur la situation

Tableau III.7 : Liste des paramètres contextuels de l'entité « Environnement de Conduite »

2.3 Construction de la base de connaissances

Afin de préparer l'étape d'identification des éléments contextuels explicatifs du critère observé, une base de connaissance est créée et sera exploitée pour la mise en place des transformations nécessaires à la découverte de connaissances contextuelles. Dans cette étape, les méthodes classiques de fouilles de données sont mobilisées comme outils d'analyse dans le processus de découverte interactive de connaissances pour faciliter l'observation de l'activité. Dans notre cas, nous avons utilisé la méthode de classification supervisée « Arbre de décision » de SPSS pour déterminer les règles de segmentation et de catégorisation des paramètres observés de l'activité. Dans le cas de cette analyse, deux opérations ont été effectuées pour préparer la phase d'assistance à la découverte de connaissances à partir des M-Traces. Ces opérations ont été réalisées sur la base de données prétraitées contenant 1258 trajets collectés en situation de conduite naturelle avec un véhicule Clio de la marque Renault. La première opération sur les trajets de la base de données prétraitée est la segmentation de l'activité de conduite en situation de conduite, et la deuxième opération est de catégoriser ces situations en fonction du critère observé qui est la consommation en carburant. La base de connaissance sera donc constituée de l'ensemble de règles permettant la segmentation et la catégorisation de ces séquences. Ces opérations sont détaillées dans les sections suivantes :

2.3.1 Segmentation de l'activité de conduite en sous-parties (situations de conduite) :

L'objectif de l'application de la méthode de découverte de connaissances contextuelles dans l'activité de conduite est d'identifier le contexte explicatif de la consommation en carburant. Le contexte étant spécifique à la tâche en cours (sous-partie de l'activité globale), il est important de travailler sur des segments homogènes de l'activité pour identifier quelles sont les entités impliquées dans l'activité dont les variations peuvent expliquer celle de la consommation en carburant. L'activité de conduite est segmentée en plusieurs situations de conduite comme par exemple les situations : aller tout droit, virage, changement de voie. La segmentation de l'activité de conduite permet de simplifier sa complexité lors de l'analyse car elle est représentée comme une succession de sous-parties (situations de conduite) qui seront ensuite catégorisées en fonction de la consommation en carburant. Sur les 1258 trajets de la base de données prétraitées 70h de conduite ont été codées manuellement. Le codage de l'activité de conduite consiste à visualiser les vidéos de conduite pour identifier et coder les objectifs ou actions de conduite du conducteur (aller tout droit, virage, tourner à droite et tourner à gauche) et l'environnement de conduite tels que le trafic, le type d'infrastructure, les événements et les incidents pendant la conduite. Ces trajets codés ont été analysés avec les méthodes d'apprentissage automatique (apprentissage supervisé) pour découvrir les règles de

segmentation de l'activité de conduite. Ces règles ont été utilisées pour faire la segmentation automatique de l'activité de conduite en situation de conduite sur le reste des trajets de la base de données non codée manuellement.

Dans la Figure III.17 ci-dessous, on ne s'intéresse qu'aux trois premières modalités : *AllerToutDroit*, *TournerADroite*, *TournerAGauche*. Une classification du paramètre objectif de conduite est effectuée pour identifier, à partir de l'ensemble des paramètres enregistrés décrivant les situations de conduite, les règles permettant de différencier les différentes modalités de ce paramètre. La méthode des arbres de décision est utilisée pour effectuer l'apprentissage automatique de la segmentation de l'activité de conduite. Cette méthode a été choisie pour sa rapidité d'exécution aussi bien dans la phase de construction de l'arbre que lors de la classification de nouveaux cas. De plus, les règles fournies par cette méthode sont facilement interprétables et permettent une meilleure compréhension du problème étudié. L'arbre de décision obtenu pour la classification du paramètre « ObjectifDriving » est le suivant :

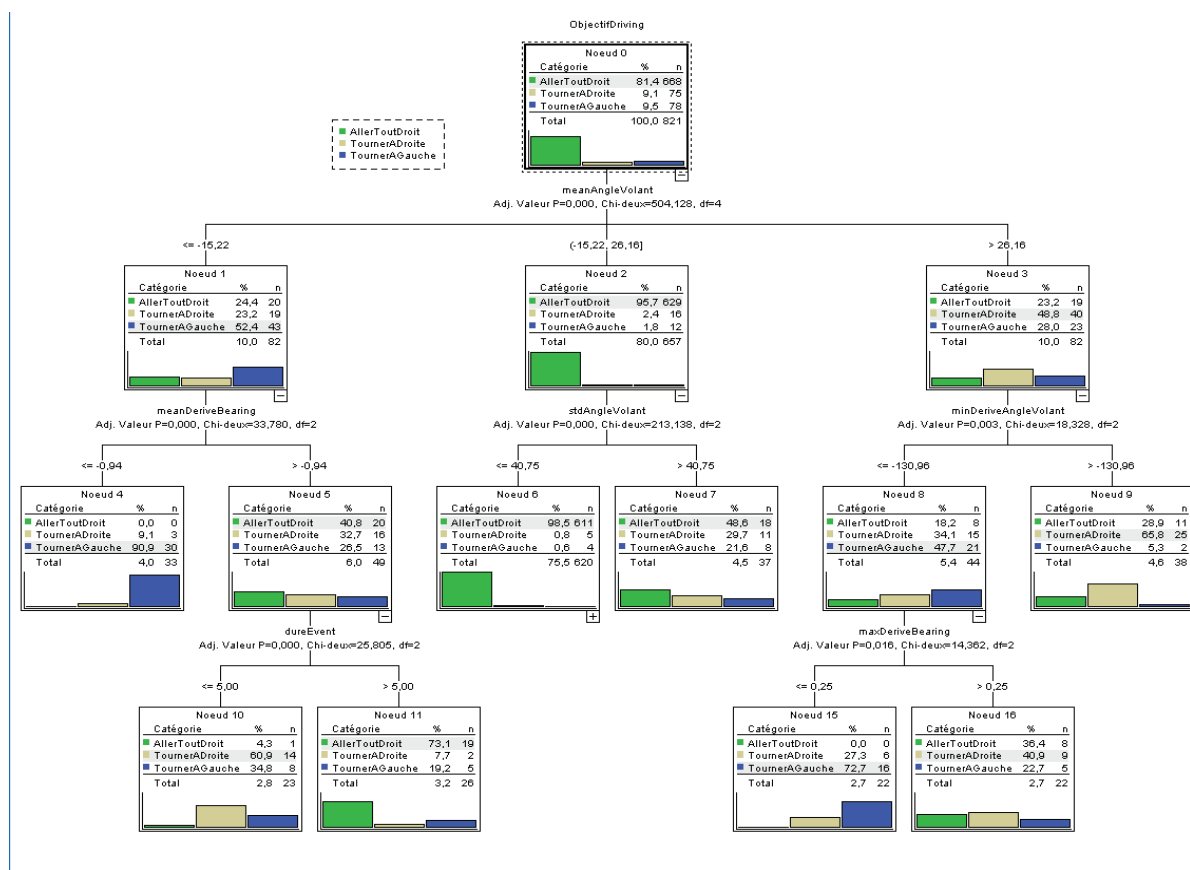


Figure III.17 : Illustration de l'arbre de décision pour la segmentation de l'activité de conduite

Avec cet arbre de décision les situations de conduite « *AllerToutDroit*, *TournerADroite*, *TournerAGauche* » sont identifiées automatiquement sur les trajets non codés. Par exemple, la

règle de segmentation suivante permet d'identifier les situations de conduite «*AllerToutDroit*» sur un trajet de l'activité de conduite :

$$R_{Noeud6} = \text{Si } -15 < \text{MeanAngleVolant} < 26 \text{ Et } \text{stdDeriveeBearing} \leq 40$$

→ **ObjectDriving** = « *AllerToutDroit* ».

Cette règle permet d'identifier les situations de conduite «*AllerToutDroit*» avec les paramètres angle du volant et l'écart type de la dérivée du cap du véhicule (angle entre l'axe longitudinal du véhicule et le nord) avec un taux de 98,5 % de situations bien classées.

Ces règles de segmentation sont validées par l'analyste pour ensuite être utilisées sur l'ensemble des trajets de la base de données. Tous les trajets sont alors segmentés en situations de conduite qui sont : « *TournerAGauche*, *TournerADroite*, *AllerToutDroit* ». La segmentation des situations de conduite a été faite avec un critère supplémentaire sur la durée des situations qui est fixée au maximum à 60 secondes de conduite. En tout, 13146 situations de conduite sont créées et leur répartition est donnée dans la figure suivante :

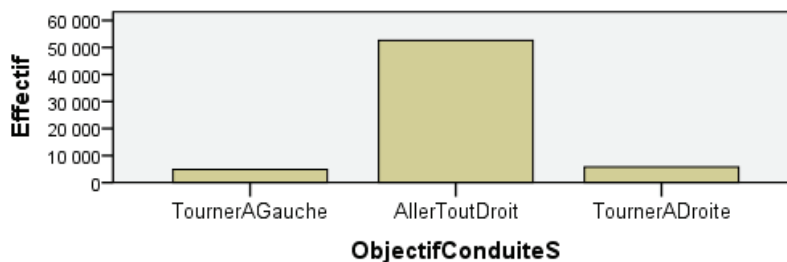


Figure III.18 : Répartition des situations de conduite après la segmentation

2.3.2 Catégorisation du critère observé des situations de l'activité de conduite :

Dans cette partie, l'objectif est de définir et préparer le critère observé. Dans cette analyse, notre critère observé de l'activité de conduite est la consommation en carburant. Pour cela, l'activité de conduite a été segmentée en plusieurs situations de conduite qui représentent des sous-parties ou tâches de l'activité de conduite. A partir de ces situations, on s'intéresse aux indicateurs de consommation en carburant pour faire leur catégorisation en fonction de la consommation en carburant. Au total, trois indicateurs de consommation en carburant sont calculés sur les situations de conduite. Le premier indicateur est la consommation en carburant réelle collectée pendant l'activité (*IND_ConsoReelle*) et il s'exprime en l/100km. La répartition de cette consommation en carburant des situations de conduite est donnée par la figure suivante :

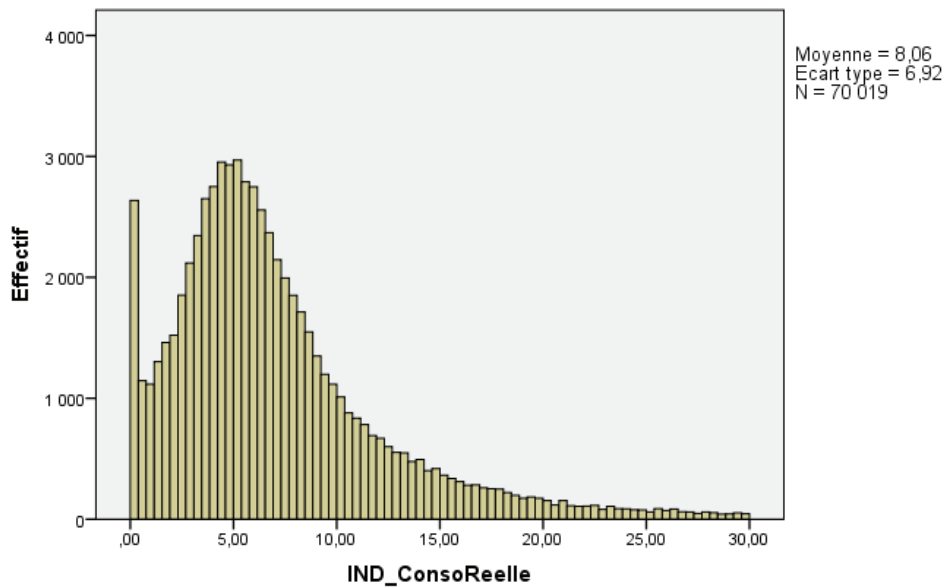


Figure III.19 : Répartition de la consommation en carburant réelle en l/100km des situations de conduite

Le deuxième et troisième indicateurs de consommation en carburant sont fournis par le modèle d'optimisation de Mensing et al., [45]. Ce modèle d'optimisation calcule un profil vitesse optimisé sur le trajet qui permet de couvrir la même distance dans le même temps mais en consommant moins. Puis, le modèle de simulation du moteur calcule une consommation en carburant (*IND_Mod_ConsoReelle*) à partir du profil vitesse réel du conducteur pendant le trajet et une consommation optimisée (*IND_Mod_ConsoOpt*) à partir du profil de vitesse optimisé.

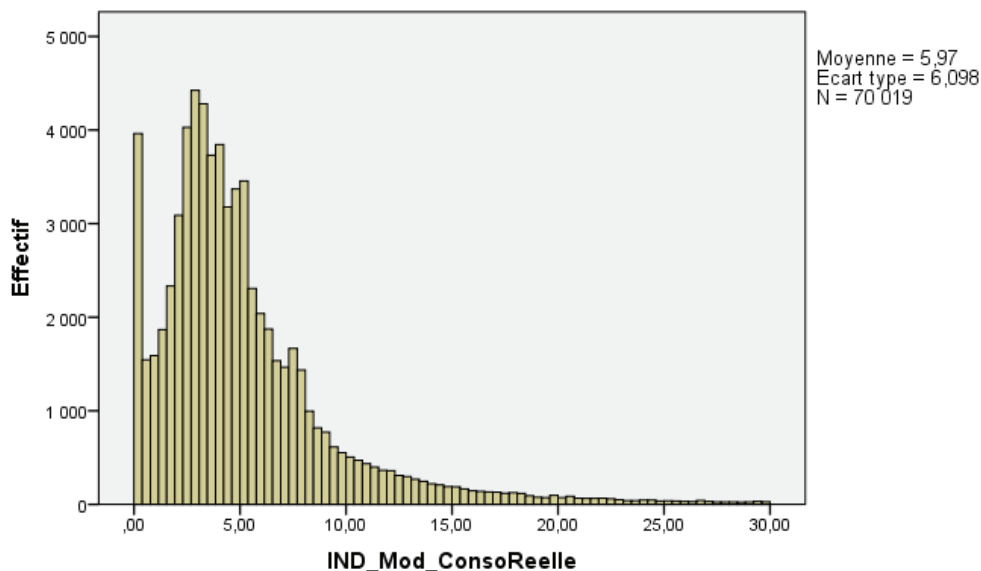


Figure III.20 : Répartition de la consommation en carburant réelle calculé par le modèle d'optimisation de Mensing et al., [45] en l/100km

En observant la courbe de la consommation en carburant réelle collectée (**Figure III.19**) par les capteurs du véhicule et la consommation en carburant réelle (**Figure III.20**) calculée par le

modèle d'optimisation à partir du profil vitesse du conducteur, nous voyons que les deux courbes ont la même allure mais des valeurs plus basses pour le modèle. Ceci est sans doute due au fait que le modèle ne prend pas en compte les pentes et la température du moteur.

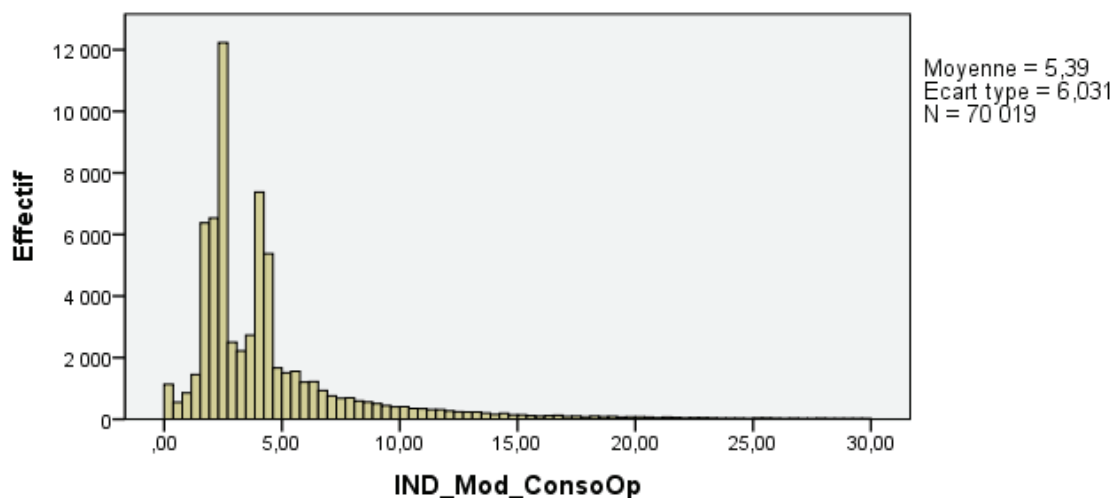


Figure III.21 : Répartition de la consommation en carburant optimisée calculée par le modèle d'optimisation de Mensing et al., [45] en l/100km

A partir de ces indicateurs de consommation en carburant calculés par le modèle d'optimisation, un indicateur de gain de consommation en carburant est calculé sur l'ensemble des situations de conduite. Cet indicateur de gain de consommation est égal à l'écart de consommation en carburant entre la consommation en carburant réelle du modèle d'optimisation et la consommation en carburant optimisée du modèle. Il permet d'identifier les situations de conduite qui sont en sous ou surconsommation par rapport à la consommation en carburant optimale du modèle d'optimisation. Il s'exprime également en l/100km et il est défini dans la figure suivante :

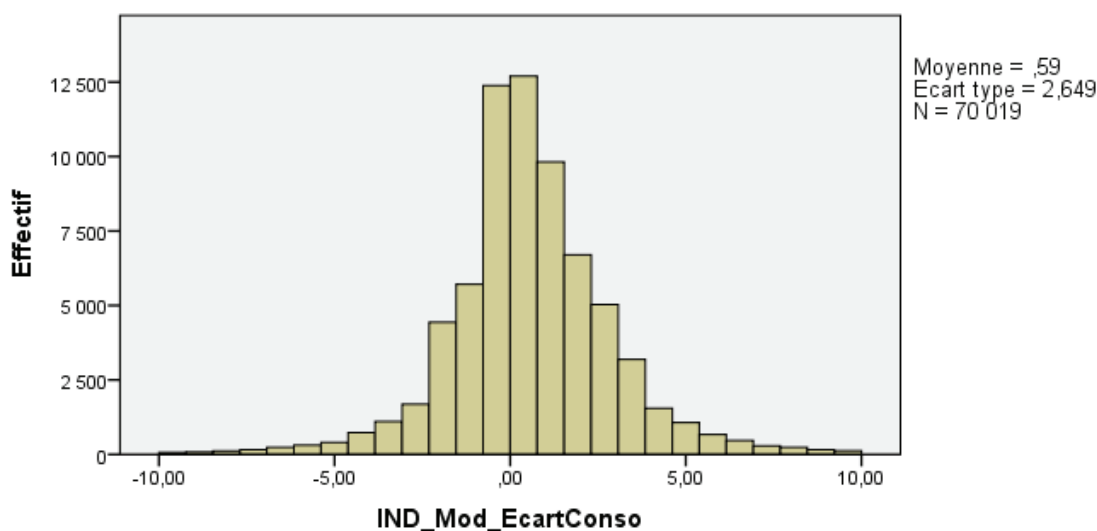


Figure III.22 : Répartition de l'écart entre les consommations en carburant réelle et optimisée du modèle d'optimisation de Mensing et al., [45]

A partir de ces indicateurs de consommation en carburant sur les différentes situations de conduite, deux critères de consommation vont être calculés pour catégoriser ces situations. Le premier critère observé décrit la consommation en carburant réelle collectée pendant l'activité (*CA_ConsoReelle*) et le second critère décrit l'écart entre les consommations en carburant calculées par le modèle de consommation en carburant sur des profils de vitesse réel et optimisé (*CA_Mod_EcartConsoOpt*). La catégorisation du premier critère est effectuée à partir de l'indicateur *IND_ConsoReelle* qui est un paramètre numérique. Il est donc nécessaire de faire une discrétisation pour obtenir les différentes classes de consommation en carburant. Après avoir essayé et analysé plusieurs méthodes de discrétisation, celle des quantiles a été retenue. Cette méthode consiste à faire une équipartition des éléments dans les différentes classes. Puis, le choix a été fait de faire une répartition permettant de se concentrer sur les deux extrêmes de la population (les situations de conduite qui ont une consommation faible ou forte en carburant). Pour cela, trois catégories de consommation en carburant ont été créées de la manière suivante en utilisant le tableau des percentiles de l'indicateur *IND_ConsoReelle* (voir Annexe 2 : Tableau A.7) :

Catégorie ou classe de consommation en carburant <i>CA_ConsoReelle</i>	Segmentation de l'indicateur <i>IND_ConsoReelle en l/100km</i>	Pourcentage d'individus
Faible	$IND_ConsoReelle \leq 3,76$	25% des situations
Moyenne	$3,76 < IND_ConsoReelle \leq 9,78$	50% des situations
Forte	$IND_ConsoReelle > 9,78$	25% des situations

Tableau III.8 : Catégorisation de la consommation en carburant des situations de conduite

Le second critère observé décrit le potentiel gain de consommation en carburant obtenu avec le modèle de d'optimisation de la consommation en carburant à partir d'un profil de vitesse optimisé de Mensing et al., [45]. Ce critère est mise en place avec l'indicateur *IND_Mod_EcartConsoOpt*. A partir du tableau des percentiles de cet indicateur (voir annexe 2 : Tableau A.8), l'objectif est de créer quatre catégories de situation de conduite avec la méthode de discrétisation. La classe « Sous-consommation » correspond aux situations de conduite dont la consommation en carburant réelle calculée est plus petite que la consommation en carburant optimisée proposée par le modèle d'optimisation. Il y a donc une sous-consommation en carburant par rapport à l'optimisation mise en place. La classe « Conso0 » correspond aux situations de conduite où la consommation réelle et optimisée du

modèle sont à peu près égales avec au maximum un gain de consommation entre (-0,55 et 0,46 l/100km donc proche de 0). Les deux classes de surconsommation « SurConso+0.5L et SurConso+1.5L » correspondent aux situations de conduite sur lesquelles il y a un gain de consommation potentiel de plus de 0,5 l/100km pour la première classe et la seconde avec gain de plus 1,77 l/100km. La répartition équitable des classes est donnée dans le tableau suivant :

Classe de gain de consommation en l/100km <i>CA_Mod_EcartConsoOpt</i>	Segmentation de l'indicateur <i>IND_Mod_EcartConsoOpt en l/100km</i>	Pourcentage d'individus
SousConso	$IND_Mod_EcartConsoOpt \leq -0,54$	25% des situations
Conso0	$-0,54 < IND_Mod_EcartConsoOpt \leq 0,46$	25% des situations
SurConso+0.5L	$0,46 < IND_Mod_EcartConsoOpt \leq 1,76$	25% des situations
SurConso+1.5L	$IND_Mod_EcartConsoOpt > 1,76$	25% des situations

Tableau III.9 : Catégorisation du gain de consommation en carburant des situations de conduite

2.3.3 Classification des situations de conduite en fonction du critère observé « consommation en carburant »

La classification des critères observés a pour objectif d'utiliser **uniquement les paramètres non contextuels** pour construire des groupes de situations de conduite qui appartiennent à la même catégorie de critère. Les techniques de classification sont mobilisées également ici pour atteindre cet objectif et extraire les règles permettant de classer les situations de conduite à partir des paramètres non contextuels pour prédire le critère observé. Les règles de classification découvertes ici seront utilisées dans l'étape suivante pour préparer l'identification des paramètres contextuels explicatifs de la consommation en carburant.

La classification supervisée a pour objectif de construire des groupes de situation de conduite homogènes selon une variable de classe. La variable de classe étant la consommation en carburant, l'objectif est de faire la classification des situations de conduite homogènes en termes de consommation en carburant. La situation de conduite analysée est « *AllerToutDroit* ». La variable de classe possède trois modalités qui sont : *Faible*, *Moyenne* et *Forte*. Cette classification doit expliquer et prédire la consommation en carburant des situations de conduite « *AllerToutDroit* » à partir des paramètres indépendants non

contextuels. Les arbres de décisions sont mobilisés dans cette partie. La classification est effectuée de la manière suivante : la base des situations de conduite a été divisée en deux bases, une base d'apprentissage pour la découverte de connaissances et une deuxième pour la validation des connaissances découvertes qui sera réalisée dans l'étape 5. La base d'apprentissage est composée de 7724 situations de conduite « *AllerToutDroit* » dont 13% ont une faible consommation en carburant (3l/100km), 72 % ont une consommation en carburant moyenne avec (6l/100km) et 15% ont une consommation en carburant forte avec plus de 6l/100km. L'arbre de décision suivant donne la classification des situations de conduite « *AllerToutDroit* » en termes de consommation en carburant. L'arbre de décision est mis en place avec l'outil d'analyse de données SPSS. L'objectif est de constituer des groupes homogènes en termes de consommation en carburant c'est-à-dire de construire des nœuds qui ne contiennent qu'un seul type de consommation en carburant ou bien de réduire le taux d'erreur dans chaque nœud. Ici nous allons utiliser une méthode de classification supervisée (Arbre de décision CHAID, CRT) [103], [104]. La méthode d'arbre de décision été choisie car elle a l'avantage de fournir des règles décrivant le gain à partir des paramètres dépendants.

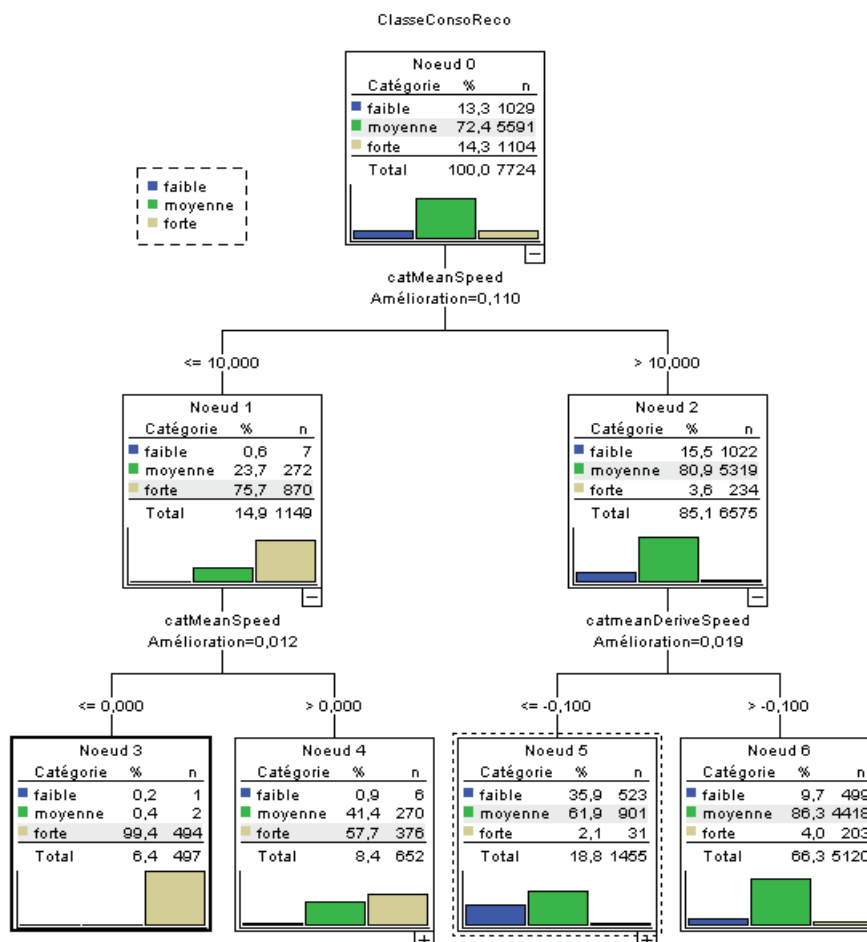


Figure III.23 : Arbre de décision de la classification de la consommation en carburant avec les paramètres non contextuels.

Avec cette première partie de l'arbre de décision de la classification de la consommation en carburant avec les paramètres non contextuels, le nœud 3 permet de classer les situations de conduite « *AllerToutDroit* » qui ont une consommation en carburant « *Forte* » avec un taux de bonne prédiction de 99,4 % soit un taux d'erreur de 0,6 %. Les situations de ce nœud correspondent à des situations proches de l'arrêt (si *catMeanSpeed* vaut zéro, cela veut dire que la vitesse moyenne était comprise entre 0 et 5 km/h). Il est donc normal que la consommation pour 100km/h soit très importante. Le nœud 6 classe celles qui ont une consommation « *Moyenne* » avec un taux de prédiction de 86,3 % soit un taux d'erreur de 13,7%. Les situations de ce nœud correspondent à des situations de consommation moyenne dans lesquelles le véhicule roule (si *catMeanSpeed* est supérieur à 10, cela veut dire que la vitesse moyenne était supérieure à 15 km/h) et il n'y a pas de décélération, contrairement à celles du nœud 5 avec décélération donc avec une consommation plus faible. Le nœud 4 a un taux de prédiction de 56,7 % de la classe « *Forte* » avec un taux d'erreur de 45,3 % sur des situations dans lesquelles le véhicule roule à faible vitesse et accélère, tandis que le nœud 5 permet de classer les éléments avec une consommation « *Moyenne* » en carburant avec un taux d'erreur de 38 % sur des situations dans lesquelles le véhicule roule et décélère.

Les résultats des nœuds 3 et 6 sont considérés comme finaux et aucun autre paramètre n'a permis d'obtenir un meilleur taux de prédiction. Ce n'est pas le cas des nœuds 4 et 5. En effet, les éléments mal classés du nœud 4 sont essentiellement des « *AllerToutDroit* » avec une consommation en carburant « *Moyenne* », ce nœud classe à tort 50% des éléments dans la classe « *Forte* » alors qu'ils sont de la classe « *Moyenne* ». De même, pour le nœud 5, qui classe à tort dans la classe « *Faible* » une fois sur deux des éléments de la classe « *Moyenne* ». Le reste de l'arbre de classification de ces deux nœuds est détaillé dans l'Annexe 3.

Les nœuds finaux de cet arbre étant très hétérogènes, notre objectif dans l'étape 2.4 sera de construire des groupes homogènes de situations « *AllerToutDroit* » grâce aux paramètres contextuels.

2.3.4 Stockage des connaissances issues de la fouille de données dans la base de connaissances

En résumé, cette fouille de données a identifié des règles qui permettent de faire des segments temporels homogènes de l'activité de conduite en fonction de l'objectif de conduite tels que : *AllerToutDroit*, *TournerADroite*, *TournerAGauche* et *ChangementDeVoie*. Elle a permis également d'identifier des règles qui permettent d'étiqueter et de classer les segments de conduite en fonction de la consommation en carburant en utilisant les paramètres liés au

véhicule. Les règles issues de cette phase sont stockées dans la base de connaissances, et seront utilisées dans l'étape 2.4 pour identifier le contexte explicatif de la consommation.

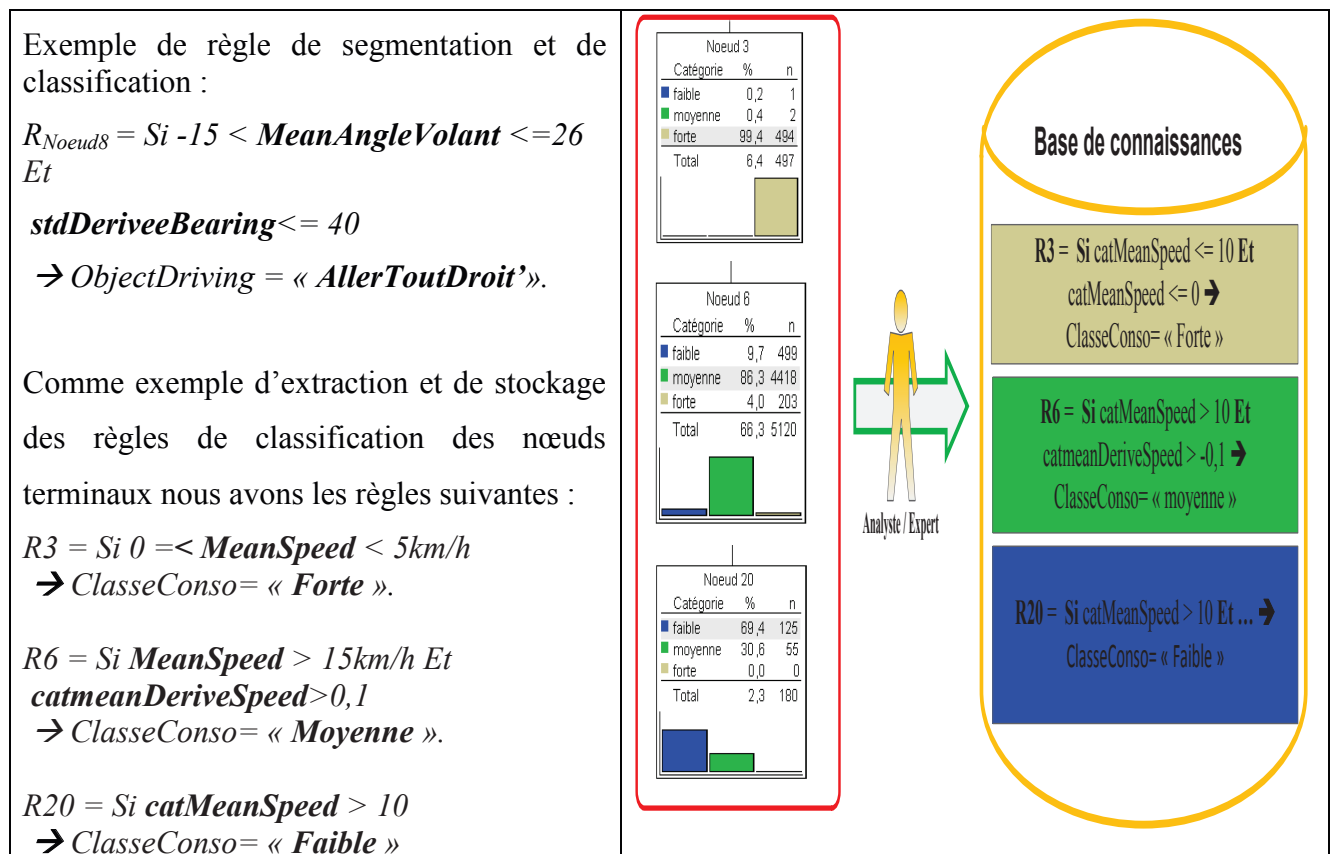


Figure III.24 : Construction de la base de connaissances à partir des arbres de décision

A la fin de cette étape, l'objectif était de créer une base de connaissances initiale exploitable pour l'étape de la découverte de connaissances contextuelles. Elle a permis de mettre en place :

- Les règles de segmentation automatique de l'activité de conduite en situation de conduite
- De définir des critères objectifs de la consommation en carburant que l'on cherche à expliquer à partir des indicateurs de consommation calculés sur les situations.
- De mettre en place les règles de classification des situations de conduite en fonction du critère observé de la consommation en carburant avec des paramètres non contextuels.

Les règles arrivent donc à expliquer de manière logique avec les paramètres non contextuels 82 % des situations de conduite. La prochaine étape est donc maintenant de trouver quels sont les paramètres contextuels qui pourraient expliquer pourquoi les 18 % restants ont été mal classés. Ces règles de segmentation et de classification des situations de conduite seront utilisées pour préparer les données avant l'étape d'identification des paramètres contextuels explicatifs de la consommation en carburant dans l'étape suivante.

2.4 Processus d'assistance à la découverte de connaissances contextuelles

Cette étape a pour objectif d'identifier parmi les paramètres candidats au contexte de l'activité de conduite ceux qui sont explicatifs de la consommation en carburant. Pour cela le principe est simple, à partir de la base de connaissances, l'activité de conduite est segmentée en plusieurs situations de conduite. Chaque situation de conduite est étiquetée d'un indicateur de consommation réelle et d'un indicateur de prédiction de la consommation en carburant mis en place grâce aux règles de classification extraites de la phase précédente. Les deux indicateurs sont alors comparés pour identifier les éléments mal classés qui seront mobilisés dans cette étape afin de découvrir le contexte explicatif pouvant expliquer ces erreurs de classification. Pour cela, deux étapes d'identification des paramètres contextuels sont utilisées, la première consiste à observer et comparer les *obsels* d'une même situation de conduite qui n'ont pas les mêmes consommations en carburant. Dans cette étape, les M-Traces permettent l'identification des contextes qui pourraient expliquer les différences de consommation et aide l'analyste à identifier des nouveaux paramètres contextuels à créer. Puis la deuxième étape consiste à généraliser les observations faites sur les M-Traces à l'ensemble des trajets par des méthodes de classification supervisée.

2.4.1 Représentation de l'activité de conduite par les M-Traces

Dans cette étape l'outil ATraDeCC est utilisé pour la mise des M-Traces. A partir des données collectées de l'activité de conduite la M-Trace première est créée. Puis les règles de la base de connaissances sont utilisées pour créer, à partir de la M-Trace collectée, les transformations nécessaires pour la segmentation de l'activité de conduite et leur catégorisation en termes de consommation en carburant.

Nous avons défini une transformation pour créer des obsels décrivant les situations de conduite particulières à l'aide de paramètres comme par exemple, la position du volant. Cette transformation associera des attributs pour caractériser ces obsels comme par exemple, les valeurs minimales moyennes et maximales de l'angle volant et la consommation de carburant. Un exemple de transformation est donné dans la figure ci-dessous.

Transformation de la M-trace collectée: Trajte1 Visualiser la M-Trace

Nom du type observé :

Liste des attributs:

Liste type des attributs:

Fonction des attributs dans la clause :

Clause de Transformation(Where):

Commentaire observé:

Niveau de l'observé:

Voir les attributs d'un type d'observé de la M-Trace:

Selectionnez un Observé:

Figure III.25 : Page de création d'une transformation d'une M-Trace

Cette transformation est alors résumée dans la figure suivante.

Les Transformations de la M-Trace : Trajte1 Visualiser la M-Trace Effectuer une Transformation Créer une séquence de type d'obsels

Liste des transformations de la M-trace Trajte1

N°	Type Obsel	Liste des attributs	Règles de transformations	Commentaire
1	AngleVolantToutDroit	MeanAngleVolant, MaxAngleVolant, MinAngleVolant, MaxBearing, MinBearing, Distancetravel, MeanSpeed, FuelConsumption	InfoDrive.SteeringWheelAngle BETWEEN -10 AND 10	Angle volant virrage moyen.Sequence angle volant entre -10 et 10.

Voir les obsels des transformations de la M-Trace: Trajte1

Selectionnez un type d'obsel transformé:

Figure III.26 : Page de résultat de la transformation d'une M-Trace

Après son application, nous avons une liste d'obsels issus de la transformation avec leurs attributs présentés dans la figure suivante.

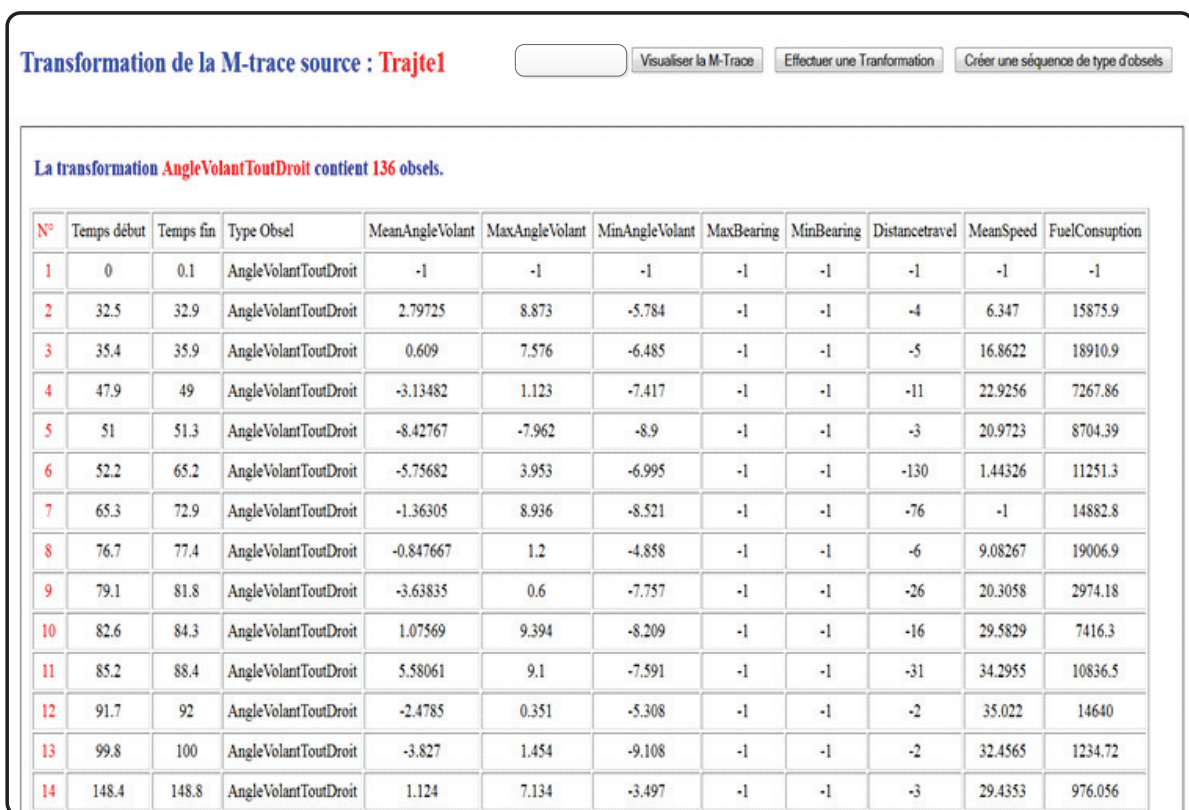


Figure III.27 : Liste des obsels issues d'une transformation

Les règles de classification avec les paramètres non contextuels sont implémentées avec les M-Traces par les transformations correspondantes pour vérifier le résultat de ces classifications tout au long d'un trajet. Comme exemple d'usage des transformations de M-Traces, nous proposons l'usage qui permet d'isoler les situations de conduite mal classées par la classification des paramètres non contextuels.

Nous avons visualisé les différentes traces obtenues en distinguant les situations mal classifiées et le contexte de ces situations, éventuellement source d'explication d'un comportement mal classifié par les règles non contextuelles, qui seraient donc incomplètes pour ces situations.

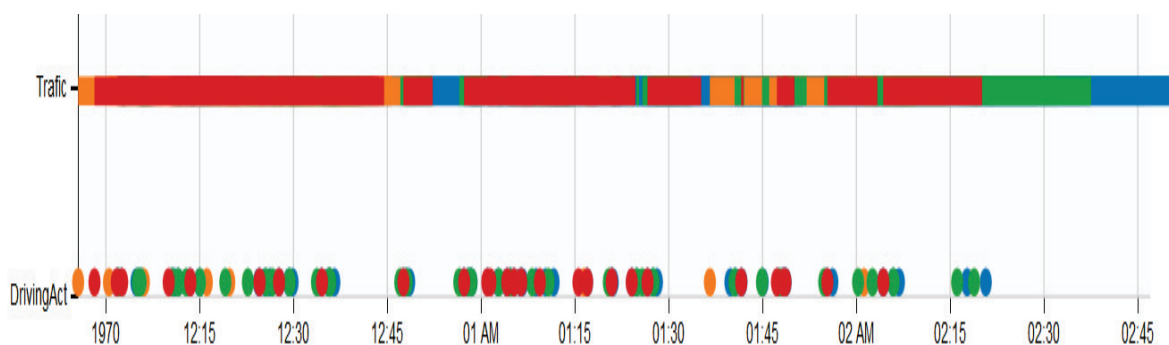


Figure III.28 : Visualisation des obsels mal classés pour identifier les paramètres contextuels

Nous avons alors observé les différentes M-Traces pour identifier quels paramètres contextuels pourraient expliquer les différences de consommations. Les M-Traces contenant les obsels mal classés sont observées et comparées entre elles pour expliquer la différence de classe de consommation en carburant entre la classe réelle et celle prédite par la classification. L'observation est effectuée obsel par obsel et nous avons constaté que les paramètres contextuels qui expliquent ces différences d'étiquette de consommation en carburant sont souvent des paramètres qui décrivent le trafic, le conducteur et l'infrastructure. Ces paramètres contextuels sont des paramètres qui caractérisent les composants contexte de Bazire et al., [41]. Le trafic et l'infrastructure décrivent le composant « *environnement* » et les informations des conducteurs décrivent le composant « *conducteur* ». Il manque le composant « *véhicule* » qui est représenté par le même type de véhicule et la marque de véhicule. En plus de l'identification de ces paramètres contextuels explicatifs de la consommation en carburant, cette observation nous a permis de montrer aussi l'influence des actions avant et après la situation de conduite analysée sur la consommation en carburant. En effet, les types d'infrastructure avant et après une situation de conduite apparaissent également comme pertinents pour l'analyse de la consommation en carburant car cela explique dans certaines conditions une différence de classe de consommation entre la classe réelle et celle prédite. Ceci montre l'importance de prendre en compte la temporalité de l'activité de conduite. L'observation des traces a donc permis à l'analyste d'identifier de nouveaux attributs caractérisant les aspects temporels de l'activité de conduite, l'un pour décrire le type d'infrastructure précédant et l'autre pour le type d'infrastructure suivant. Une fois ces attributs identifiés, une itération de type « Révision d'observés bruts » a permis de les inclure dans le modèle de contexte et de les renseigner.

L'ensemble de tous les attributs contextuels sont alors mobilisés pour la généralisation des connaissances découvertes avec les M-Traces. Cette généralisation est effectuée en mettant l'accent sur les composants contexte identifiés à travers les paramètres contextuels tels que le trafic, l'infrastructure et les informations relatives aux conducteurs.

2.4.2 Découverte des règles de classification pour identifier les paramètres contextuels explicatifs de la consommation en carburant

Après avoir isolé les situations de conduite mal classées et identifié quels paramètres contextuels pouvaient expliquer ces erreurs de classement avec les M-Traces, l'objectif est d'identifier les règles sur les paramètres contextuels permettant de corriger les erreurs de classification pour améliorer le taux de bonne prédiction de la consommation en carburant.

Suivant le modèle des composants contextes utilisé dans cette thèse, les paramètres candidats au contexte de l'activité de conduite sont mobilisés pour la classification des situations mal classées. Dans la base d'apprentissage, il y a en tout 1321 éléments mal classés. A chaque niveau de la classification les paramètres caractérisant les différents composants contexte de l'activité de conduite sont mobilisés et celui qui a le meilleur taux de prédiction est retenu. Puis les autres composants sont mobilisés, tout à tour, jusqu'à avoir un taux de classification satisfaisant pour l'analyste.

Classification de l'entité « Environnement de conduite - trafic »

Les premiers paramètres qui ont été identifiés font partie du composant contexte « Environnement de conduite » d'après le modèle des composants de Bazire et al., [41]. La classification avec les paramètres caractérisant le composant contexte « Environnement » est donnée par l'arbre de décision suivant :

Dans cet arbre les paramètres contextuels explicatifs de la consommation en carburant sont : ***PRC_TimeToCollision*** et ***TrafficBySpeedLimit***. D'après l'arbre de décision, plus le temps de collision avec le véhicule de devant est faible (inférieur à 4,6 seconde) plus la consommation en carburant de la situation de conduite « *AllerToutDroit* » est « *Forte* » avec un taux de prédiction de 35% sinon elle est « *Faible* » à 73 % (temps de collision supérieur à 4,6 secondes). Quand le temps de collision est faible et que le trafic est libre (fluide), la consommation en carburant est « *Faible* » avec 59 % de taux de bonne prédiction. Quand le trafic est « *contraint* ou *très contraint* » et le temps de collision inférieur à 2,1 secondes, la consommation est « *Forte* » avec 51 % de bonne prédiction sinon la consommation en carburant est « *Moyenne* » avec un taux de prédiction de 33 %. Cet arbre permet de classer 64 % des éléments mal classés soit un taux d'erreur de 36 %.

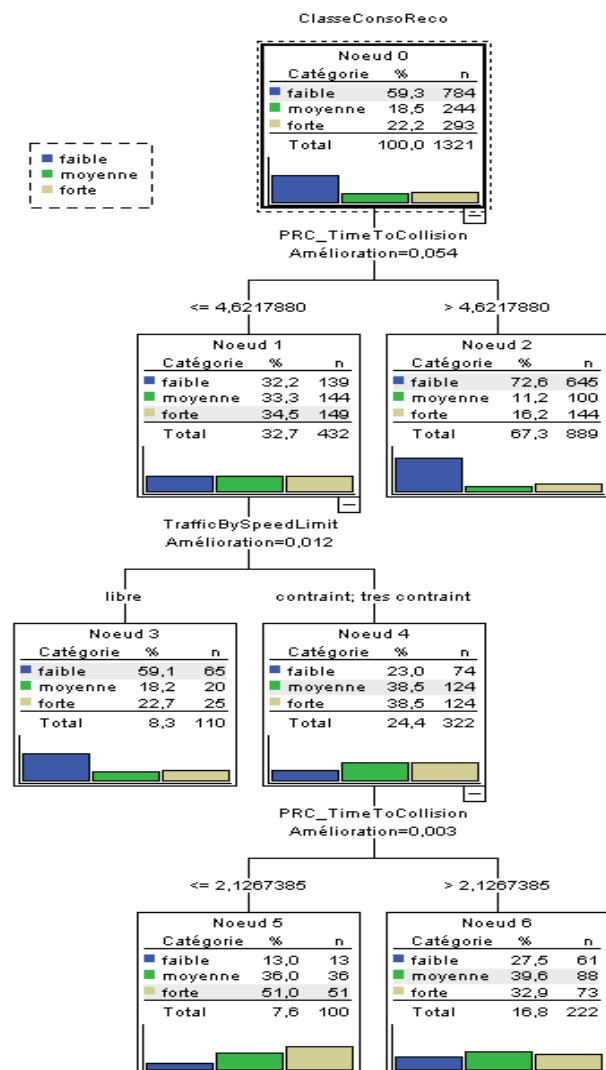


Figure III.29 : Arbre de décision avec les paramètres contextuels du composant contexte « Environnement de conduite »

Nous avons donc les règles de classification suivantes :

$R_{Trafic_{noeud2}} = Si \text{ PRC_TimeToCollision} > 4.6$
 $\rightarrow \text{ClasseConso} = \ll \text{Faible} \gg.$

$R_{Trafic_{noeud3}} = Si \text{ TrafficBySpeedLimit} = \text{'libre'}$
 $Et \text{ PRC_TimeToCollision} \leq 4.6$
 $\rightarrow \text{ClasseConso} = \ll \text{Faible} \gg.$

$R_{Trafic_{noeud5}} = Si \text{ TrafficBySpeedLimit} = \text{'contraint ou très contraint'}$
 $Et \text{ PRC_TimeToCollision} \leq 2.1$
 $\rightarrow \text{ClasseConso} = \ll \text{Forte} \gg.$

$R_{Trafic_{noeud6}} = Si \text{ TrafficBySpeedLimit} = \text{'contraint ou très contraint'}$
 $Et 2.1 < \text{PRC_TimeToCollision} \leq 4.6$
 $\rightarrow \text{ClasseConso} = \ll \text{Moyenne} \gg.$

Sur les 1321 éléments mobilisés dans cette étape d'identification des paramètres contextuels 845 éléments sont maintenant bien classés grâce aux paramètres contextuels *temps de collision* et *trafic*. Cette classification ayant le meilleur taux de prédiction est alors enregistrée et utilisée dans les prédictions suivantes.

Classification de l'entité « Conducteur »

La deuxième classification avec le composant « *Conducteur* » est effectuée en mobilisant les nœuds terminaux de la première classification et les paramètres contextuels caractérisant le composant « *Conducteur* ». Cette classification donne l'arbre de décision suivant :

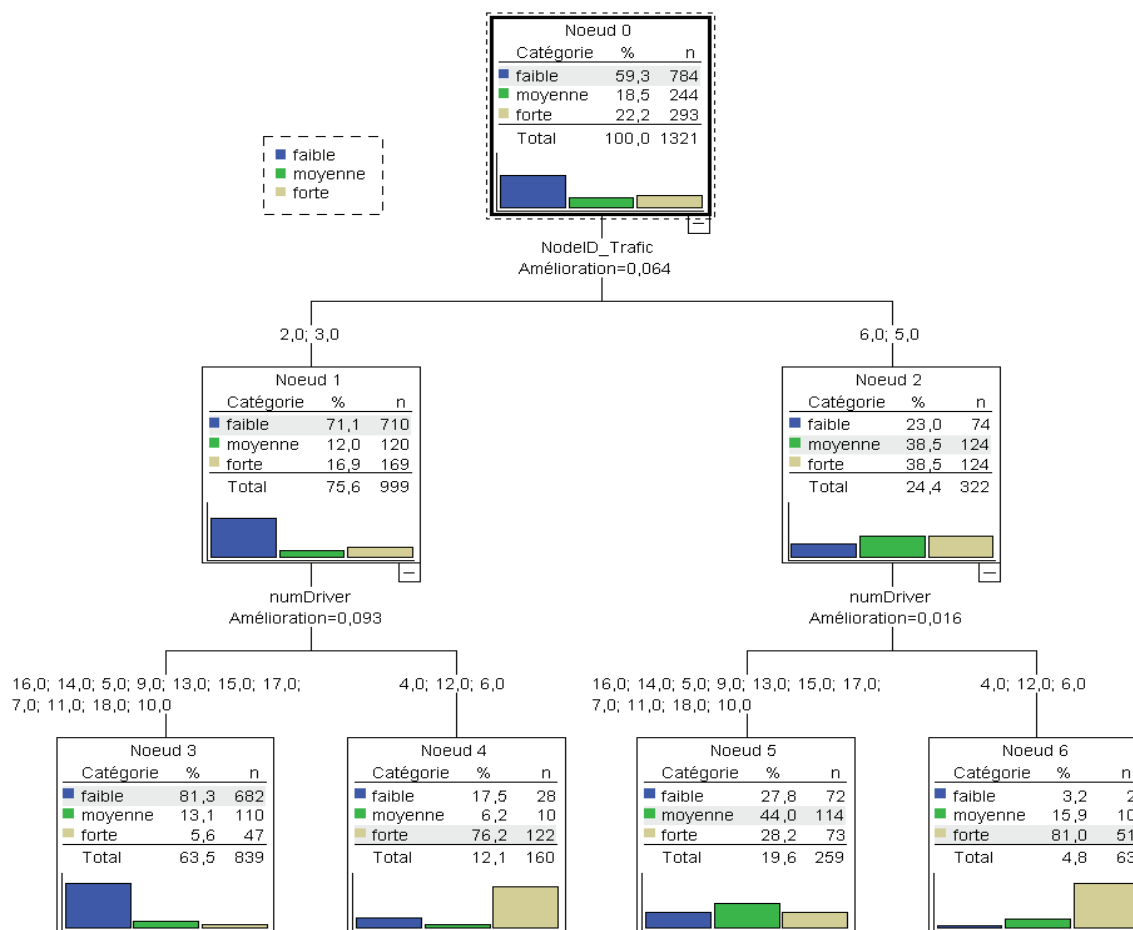


Figure III.30 : Arbre de décision avec les paramètres contextuels du composant contexte « Conducteur »

Cet arbre de décision permet de séparer les situations qui ont une consommation en carburant « Faible » et « Forte » grâce aux paramètres contextuels relatifs au composant « Conducteur » et de l'identifier des nœuds terminaux de la classification du composant « Environnement de conduite (le trafic) ». Parmi ces paramètres, seuls deux sont mobilisés dans cette classification : *NodeID_Trafic* et *numDriver*. Le premier paramètre (*NodeID_Trafic*) correspond aux nœuds terminaux de la classification des situations de conduite avec les paramètres contextuels relatifs au trafic. Parmi les paramètres candidats au contexte de l'activité de conduite du composant « Conducteur », seul l'identifiant du conducteur (*numDriver*) permet d'améliorer le taux de prédiction de la situation de conduite « AllerToutDroit ». Ce paramètre est donc identifié comme un paramètre contextuel de l'activité de conduite. L'arbre de décision met en évidence l'influence des comportements du conducteur sur la consommation en carburant conformément à ce qui est décrit dans la littérature. Il permet de faire deux groupes de conducteurs, un groupe qui a une consommation

« *Faible et Moyenne* » nœuds 3 et 5 et l'autre une consommation en carburant « *Forte* » avec les nœuds 4 et 6.

Les règles de classification correspondant au premier groupe de consommateur de carburant sont:

$RConducteur_{noeud3} = Si \text{ NodeID_Trafic} = 2 ; 3 \text{ Et } numDriver \neq 4 ; 6 ; 12$
 $\rightarrow \text{ClasseConso} = \text{« Faible »}$

$RConducteur_{noeud5} = Si \text{ NodeID_Trafic} = 5 ; 6 \text{ Et } numDriver \neq 4 ; 6 ; 12$
 $\rightarrow \text{ClasseConso} = \text{« Moyenne »}$

A partir du nœud 1 qui correspond aux situations « *AllerToutDroit* » avec peu de trafic et un temps de collision élevé, le paramètre **numDriver** permet d'isoler un premier groupe dans le nœud 3. Ce nœud classe les conducteurs ayant une consommation en carburant « *Faible* » avec un taux de bonne prédiction de 81,3 % et un taux d'erreur de 18,7 % dont 13,1 % de la classe « *Moyenne* » et 5,6 % de la classe « *Forte* ». Grâce au paramètre contextuel **numDriver** le taux de bonne prédiction de la classe « *Faible* » est augmenté de 10,2 % soit un taux de bonne prédiction de 81,3 %.

A partir de ce nœud 2 qui correspond aux situations « *AllerToutDroit* » avec du trafic (contraint ou très contraint) et un temps de collision faible, le paramètre contextuel **numDriver** permet d'isoler un deuxième groupe de conducteurs ayant une consommation en carburant « *Moyenne* » dans le nœud 5 avec un taux de bonne prédiction 44 % soit un taux d'erreur de 54 % dont 27,8 % de classe « *Faible* » et 28,2 % de la classe « *Forte* ». Par rapport au nœud 2 le taux de prédiction de la classe « *Moyenne* » est augmenté de 9,5 %.

Pour ces conducteurs, nous voyons que dans les situations avec peu de trafic ($R\text{Trafic } 2 \text{ ou } 3$), ils consomment faiblement ($RConducteur_{noeud3}$) tandis qu'avec du trafic plus contraint, ils consomment plus ($RConducteur_{noeud5}$).

Le deuxième groupe de conducteurs correspond au nœud 4 et nœud 6 qui classifient les conducteurs ayant une « *Forte* » consommation en carburant. Les règles pour accéder à ces nœuds sont :

$RConducteur_{noeud4} = Si \text{ NodeID_Trafic} = 2 ; 3 \text{ Et } numDriver = 4 ; 6 ; 12$
 $\rightarrow \text{ClasseConso} = \text{« Forte »}$

$RConducteur_{noeud6} = \text{Si } NodeID_Trafic = 5 ; 6 \text{ Et } numDriver = 4 ; 6 ; 12$
 $\rightarrow ClasseConso = \langle \text{Forte} \rangle$

Le nœud 4 permet avec la règle de classification ci-dessus de classer 76,2 % d'éléments qui ont une « Forte » consommation en carburant grâce au paramètre contextuel *numDriver*. Par rapport au nœud 1 qui avait un très faible taux de bonne prédiction (16,9 %), le nœud 4 a un taux de prédiction de 76,2 % de la classe « Forte ».

Le nœud 6 a un taux de bonne prédiction de 81,1% de la classe « Forte », 15,9% de la classe « Moyenne » et 3,2% de la classe « Faible » grâce au paramètre contextuel *numDriver*.

Ces 2 nœuds montrent que quel que soit le trafic si l'identifiant du conducteur est 4, 5 ou 6 la consommation est « Forte ». À partir de ces règles qui identifient les conducteurs ayant une consommation « Forte » une règle générale peut être créée à la place des deux règles $RConducteur_{noeud4}$ et $RConducteur_{noeud6}$. Cette règle est la suivante :

$RConducteur_{noeud4et6} = \{NodeID_Trafic = ?\} \text{ Si } numDriver = 4 ; 6 ; 12$
 $\rightarrow ClasseConso = \langle \text{Forte} \rangle$.

Cette règle permet d'identifier les conductions qui ont une forte consommation en carburant quel que soit le trafic. Cela peut être caractérisé par une conduite plus agressive.

Sur les 1321 mal classés 964 éléments sont bien classifiés dont 845 avec les paramètres du composant contexte « Environnement de conduite » et 119 par paramètres du composant « *Conducteur* ». Cette classification a permis d'isoler les conducteurs ayant une consommation « Moyenne » en carburant dans le nœud 5, les conducteurs ayant une consommation « Forte » en carburant quel que soit le trafic dans le nœud 4 et le nœud 6 et ceux qui ont une « Faible » consommation dans le nœud 3. Sur l'ensemble des éléments mal classés les paramètres *NodeID_Trafic* et *numDriver* permettent d'en classer au total 73,4 % soit une augmentation de 9,1 %. Pour être utilisé dans les classifications suivantes, les nœuds terminaux sont enregistrés.

Classification de l'entité « Environnement de conduite- Infrastructure »

Après la classification prenant en compte les paramètres contextuels des composants contexte « *Conducteur* » et « *Environnement de conduite - le trafic* », on introduit les paramètres relatifs à l'environnement de conduite caractérisant l'infrastructure. L'infrastructure est

caractérisée par exemple par le type de la route (urbain, autoroute et campagne), la présence ou non de feux de signalisation, le type d'infrastructure (par exemple, ligne droite, rond-point, intersection). De plus, l'analyse des M-Traces nous ayant montré que les infrastructures précédente et suivante étaient susceptibles d'avoir de l'influence sur la consommation en carburant des situations de conduite, des paramètres infrastructures précédente et suivante ont alors été créés et mobilisés pour cette classification. Elle est fournie avec l'arbre de décision suivant :

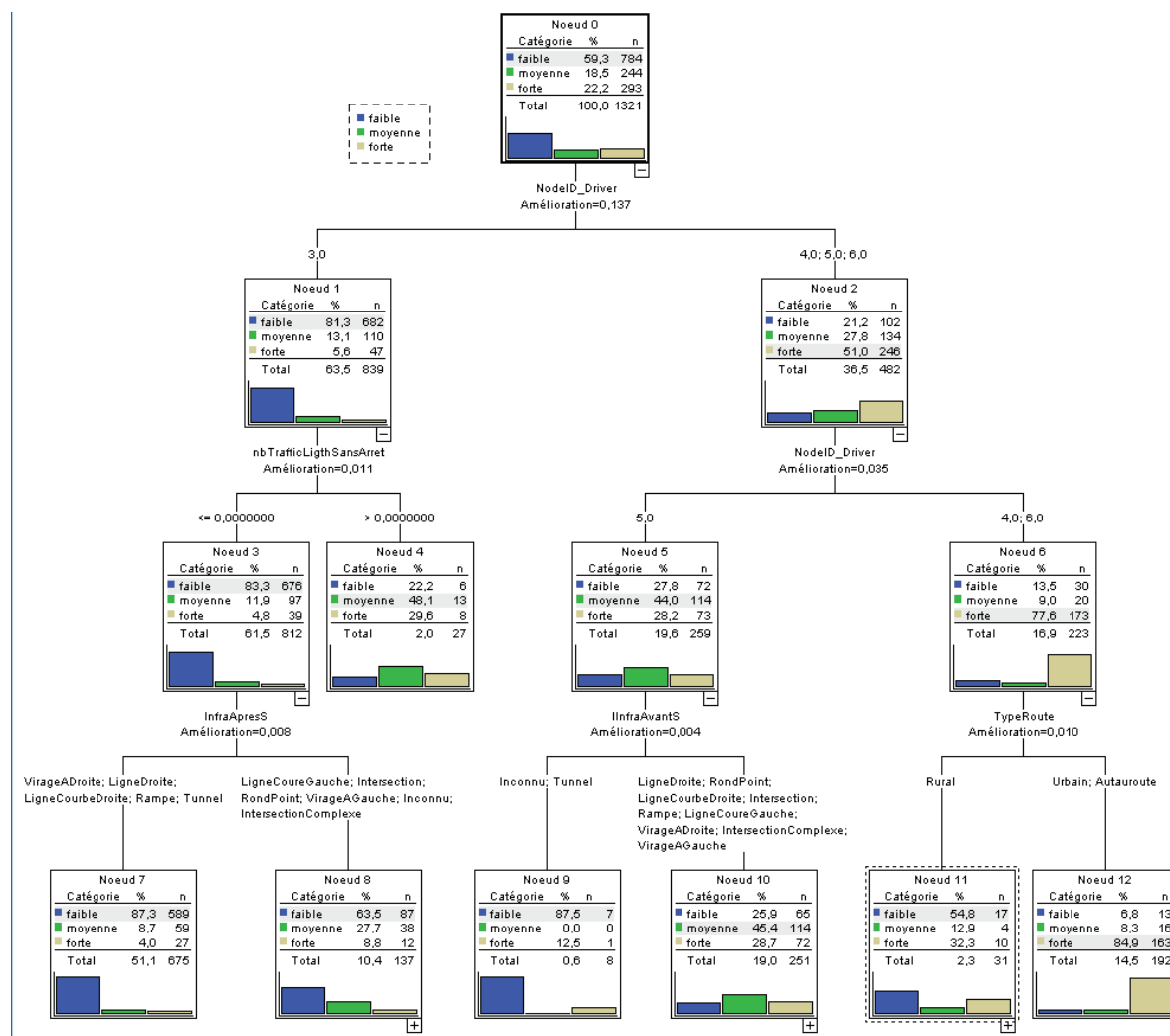
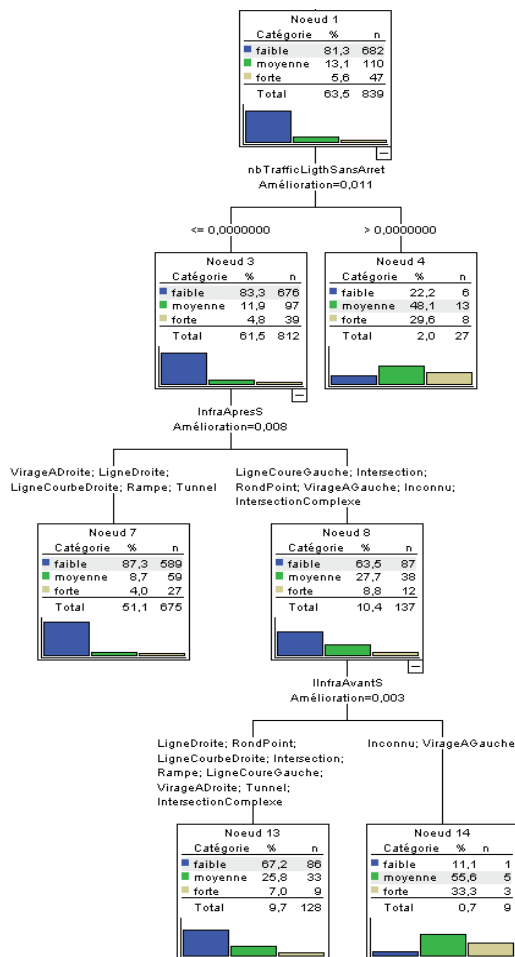


Figure III.31 : Arbre de décision avec les paramètres contextuels du composant contexte « Environnement de conduite- Infrastructure »

Les paramètres impliqués dans cette classification sont : *NodeID_Driver*, *nbTrafficLigthSansArret*, *typeRoute*, *Map_speedLimit*, *MapNumofLanes*, *densityTrafficLigthByKm*, *InfraAprès*, *InfraAvant*. A partir des nœuds terminaux de la classification du composant « **Conducteur** », nous considérerons les situations qui ont été classifiées différemment avec les paramètres caractérisant le composant contexte « **Environnement de conduite – Infrastructure** ». Plus précisément, nous analyserons un par

un les nœuds 1, 5 et 6 de la Figure III.31, correspondant aux trois classes « *Faible, Moyenne et Forte* » de la classification du composant « *Conducteur* ».

Nous commençons par le nœud 1 qui correspond aux consommations faibles et nous étudions uniquement les nœuds enfants qui ont une classification moyenne ou forte. Dans l'arbre de décision (Figure III.31), les situations du nœud 1 sont les situations avec les conducteurs ne consommant pas beaucoup avec peu de trafic, ce nœud classe la classe « *Faible* ».



Dans cet arbre de décision (Figure III.33), si des situations classées en « *Faible* » par la classification précédente sont précédées par une situation de virage à gauche et sont suivies par une situation d'intersection (rond-point, intersection complexe) ou de virage à gauche (courbe ou virage), alors les situations sont classées comme « *Moyenne* » (nœud 14). Ceci peut s'expliquer par le fait, que les situations avant et après obligent le conducteur à rouler moins vite et ont une influence sur la situation « *AllerToutDroit* ». De même, les situations dans lesquelles, il y a des feux de signalisation et pour lesquelles il n'y a pas eu d'arrêt, sont aussi reclassées en « *Moyenne* » (nœud 4). Ceci pourrait s'expliquer par des ralentissements au niveau de ces feux.

Figure III.32 : Arbre de décision à partir du nœud1 avec les paramètres relatif à l'infrastructure

$R_{Infrastructure_{nœud4}} = Si NodeID_Conducteur = 3 Et nbTrafficLigthSansArret > 0$
 $\rightarrow ClasseConso = \ll Moyenne \gg ;$

$R_{Infrastructure_{nœud14}} = Si NodeID_Conducteur = 3 Et InfraAprèsS =$
 'LigneCourbeAGauche ; Intersection ; IntersectionComplexe ; RondPoint ;
 VirageAGauche' Et InfraAvantS = 'VirageAGauche'
 $\rightarrow ClasseConso = \ll Moyenne \gg$

$R_{Infrastructure_{nœud1}} = Si NodeID_Conducteur = 3$

Et $RInfrastructure_{noeud4} = Faux$ Et $RInfrastructure_{noeud14} = Faux$

→ ClasseConso= « **Faible** »

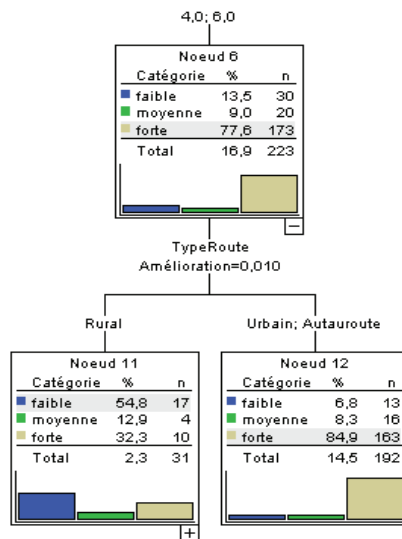


Figure III.34 : Extrait arbre de décision par le paramètre contextuel *type de route*

Dans l'arbre de décision de la Figure III.35, à partir du nœud 6 qui classe les éléments qui ont une consommation « Forte », nous étudions uniquement les nœuds enfants qui ont une classification « Faible » ou « Moyenne ».

Si des situations classées en consommation « Forte » par la classification précédente sont dans un contexte de conduite « rural » alors les situations sont classées comme « Faible » (nœud 11). Ceci pourrait s'expliquer par le fait, que l'on consomme plus en urbain ou à haute vitesse sur autoroute qu'en rural.

$RInfrastructure_{noeud11} = Si NodeID_Conducteur = 4 ; 6 Et TypeRoute = 'Rural'$

→ ClasseConso= « **Faible** » ;

$RInfrastructure_{noeud6} = Si NodeID_Conducteur = 4 ; 6 Et TypeRoute = !'Rural'$

→ ClasseConso= « **Forte** »

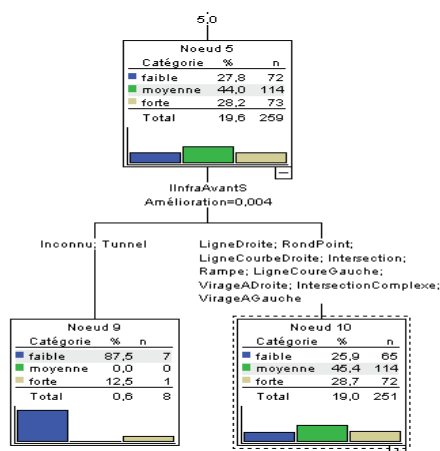


Figure III.36 : Arbre de décision par le paramètre contextuel *infrastructure avant*

Pour le nœud 5 (Figure III.37), qui correspond aux consommations fortes, nous étudions uniquement les nœuds enfants qui ont une classification moyenne ou faible. Si des situations classées en « Moyenne » par la classification précédente, sont après une situation dans un tunnel, les situations sont classées comme « Faible » (nœud 9). Ceci pourrait s'expliquer par le fait, que l'on roule moins vite en sortie d'un tunnel.

RInfrastructure_{noeud9} = *Si NodeID_Conducteur = 5 Et InfraAvant = 'Tunnel*

→ ClasseConso = « **Faible** »

RInfrastructure_{noeud5} = *Si NodeID_Conducteur = 5 Et InfraAvant = ! 'Tunnel*

→ ClasseConso = « **Moyenne** »

Sur l'ensemble des éléments mal classés les paramètres *NodeID_Driver* et paramètres caractérisant l'infrastructure tels que *nbTrafficLigthSansArret*, *typeRoute*, *Map_speedLimit*, *MapNumofLanes*, *densityTrafficLigthByKm*, *InfraAprès*, *InfraAvant* permettent un total de classification avec les paramètres contextuels de 78,7 % soit une augmentation de 5,3 % par rapport à la classification du composant « *Conducteur* ».

En fin d'analyse des paramètres contextuels, sur les 1321 situations de conduite « *AllerToutDroit* » mal classées par les paramètres non contextuels, 78,7 % sont maintenant bien classées soit 1040 de bien classées sur l'ensemble des mal classées. Par rapport à l'ensemble des éléments de l'échantillon d'apprentissage une augmentation du taux de prédiction de 13,4 % est obtenue grâce aux paramètres contextuels soit un taux de bonne prédiction de 96,4 %, ce qui nous fait un taux d'erreurs de classification de 3,7 % sur l'ensemble des situations d'« *AllerToutDroit* ». Le taux d'erreurs de classification avec l'ensemble des paramètres contextuels et non contextuels étant inférieur à 5%, la mise en place d'un modèle de prédiction de la consommation en carburant à partir des règles découvertes peut-être envisager.

2.5 Évaluation et validation des connaissances contextuelles

La validation des connaissances découvertes est effectuée en reprenant les arbres de décision, mais avec une méthode d'échantillonnage. En début d'analyse, les données ont été séparées en deux échantillons, l'échantillon d'apprentissage (60 % des données) et l'échantillon de test (40 % des données). L'échantillon d'apprentissage a été utilisé pour la mise en place des règles de classification de la consommation en carburant dans l'étape précédente. Ces règles sont maintenant évaluées et validées sur l'échantillon de test dans cette étape. Cet échantillon de test est constitué d'éléments n'ayant pas participé à la phase d'apprentissage des règles de classification. Les règles seront validées si les taux de bonne prédiction sur les deux échantillons sont proches : plus ce taux est élevé, plus les règles de classification sont de bonne qualité.

Ci- dessous, le tableau de classification de la consommation en carburant avec les paramètres dynamiques de l'activité de conduite, c'est-à-dire les paramètres non contextuels, indique les différents taux de prédiction des deux échantillons (apprentissage et test). Ensuite, les tableaux de toutes les classifications de la consommation en carburant avec les différents paramètres contextuels sont fournis.

Classification					
Echantillon	Observations	Prévisions			
		faible	moyenne	forte	Pourcentage correct
Apprentissage	faible	245	783	1	23,8%
	moyenne	118	5347	126	95,6%
	forte	0	293	811	73,5%
	Pourcentage global	4,7%	83,2%	12,1%	82,9%
Test	faible	151	516	1	22,6%
	moyenne	75	3403	93	95,3%
	forte	0	199	460	69,8%
	Pourcentage global	4,6%	84,1%	11,3%	82,0%

Méthode de développement : CRT
Variable dépendante : ClasseConsoReco

Tableau III.10 : Classification de la consommation en carburant avec les deux échantillons sans les paramètres contextuels

Ce premier tableau est la classification de la consommation en carburant avec les paramètres non contextuels de l'activité de conduite. L'échantillon d'apprentissage contient 7724 situations de conduite « *AllerToutDroit* » et celui de l'échantillon de test contient 4898 situations « *AllerToutDroit* ». Le tableau nous donne les pourcentages globaux de bonne prédiction des deux échantillons (apprentissage et test), ainsi que les pourcentages de bonne prédiction de chacune des modalités du critère Dans ce tableau, les deux échantillons ont presque le même taux de bonne prédiction de 82,9 % pour l'échantillon d'apprentissage et 82,2 % pour l'échantillon de test. Les deux échantillons classifient bien la classe « *Moyenne* » à plus de 95% de bonnes prédictions. Ils classifient la classe « *Forte* » à 73,5 % de bonnes prédictions pour l'échantillon d'apprentissage et 69,8 % pour l'échantillon de test. La classe « *Faible* » est la moins bien prédite avec 23,8 % pour l'échantillon d'apprentissage et 22,6 % pour l'échantillon de test.

A partir des éléments mal classés avec les paramètres non contextuels, d'autres classifications ont été effectuées pour bien reclasser les éléments mal prédits. Ces tableaux sont les suivant : Le premier tableau de classification est celui de la classification mettant en évidence l'influence de l'environnement de conduite sur la consommation en carburant à travers les paramètres relatifs au *Trafic* (*PRC_TimeToCollision* et *TrafficBySpeedLimit*).

Classification

Echantillon	Observations	Prévisions			Pourcentage correct
		faible	moyenne	forte	
Apprentissage	faible	710	61	13	90,6%
	moyenne	120	88	36	36,1%
	forte	169	73	51	17,4%
	Pourcentage global	75,6%	16,8%	7,6%	64,3%
Test	faible	476	31	10	92,1%
	moyenne	95	54	19	32,1%
	forte	117	43	39	19,6%
	Pourcentage global	77,8%	14,5%	7,7%	64,4%

Méthode de développement : CRT
 Variable dépendante : ClasseConsoReco

Tableau III.11 : Classification de la consommation en carburant avec les deux échantillons avec les paramètres contextuels relatifs au trafic

Dans ce tableau, les deux échantillons ont presque le même taux de bonne prédiction de 64,3 % pour l'échantillon d'apprentissage et 64,4 % pour l'échantillon de test sur les éléments mal prédites de la classification avec les paramètres non contextuels. Les deux échantillons classifient la classe « *Faible* » à plus de 90 % de bonne prédiction. Les classes de prédiction « *Moyenne* » et « *Forte* » ne sont pas très bien prédites par les paramètres décrivant le trafic. Ils ont respectivement un taux de bonnes prédictions de la classe « *Moyenne* » de 36,1 % pour l'échantillon d'apprentissage et 32,1 % pour l'échantillon de test. La classe « *Forte* » est la moins bien prédite avec seulement 17,4 % pour l'échantillon d'apprentissage et 19,6 % pour l'échantillon de test.

Le deuxième tableau de classification est celui de la classification mettant en évidence l'influence du conducteur sur la consommation en carburant à travers les paramètres relatifs au *conducteur* (*NodeID_Trafic* et *NumDriver*).

Classification

Echantillon	Observations	Prévisions			Pourcentage correct
		faible	moyenne	forte	
Apprentissage	faible	682	72	30	87,0%
	moyenne	110	114	20	46,7%
	forte	47	73	173	59,0%
	Pourcentage global	63,5%	19,6%	16,9%	73,4%
Test	faible	451	41	25	87,2%
	moyenne	79	66	23	39,3%
	forte	31	47	121	60,8%
	Pourcentage global	63,5%	17,4%	19,1%	72,2%

Méthode de développement : CRT
 Variable dépendante : ClasseConsoReco

Tableau III.12 : Classification de la consommation en carburant avec les deux échantillons avec les paramètres contextuels relatifs au conducteur

Dans ce tableau, les deux échantillons ont presque le même taux de bonnes prédictions global de 73,4 % pour l'échantillon d'apprentissage et 72,2 % pour l'échantillon de test sur les éléments mal prédites de la classification avec les paramètres non contextuels. Les deux échantillons classifient la classe « *Faible* » à plus de 87 % de bonne prédiction. Pour les classes de prédiction « *Moyenne* » et « *Forte* », par rapport à la classification avec les paramètres décrivant le Trafic, il y a une légère amélioration avec les paramètres décrivant le conducteur avec un taux prédiction à 46,7 % de la classe « *Moyenne* » pour l'échantillon d'apprentissage et 39,3 % de la classe « *Moyenne* » l'échantillon de test. Pour la classe « *Forte* », une augmentation du taux de classification est aussi observée avec une prédiction de 59 % pour l'échantillon d'apprentissage et 60,8% pour l'échantillon de test.

La troisième classification est celle qui met en évidence l'influence de l'environnement de conduite sur la consommation en carburant à travers les paramètres relatifs à l'infrastructure (*NodeID_Driver, InfraAvant, InfraAprès* et *nbTrafficLightSansArret*).

Classification					
Echantillon	Observations	Prévisions			Pourcentage correct
		faible	moyenne	forte	
Apprentissage	faible	730	32	22	93,1%
	moyenne	105	109	30	44,7%
	forte	49	44	200	68,3%
	Pourcentage global	66,9%	14,0%	19,1%	78,7%
Test	faible	465	28	24	89,9%
	moyenne	92	41	35	24,4%
	forte	44	28	127	63,8%
	Pourcentage global	68,0%	11,0%	21,0%	71,6%

Méthode de développement : CRT
Variable dépendante : ClasseConsoReco

Tableau III.13 : Classification de la consommation en carburant avec les deux échantillons avec les paramètres contextuels relatifs à l'infrastructure

Dans ce tableau, le taux global de bonne prédiction de l'échantillon d'apprentissage est 78,7 % et celui de l'échantillon de test est de 71,6 %. Une petite différence de 7,1 % est constatée entre les deux échantillons sur les éléments mal prédits de la classification avec les paramètres non contextuels. Les deux échantillons classifient bien la classe « *Faible* » à 93,1 % pour l'échantillon d'apprentissage et 89,9 % pour échantillon de test. La prédiction de la classe « *Faible* » de la classification avec les paramètres contextuels du conducteur est ainsi améliorée grâce à la classification avec les paramètres contextuels de l'infrastructure de 6,1 %. Les classes de prédiction « *Moyenne* » et « *Forte* » sont également améliorées par les paramètres contextuels décrivant l'infrastructure avec un taux de bonne prédiction de la classe « *Moyenne* » respectivement de 44,7 % et de 24,4 % pour l'échantillon d'apprentissage et de test. La classe « *Forte* » est aussi améliorée avec un taux de bonne prédiction de 68,3 % pour

l'échantillon d'apprentissage et 63,8 % pour l'échantillon de test au lieu d'un taux de prédiction de 59 % pour l'échantillon d'apprentissage et 60,8 % pour l'échantillon de test. En tout grâce à ces classifications avec les paramètres contextuels 78,7 % des éléments mal classés par la classification avec les paramètres non contextuels sont reclassés correctement, ce qui correspond à 13,45 % sur l'ensemble des données.

En conclusion, le contexte a permis d'identifier des règles qui permettent d'expliquer presque 90 % de la variabilité de la consommation en carburant pour les situations de la classe « *Faible* », et 70 % de cette variabilité pour les situations de la classe « *Forte* ». De plus, nous constatons que les règles découvertes sont assez robustes car les taux de prédiction sont semblables sur les deux échantillons d'apprentissage et de test. Par contre, nous n'expliquons pas bien cette variabilité pour les situations de la classe « *Moyenne* ».

2.6 Itérations entre les différentes phases

Dans l'application de la méthode de découverte de connaissances contextuelles à l'analyse de l'activité de conduite automobile, plusieurs types d'itération ont été nécessaires jusqu'à l'obtention des résultats satisfaisants. Par exemple, une itération « Révision d'observés bruts » de l'étape 4 à l'étape 1 a permis de décrire les paramètres trafic et la présence de feux de signalisation. De même, les types d'infrastructure des situations de conduite juste avant et après une situation de conduite donnée ont permis de décrire les aspects temporels de l'activité de conduite. Un autre exemple de l'itération de ce type (passage de l'étape 2 à l'étape 1) a permis de rajouter le critère de consommation optimisée. Nous avons alors ajouté à chaque fois de nouveaux attributs au composant contexte concerné puis identifié les données à utiliser pour renseigner ces attributs et pour dérouler les étapes 2, 3 et 4. Comme exemple d'itération de type « Révisions d'indicateurs calculés », nous avons la création des indicateurs décrivant les situations de conduite à l'aide des fonctions d'agrégation de paramètres continus telles que le pourcentage de temps d'utilisation d'un rapport de boîte de vitesse ou la densité de feux de signalisation par km sur une séquence de conduite. Dans le cadre de notre application, les itérations de ces deux types sont les plus utilisées car cette analyse avait pour but d'identifier le contexte explicatif de la consommation en carburant de l'activité de conduite. Nous nous sommes donc focalisés sur les paramètres contextuels qui ont été créés au fur et à mesure de l'analyse dans sa globalité et sur la validation de la pertinence des connaissances produites.

2.7 Application de la méthode sur le critère d'écart de consommation

Une deuxième analyse a été menée pour expliquer le second critère observé : l'écart de consommation entre la consommation en carburant sur le profil de vitesse réel et sur le profil de vitesse optimisé du modèle de Mensing et al., dans [100]. Ce paramètre, décrit dans la partie 2.2.4.a, représente le gain de consommation réalisable avec le modèle d'optimisation. A partir de ce paramètre, trois groupes de gain de consommation ont été mis en place. Le premier groupe correspond aux situations de conduite « *AllerToutDroit* » dont la consommation réelle est équivalente à la consommation optimisée (écart des consommations réelle et optimisée égal à 0). Le second groupe correspond aux situations dont la consommation réelle est plus grande que la consommation optimisée de moins de 1,5 l/100 km (l'écart entre la consommation réelle et optimisée est compris entre 0,5 et 1,5l l/100 km). Et le dernier correspond aux situations qui ont un écart entre la consommation réelle et optimisée plus grand que 1,5 l/100 km. Une fois ces groupes créés, notre objectif est de faire la modélisation et l'identification des contextes explicatifs de ce nouveau critère observé. Pour cela, les étapes 1, 2 et 3 de la méthode de découverte de connaissances contextuelles sont les mêmes que pour l'analyse de la consommation en carburant réelle collectée effectuée dans le Chapitre III. Par contre, les étapes 4 et 5, pour identifier les paramètres contextuels explicatifs de ce nouveau critère observé, sont différentes. L'ensemble du processus de découverte des connaissances contextuelles et des règles extraites de cette analyse sont détaillées dans l'annexe 4.

Les principaux résultats de cette analyse sont les suivants. Les paramètres non contextuels explicatifs de l'écart de consommation en carburant entre la consommation réelle et la consommation optimisée par Mensing et al., dans [100] ont été identifiés dans un premier temps par la classification avec les paramètres non contextuels. Ces paramètres sont : la moyenne de la vitesse, le rapport de boîte de vitesse le plus utilisé durant la situation de conduite et le régime moteur le plus utilisé durant la situation de conduite, ils démontrent bien l'influence de la dynamique du véhicule sur la consommation en carburant. Après cette classification avec les paramètres non contextuels, l'objectif était de reclasser correctement les situations de conduite mal classées par cette première classification en utilisant les paramètres contextuels. Les paramètres décrivant l'infrastructure de la route et le conducteur ont été mobilisés pour cela, ils ont permis de reclasser correctement plus 55,1 % des situations mal classées par les paramètres non contextuels. En plus de cela, certains contextes explicatifs d'un écart important de consommation en carburant entre les consommations réelle et optimisée ont été identifiés. Par exemple, les situations de conduite sur autoroute dans la nuit

qui était classée dans la classe « *Sur consommation* » ont été reclassées dans la classe « *Sur consommation forte* » par les paramètres contextuels type de route et moment de conduite. Ces paramètres décrivant ce contexte de conduite en autoroute la nuit peut être considéré comme une explication de l'écart important de consommation en carburant entre les consommations réelle et optimisée du modèle d'optimisation. Un autre exemple de contexte explicatif d'une forte consommation en carburant sur autoroute est le nombre de voies. Certaines règles de classification avec les paramètres montre bien l'influence de ce paramètre sur la consommation, plus le nombre de voies est élevé plus la consommation en carburant sur autoroute est élevée. Ces paramètres contextuels permettant de reclasser correctement les situations de conduite mal classées par les paramètres non contextuels représentent le contexte explicatif de la consommation en carburant, ce sont des connaissances contextuelles exploitables pour l'explication et la compréhension de certaines situations de conduite nécessitant une forte consommation en carburant. Ces connaissances contextuelles peuvent être exploitées pour assister les conducteurs dans ces contextes de conduite identifiés par la méthode de découverte de connaissances contextuelles pour améliorer leur consommation en carburant lors de la conduite automobile.

3 Synthèse du chapitre application de la méthode de découverte de connaissances contextuelles dans le domaine du transport

Dans ce chapitre, l'objectif était de **mettre en œuvre une application de la méthode proposée** pour valider le processus de découverte de connaissances contextuels défendu. Une application de la méthode est effectuée pour expliquer la consommation en carburant de l'activité de conduite en tenant compte du contexte de l'activité de conduite. Les différentes étapes ont été appliquées sur les situations de conduite « *AllerToutDroit* » d'une durée de 60 secondes. Les résultats obtenus sont très encourageants car ils sont conformes à ceux de la littérature ce qui prouve la **fiabilité des connaissances** identifiées par cette méthode.

L'utilisation de cette méthode a permis **d'identifier les paramètres explicatifs** de la consommation en carburant comme étant le contexte explicatif des situations de conduite « *AllerToutDroit* ». Pour cela, une première classification a été nécessaire pour identifier les éléments mal classés de la classification avec les paramètres non contextuels de la consommation en carburant. Ces éléments mal classés ont été observés dans un premier temps avec les M-Traces pour identifier les paramètres contextuels expliquant la différence d'étiquette de consommation en carburant réelle et prédite des situations de conduite « *AllerToutDroit* » et pour aider l'analyste à créer de nouveaux paramètres. Les paramètres

identifiés sont relatifs à l'environnement de conduite et au conducteur comme le trafic, le type de route, le nombre de voies, le type d'infrastructure, le type d'infrastructure avant et après la situation « *AllerToutDroit* » et l'identifiant du conducteur.

Pour extraire les règles de classification et généraliser les connaissances contextuelles découvertes avec les M-Traces à partir des éléments mal classés, trois classifications ont été nécessaires pour déterminer, parmi les paramètres contextuels candidats au contexte, ceux qui sont capables de bien les reclasser. La première classification a permis d'identifier dans le composant contexte « *Environnement - trafic* », deux attributs pertinents : le **trafic et le temps de collision avec le véhicule précédent**. Ces paramètres contextuels mettent en évidence l'influence du trafic sur la consommation en carburant. La seconde classification a permis d'identifier dans le composant contexte « *Conducteur* », un seul attribut pertinent : **l'identifiant du conducteur** mettant en évidence l'influence du conducteur sur la consommation en carburant. Enfin, la troisième classification a identifié dans le composant contexte « *Environnement - infrastructure* » quatre attributs pertinents : le **type de route, la densité de feux tricolores, les infrastructures précédant et suivant la situation de conduite**, mettant ainsi en évidence l'influence de l'infrastructure en général sur la consommation en carburant. L'ensemble de ces paramètres relatifs au trafic, à l'infrastructure et au conducteur ont donc été identifiés comme les paramètres contextuels explicatifs de la consommation en carburant.

Ces paramètres contextuels ont permis de bien reclasser 78,7% des éléments mal classés par la classification avec les paramètres non contextuels, ce qui correspond à 13,45% sur l'ensemble des données. De plus, ces taux ont été validés sur un ensemble de données de test. Grâce aux paramètres contextuels, un **taux de bonne prédiction de 96.35 %** est réalisé sur l'ensemble des situations « *AllerToutDroit* » au lieu des 82,9 % réalisé avec uniquement les paramètres non contextuels.

Une deuxième analyse (voir annexe 4) a aussi été réalisée sur un deuxième critère : écart de consommation en carburant entre la consommation réelle et la consommation optimisée par Mensing et al. dans [100]. Ce critère permet de savoir pour une situation donnée, s'il existe une autre façon de conduire qui serait plus économe en tenant compte des contraintes liées aussi bien au trafic qu'aux limites de vitesse autorisée. Cette analyse a permis **d'identifier dans quelles situations et pour quels type de conducteur, le gain potentiel de consommation pouvait être le plus important** (par exemple, les situations « *AllerToutdroit* » sur autoroute avec plus de 2 voies, les trajets nocturnes sur autoroute ou les situations « *AllerToutdroit* » après une intersection complexe). Ces connaissances

contextuelles peuvent être exploitées pour la mise en place d'un système d'aide à la conduite écologique qui tient compte du contexte de conduite. En effet, elles peuvent aider le concepteur du système à définir comment assister les conducteurs suivant leur style de conduite dans les contextes de conduite spécifiques identifiés par la méthode de découverte de connaissances contextuelles.

En conclusion, la méthode de découverte de connaissances contextuelles a permis d'une part, **d'identifier les paramètres contextuels explicatifs** de la consommation en carburant sur deux critères et d'autre part, **d'améliorer le taux de classification ou de prédiction de ces critères**. Cette application de la méthode de découverte de connaissances contextuelles pour l'analyse de l'activité de conduite dans le domaine du transport démontre la **faisabilité de cette méthode** et les résultats obtenus pour expliquer la consommation en carburant avec les paramètres contextuels de l'activité de conduite sont encourageants.

Chapitre IV. Discussion

La méthodologie proposée dans cette thèse a permis de mettre en place une nouvelle manière d'analyser une activité en mettant au cœur de l'analyse le *contexte*. Le contexte étant défini par Dey dans [35] comme étant « *toute information pouvant être utilisée pour caractériser les entités d'une activité. Les entités étant les éléments impliqués dans la réalisation de cette activité* », l'objectif est de considérer ces informations susceptibles d'expliquer une situation ou un phénomène observé lors de l'analyse d'une activité comme la consommation en carburant pour l'activité de conduite. L'application de cette méthode a été effectuée dans le domaine du transport pour expliquer la consommation en carburant en tenant compte du contexte de l'activité de conduite. Suivant la définition du contexte de Dey et du modèle des composants contexte de Bazire et al., dans [41], nous avons défini un modèle de contexte pour l'analyse de l'activité de conduite décrit par la figure suivante :

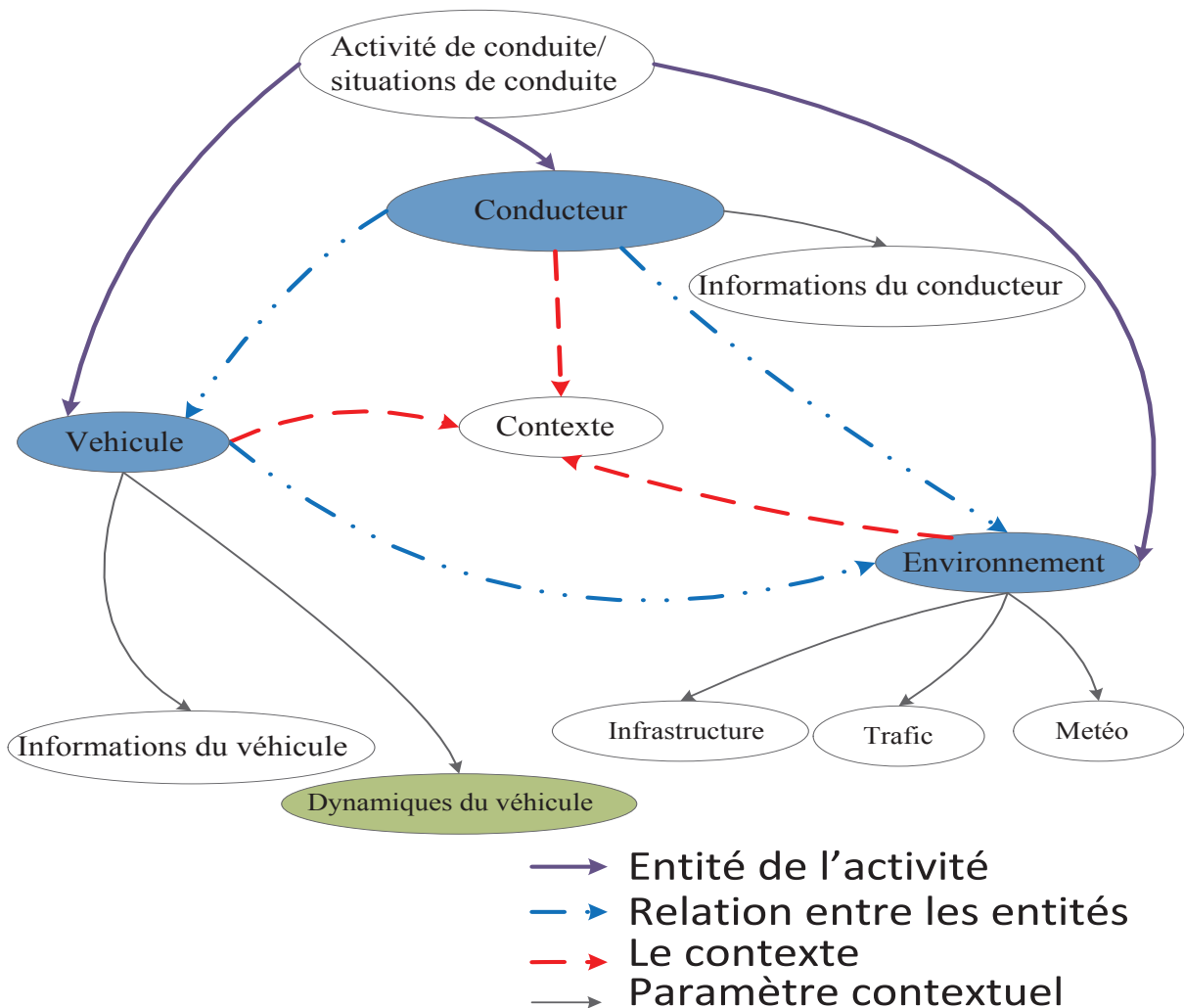


Figure IV.1 : Le contexte au cœur de l'analyse de l'activité de conduite

L'activité de conduite est composée de trois entités qui interagissent ensemble pour la réalisation de la conduite automobile. Ces entités sont : le «*Conducteur*», le «*Véhicule*» et l'«*Environnement de conduite*» et chacune d'entre elles est caractérisée par un ensemble d'informations. Ces informations représentent le contexte de l'activité de conduite et peuvent être exploitées pour expliquer une situation de conduite ou un critère observé comme la consommation en carburant. Nous les avons alors mobilisées pour reclasser les situations de conduites homogènes mal classées par les paramètres non contextuels. Sans l'utilisation des paramètres contextuels, nous avons réussi à classer correctement 82,9 % des situations. Puis, avec cette méthode de découverte de connaissances contextuelles, 78,7 % des situations de conduite mal classées (par les paramètres non contextuels) ont pu être expliquées, soit une augmentation de taux de bonne classification de plus de 13,45 %. Ces connaissances permettent donc **d'expliquer une grande part de la variabilité de la consommation due au contexte.**

Cette méthode a permis, non seulement d'augmenter le taux de bonne classification, mais aussi **d'identifier les paramètres contextuels** explicatifs d'une forte consommation en carburant sur les mêmes situations de conduite. Par exemple, la consommation en carburant des situations de conduite «*AllerToutDroit*» est différente selon le type de **trafic** (fluide, contraint ou bouchon). Nous avons aussi montré que la consommation en carburant de ces situations de conduite dépend aussi des paramètres contextuels du **conducteur** et de l'**infrastructure de la route**. Par exemple, quel que soit le type de trafic, certains conducteurs ayant une conduite agressive ont une consommation en carburant forte sur les situations de conduite «*AllerToutDroit*». Dans cette étude, 3 conducteurs ont été identifiés comme des grands consommateurs de carburant. Puis, nous avons aussi montré l'influence des paramètres contextuels relatifs à l'«*infrastructure*» tels que le *type de route*, la *densité de feux de signalisation*, le *type d'infrastructure avant et après la situation* sur la consommation en carburant. En effet, la consommation en carburant des situations de conduite «*AllerToutDroit*» est différente selon le *type de route*, elle est plus forte sur autoroute et en urbain, et elle est plus faible en rural. Enfin d'autres paramètres caractérisant l'infrastructure comme, la *densité de feux de signalisation*, *infrastructure avant et après* de la situation de conduite «*AllerToutDroit*» ont aussi une influence sur la consommation en carburant. Dans la littérature, l'influence du type de véhicule et des caractéristiques de la route telle que le type de route, l'infrastructure de la route et la densité des feux de signalisation sont démontrés par Brundell-Freij et al., dans [97] de manière générale sur toutes les situations sans distinction. L'influence du mode de conduite du conducteur sur la consommation en carburant

a été aussi mise en évidence par d'Ericsson dans [16]. Les connaissances contextuelles de l'activité de conduite découvertes par la méthode de découverte de connaissances contextuelles proposée dans cette thèse correspondent donc à ceux présentés dans les travaux existants du domaine [16], [97]. Le fait d'avoir des résultats concordants avec la littérature et les mêmes constats que les études existantes du domaine sur la consommation en carburant démontrent la validité de l'approche proposée dans cette thèse.

Les paramètres identifiés par notre méthode sont bien connus pour avoir une influence sur la consommation en carburant mais ils ne sont pas considérés comme le contexte explicatif de la consommation en carburant car leurs influences ne sont pas associées à des situations particulières. L'apport de notre méthode a été **d'identifier dans quelles situations bien précises ces paramètres contextuels sont influents**. Par exemple, de façon générale les conducteurs ayant une forte consommation en carburant consomment faiblement dans les situations de conduite « *AllerToutDroit* » en milieu rural. Dans cette situation précise, la prédiction d'un conducteur classé comme gros consommateur est corrigée. Un autre exemple pertinent de l'influence des paramètres contextuels, est le type d'infrastructure avant et après la situation « *AllerToutDroit* ». Selon leurs valeurs la consommation varie : une situation « *AllerToutDroit* » après une intersection (virage) consomme plus qu'une situation « *AllerToutDroit* » après un tunnel. Ces exemples démontrent le niveau de détail de notre approche en fonction des paramètres contextuels. En effet, les connaissances contextuelles découvertes précisent les connaissances du domaine grâce à l'augmentation du taux de bonne classification de la consommation en carburant des situations de conduite « *AllerTouDroit* » avec les paramètres contextuels *Trafic*, *Conducteur* et *Infrastructure*. De plus, elles identifient les paramètres appartenant au contexte explicatif de la consommation en carburant sur ces situations particulières.

La méthode utilisée est **interactive et itérative** et elle mobilise les méthodes de fouille de données pour la classification du paramètre observé. L'interactivité de cette méthode est due à l'utilisation des traces d'interaction permettant à l'analyste d'observer l'activité. Elle lui permet d'introduire des hypothèses de causalité entre un contexte et le critère observé dans le cycle de la découverte de connaissance. Cette méthode permet aussi de tenir compte de la temporalité de l'activité analysée et du contexte de l'activité. Les atouts des M-Traces pour l'analyse de l'activité de conduite, ont déjà fait leurs preuves pour l'étude comportementale du conducteur et la segmentation de certaines situations de conduite à travers les travaux présentés dans [4]–[6]. L'apport de notre méthode est **d'intégrer dans cette approche M-**

Traces la notion de contexte. En effet, elle permet d'identifier, de modéliser et d'exploiter le contexte explicatif d'un critère observé à partir de notre modèle de contexte, basé sur la définition du contexte de Dey [35] et sur le modèle des composants contexte de Bazire et al., [41]. De plus, cette méthode est itérative, il est donc possible en fonction des besoins de l'analyse de revenir d'une étape à une étape précédente. Par exemple, à partir de l'étape 4, il est possible de revenir à l'étape 2 ou à l'étape 3 pour ajouter de nouveaux paramètres selon les besoins et les observations de l'analyste. Enfin, l'analyste a la possibilité d'itérer autant de fois qu'il le souhaite jusqu'à l'obtention de résultats satisfaisants pour lui.

Enfin, cette méthode se révèle une méthode d'ingénierie de connaissance innovante car elle met en œuvre notre proposition d'évolution du cycle de découverte de connaissances basé sur celui de Fayyad dans [7], [98]. Comme le cycle de Fayyad, elle est composée de cinq étapes. La première étape de notre méthode consiste à définir les composants contexte de l'activité analysée alors que celle de Fayyad se concentre sur la sélection des données. Les deux étapes suivantes de Fayyad « Preprocessing » et « Transformation » sont groupées dans l'étape 2 « Collecte et préparation de données » de notre méthode. Mais nous avons découpé l'étape 4 de « Data mining » en deux étapes, une de « Construction de la base de connaissances » puis une de « Découverte de connaissances contextuelles » basée sur l'exploitation des M-Traces. Comme le cycle de Fayyad, la cinquième phase est la phase d'évaluation et de validation des connaissances découvertes. Contrairement au cycle de Fayyad dont l'objectif est de découvrir de nouvelles connaissances à partir de l'analyse d'un grand volume de données, notre méthode permet de découvrir les connaissances contextuelles à partir de l'analyse tracée de l'activité analysée, en associant l'analyste et ses connaissances à toutes les étapes. Cette méthode manipule les M-Traces avec un SBT (Système à Base de Traces) développé dans cette thèse sur le même principe que celui du LIRIS, KTBS [93].

Cette méthode de découverte de connaissances contextuelles répond aux différentes questions de recherche que nous avons identifiées dans le Chapitre I.4. Ces questions de recherche sont fournies dans le (Tableau IV.1) avec la méthode proposée dans cette thèse.

Travaux	Contexte			M-Traces		Consommation en carburant		Méthode proposée par cette thèse
	Dey et al., [35]	Bazire et al., [41]	Ngo et al., [73]	Gorgeron et al., [4]	Mathem et al., [5], [6]	Ericsson [16]	Brundell-Freij et al., [97]	
Questions de recherches	Non	Non	Non	Oui	Oui	Oui	Oui	Traoré et al., [105]
L'activité de conduite est-elle Analyisée dans ce travail ?	Non	Non	Oui	Oui	Oui	Non	Oui	Oui
Est-ce que les méthodes d'analyse de l'activité proposent un cycle itératif de construction des connaissances ?	Non	Non	Non	Oui	Oui	Non	Non	Oui
Il y a-t-il une prise en compte de la temporalité de l'activité ?	Non	Non	Non	Oui	Oui	Non	Non	Oui
Il y a-t-il une prise en compte du contexte lors de l'analyse ?	Oui	Oui	Oui	Non	Non	Non	Non	Oui
Est-ce qu'il y a la modélisation du contexte lors de l'analyse de l'activité ?	Non	Non	Oui	Non	Non	Non	Non	Oui
Est-ce que le contexte est explicitement considéré comme un type de connaissance spécifique ?	Non	Non	Oui	Non	Non	Non	Non	Oui
Est-ce que la M-Trace est utilisée lors de l'analyse de l'activité ?	Non	Non	Non	Oui	Oui	Non	Non	Oui
Est-ce que l'approche SBT est utilisée pour l'exploration de variables contextuelles dans les travaux cités ?	Non	Non	Non	Non	Non	Non	Non	Oui

Tableau IV.1 : Liste des questions de recherche avec la méthode proposée

La méthode de découverte de connaissances contextuelles proposée a permis d'analyser l'activité de conduite en prenant en compte de sa dimension temporelle et de son contexte. En effet, l'analyse d'une activité en tenant compte de son évolution dans le temps et des connaissances de l'analyste, est possible grâce aux M-Traces stockées dans un SBT. De plus, après une phase de modélisation du contexte, cette méthode permet de définir les éléments candidats au contexte de l'activité dans le but d'identifier ceux qui sont ou non du *contexte explicatif* du critère observé. Pour identifier les éléments contextuels explicatifs d'une situation analysée, les méthodes de fouille de données sont mobilisées par le processus de découverte de connaissances candidates contextuelles sur les traces modélisées (M-Traces) de l'activité de conduite. L'utilisation des M-Traces permet à la méthode d'être interactive et itérative.

La principale limite de cette méthode est qu'il faut travailler sur des sous-parties ou des séquences de l'activité homogènes pour la mise en évidence du contexte. En effet, pour pouvoir identifier correctement quel contexte peut influencer un critère, il est nécessaire de comparer des situations avec les mêmes caractéristiques de l'activité, mais avec des valeurs de critère différentes. Ces différences sont dues aux caractéristiques dynamiques et évolutives du contexte.

Une autre limite de cette méthode de découverte de connaissances contextuelles est la nécessité d'avoir un grand volume de données de l'activité analysée. En effet, c'est à partir des données collectées de l'activité que les connaissances contextuelles sont mises en place ainsi que les M-Traces. Les méthodes de fouille de données étant mobilisées pour la découverte des règles de classification, un volume important de données est nécessaire pour la mise en place de cette méthode de découverte de connaissances contextuelles.

Conclusion et Perspectives

L'expérimentation de la méthode de **découverte interactive de contexte à partir de l'analyse des traces d'activités** a été réalisée à l'aide d'un outil numérique d'analyse de **traces modélisées**. Cet outil a été développé spécifiquement pour s'appliquer au contexte de l'activité de conduite. Avec cette approche, nous avons démontré les qualités méthodologiques, la validité de la méthode et l'intérêt de l'instrumentation du processus d'analyse. Le chercheur associé à l'expérimentation qui a suivi ce travail en a apprécié les capacités d'itérations et les capacités d'assistance à la validation des variables *candidates* au contexte pour expliquer tel ou tel comportement. Des résultats intéressants et originaux ont été obtenus sur les contextes explicatifs des comportements de conduite qui provoquent une surconsommation de carburant, question très importante dans la communauté de recherche sur le transport.

Le contexte étant très dynamique et évolutif, sa modélisation et son identification sont des tâches très difficiles qui demandent la prise en compte de différents facteurs décrivant : l'aspect temporel, l'environnement dans lequel l'activité est réalisée et la personne qui l'a effectuée. La méthode et l'outil associé ont permis d'identifier et de modéliser le contexte d'une activité malgré cette complexité intrinsèque. Le modèle des composants contexte proposé par Bazire et al., dans [41] a structuré la manière de considérer les différents types de contexte. Ce modèle a facilité la découverte interactive des variables explicatives en ciblant des observations adéquates.

L'expérimentation qui a été réalisée a permis de conforter les hypothèses sous-jacentes aux deux contributions principales qui relèvent, d'une part, de l'ingénierie des connaissances, et d'autre part, des méthodes d'analyse dans le domaine du transport : la clarification d'une démarche de **découverte interactive d'éléments contextuels** dans l'observation d'une activité représentée par des Traces modélisées. Cette expérimentation débouche sur un ensemble de principes, une méthode et un environnement de découverte interactive de connaissances contextuelles. Sur la base de la contribution principale, une démarche d'analyse originale dans le domaine de la conduite automobile a été mise en place. Les premiers résultats très encourageants mettent en évidence des éléments de contexte qu'il serait difficile d'établir sans l'assistance de la méthode de découverte de connaissances contextuelles proposée par cette thèse. La notion de découverte interactive de contextes avec l'aide d'une représentation de l'activité par des **traces modélisées** ouvre des perspectives particulièrement intéressantes pour toutes les communautés qui s'intéressent, d'une manière dynamique et adaptative à la

caractérisation du contexte. On peut citer naturellement le **domaine des activités se déroulant avec le Web**, permettant la personnalisation, l'adaptation dynamique à l'évolution des situations, les assistances à l'appropriation, le guidage pour des actions collaboratives dans des contextes différents, pour le développement d'applications *responsives* en particulier dans les situations de handicap, etc. Ces questions font l'objet de recherches actives, en particulier dans la communauté de recherche sur les Environnements Informatiques pour l'Apprentissage Humain (EIAH) et naturellement dans le domaine du WEB. Les équipes de recherche TWEAK et SICAL (Situating Interaction, Collaboration, Adaptation and Learning) du LIRIS sont particulièrement intéressées pour développer ces recherches dans d'autres domaines que le transport. Enfin, la communauté de recherche spécifiquement attachée à l'étude des concepts liés à la notion de CONTEXT pourrait associer ce type de travaux à tout ce qui concerne spécifiquement l'ingénierie des connaissances contextuelles, en articulation avec la communauté de recherche sur l'ingénierie des connaissances. De même dans la robotique, cette méthode peut être utilisée pour rendre sensible des robots aux *Contextes* en apprenant à identifier les contextes interactionnels avec la méthode de découverte interactive de connaissances contextuelles, en particulier lors des interactions homme-machine.

Les résultats obtenus de cette application dans le domaine du transport pourraient être utilisés pour la mise en place d'un système d'aide à la conduite écologique qui permet une réduction de la consommation en carburant à partir d'une situation de conduite donnée. Ce type de système pourrait fournir des conseils de conduite en temps réels à l'utilisateur en fonction de la situation de conduite, de l'environnement de conduite et des comportements au volant du conducteur. Ces conseils seront établis sur des situations de conduite réelles analysées avec la méthode de découverte de connaissances contextuelles. Les règles de classification identifiant les contextes explicatifs d'une forte ou sous consommation en carburant pourraient alors être utilisées par le système écologique dans des situations de conduite similaires pour conseiller le conducteur dans le but d'avoir une consommation en carburant optimale en fonction de l'environnement de conduite comme (le type de route, le trafic et l'infrastructure mais aussi le profil du conducteur). Il s'agirait alors de généraliser l'approche en s'intéressant à toutes les situations de conduite comme cette thèse a montré qu'il était intéressant de le faire pour la situation « *AllerToutDroit* ». Une des perspectives intéressantes de cette thèse est donc d'utiliser la méthode de découverte de connaissances contextuelles sur toutes les autres situations de conduite comme les virages, les changements de voies et les stationnements en parking afin de construire une grande base de connaissances contextuelles de la consommation en carburant de l'activité de conduite. Elle pourra alors être exploitée pour

avoir une meilleure connaissance de la consommation en carburant sur l'ensemble des situations de conduite. Une identification des contextes explicatifs de la consommation en carburant sur l'ensemble des situations de conduite permettra alors de construire un système d'aide à la conduite sensible au contexte de conduite (environnement de conduite). Cela pourra donner lieu à des véhicules conscients de l'environnement et de son évolution dans le temps. Une autre application possible de cette méthode de découverte de connaissances contextuelles dans le monde du transport pourrait être la détection et la compréhension des incidents en identifiant des contextes explicatifs ou propices aux incidents de la route. Cette méthode de découverte de connaissance pourrait aussi être utilisée pour étudier d'autres aspects de l'activité de conduite que la consommation comme la compréhension des comportements des conducteurs (décision de doubler, réduction de vitesse, ...). Ce type d'information pourrait être très pertinent pour les véhicules autonomes car la prise en compte du contexte est très complexe.

Les perspectives immédiates de la thèse concernent l'intégration de l'approche dans les méthodes et outils de l'Ifsttar et au-delà dans la communauté de recherche du domaine des transports. L'intégration de l'outil d'analyse interactive dans une chaîne d'observation-analyse itérative et interactive reste à réaliser. Les travaux à venir pourront bénéficier d'une expérimentation en vraie grandeur ayant mis en œuvre plusieurs outils en les articulant explicitement dans la démarche itérative et interactive : préparation des données, fouille des données pour obtenir des variables candidates, mise en évidence des variables explicatives sur les traces modélisées.

A moyen terme, les perspectives vont intéresser les communautés de l'ingénierie des connaissances et du contexte qui peuvent développer des approches originales et efficaces pour établir clairement ce qui fait contexte ou non. Des études pourront s'intéresser à comment exploiter ces connaissances contextuelles pour assister, adapter, personnaliser, concevoir des environnements interactionnels dynamiques pour (analyser, partager, apprendre) une activité.

A plus long terme, c'est dans le domaine des agents robotiques que les applications pourraient se révéler intéressantes lorsque des agents autonomes doivent faire face à des situations très variées et évoluant rapidement pour la réalisation de tâches incomplètement définies pour n'importe quel contexte.

Bibliographie

- [1] J.-C. Duplessy, “État des connaissances et incertitudes sur le changement climatique induit par les activités humaines,” *Comptes Rendus Académie Sci.-Ser. IIA-Earth Planet. Sci.*, vol. 333, no. 12, pp. 765–773, 2001.
- [2] Commissariat général du développement Durable, “Chiffres clés du transport - Edition 2015,” Mar-2015. [Online]. Available: http://www.statistiques.developpement-durable.gouv.fr/fileadmin/documents/Produits_editoriaux/Publications/Reperes/2015/reperes-transport-ed2015-b.pdf. [Accessed: 15-Jan-2016].
- [3] B. Beusen, S. Broekx, T. Denys, C. Beckx, B. Degraeuwe, M. Gijssbers, K. Scheepers, L. Govaerts, R. Torfs and L. Int Panis, “Using on-board logging devices to study the longer-term impact of an eco-driving course,” *Transp. Res. Part Transp. Environ.*, vol. 14, no. 7, pp. 514–520, Oct. 2009.
- [4] O. L. Georgeon, A. Mille, T. Bellet, B. Mathern, and F. E. Ritter, “Supporting activity modelling from activity traces,” *Expert Syst.*, vol. 29, no. 3, pp. 261–275, 2012.
- [5] B. Mathern, A. Mille, and T. Bellet, “An Interactive Method to Discover a Petri Net Model of an Activity,” 2010. [Online]. Available: <https://hal.archives-ouvertes.fr/hal-00544388/>. [Accessed: 12-Oct-2015].
- [6] B. Mathern, “Découverte interactive de connaissances à partir de traces d’activité : Synthèse d’automates pour l’analyse et la modélisation de l’activité,” Thèse de doctorat en informatique, Université Claude Bernard-Lyon I, 2012.
- [7] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, “From data mining to knowledge discovery in databases,” *AI Mag.*, vol. 17, no. 3, p. 37, 1996.
- [8] P. Cornut, “La convention Climat,” *Les cahiers de Global Chance*, no. 8, pp. 53–56, 1997.
- [9] J. Jouzel and C. Lorius, “Évolution du climat: du passé récent vers le futur,” *Comptes Rendus Académie Sci.-Ser. IIA-Earth Planet. Sci.*, vol. 328, no. 4, pp. 229–239, 1999.
- [10] GIEC (Groupe d’Experts International sur l’Evolution du Climat), “Préface de l’Aperçu général du GIEC,” 1990.
- [11] GIEC (Groupe d’Experts International sur l’Evolution du Climat), “Changement de climat 1995,” 1995.
- [12] Réseau Action Climat France (RACF), “Changement Climatique et Transports par Réseau Action Climat-France (RAC-F).” Février-2007.
- [13] M. Ministère de l’Ecologie, du Développement Durable et de l’Energie, “Transports - Ministère de l’Ecologie, du Développement Durable et de l’Energie,” Juin-2014.

- [Online]. Available: <http://www.developpement-durable.gouv.fr/Transports,34304.html>.
[Accessed: 17-Sep-2015].
- [14] I. DE Vlieger, "On board emission and fuel consumption measurement campaign on petrol-driven passenger cars," *Atmos. Environ.*, vol. 31, no. 22, pp. 3753–3761, 1997.
- [15] E. Ericsson, K. Brundell-Freij, "Influence of street characteristics, driver category and car performance on urban driving patterns," *Transp. Res. Part Transp. Environ.*, vol. 10, no 3, pp. 213–229., 2005.
- [16] E. Ericsson, "Independent driving pattern factors and their influence on fuel-use and exhaust emission factors," *Transp. Res. Part Transp. Environ.*, vol. 6, no 5, pp. 325–345, 2001.
- [17] L. Smith, "Reducing the environment impact of driving - Effectiveness of Driver Training - Transport Research International Documentation - TRID," Austria, 1999.
- [18] A. Rakotonirainy, N. Haworth, G. Saint-Pierre, and P. Delhomme, "Research issues in Eco-driving," *Qld. Univ. Technol. Fr. Inst. Sci. Technol. Transp.*, 2011.
- [19] J. McCarthy, "Notes on formalizing context," 1993.
- [20] V. Akman and M. Surav, "Steps toward formalizing context," *AI Mag.*, vol. 17, no. 3, p. 55–72, 1996.
- [21] R. V. Guha, "Contexts: A Formalization and Some Applications," Stanford University Computer Science Department, 1991.
- [22] D. Lenat, "The dimensions of context-space," *Available Online URL <http://www.Casbah.Org/resources/cycContextSpace.Shtml>*, 1998.
- [23] P. Brézillon, "Modeling and Using Context: Past, Present and Future," *Decis. Support Thought Knowl. Manag.*, pp. 301–320, 2000.
- [24] F. Giunchiglia and P. Bouquet, "Introduction to contextual reasoning: an artificial intelligence perspective," in *Perspectives on Cognitive Science*, vol. 3, Sofia: B Konikof, 1996, pp. 138–159.
- [25] P. Bouquet, C. Ghidini, F. Giunchiglia, and E. Blanziero, "Theories and uses of context in knowledge representation and reasoning," *Journal of pragmatics*, vol. 35, no. 3, pp. 455–484, 2003.
- [26] B. Schilit, N. Adams, and R. Want, "Context-Aware Computing Applications," *In Mobile Computing Systems and Applications*, pp. 85–90, 1994.
- [27] B. N. Schilit and M. M. Theimer, "Disseminating active map information to mobile hosts," *IEEE Netw.*, vol. 8, no. 5, pp. 22–32, 1994.
- [28] P. J. Brown, "The stick-e document: a framework for creating context-aware applications," *Electron. Publ.*, vol. 8, pp. 259–272, 1995.

- [29] A. Ward, A. Jones, and A. Hopper, "A new location technique for the active office," *IEEE Pers. Commun.*, vol. 4, no. 5, pp. 42–47, 1997.
- [30] P. J. Brown, J. D. Bovey, and X. Chen, "Context-aware applications: from the laboratory to the marketplace," *IEEE Pers. Commun.*, vol. 4, no. 5, pp. 58–64, 1997.
- [31] N. Ryan, J. Pascoe, and D. Morse, "Enhanced reality fieldwork: the context aware archaeological assistant," *Bar Int. Ser.*, vol. 750, pp. 269–274, 1999.
- [32] P. Brézillon and J.-C. Pomerol, "Contextual knowledge sharing and cooperation in intelligent assistant systems," in *Le travail humain*, vol. 62, 3 vols., 1999, pp. 223–246.
- [33] D. Thevenin and J. Coutaz, "Plasticity of user interfaces: Framework and research agenda," in *Proceedings of INTERACT*, 1999, vol. 99, pp. 110–117.
- [34] S. Greenberg, "Context as a dynamic construct," *Hum.-Comput. Interact.*, vol. 16, no. 2, pp. 257–268, 2001.
- [35] A. K. Dey, "Understanding and Using Context," *Pers. Ubiquitous Comput.*, vol. 5, pp. 4–7, 2001.
- [36] K. Henriksen, J. Indulska, and A. Rakotonirainy, "Modeling context information in pervasive computing systems," in *International Conference on Pervasive Computing*, 2002, pp. 167–180.
- [37] G. Rey, J. Coutaz, and J. L. Crowley, "The contextor: a computational model for contextual information," *Workshop Build. Bridg. Interdiscip. Context-Sensitive Comput.*, 2002.
- [38] T. B. Mena, N. B.-B. Saoud, M. B. Ahmed, and B. Pavard, "Towards a methodology for context sensitive systems development," in *International and Interdisciplinary Conference on Modeling and Using Context*, 2007, pp. 56–68.
- [39] M. Miraoui and C. Tadj, "A service oriented definition of context for pervasive computing," in *Proceedings of the 16th International Conference on Computing, IEEE computer society press (to appear), Mexico city, Mexico*, 2007.
- [40] W. J. Clancey, "The conceptual nature of knowledge, situations, and activity.," *Hum. Mach. Expert. Context*, pp. 247–291, 1997.
- [41] M. Bazire and P. Brézillon, "Understanding Context Before Using It," in *Modeling and using context*, Berlin Heidelberg: Springer, 2005, pp. 29–40.
- [42] A. Traoré, H. Tattegrain, and A. Mille, "A Trace Analysis Based Approach for Modeling Context Components," in *Modeling and Using Context*, Springer Berlin Heidelberg, 2013, pp. 371–380.
- [43] T. Strang and C. Linnhoff-Popien, "A context modeling survey," in *Workshop Proceedings*, 2004.

- [44] C. Bettini, O. Brdiczka, K. Henriksen, J. Indulska, D. Nicklas, A. Ranganathan and D. Riboni, "A survey of context modelling and reasoning techniques," *Pervasive Mob. Comput.*, vol. 6, no. 2, pp. 161–180, Apr. 2010.
- [45] C. Bolchini, C. A. Curino, E. Quintarelli, F. A. Schreiber, and L. Tanca, "A data-oriented survey of context models," *ACM Sigmod Rec.*, vol. 36, no. 4, pp. 19–26, 2007.
- [46] A. Held, S. Buchholz, and A. Schill, "Modeling of context information for pervasive computing applications," *Proc. SCI*, pp. 167–180, 2002.
- [47] J. Pascoe, "The stick-e note architecture: extending the interface beyond the user," in *Proceedings of the 2nd international conference on Intelligent user interfaces*, 1997, pp. 261–264.
- [48] L. Kagal, V. Korolev, H. Chen, A. Joshi, and T. Finin, "Centaurus: A Framework for Intelligent Services in a Mobile." Distributed Computing Systems Workshop, 2001 International Conference on. IEEE, 2001
- [49] J. Bauer, R.-D. Kutsche, and R. Ehrmanntraut, "Identification and modeling of contexts for different information scenarios in air traffic," *Tech. Univ. Berl. Diplomarb.*, 2003.
- [50] S. Kubicki, S. Lepreux, and C. Kolski, "Distributed UI on Interactive Tabletops: Issues and Context Model," in *Distributed User Interfaces: Usability and Collaboration*, Springer London, 2013, pp. 27–38.
- [51] M. Baas, S. Bernonville, N. Bricon-Souf, S. Hassler, C. Kolski, and G. A. Boy, "Towards a Context Model for Human-Centered Design of Contextual Data Entry Systems in Healthcare Domain," in *International Conference on Engineering Psychology and Cognitive Ergonomics*, International Conference on Engineering Psychology and Cognitive Ergonomics., 2014, pp. 223–233.
- [52] K. Henriksen and J. Indulska, "Modelling and using imperfect context information," in *Pervasive Computing and Communications Workshops, 2004. Proceedings of the Second IEEE Annual Conference on*, 2004, pp. 33–37.
- [53] K. Cheverst, N. Davies, K. Miltchel, A. Friday, and C. Efstratiou, "Developing a context-aware electronic tourist guide: some issues and experiences," in *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, 2000, pp. 17–24.
- [54] T. Hofer, W. Schwinger, M. Pichler, G. Leonhartsberger, and J. Altmann, Eds., "Context-Awareness on Mobile Devices - the Hydrogen Approach," in *System Sciences*, Proceedings of the 36th Annual Hawaii International Conference, 2003, p. 10–pp.
- [55] J. Ye, S. Dobson, and S. McKeever, "Situation identification techniques in pervasive computing: A review," *Pervasive Mob. Comput.*, vol. 8, no. 1, pp. 36–66, Feb. 2012.

- [56] L. Chen, C. Nugent, M. Mulvenna, D. Finlay, and X. Hong, “Semantic smart homes: towards knowledge rich assisted living environments,” in *Intelligent Patient Management*, Springer, 2009, pp. 279–296.
- [57] H. L. Chen, “An intelligent broker architecture for pervasive context-aware systems.,” Thèse de doctorat, Baltimore County : Department of CSEE, University of Maryland, United States, 2004.
- [58] D. Zhang, T. Gu, and X. H. Wang, Eds., “Enabling context-aware smart home with semantic web technologies,” *Int. J. Hum.-Friendly Welf. Robot. Syst.*, vol. 6, no. 4, 2005.
- [59] M. Miraoui, “Architecture logicielle pour l’informatique diffuse: Modélisation du contexte et adaptation dynamique des services,” Thèse de doctorat en Génie, École de technologie supérieure, Québec, 2009.
- [60] T. Strang, C. Linnhoff-Popien, and K. Frank, “CoOL: A context ontology language to enable contextual interoperability,” in *Distributed applications and interoperable systems*, 2003, pp. 236–247.
- [61] P. Brézillon, L. Pasquier, and J.-C. Pomerol, “Reasoning with contextual graphs,” *Eur. Journal Oper. Res.*, vol. 136, no. 2, 2002.
- [62] A. Aroua, “Conception et réalisation d’un consultant basé sur le contexte: application en histopathologie pour la gradation du cancer du sein,” Thèse de doctorat en informatique, Université Paris 6, France, 2014.
- [63] P. Brezillon, “Représentation des procédures et pratiques par les graphes contextuels.” Available Online at <ftp://nephtys.lip6.fr/lip6/reports/2003/lip6.2003.013.pdf>, 2003.
- [64] J. Sun, Y. Zhang, and K. He, “A petri-net based context representation in smart car environment,” in *International Conference on Grid and Pervasive Computing*, 2010, pp. 162–173.
- [65] I. Riahi, “Vers une approche de spécification et de génération d’interfaces sensibles au contexte: Application au domaine médical,” Thèse de doctorat en informatique, Ecole Nationale des Sciences de l’Informatique, Tunis, 2016.
- [66] J. C. Augusto, J. Liu, P. McCullagh, H. Wang, and J.-B. Yang, “Management of uncertainty and spatio-temporal aspects for monitoring and diagnosis in a smart home,” *Int. J. Comput. Intell. Syst.*, vol. 1, no. 4, pp. 361–378, 2008.
- [67] B. Gottfried, H. W. Guesgen, and S. Hübner, “Spatiotemporal reasoning for smart homes,” in *Designing Smart Homes*, Springer, 2006, pp. 16–34.
- [68] J. F. Allen, “Maintaining knowledge about temporal intervals,” *Commun. ACM*, vol. 26, no. 11, pp. 832–843, 1983.

- [69] K. Henriksen, “Towards a hybrid approach to context modelling, reasoning and interoperation,” in *Proceedings of the First International Workshop on Advanced Context Modelling, Reasoning And Management, in conjunction with UbiComp*, pp. 1–8, 2004.
- [70] A. Agostini, C. Bettini, and D. Riboni, “Hybrid reasoning in the care middleware for context awareness.,” *Int. J. Web Eng. Technol.*, vol. 5, no. 1, pp. 3–23.
- [71] M. Baldauf, S. Dustdar, and F. Rosenberg, “A survey on context-aware systems,” *Int. J. Ad Hoc Ubiquitous Comput.*, vol. 2, no. 4, pp. 263–277, 2007.
- [72] G. Castelli, M. Mamei, and F. Zambonelli, “Engineering executable agents using multi-context systems,” in *Engineering Environment-Mediated Multi-Agent Systems*, vol. 5049, *Journal of Logic and Computation*, 2008, pp. 223–239.
- [73] K. A. P. Ngoc, Y.-K. Lee, and S.-Y. Lee, “Context knowledge discovery in ubiquitous computing,” in *On the Move to Meaningful Internet Systems 2005: OTM 2005 Workshops*, 2005, pp. 33–34.
- [74] V. Milea, F. Frasinca, and U. Kaymak, “Knowledge Engineering in a Temporal Semantic Web Context,” *Eighth International Conference on: IEEE*, 2008, pp. 65–74.
- [75] R. Mullins, T. Carsten Pils, D. I. Roussaki, and D. NTUA, “Context and Knowledge Management,” *Mob. Serv. Platf. Clust. White Pap. June*, pp. 1–47, 2008.
- [76] A. Quintart, “Sur les traces de l’Homme : enquête sur la préhistoire,” *Available Online at <https://www.ulb.ac.be/ccs/docs/Cahierpedagogique.pdf>*, pp. 1–52, Sep. 2002.
- [77] B. Galinon-Méléneac and S. Zlitni, “L’Homme-trace, producteur de traces numériques,” *Traces Numér. Prod. À L’interprétation*, pp. 7–19, 2013.
- [78] A. Mille, *De la trace à la connaissance à l’ère du Web*, vol. 1 no. 59, *Intellectica*, pp. 7–28, 2013.
- [79] K. Lund and A. Mille, “Traces, traces d’interactions, traces d’apprentissages: définitions, modèles informatiques, structurations, traitements et usages,” *Anal. Traces Personnal. Environ. Inform. Pour Apprentiss. Hum. Hermès*, pp. 21–66, 2009.
- [80] G. Dyke, “Un modèle pour la gestion et la capitalisation d’analyses de traces d’activité en interaction collaborative,” *Ecole Nationale Supérieure des Mines de Saint-Etienne*, 2009.
- [81] C. Choquet and S. Iksal, “Modeling Tracks for the Model Driven Reengineering of a TEL System,” *JILR J. Interact. Earn. Res. Spec. Issue Usage Anal. Learn. Syst. Exist. Approaches Sci. Issues*, vol. 18, no. 2, pp. 161–184, 2007.
- [82] J. Laflaquière, L. S. Settouti, Y. Prié, and A. Mille, “A Trace-Based Framework for Experience Management and Engineering,” in *Second International Workshop on Experience Management and Engineering in 10th International Conference on*

- Knowledge-Based Intelligent Information and Engineering Systems 2006, UK, October 9-11, 2006*, 2006, vol. 4251, pp. 1171–1178.
- [83] G. M. Olson, J. D. Herbsleb, and H. H. Rueter, “Characterizing the sequential structure of interactive behaviors through statistical and grammatical techniques,” *Hum-Comput Interact*, vol. 9, no. 4, pp. 427–472, 1994.
- [84] A. Mille, G. Caplat, and M. Philippon, “Faciliter les activités des utilisateurs d’environnements informatiques: quoi, quand, comment,” *Intellectica*, vol. 2, no. 44, pp. 121–143, 2006.
- [85] R. Zarka, “Trace-based reasoning for user assistance and recommendations,” Lyon, INSA, 2013.
- [86] A. R. Cooper, “Learning Analytics Interoperability-a survey of current literature and candidate standards.” Available Online at <http://blogs.cetis.ac.uk/adam/2013/05/03/learning-analytics-interoperability>, accessed, pp. 01-19, 2014.
- [87] T. Hoel and W. Chen, “Learning Analytics Interoperability—looking for Low-Hanging Fruits,” in *Workshop proceedings of the 22nd International Conference on Computers in Education*, 2014, pp. 253–263.
- [88] A. Lebis, “Vers une capitalisation des processus d’analyse de traces,” in *Rencontres Jeunes Chercheurs en EIAH (RJC-EIAH 2016)*, 2016.
- [89] J.-M. Gilliot, S. Garlatti, I. Rebai, and M. Belen-Sapia, “Le concept de iMoc pour une ouverture maîtrisée,” in *EIAH 2013*, 2013.
- [90] S. Iksal, C. Choquet, and D. P. T. Ngoc, “A Generic Modeling of Indicator with UTL - The Collaborative Action Function Example.,” in *CSEDU 2010*, Valencia, Spain, 2010.
- [91] O. L. Georgeon, “Analyses de traces d’activité pour la modélisation cognitive : application à la conduite automobile,” Thèse de doctorat en psychologie, Université Lyon2, Lyon, 2008.
- [92] D. Cram, D. Jouvin, and A. Mille, “Visualisation interactive de traces et réflexivité: application à l’EIAH collaboratif synchrone eMédiathèque,” *Rev. Sticef Org*, vol. 21, no. 07, pp. 1–22, 2007.
- [93] P.-A. Champin, Y. Prié, and A. Mille, “Musette: Modelling uses and tasks for tracing experience,” *ICCBR*, vol. 3, pp. 279–286, 2003.
- [94] L. S. Settouti, Y. Prié, A. Mille, and J.-C. Marty, “Systèmes à base de trace pour l’apprentissage humain,” in *Colloque international TICE*, 2006.
- [95] Settouti, L. S., “Systèmes à Base de Traces Modélisées : Modèles et Langages pour l’exploitation des traces d’Interactions.,” Thèse de doctorat Informatique, Université Claude Bernard Lyon 1, 2011.

- [96] J. Laflaquière, Y. Prié, and A. Mille, “Ingénierie des traces numériques d’interaction comme inscriptions de connaissances,” in *19es Journées Francophones d’Ingénierie des Connaissances (IC 2008)*, Nancy, France, 2008, pp. 183–195.
- [97] K. Brundell-Freij and E. Ericsson, “Influence of street characteristics, driver category and car performance on urban driving patterns,” *Transp. Res. Part Transp. Environ.*, vol. 10, no. 3, pp. 213–229, May 2005.
- [98] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, “The KDD Process for Extracting Useful Knowledge from Volumes of Data,” *Commun. ACM*, vol. 39, no. 11, pp. 27–34, 1996.
- [99] U. Fayyad and P. Stolorz, “Data mining and KDD: Promise and challenges,” *Future Gener. Comput. Syst.*, vol. 13, no. 2, pp. 99–115, 1997.
- [100] F. Mensing, E. Bideaux, R. Trigui, and H. Tattegrain, “Trajectory optimization for eco-driving taking into account traffic constraints,” *Transp. Res. Part Transp. Environ.*, vol. 18, pp. 55–61, Jan. 2013.
- [101] F. Mensing, “Optimal Energy Utilization in Conventional, Electric and Hybrid Vehicles and its Application to Eco Driving,” LTE IFSTTAR, 2013.
- [102] F. Mensing, E. Bideaux, R. Trigui, J. Ribert, and J. Bruno, “Eco-driving: An economic or ecologic driving style?,” *Transp. Res. Part C Emerg. Technol.*, vol. 38, pp. 110–121, 2014.
- [103] R. Rakotomalala, “Arbres de décision,” *Rev. Modul.*, vol. 33, pp. 163–187, 2005.
- [104] M. Kamber, L. Winstone, S. Cheng, and J. Han, “Generalization and decision tree induction: efficient classification in data mining,” In *Research Issues in Data Engineering, Proceedings. Seventh International Workshop*, pp. 111–120, 1997.
- [105] A. Traoré, A. Mille, and H. Tattegrain, “Assistance à la découverte de connaissances contextuelles à partir de l’analyse des traces,” *AFIA (Association Française pour l’Intelligence Artificielle)*, Rennes, pp. 33–44, 2015.

Annexes

1 Données Collectées

Nom	Description
timeCode	Le temps écoulé depuis le début du trajet
CC_SL-State	Variable nominale indiquant l'état d'utilisation du régulateur (Cruise Control : CC) et du limiteur (Speed limiter: SL)
CC_SL-Speed	Paramètre numérique correspondant à la vitesse de régulation ou de limitation des systèmes CC et SL
CC_SL-OverSpeed	Variable binaire indiquant s'il y a un dépassement de la vitesse programmée sur le SL ou CC
Distance_Totalizer	Variable numérique indiquant la distance totale parcourue par le véhicule en km
Vehicle_Speed	Variable numérique indiquant la Vitesse instantanée du véhicule en km/h
Vehicle_Longi_Accel	Variable numérique indiquant l'accélération longitudinale du véhicule en m/s^2
Vehicle_Trans_Accel	Variable numérique indiquant l'accélération transversale du véhicule en m/s^2
Clutch_Switch_Max_Travel	Variable nominale indiquant l'utilisation de la pédale d'embrayage
Clutch_Switch_Travel	Variable nominale indiquant l'utilisation de la pédale d'embrayage lors de la conduite
Brake_Info_Status	Variable nominale indiquant l'utilisation de la pédale de frein lors de la conduite
Steering_Wheel_Angle	Variable numérique en degré qui donnant l'angle du volant lors de la conduite
Engine_RPM	Variable numérique en Tr/min donnant le régime moteur du véhicule
Fuel_Consumption	Variable numérique en mm^3 donnant la consommation en carburant instantanée du véhicule

Tableau A.1 : Liste des paramètres issus du bus CAN

DistanceToLeftLaneMark	Variable numérique en m, donnant la distance entre la voiture et la marque ligne à gauche du véhicule
DistanceToRighttLaneMark	Variable numérique en m, donnant la distance entre la voiture et la marque ligne à droite du véhicule
LeftLaneConfidence	Confiance estimé dans la valeur de DistanceToLeftLaneMark
RightLaneConfidence	Confiance estimé dans la valeur de DistanceToRighttLaneMark

Tableau A.2 : Liste des paramètres issus de la caméra multi fonction

R_RelativeVelocity	Variable numérique en km/h indiquant l'écart de vitesse entre le véhicule et celui le précédant.
R_Range	Variable numérique en m indiquant l'écart de distance entre celle du véhicule instrumenté et celle de celui le précédant.
R_Acceleration	Variable numérique en m/s ² indiquant l'écart d'accélération entre celle du véhicule instrumenté et celle de celui le précédant.
R_RelativeAzimuth	Variable numérique en degré indiquant la position angulaire du véhicule précédant le véhicule instrumenté.

Tableau A.3 : Liste des paramètres issus du Radar

GPS_Speed	Variable numérique indiquant la vitesse en km/h fournit par le GPS
GPSLatitude	Variable numérique en degré indiquant la latitude du GPS
GPSLongitude	Variable numérique en degré indiquant la longitude du GPS
GPSBearing	Variable numérique indiquant l'angle en degré entre l'orientation du véhicule et le nord

Tableau A.4 : Liste des paramètres issus du GPS

Map_MatchedLongitude	Variable numérique en degré indiquant la longitude de la position du véhicule après la phase de mise en correspondance de la valeur GPSLongitude et la base de données cartographique
Map_MatchedLatitude	Variable numérique en degré indiquant la latitude de la position du véhicule après la phase de mise en correspondance de la valeur GPSLatitude et la base de données cartographique
Map_MatchedHeading	Variable numérique indiquant l'angle en degré entre l'orientation du véhicule et le nord après la phase de mise en correspondance de la valeur GPSBearing et la base de données cartographique.
Map_NewLink	Variable nominale binaire indiquant la présence ou pas d'une nouvelle voie
Map_SpeedLimit	Variable numérique indiquant la vitesse limite fournie par le GPS en km/h
Map_isUrban	Variable nominale binaire indiquant si le véhicule roule en urbain ou pas
Map_NumofLanes	Variable nominale indiquant le nombre de voie de la route
Map_RoadType	Variable nominale indiquant le type de la route
Map_AllowedDrivingDirection	Variable nominale indiquant les sens de circulation du segment
Map_Tunnel	Variable nominale binaire indiquant la présence ou pas d'un tunnel
Map_Bridge	Variable binaire indiquant la présence ou pas d'un pont
Map_FormOfWay	Donne le type d'infrastructure d la route
Map_TrafficLight	Variable nominale indiquant la présence ou pas de feu de signalisation tricolores sur laquelle le véhicule circule
Map_StopSign	Variable binaire indiquant la présence d'un panneau

	signalisation de stop sur la voie sur laquelle le véhicule circule
Map_PrioritySign	Variable binaire indiquant la présence d'un panneau de signalisation de priorité sur la voie sur laquelle le véhicule circule
Map_JunctionType	Variable nominale indiquant le type de jonction de la voie sur laquelle le véhicule circule

Tableau A.5 : Liste des paramètres issus de la phase de mise en correspondance de la position GPS et la base de données cartographique (Map Matching)

Notation du paramètre	Description du paramètre calculé	Paramètres utilisés pour le calcul
Paramètres calculés initialement		
PC_DistanceParcours	La distance parcourue depuis le début du parcours.	Vehicle_Speed
PC_RBV	les rapports de boîte vitesse instantanés du véhicule.	Vehicle_Speed, Engine_RPM et Clutch_Switch_Travel.
PC_TypeSegmentMap_DerivedMatchedHeading	Indique le changement de segment de route pendant la conduite.	Vehicle_Speed, Map_MatchedHeading
PC_VariationAngleVolant (IndSteering_Wheel_Angle)	Indique la variation de l'angle de volant (faible en aller tout droit et élevée lors des virages).	Vehicle_Speed, Steering_Wheel_Angle,
PC_ObjectifConduite	Indique les actions de conduite effectuées par le conducteur comme aller tout droit, tourner à droite, changer de voie.	Vehicle_Speed, Steering_Wheel_Angle, IndSteering_Wheel_Angle, Map_DerivedMatchedHeading
PC_TypeInfrastructure	Type d'infrastructure de la route (exemples Ligne Droite, Intersection Complexe, Ligne Courbe à Droite)	Vehicle_Speed, PC_ObjectifConduite, Map_FormOfWay, Map_Tunnel
PC_TypeRoute	Type de route (Urbain, rural ou autoroute)	Map_RoadType
PC_TimeToCollision	Valeur du temps à la collision entre le véhicule équipé et le véhicule le précédant.	R_Range, R_RelativeVelocity
PC_DynTimeToCollision	Valeur du temps à la collision entre le véhicule équipé et le véhicule le précédant tenant compte de la différence d'accélération entre les deux véhicules.	R_Range, R_RelativeVelocity, R_Acceleration
Paramètres calculés lors d'itération « Révision d'observés bruts »		
PC_fd50m1Sig PC_fd50m30Sig PC_fd50m150Sig PC_fd100m1Sig PC_fd100m30Sig PC_fd100m150Sig	Indique à chaque instant le pourcentage de présence d'un véhicule dans les 50, 100 ou 300 mètres devant le véhicule instrumenté sur une fenêtre glissante de 2 secondes (1 seconde avant 1 seconde après) 1 minute (30 secondes avant 30	R_Range

PC_fd300m1Sig PC_fd300m30Sig PC_fd300m150Sig	secondes après) ou 5 minutes (150 secondes avant 150 secondes après).	
PC_TraficVitesseLimite	Ce paramètre calculé correspond à l'écart entre la vitesse limite et la vitesse du véhicule, plus cet écart est grand plus il y a du trafic. C'est un paramètre calculé avec les modalités [fluide = (écart <20) ; contraint (écart entre 20 et 40) ; bouchon = (écart>40)].	Vehicle_Speed, Map_SpeedLimit
PC_NbFeuTricolorParKm	Nombre de feux tricolores par km.	Map_TrafficLightSig, timeCode, Vehicle_Speed
PC_NbStopParKm	Nombre de panneaux stop par km	Map_StopSignSig, timeCode, Vehicle_Speed
PC_NbPrioriteParKm	Nombre de panneaux de priorité par km.	Map_PrioritySignSig, timecode, Vehicle_Speed
PC_StyleConduite	Indique le style de conduite du conducteur (souple, agressive, très agressive) qui a été déterminé en étudiant avec l'analyste les histogrammes de la vitesse, du régime moteur et d'utilisation des rapports de boîte de vitesse	DriverID
PC_Mod_Conso	Consommation en carburant calculée par le modèle durant la situation de conduite.	Timecode, Vehicle_Speed Engine_RPM
PC_Mod_ConsoOpt	Consommation en carburant optimisée calculée par le modèle d'optimisation durant la situation de conduite.	Timecode, Vehicle_Speed, R_Range, R_RelativeVelocity, R_Acceleration, MapSpeeLimit, Engine_RPM

Tableau A.6 : Liste des paramètres calculés et leurs descriptions

2 Répartition des consommations en carburant

Statistiques		
IND_ConsoReelle		
N	Valide	70019
	Manquante	0
Centiles	10	1,7456
	20	3,2148
	25	3,7556
	30	4,2403
	40	5,1590
	50	6,1072
	60	7,2412
	70	8,7178
	75	9,7730
	80	11,2150
	90	16,8199

Tableau A.7 : Tableau des percentiles de l'indicateur de la consommation en carburant *IND_ConsoReelle*

Statistiques		
IND Mod EcartConsoOpt		
N	Valide	70019
	Manquant	0
Percentiles	10	-2,0377
	20	-,9142
	25	-,5404
	30	-,2504
	33	-,0985
	40	,0215
	50	,4552
	60	,8787
	66	1,1970
	70	1,4375
	75	1,7622
	80	2,1881
	90	3,2949

Tableau A.8 : Tableau des percentiles de l'indicateur *IND_Mod_EcartConsoOpt*

3 Compléments pour la classification de la consommation en carburant avec les paramètres non contextuels

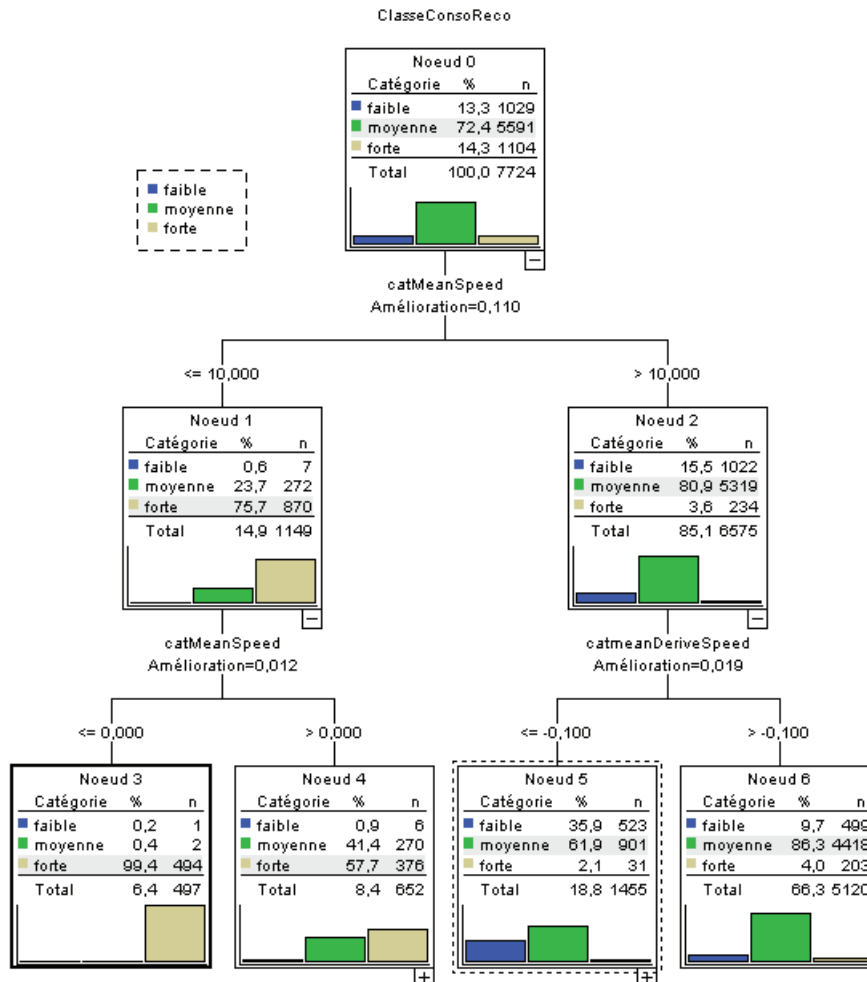


Figure A.1 : Arbre de décision de la classification de la consommation en carburant avec les paramètres non contextuels

Les résultats des nœuds 3 et 6 sont considérés comme finaux et aucun autre paramètre n'a permis d'obtenir un meilleur taux de prédiction. Ce n'est pas le cas des nœuds 4 et 5. En effet, les éléments mal classés du nœud 4 sont essentiellement des « *AllerToutDroit* » avec une consommation en carburant « *Moyenne* », ce nœud classe à tort 50 % des éléments dans la classe « *Forte* » alors qu'ils sont de la classe « *Moyenne* ». De même, pour le nœud 5, qui classe à tort dans la classe « *Faible* » une fois sur deux, des éléments de la classe « *Moyenne* ».

Nous allons donc continuer d'explorer les situations de ces deux nœuds.

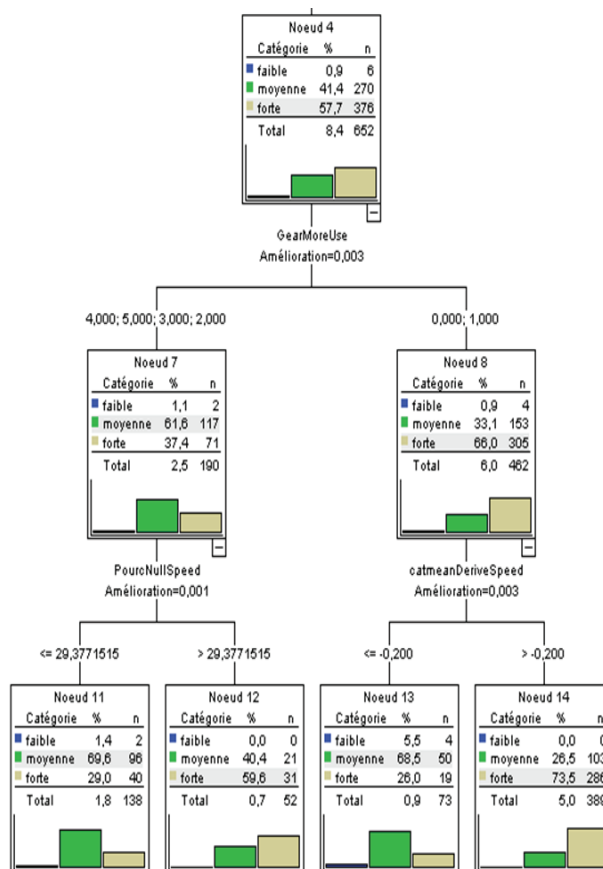


Figure A.2 : Arbre de décision de la classification à partir du nœud 4

Le nœud 4 permet de classer les éléments qui ont une « Forte » consommation en carburant avec un taux d'erreur de 42%. En effet, ce nœud est constitué de 58 % d'éléments avec une consommation « Forte », 41% d'élément en consommation « Moyenne » et 1% de consommation « Faible ». Cet arbre de décision permet de séparer les deux classes « Forte et moyenne » qui sont majoritairement représentés dans le nœud 4 grâce aux paramètres indépendants non contextuels suivant : le rapport de boîte de vitesse le plus utilisée (*GearMoreUse*), le pourcentage de vitesse nulle (*PoucNullSpeed*) et la moyenne de la dérivée de la vitesse (*CatMeanDeriveSpeed*), qui représente une accélération quand sa valeur est positif sinon une décélération. Il montre que, lorsque le rapport de boîte est faible (0 ou 1), si le véhicule décélère, il consomme moyennement (nœud 13) sinon il consomme beaucoup (nœud 14). Si le rapport de boîte est important, si le véhicule s'arrête peu (moins de 30% du temps), il consomme moyennement (nœud 11) sinon il consomme plus (nœud 12).

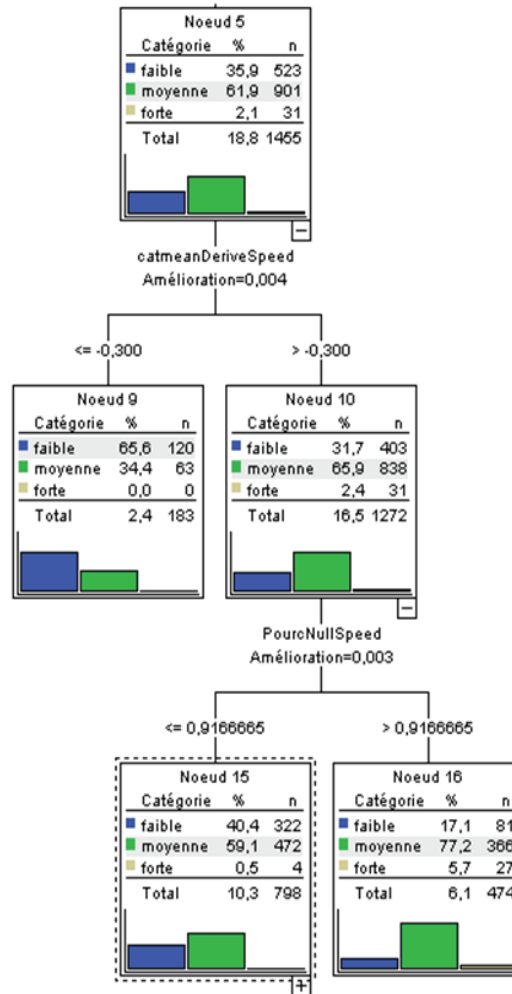


Figure A.3 : Arbre de décision de la classification à partir du nœud 5

Le nœud 5 permet de classer les éléments avec une consommation « Moyenne » en carburant avec un taux d'erreur de 38%. En effet, ce nœud est constitué de 62% d'éléments avec une consommation « Moyenne », 38% d'éléments avec une consommation en carburant « Faible » et 2% de consommation « Forte ». Cet arbre de décision permet de séparer les deux classes (« Faible » et « Moyenne ») qui sont majoritairement représentées dans le nœud 5 grâce aux paramètres indépendants non contextuels suivants : le pourcentage de vitesse nulle (*PourcNullSpeed*) et la moyenne de la dérivée de la vitesse (*CatMeanDeriveSpeed*). Le paramètre *CatMeanDeriveSpeed* correspond à une décélération quand il est négatif et une accélération quand il est positif.

Le nœud 9 permet de classer les éléments qui ont une « Faible » consommation due à une forte décélération avec un taux de bonne prédiction de 66% et un taux d'erreur de 34%. Le nœud 16 classifie les éléments avec une consommation en carburant « Moyenne » si le véhicule est à l'arrêt lors de la séquence. Le taux de bonne prédiction de ce nœud est de 77% et un taux d'erreur de 33%. Aucun autre paramètre n'a permis d'obtenir un meilleur taux de prédiction. Le nœud 15 prédit la consommation « Moyenne » si le véhicule ne s'arrête pas du

tout avec 59% de bonne prédiction, un taux d'erreur de 41% dont 40% de consommation « *Faible* » et 1% de consommation « *Forte* ».

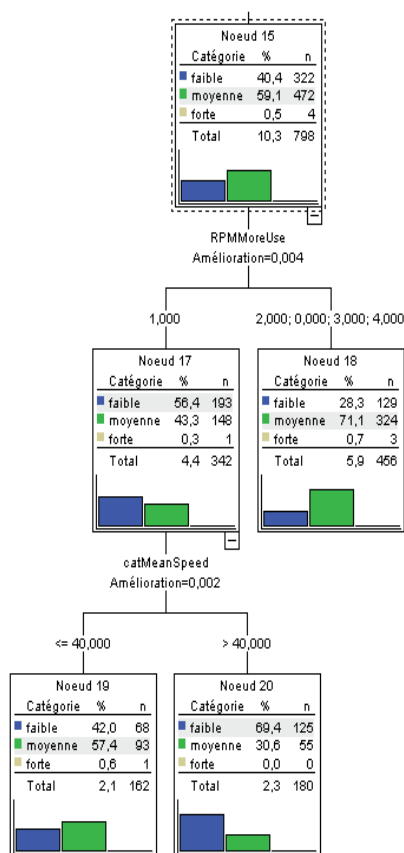


Figure A.4 : Arbre de décision de la classification à partir du nœud 15

Cet arbre de décision permet de séparer les deux classes « *Faible* et *Moyenne* » qui sont majoritairement représentées dans le nœud 15 grâce aux paramètres indépendants non contextuels suivants : le régime moteur le plus utilisé (*RPMMoreUse*) et la moyenne de la vitesse (*CatMeanSpeed*).

Il montre que lorsque le rapport de régime moteur est faible (entre 1000 et 2000), si le véhicule roule vite (supérieur à 40 km/h), il consomme faiblement (nœud 20) sinon il consomme le plus souvent moyennement (nœud 19). Si le rapport de régime moteur est important, il consomme moyennement (nœud 18).

4 Application de la méthode de découverte de connaissances contextuelles sur l'écart entre la consommation en carburant réelle et optimisée.

La classification avec les paramètres non contextuels est donnée par l'arbre de décision suivant :

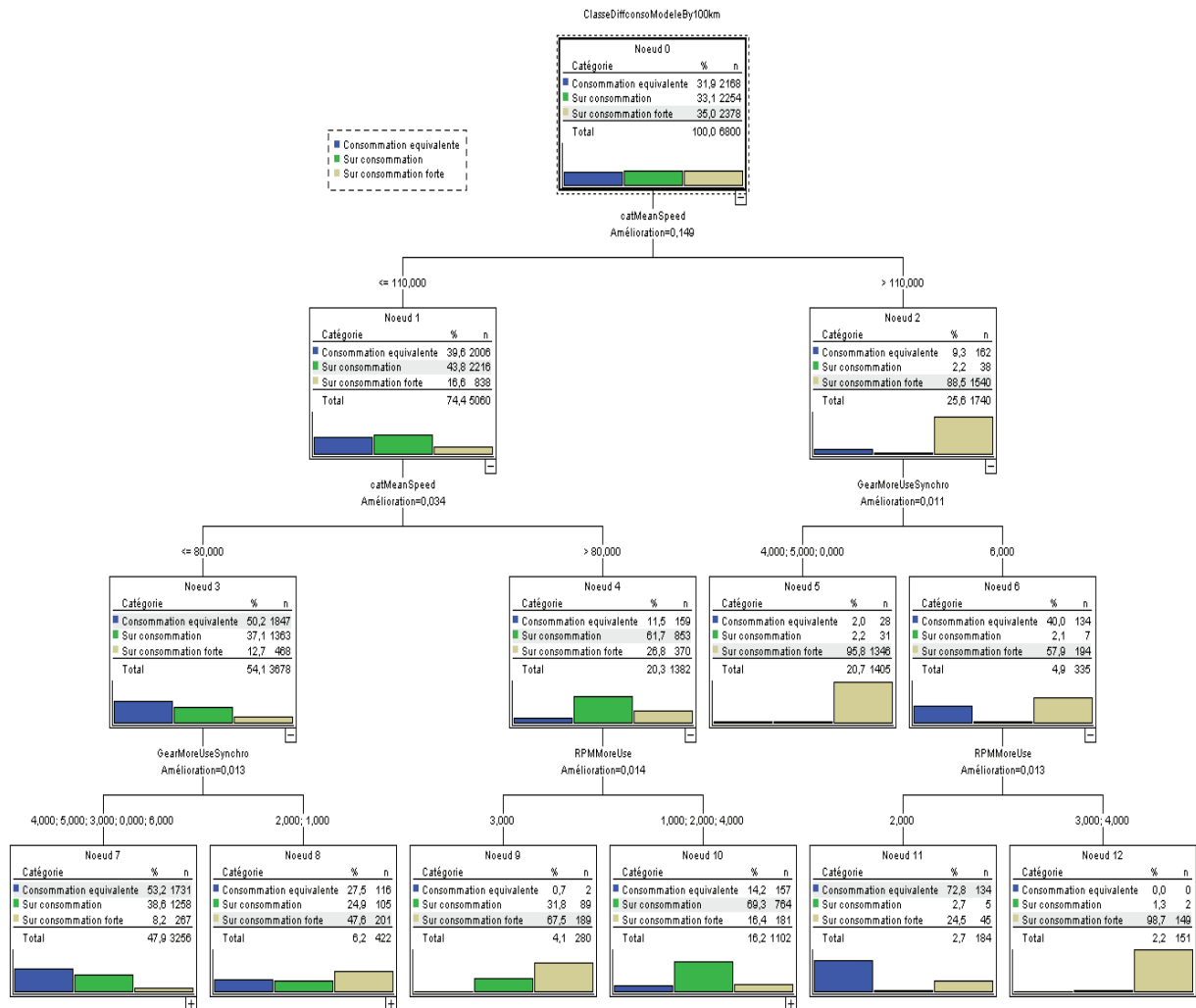


Figure A.5 : Arbre de décision de la classification avec les paramètres non contextuels

Voici quelques règles de classification extraites de l'arbre de décision :

$R1_{noeud7} = Si \text{ catMeanSpeed} \leq 80 \text{ Et } \text{GearMoreUseSyncro} = 3 ; 4 ; 5 ; 6$

$\rightarrow \text{ClasseDiffConso} = \langle\langle \text{Consommation équivalente} \rangle\rangle$

La règle $R1$ fournie par le nœud 7 de l'arbre de décision caractérise les situations «*AllerToutDroit*» ayant une consommation en carburant équivalente avec le modèle d'optimisation quand la moyenne de la vitesse est inférieure à 80km/h et que le rapport de

boîte de vitesse utilisé est élevé (3 ; 4 ; 5 ; 6). Le conducteur utilisant déjà un rapport de boîte assez élevé, le modèle ne propose pas de solution plus économe.

$R2_{noeud8} = Si \text{ catMeanSpeed} \leq 80 \text{ Et } \text{GearMoreUseSyncro} = 2 ; 1 \rightarrow \text{ClasseDiffConso} = \text{«Sur consommation forte»}$

La règle $R2$ fournie par le nœud 8 de l'arbre de décision caractérise les situations «*AllerToutDroit*» en sur consommation forte par rapport au modèle d'optimisation quand la moyenne de la vitesse est inférieure à 80km/h et que le rapport de boîte de vitesse utilisé est bas (1 ; 2). Dans ce cas, le conducteur n'utilisant pas un rapport de boîte assez élevé, le modèle propose une solution plus économe qui augmente le rapport boîte de vitesse, d'où un écart important de consommation en carburant entre la consommation optimisée et la consommation réelle.

$R3_{noeud9} = Si \ 80 < \text{catMeanSpeed} \leq 110 \text{ Et } \text{RPMMoreUse} = 3 \rightarrow \text{ClasseDiffConso} = \text{«Sur consommation forte»}$

La règle $R3$ fournie par le nœud 9 de l'arbre de décision caractérise les situations «*AllerToutDroit*» en sur consommation forte par rapport au modèle d'optimisation quand la moyenne de la vitesse est comprise entre à 80km/h et 110km/h et quand le régime moteur le plus utilisé sur le segment de conduite est de la catégorie 3 (entre 2500 et 3500 tours par minute). Dans ce cas, le régime moteur est élevé pour une vitesse peu élevée, le conducteur n'utilise pas un rapport de boîte assez élevé. Le modèle propose une solution plus économe qui augmente le rapport boîte de vitesse d'où un écart élevé de consommation en carburant entre les consommations réelle et optimisée.

$R5_{noeud5} = Si \ \text{catMeanSpeed} > 110 \text{ Et } \text{GearMoreUseSyncro} = 4 ; 5 ; \rightarrow \text{ClasseDiffConso} = \text{«Sur consommation forte»}$

La règle $R5$ fournie par le nœud 5 de l'arbre de décision caractérise les situations «*AllerToutDroit*» en sur consommation forte par rapport au modèle d'optimisation quand la moyenne de la vitesse est supérieure à 80km/h et quand le rapport de boîte de vitesse le plus utilisé sur le segment de conduite est 4 ou 5 alors que le modèle d'optimisation propose d'utiliser le sixième rapport.

$R6_{noeud11} = Si \ \text{catMeanSpeed} > 110 \text{ Et } \text{GearMoreUseSyncro} = 6 \text{ Et } \text{RPMMoreUse} = 2 \rightarrow \text{ClasseDiffConso} = \text{«Consommation équivalente»}$

La règle *R6* fournie par le nœud 11 de l'arbre de décision caractérise les situations «*AllerToutDroit*» en consommation équivalente par rapport au modèle d'optimisation quand la moyenne de la vitesse est supérieure à 110 km/h avec le rapport de boîte de vitesse égal à 6 et quand le régime moteur le plus utilisé sur le segment de conduite est de la catégorie 2 (entre 1500 et 2500 tours par minute). Le régime moteur étant assez bas pour une vitesse importante, le modèle d'optimisation ne peut faire mieux que le conducteur.

$R7_{noeud12} = Si \text{ catMeanSpeed} > 110 \text{ Et } GearMoreUseSyncro = 6 \text{ Et } RPMMoreUse = 3 ; 4 \rightarrow$
ClasseDiffConso = «**Sur consommation forte**»

La règle *R6* fournie par le nœud 12 de l'arbre de décision, caractérise les situations «*AllerToutDroit*» en consommation équivalente par rapport au modèle d'optimisation quand la moyenne de la vitesse est supérieure à 110km/h avec le rapport de boîte de vitesse égale 6 et quand le régime moteur le plus utilisé sur le segment de conduite est de la catégorie 3 et 4 (entre 2500 et 3500 tours par minute). Dans ce cas, avec un régime moteur très haut, il est probable que le conducteur soit en survitesse par rapport au modèle.

Classification					
		Prévisions			
Echantillon	Observations	Consommation équivalente	Sur consommation	Sur consommation forte	Pourcentage correct
Apprentissage	Consommation équivalente	1639	442	87	75,6%
	Sur consommation	697	1356	201	60,2%
	Sur consommation forte	228	261	1889	79,4%
	Pourcentage global	37,7%	30,3%	32,0%	71,8%
Test	Consommation équivalente	1001	303	47	74,1%
	Sur consommation	469	871	131	59,2%
	Sur consommation forte	157	166	1175	78,4%
	Pourcentage global	37,7%	31,0%	31,3%	70,5%

Méthode de développement : CRT

Variable dépendante : ClasseDiffconsoModeleBy100km

Tableau A.9 : Tableau de classification du critère écart entre la consommation réelle et optimisée du modèle d'optimisation avec les paramètres non contextuels

La classification avec les paramètres non contextuels permet de classer correctement 71,8 % des situations de conduite, soit un taux d'erreur de 28,2 %. Les paramètres mobilisés pour cette classification sont : la moyenne de la vitesse, le rapport de boîte de vitesse le plus utilisé durant la situation de conduite et le régime moteur le plus utilisé durant la situation de conduite. Ces règles montrent bien l'influence de la vitesse, le choix du rapport de boîte de vitesse et du régime moteur sur la consommation en carburant, ce qui est conforme aux connaissances des ingénieurs en motorisation.

Pour l'identification des paramètres contextuels, nous nous concentrons sur les situations de conduite qui n'ont pas été classifiées correctement par les paramètres non contextuels. Ils correspondent aux 28,2 % mal classés, la première classification avec les paramètres contextuels est effectuée avec les paramètres décrivant l'infrastructure de la route. En tout six paramètres sont mobilisés : la vitesse limite de route, la densité de feux tricolores par kilomètre, le type de la route, le nombre de voies, le moment de la conduite, le type d'infrastructure avant et le type d'infrastructure après la situation de conduite. Parmi les situations mal classées, 27,6 % sont de la classe « *Consommation équivalente* », 46,9 % dans la classe « *Sur consommation* » et 25,5 % dans la classe « *Sur consommation forte* ». Avec cette répartition, ces situations sont classées comme étant de la classe « *Sur consommation* » par arbre de décision au début de la classification. La majorité des situations mal classées étant de la classe « *Sur Consommation* », nous cherchons les règles qui permettent d'identifier les deux autres classes (« *Consommation équivalente* » et « *Sur consommation forte* »).

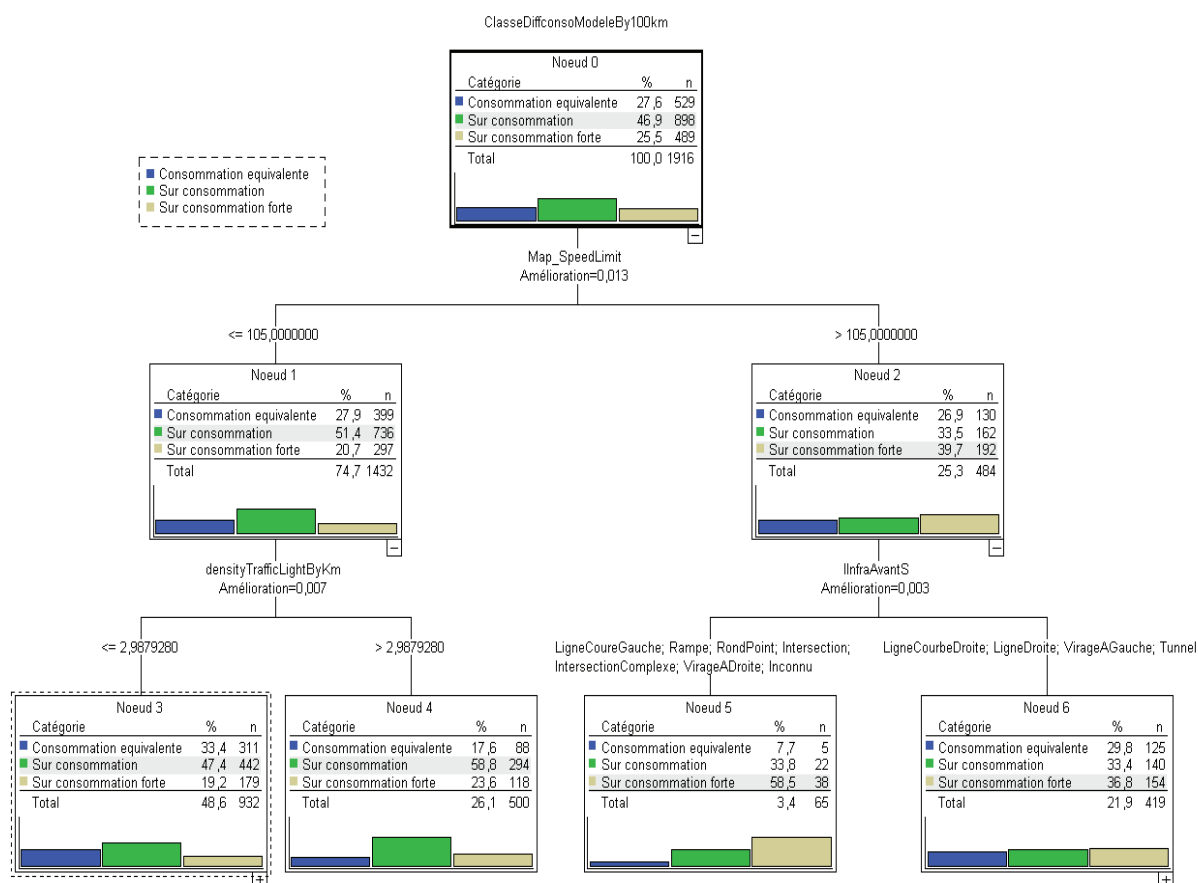


Figure A.6 : Arbre de classification décrivant l'infrastructure de la route.

La classification avec les paramètres décrivant l'infrastructure de la route nous permet d'avoir les règles de classification suivante grâce aux nœuds terminaux. Voici quelques exemples de

règles de classification identifiant les classes « *Consommation équivalente* » et « *Sur consommation forte* » :

$R_{\text{Infra}1_{\text{nœud}5}} = \text{Si } \mathbf{Map_SpeedLimit} > 105 \text{ Et } \mathbf{InfraAprèsS} = \text{'Rampe, RondPoint, Intersection, VirageADroite, IntersectionComplexe'}$

$\rightarrow \text{ClasseDiffConso} = \text{«Sur consommation forte»}$.

La règle *RInfra1* est donnée par le nœud final 5 de la Figure A.6, elle permet le reclassement des situations de conduites grâce aux paramètres vitesse limite de la route et type d'infrastructure avant la situation de conduite « *AllerToutDroit* ». Cette règle permet de reclasser certaines situations de conduite en « *Sur consommation forte* » au lieu de la classe « *Sur consommation* » car la consommation optimisée est plus basse que la consommation réelle. Dans cette règle, une optimisation de la consommation est possible lorsque le conducteur arrive par exemple sur une intersection à la fin d'une voie rapide (la sortie d'autoroute). Cette règle identifie donc cette situation de conduite dans laquelle un système d'éco-conduite pourrait être utile pour gérer les sorties d'autoroute ou l'arrivée sur une intersection ou un rondpoint à la suite d'une voie rapide.

Nous continuons l'exploration de l'arbre de classification des paramètres décrivant l'infrastructure de la route avec le nœud non terminal 6 de la Figure A.6. Pour être un élément du nœud 6 de la Figure A.6, les situations de conduite doivent respecter la règle suivante :

$R_{\text{Infra}2_{\text{nœud}6}} = \text{Si } \mathbf{Map_SpeedLimit} > 105 \text{ Et } \mathbf{InfraAvantS} = \text{'LigneDroite, LigneCourbeDroite, VirageAGauche, Tunnel'}$ $\rightarrow \text{IdNœud} = \text{«6»}$.

Elle caractérise les situations « *AllerToutDroit* » sur ligne droite avec une grande vitesse, potentiellement sur autoroute. A partir de ce nœud, nous avons l'arbre de décision suivant :

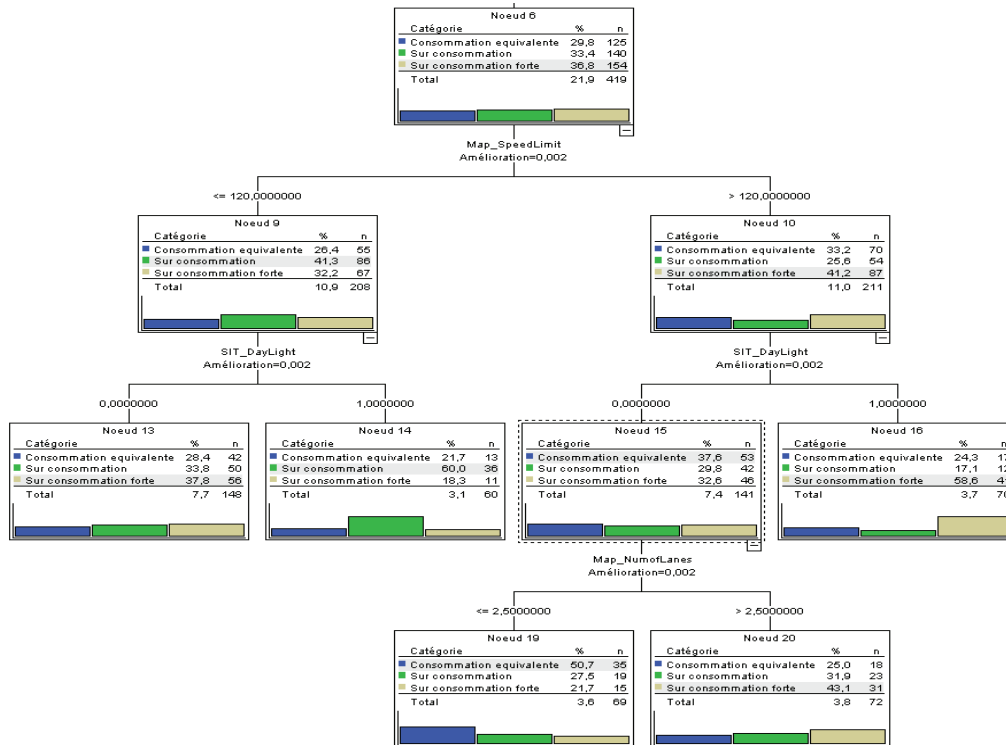


Figure A.7 : Suite arbre de décision à partir du nœud 6 de la classification décrivant l'infrastructure de la route

A partir du nœud 6, nous avons trois nœuds terminaux qui permettent le reclassement des situations de conduite. Il s'agit des nœuds 16, 19 et 20.

$$RInfra3_{noeud16} = Si \ RInfra2_{noeud6} \ Et \ Map_SpeedLimit > 120 \ Et \ Sit_DayLight = 1 \\ \rightarrow ClasseDiffConso = \langle \langle \text{Sur consommation forte} \rangle \rangle.$$

La règle *RInfra3* nœud 16 de la Figure A.7 reclasse les situations de conduite de la nuit sur autoroute dans la classe « *Sur consommation forte* » au lieu de « *Sur consommation* ». Les paramètres type de route (Autoroute) et moment de conduite (Nuit) représentent donc un contexte de sur consommation élevée par rapport au modèle d'optimisation.

$$RInfra4_{noeud19} = Si \ RInfra2_{noeud6} \ Et \ Map_SpeedLimit > 120 \ Et \ Sit_DayLight = 0 \ Et \\ Map_NumofLanes \leq 2 \\ \rightarrow ClasseDiffConso = \langle \langle \text{Consommation équivalente} \rangle \rangle.$$

La règle *RInfra4* du nœud 19 de la Figure A.7 reclasse les situations de conduite de jour sur autoroute avec peu de voies (nombre de voies inférieur ou égale à 2) dans la classe « *Consommation équivalente* » au lieu de « *Sur consommation* ». En plus des paramètres type de route (*Autoroute*) et moment de la conduite (*Jour*), le nombre de voies aussi peut être un indicateur d'une forte consommation en carburant par rapport au modèle d'optimisation. Dans

une situation de conduite sur autoroute en journée avec peu de voies de circulation, la consommation réelle et la consommation optimisée du modèle d'optimisation sont équivalente.

$RInfra5_{noeud20} = Si \ RInfra2_{noeud6} \ Et \ Map_SpeedLimit >120 \ Et \ Sit_DayLight =0 \ Et \ Map_NumofLanes > 2$

→ *ClasseDiffConso* = « **Sur consommation forte** ».

La règle *RInfra5* nœud 20 de la Figure A.7 reclasse les situations de conduite de jour sur autoroute avec un nombre de voies important (supérieur à 2) dans la classe « *Sur consommation forte* » au lieu de « *Sur consommation* ». En plus des paramètres type de route (*Autoroute*) et moment de la conduite (Jour), le nombre de voies supérieur à 2 aussi est un indicateur de « sur consommation forte ». C'est donc un contexte de sur consommation élevée par rapport au modèle d'optimisation.

Les règles de classification identifiées ici grâce aux paramètres contextuels caractérisant l'infrastructure de la route permettent, d'une part de reclasser correctement les situations « *AllerToutDroit* » mal classées par la classification avec les paramètres non contextuels. Et d'autre part, elles permettent d'identifier les contextes de conduites nécessitant une forte sur consommation en carburant par rapport au modèle d'optimisation. Ces situations précises de sur consommation forte peuvent être détectées automatiquement pour mettre en place un système écologique de conduite qui conseillera le conducteur dans ce contexte précis. Le tableau de classification avec les paramètres décrivant l'infrastructure de la route est donné par le tableau suivant :

		Classification			
		Prévisions			
Echantillon	Observations	Consommation équivalente	Sur consommation	Sur consommation forte	Pourcentage correct
Apprentissage	Consommation équivalente	123	324	82	23,3%
	Sur consommation	86	705	107	78,5%
	Sur consommation forte	36	287	166	33,9%
	Pourcentage global	12,8%	68,7%	18,5%	51,9%
Test	Consommation équivalente	70	219	61	20,0%
	Sur consommation	67	437	96	72,8%
	Sur consommation forte	22	193	108	33,4%
	Pourcentage global	12,5%	66,7%	20,8%	48,3%

Méthode de développement : CRT

Variable dépendante : *ClasseDiffconsoModeleBy100km*

Tableau A.10 : Tableau de classification du critère écart de consommation avec les paramètres contextuels décrivant l'infrastructure de la route

La classification avec les paramètres de l'infrastructure de la route permet de reclasser 51,9 % des situations de la base d'apprentissage et 48,3 % de la base de test. Après la classification

avec les paramètres contextuels caractérisant l'infrastructure, ceux décrivant le conducteur a été introduit pour reclasser les situations de conduite mal classées.

La classification est donnée par l'arbre de décision suivant :

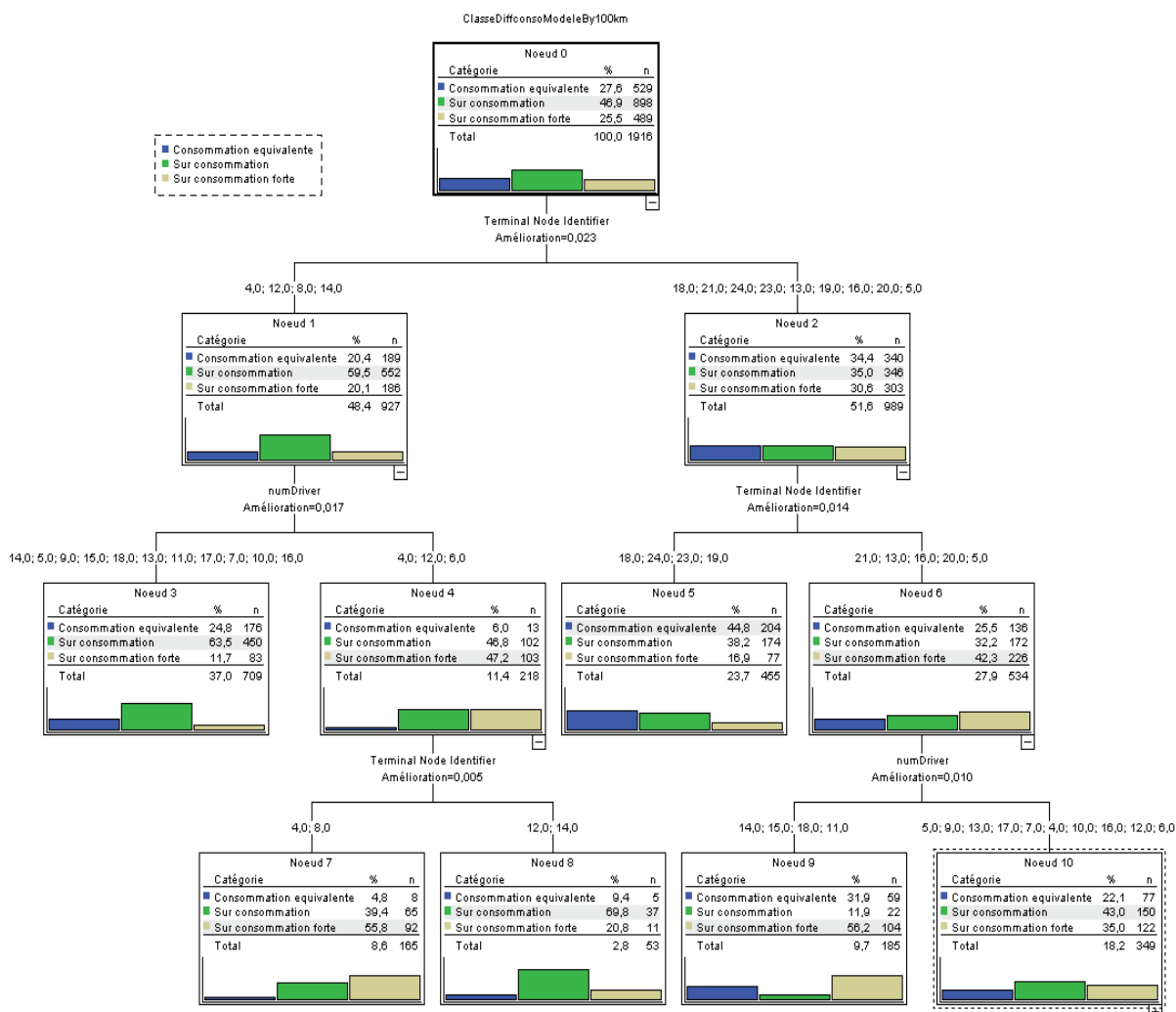


Figure A.8 : Arbre de décision de la classification avec les paramètres contextuels décrivant le conducteur

Cet arbre de décision nous permet d'avoir les règles de classification suivante :

$R_{Driver1_{noeud7}} = Si IDNodeInfra = 4 ; 8 Et NumDriver = 4 ; 12 ; 6$

$\rightarrow ClasseDiffConso = \ll \text{Sur consommation forte} \gg.$

La règle $R_{Driver1}$ du nœud 7 de l'arbre de décision permet de reclasser les situations « *AllerToutDroit* » en sur consommation forte grâce aux paramètres $IDNodeInfra$ et identifiant du conducteur ($NumDriver$). Le paramètre $IDNodeInfra$ correspond aux situations de conduite respectant les règles menant aux nœuds 4 et 8 de l'arbre de décision avec les

paramètres décrivant l'infrastructure de la route. La règle de classification menant au nœud 4 est la suivante :

$$RInfra_{noeud4} = Si \textit{Map_SpeedLimit} \leq 105 \textit{ Et densityTrafficLigthBykm} > 3 \\ \rightarrow IDNodeInfra = 4.$$

Cette règle caractérise les situations de conduite en ville avec une forte densité de feux tricolores.

$$RInfra_{noeud8} = Si \textit{Map_SpeedLimit} \leq 105 \textit{ Et densityTrafficLigthBykm} \leq 3 \textit{ Et InfraAvantS} \\ = \textit{'Rampe, Intersection, VirageADroite, VirageAGauche, Tunnel'}$$
$$\rightarrow IDNodeInfra = 8.$$

La règle *RInfra* du nœud 8 de l'arbre de décision de la classification avec les paramètres de l'infrastructure caractérise les situations de conduite en ville avec peu de feux tricolores précédée par exemple par une intersection ou un tunnel.

A partir de ces nœuds de la classification avec les paramètres décrivant l'infrastructure, la règle *RDriver* 1 nous permet de reclasser les situations dans la classe « *Sur consommation forte* » pour certains conducteurs, dont l'identifiant est égal à : 4, 6, 12. L'identifiant du conducteur est un paramètre explicatif de la sur consommation élevée par rapport au modèle d'optimisation car il permet de reclasser correctement certaines situations de conduite.

$$RDriver2_{noeud9} = Si \textit{IDNodeInfra} = 5 ; 13 ; 16 ; 20 ; 21 \textit{ Et NumDriver} = 11 ; 14 ; 15 ; 18 \rightarrow \\ \textit{ClasseDiffConso} = \textit{\langle\langle Sur consommation forte \rangle\rangle}.$$

La règle *RDriver2* du nœud 9 classe les situations de conduite dans les nœuds 5, 13, 16, 20 et 21 de la classification avec les paramètres décrivant l'infrastructure de la route avec les conducteurs 11, 14, 15, 18 sont reclassées en sur consommation forte par les paramètres *IDNodeInfra* et *NumDriver*.

Avec les paramètres décrivant le conducteur, le taux de classification global est de 55,1 % sur la base d'apprentissage et 52,5 % sur la base test. Une petite augmentation de 3,2 % pour la base d'apprentissage est constatée avec les paramètres décrivant le conducteur. Pour la base de test, une augmentation de 4,2 % est constatée grâce aux paramètres caractérisant le conducteur.

Le tableau de classification avec les paramètres du conducteur est donné par le tableau suivant :

Classification					
		Prévisions			
		Consommation équivalente	Sur consommation	Sur consommation forte	Pourcentage correct
Echantillon	Observations				
Apprentissage	Consommation équivalente	204	243	82	38,6%
	Sur consommation	174	604	120	67,3%
	Sur consommation forte	77	164	248	50,7%
	Pourcentage global	23,7%	52,8%	23,5%	55,1%
Test	Consommation équivalente	131	167	52	37,4%
	Sur consommation	112	377	111	62,8%
	Sur consommation forte	48	115	160	49,5%
	Pourcentage global	22,9%	51,8%	25,4%	52,5%

Méthode de développement : CRT

Variable dépendante : ClasseDiffconsoModeleBy100km

Tableau A.11 : Tableau de classification du critère écart de consommation avec les paramètres contextuels décrivant le conducteur

Liste des Acronymes

AEE	Agence Européenne de l'Environnement
ASC	Aspect-Scale Context
CAN	Controller Area Network
CC	Cruise Control
CC/PP	Composite Capabilities/ Preference Profile
CEESAR	Centre européen d'Etude de Sécurité et d'analyse des Risques
CCML	Centaurus Capability Markup Language
CML	Context Modeling Language
CoOL	Context Ontology Language
CONON	CONtext ONtology
CSCP	Comprehensive Structure Context Profile
EIAH	Environnements Informatiques pour l'Apprentissage Humain
EuroFOT	European Field Operationnal Test
GES	Gaz à effet de serre
GIEC	Groupe Intergouvernemental d'Etude du Climat
GPS	Global Positioning System
IC	Ingénierie des Connaissances
Ifsttar	Institut Français des Sciences et Technologie des Transports, de l'Aménagement et des Réseaux
KTBS	Kernel for Trace-Based Systems
Lescot	Laboratoire Ergonomie et Sciences Cognitives pour les Transports
LIRIS	Laboratoire d'Informatique en Images et systèmes d'Information
MEDDE	Ministère de l'Ecologie du Développement Durable et de l'Energie
MOOCS	Massively Open Online Courses
M-Trace	Trace Modélisée
ORM	Object-Role Modeling
OWL	Ontology Web Language
OWL-DL	Ontology Web Language Description Logis
PPDL	Pervasive Profile Description Language
RDF	Resource Description Framework
RDF/S	Resource Description Framework Schema
SBT	Système à Base de Traces
SBTm	Système à Base de Traces modélisées

SICAL	Situated Interaction, Collaboration, Adaptation and Learning
SID	Système Informatique Diffus
SGBT	Système de Gestion de Base de Traces
SGML	Standard Generic Markup Language
SL	Speed limiter
SOUPA	Standard Ontology for Ubiquitous and Pervasive Applications
TWEAK	Traces, Web, Education, Adaptation, Knowledge
UML	Unified Modelling Language
URL	Uniform Resource Locator
XML	Extensible Markup Language