Decision making strategy for antenatal echographic screening of foetal abnormalities using statistical learning

Rémi Besson

NNT : 2019SACLX037

THESE DE DOCTORAT
DE L'UNIVERSITE PARIS-SACLAY

préparée à

L'ÉCOLE POLYTECHNIQUE

ÉCOLE DOCTORALE N°574
École doctorale de mathématiques Hadamard (EDMH)

Spécialité de doctorat : Mathématiques

par

Rémi Besson

Decision support for prenatal ultrasound screening of the fetus anomalies using statistical learning

Version du 21/10/2019

Thèse présentée et soutenue à Palaiseau, le 1er Octobre 2019.

Après avis des rapporteurs:

Tristan Cazenave	(Pr., LAMSADE, Université Paris-Dauphine)
Jean-Michel Loubes	(Pr., IMT, Université Toulouse Paul Sabatier)

Composition du jury:

Stéphanie Allassonnière	(Pr., Ecole de Médecine-Université Paris-Descartes)	Directrice de thèse
Tristan Cazenave	(Pr., LAMSADE,Université Paris-Dauphine)	Rapporteur
Antoine Chambaz	(Pr., MAP5, Université Paris-Descartes)	Président du jury
Erwan Le Pennec	(Pr., CMAP, École Polytechnique)	Directeur de thèse
Jean-Michel Loubes	(Pr., IMT, Université Toulouse Paul Sabatier)	Rapporteur
Karim Lounici	(Pr., CMAP, Ecole Polytechnique)	Examinateur
Julien Stirnemann	(MCU-PH , Hôpital Necker-Enfants malades)	Examinateur

Remerciements

> Quand tu veux te réjouir, réfléchis aux vertus de ceux qui vivent avec toi, à l'activité de celui-ci, à la modestie de celui-là, à la libéralité d'un troisième, à telle autre qualité pour chacun. Rien ne nous réjouit autant que de nous représenter les vertus qui brillent dans la vie de ceux qui nous entourent et de les voir se rencontrer presque en foule. Il faut donc être toujours prêt à te les rappeler.
>
> Pensée 48, Livre VI, Marc Aurèle,
> Pensées pour moi-même.

Mes premiers remerciements vont à mes deux encadrants de thèse Erwan et Stéphanie. Cela faisait déjà plusieurs années que je voulais faire des mathématiques appliquées à la médecine. Je dois cependant bien avouer que si j'ai choisi ce projet, c'est aussi parce que c'étaient eux qui l'encadreraient. Un grand merci à Erwan pour son enthousiasme, son sens de la diplomatie, sa bienveillance et sa vaste culture mathématiques que j'ai eue l'occasion d'éprouver durant ces trois années. Ce travail de recherche a donné naissance à une start-up qui fait actuellement ses premiers pas et cela n'aurait pas été possible sans le dynamisme, le soutien sans faille et l'ambition de Stéphanie, merci à toi !

Merci à Tristan Cazenave et Jean-Michel Loubes d'avoir accepté de bonne grâce cette tâche, ô combien ingrate, mais indispensable pour moi, de rapporter ma thèse.

Merci à Antoine Chambaz et Karim Lounici d'avoir accepté de faire partie de mon jury.

Merci à tous ceux qui ont participé à ce projet et continuent l'aventure avec cette start-up. Merci à Julien, Yves et Emmanuel qui ont réussi à dégager du temps pour faire avancer ce projet même entre deux opérations. Merci à Antoine de faire habilement le lien avec les médecins à l'heure de modifier les bases de données.

Merci à Camille et Capucine, les deux sages-femmes qui ont testé de jour comme de nuit l'algorithme que je leur fournissais.

Je n'oublie pas Anne-Sophie Jannot qui m'a, la première, parlé de ce projet. A Christophe Giraud qui m'a donné la possibilité de suivre son Master Mathématiques pour les sciences du vivant qui fût probablement le meilleur tremplin de mes études et pour sa disponibilité à l'heure de conseiller les étudiants de M2. A la FMJH de m'avoir donné la possibilité avec une bourse de faire ce master. J'ai également une pensée pour les professeurs qui m'ont marqué durant mes études : Madame Tierno, Monsieur Ayache, Monsieur Leconte, Monsieur Adoumié, el señor López Blázquez et j'en oublie certainement...

Une thèse, c'est aussi un laboratoire et des collègues qui deviennent des amis. Mention spéciale à Frédéric et Kevish. A Frédéric qui a souvent eu à supporter mes explications approximatives d'une idée qui n'est pas encore formée, qui a relu la plupart de mes soumissions mais aussi et surtout pour être un camarade de débat politique, d'escalade, de voyages, de blagues de plus ou moins bon goût. Merci à Kevish pour sa bonne humeur immuable, pour nos discussions souvent perchées, j'espère bien te rendre visite au Colorado! Et désolé, mais je crois que je suis meilleur que toi au ping-pong.

A Thomas pour nos longues balades aux quatre coins de l'île-de-France et du Cantal. Eh fra eh j'tais à saturation, même quand yavait plus dégun t'y étais là pour me rappeler le terre-terre. Dans un an, tu sais où me trouver, on prendra le thé à la menthe à Noailles avant de piquer une tête à Sormiou le zin.

Merci à Rémy de m'avoir si bien accueilli à Annecy et d'avoir été un modèle par son abnégation face à la maladie. Merci à Frédérique, malgré le fait que tu m'aies donné un sacré coup de vieux en annonçant ton mariage.

Merci aux membres du bureau 2015 : Corentin C., Corentin H., Behlal pour sa faculté à faire le zouave pour être une fraction de seconde plus tard être entièrement absorbé à sa tâche, Aude pour sa bonne humeur et ses fameux boxplots, et ceux partis trop tôt Geneviève, Luca. Merci aux doctorants du CMAP qui font de ce lieu un laboratoire où il fait bon travailler Fedor, Mathieu, Othmane M., Heythem, Antoine, Juliette, Vianney, Céline B.

Merci à l'équipe administrative du CMAP, Nasséra, Alexandra L., Alexandra N. et Maud aux petits soins pour nous faciliter la vie. Merci à Elodie que j'ai monopolisé une bonne partie de l'année pour d'obscures démarches administratives de doctorat conseil. Merci à Pierre pour son efficacité redoutable à l'heure de régler les petits tracas informatiques.

A Marie pour nos sorties qui ont été comme des bouffés d'oxygène au moment de clôturer cette thèse. A Delia y a su familia que me han acompañado durante una buena parte de este viaje. A Liz Pepin pour son oreille attentive. A Céline dont l'influence récente est loin d'être négligeable.

Merci à Othmane Z. d'avoir repris avec brio l'organisation de l'équipe de foot LMS-CMAP et à toute l'équipe, dont je compte bien continuer à faire partie. La saison

2019/2020 sera, je l'espère, encore plus belle que la précédente.

Merci à Notre-Dame-de-la-Garde qui veille sur les Marseillais en exil. A la Méditerranée, à la garrigue et à la caresse du soleil qui m'ont tant manqué lorsque les semaines de crachin parisiens s'enchaînaient.

Enfin, je souhaiterais remercier ma famille. Merci à mes parents qui m'ont toujours soutenu dans mes études et dans ma vie en générale, cette thèse c'est aussi un peu la vôtre. A ma petite sœur Eva qui fait une brillante philosophe, à moins que cela soit une grande actrice en devenir (je compte sur toi si ça marche). J'ai une pensée pour mon petit frère Théo et ma cousine Sarah qui commencent leur thèse. Aux Grauby, Clémentine, Christine et Olivier pour les vacances en camping. A mes grands-parents Pierre et Liliane, chez qui j'ai passé tant de bons moments.

Contents

NOTATIONS

B	Symptom	Symptôme
D	Disease	Maladie
$\mathcal{B}(D)$	Set of symptom typical of disease D	Ensemble des symptômes typiques de D
$\mathbb{B}_i$	Set of symptoms typical of at least one disease in common with diseases for which B_i is typical	Ensemble des symptômes typiques d'au moins une maladie commune avec les maladies dont B_i est typique
H	Entropy	Entropie
$\mathbb{KL}$	Kullback-Leibler divergence	Divergence de Kullback-Leibler
π	Diagnostic strategy	Stratégie de diagnostic
μ	Behaviour policy	Stratégies générant les données
$\mathbb{S}$	State space	Espace des états
$\mathbb{A}$	Action space	Espace des actions
s_t	State visited at time t	Etat visité au temps t
a_t	Action taken at time t	Action prise à l'instant t
r_t	Instantaneous reward received at time t	Récompenses instantanée reçu au temps t
$V_\pi(s)$	Value function	Fonction de valeur
$Q_\pi(s,a)$	State-action value	Valeur action-état
$s_{(i)}$	State with presence of symptom B_i and no other information	Etat avec présence du symptôme B_i et aucune autre information
S_t	Random variable of which s_t is a realization	Variable aléatoire dont s_t est une réalisation
I	Number of inquiries before reaching a terminal state	Nombre de question à poser avant d'atteindre un état terminal
$\mathcal{P}$	Law of the environment	Loi de l'environnement

Chapter 1

INTRODUCTION EN LANGUE FRANCAISE

Résumé: *Afin de planter le décor, nous commencons par un bref tour d'horizon du monde du diagnostic prénatal en évoquant les enjeux, les acteurs et les techniques mobilisées. Nous évoquons ensuite les données à disposition et les informations que nous avons pu tirer des différentes bases existantes en libre accès pour les maladies rares. Nous continuons en explicitant le besoin actuel des échographistes pour un outil d'aide à la décision précisant par la même occasion les propriétés désirables que doit posséder notre algorithme final. Nous finissons en essayant de prendre de la hauteur et de formuler les questions mathématiques plus générales qui sont abordées dans cette thèse.*

Contents

Les travaux présentés dans cette thèse sont le résultat de mes trois ans de recherche sous la direction de mes deux directeurs de thèse, Erwan Le Pennec et Stéphanie Allassonnière, sur un sujet à l'interface mathématiques/médecine.

L'objectif de ce projet est de fournir aux praticiens du diagnostic prénatal un outil d'aide à la décision lors de la recherche de maladie rare par échographie.

Mon travail durant ces trois années s'est inscrit au sein d'une équipe pluridisciplinaire regroupant des mathématiciens, Erwan, Stéphanie et moi-même, des médecins obstétriciens, Julien Stirnemann, Emmanuel Spaggiari et Yves Ville, ainsi qu'un médecin spécialiste de santé publique, Antoine Neuraz.

1.1 Le dépistage et le diagnostic prénatal

1.1.1 Enjeux du diagnostic prénatal

La France enregistre chaque année environ 780.000 naissances dont 39.000 pour la seule région parisienne et on estime que 3 à 4 pour cent de ces grossesses sont affectées d'au moins une anomalie congénitale (voir Khoshnood B [2016]).

Les anomalies congénitales sont une cause importante de mortalité infantile: 21 % des décès dans la première année de la vie sont imputables à de telles anomalies d'après l'Institut de veille sanitaire [1]. Elles représentent également une partie conséquente des handicaps.

Il est donc clair que l'ensemble des malformations congénitales représente un enjeu de santé publique. Cependant si l'existence d'une anomalie congénitale est un évènement relativement fréquent, le grand nombre de pathologies différentes possibles dont un grand nombre sont des maladies rares, rend la tâche de l'obstétricien difficile.

1.1.2 Les outils mobilisables pour le diagnostic prénatal

L'échographie, parce qu'elle constitue une technique non-invasive et peu coûteuse, reste l'outil privilégié du dépistage prénatal. En France, la femme enceinte est tenue de suivre trois examens échographiques de contrôle à 12, 21, puis 33 semaines d'aménorrhée (absence

1. http://invs.santepubliquefrance.fr//Dossiers-thematiques/Maladies-chroniques-et-traumatismes/Malformations-congenitales-et-anomalies-chromosomiques. Accessed [04/03/2019]

de règles). Si une anomalie est détectée ou suspectée, la patiente est orientée vers un médecin spécialiste pour une échographie de seconde intention, dite aussi "échographie de diagnostic".

Si un risque significatif d'anomalie chromosomique (par exemple une trisomie 21) est identifié, une amniocentèse doit être proposée à la mère. L'amniocentèse consiste à prélever du liquide amniotique par ponction abdominale. L'étude du caryotype des cellules fœtales doit permettre de diagnostiquer avec certitude l'anomalie chromosomique suspectée. Elle est effectuée vers la 17e semaine d'aménorrhée. Cette technique intrusive a longtemps été considérée comme comportant des risques pour le foetus, le chiffre de 1% d'augmentation de risque de fausse couche attribué à l'amniocentèse a longtemps été avancé. Cependant, une étude récente de médecins de Necker a montré, lors d'essais randomisés sur 2051 femmes, qu'en utilisant les techniques actuelles il n'existe pas d'augmentation statistiquement significative du risque de fausse couche Malan et al. [2018]. Il n'en reste pas moins que cet examen est plus coûteux et complexe qu'une échographie et n'est donc pas utilisé en routine pour du dépistage.

De plus, insistons ici sur le fait que l'amniocentèse permet de détecter uniquement les syndromes dont les anomalies du caryotype sont bien connus ce qui est loin d'être le cas de toutes les maladies rares. Ces considérations expliquent le choix des syndromes retenus pour notre base de données (voir section 1.2): il doit s'agir de syndrome pour lequel il n'existe pas déjà un moyen fiable et sans risque de les identifier.

Enfin citons l'existence des examens d'imagerie par résonance magnétique (IRM) et de tomodensitométrie (TDM). Ces examens sont réalisés à la suite d'une échographie de diagnostic lorsque celle-ci a mis en évidence une anomalie. Ils doivent permettre "d'étudier certains organes ou certaines structures du foetus ou de ses annexes de façon différente et complémentaire à l'échographie. Les régions anatomiques les plus souvent explorées en IRM sont le système nerveux central, le thorax et l'abdomen du fœtus. La TDM explore le squelette fœtal." [2]

Notons donc la place particulièrement importante de l'échographie dans le dépistage prénatal puisqu'utilisée systématiquement en routine. Celle-ci constitue également la meilleure solution pour diagnostiquer un certains nombres de malformations. En effet, comme le rappelle l'arrêté du 20 Avril 2018 sur les bonnes pratiques dans le diagnostic prénatal: "l'échographie obstétricale et fœtale est dans certains cas le seul examen permettant d'identifier un risque pour le fœtus ou la grossesse en cours, de diagnostiquer une pathologie fœtale ou obstétricale et de la surveiller".

2. Extrait de l'annexe 1 de l'arrêté du 20 avril 2018 fixant les recommandations de bonnes pratiques relatives aux modalités de réalisation des examens d'imagerie concourant au diagnostic prénatal.

1.1.3 Les acteurs

Comme précisé dans l'article 1 du décret du 2 mai 2017 (2017-702) seuls les "sages-femmes et médecins disposant de compétences reconnues par un diplôme en attestant" sont habilités à réaliser les échographies obstétricales et foetales. Les prérogatives de ces deux corps de métiers ne sont cependant pas identiques.

Les sages-femmes sont titulaires d'un diplôme d'Etat les habilitant à mener les échographies de dépistage (de contrôle). Les échographies dites de diagnostic ne font, à l'inverse, pas partie de leur champ de compétences. Leur profession est encadrée par les articles L4151 du code de la santé publique.

Les médecins habilités à suivre la grossesse d'une femme sont les obstétriciens. Lorsqu'une anomalie congénitale grave est suspectée, les patients sont orientés vers les Centres Pluridisciplinaires de Diagnostic Prénatal (CPDPN), qui étaient au nombre de 49 sur le territoire français en 2015. Le diagnostic final précisant la nature et le niveau de gravité de la malformation est posé de manière collégiale par au moins deux médecins spécialistes de diagnostic prénatal.

1.1.4 Les solutions thérapeutiques possibles

Lorsqu'une malformation congénitale est diagnostiquée, la question de la prise en charge est alors évoquée. La prise de décision est toujours faite de manière collégiale par des médecins de différentes spécialités et en lien avec les Centres Pluridisciplinaires de Diagnostic Prénatal. La décision finale revenant toujours, bien sûr, aux parents.

Si l'affection de l'enfant à naître est identifiée comme étant d'une particulière gravité, c'est-à-dire impliquant un handicap lourd sans solution thérapeutique existante, une interruption médicale de grossesse (IMG) est envisageable et ce jusqu'au dernier jour de grossesse. Il s'agit de l'Article L2213-1 du Code de la santé publique:

> L'interruption volontaire d'une grossesse peut, à toute époque, être pratiquée si deux médecins membres d'une équipe pluridisciplinaire attestent, après que cette équipe a rendu son avis consultatif, soit que la poursuite de la grossesse met en péril grave la santé de la femme, soit qu'il existe une forte probabilité que l'enfant à naître soit atteint d'une affection d'une particulière gravité reconnue comme incurable au moment du diagnostic.

Notons qu'il n'est pas rare que l'IMG soit refusée par les parents alors que le prognostic vital néonatal est très pessimiste Guibet Lafaye [2009]. Dans de tels cas le diagnostic prénatal aura permis d'envisager en amont les soins palliatifs éventuels à fournir au nouveau-né dès sa naissance Charlot [2011].

Il convient ici d'embellir quelque peu ce sombre tableau et de souligner les progrès récents de la chirurgie foetale source d'espoir pour l'avenir. Ainsi le prognostic vital des foetus atteint du syndrome de transfuseur-transfusé pour des grossesses multiples a été

grandement amélioré avec les avancées récentes de la chirurgie in foeto. Sans intervention la mortalité périnatale est de 90 % et des séquelles neurologiques graves sont retrouvées chez 20 à 40 % des survivants. Les techniques récentes permettent la survie d'un des deux jumeaux dans près de 80% des cas et un risque de séquelle chez les survivants de moins de 10 % Salomon and Ville [2008]. Citons également les opérations in foeto possible pour les malformations de type spina-bifida: une anomalie de fermeture du tube neural dont la prévalence est d'une naissance sur 2000. Il a été prouvé que cette opération améliore de manière notable le prognostic du handicap de l'enfant à naître par rapport à la solution alternative d'opérer en néonatal (voir Adzick et al. [2011]). De manière générale, c'est seulement lorsqu'une telle démonstration a été faite que la chirurgie foetale est considérée comme une alternative pleinement justifiée puisque, bien sûr, celle-ci n'est pas sans risque pour la mère et l'enfant qu'elle porte. Cette technique est cependant suffisemment sûre aujourd'hui pour être proposée même dans le cas de malformations non létales (A Deprest et al. [2010]).

1.1.5 Les questions de responsabilités pénales

Nous ne prétendons pas, bien sûr, donner un aperçu exhaustif des questions de responsabilité pénale que fait émerger le diagnostic prénatal. Le lecteur imaginera sans peine combien ces questions peuvent être complexes et sans cesse en évolution.

En France, deux organismes d'Etat distincts ont été créés pour régler les contentieux entre patient et soignants afin d'offrir une alternative rapide et gratuite face à la voie judiciaire. Il s'agit de la commission de conciliation et d'indemnisation des accidents médicaux (CCI) et de l'Office National d'Indemnisation des Accidents Médicaux (ONIAM). La CCI a pour fonction de fournir une expertise pour l'établissement d'une responsabilité éventuelle du soignant ou de l'établissement de santé. L'ONIAM a plus spécifiquement pour mission d'organiser le dispositif d'indemnisation.

Si une responsabilité est établie par la CCI, l'assureur du professionnel de santé doit faire une proposition d'indemnisation qui sera acceptée ou refusée par la victime, en cas de refus celle-ci devra se tourner vers les tribunaux compétents.

Le nombre de réclamations envoyées à la CCI est en constante augmentation depuis les années 2000 (voir figure 1.1). La gynécologie obstétrique fait partie des spécialités pour lesquelles les coûts d'indemnisation judiciaire sont les plus élevés. La Mutuelle d'assurances du corps de santé français (MACSF) publie chaque année son rapport sur les procédures d'indemnisation et, en 2017, la gynécologie obstétrique arrive en troisième place avec presque 6 millions d'euros d'indemnisation (voir figure 1.2).

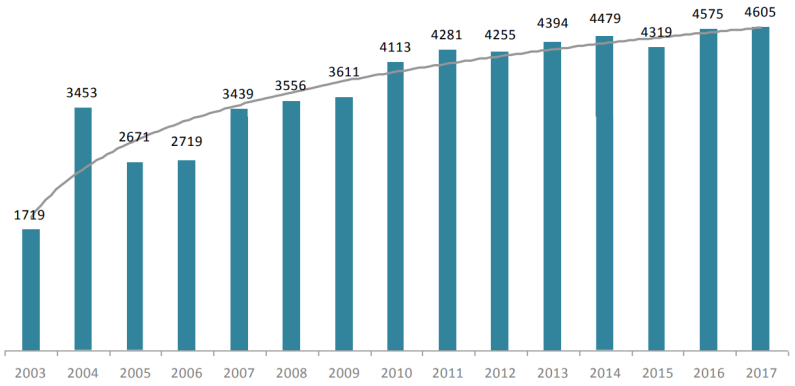


Figure 1.1 – Evolution du nombre d'entrées dans le dispositif de la commission de conciliation et d'indemnisation des accidents médicaux (source: Office National d'Indemnisation des Accidents Médicaux rapport annuel de 2017).

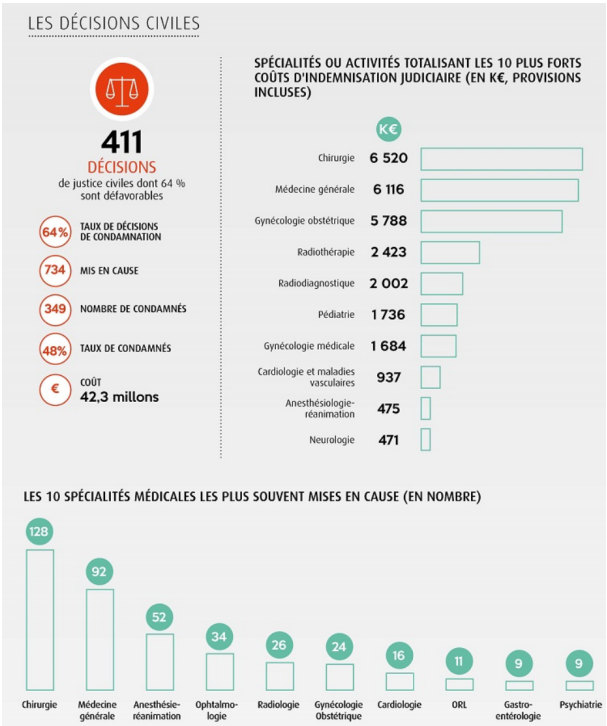


Figure 1.2 – Les décisions aux tribunaux civils concernant les litiges en médecine pour l'année 2017 selon les différentes spécialités (source: MACSF - Le Sou Médical, rapport annuel).

1.2 Les données à disposition

1.2.1 Une première base experte

Nous disposons d'une liste de maladies et pour chaque maladie d'une liste de symptôme généralement associés (que nous appellerons symptômes typiques). Nous écrivons :

$$B_i = \begin{cases} 1 & \text{si le patient a le symptôme de type i} \\ 0 & \text{sinon}\,. \end{cases}$$

Nous notons D la variable aléatoire associée à la maladie du patient : $D \in \{d_1, ..., d_k\}$. Nous faisons l'hypothèse que le patient ne présente qu'une maladie à la fois ce qui est une hypothèse raisonnable pour les maladies rares. $\{D = d_k\}$ représente l'évènement pour lequel le patient est sain ou est atteint d'une maladie qui n'est pas dans la base de données.

Nous disposons d'une estimation de la prévalence de chaque maladie, ainsi que de la probabilité de présenter chaque symptôme typique sachant la maladie.

Nous connaissons donc $\mathbb{P}[D = d_j]$ abrégé en $\mathbb{P}[D_j]$ et également $\mathbb{P}[B_i = 1 \mid D = d_j]$ abrégé en $\mathbb{P}[B_i \mid D_j]$ lorsque B_i est un symptôme typique de D_j. Ajoutons qu'il est possible d'avoir une maladie et de présenter un symptôme atypique mais nous n'avons pas d'estimation pour de tels évènements. Nous supposerons que les symptômes atypiques se manifestent avec une faible probabilité (fixée à 10^{-5}) et indépendamment des autres symptômes (conditionnellement à la maladie).

Un élément important est que nous ne disposons pas des lois jointes des symptômes sachant la maladie mais seulement des marginales. Nous abordons cette difficulté dans le chapitre 4.

Actuellement, notre base de données référence 81 maladies et 220 symptômes différents. La maladie qui présente le plus grand nombre de symptômes associés est le syndrome VACTERL, avec 19 symptômes possibles.

Toutes ces informations, que nous désignons dans la suite sous le terme de données expertes, nous ont été fournies par des médecins de l'hôpital Necker-Enfants malades en se basant sur la littérature disponible.

Dans la suite nous utiliserons de manière indistincte les termes de syndrome, pathologie ou maladie. De même les termes symptômes, anomalies et signes seront également utilisés de manière équivalente dans ce travail. Conformément à la définition du Larousse nous entendons par syndrome un ensemble de plusieurs symptômes en rapport avec un état pathologique donné et permettant, par leur groupement, d'orienter le diagnostic.

Il s'agit là de la manière classique de représenter informatiquement le concept de maladie bien que cette approche ne soit pas sans difficulté notamment lorsque la frontière entre le sain et le pathologique est brouillée Fagot-Largeault [1985].

1.2.2 Confrontation à des bases plus générales et extraction de l'arborescence des signes

A notre connaissance, il n'existe pas de base telle que la notre en libre accès, c'est-à-dire une base composée d'une liste de syndromes rares avec la liste des signes associés observables à l'échographie.

La base en libre accès la plus proche de ce que nous désirons est sans doute celle d'OrphaData [3]. Cette base de données sur les maladies rares et orphelines est mise à disposition par Orphanet, un organisme créé en France par l'INSERM en 1997 et aujourd'hui soutenu par l'Union Européenne avec une quarantaine de pays partenaires.

Cette base de données n'est toutefois pas directement exploitable en tant que telle pour notre problème car elle ne se limite pas à l'obstétrique. Ainsi de nombreux signes de la base ne sont pas observables à l'échographie et beaucoup de syndromes ne sont observables qu'en néonatal.

Nous avons commencé par mettre les symptômes de notre base initiale en correspondance avec ceux de la base Human Phenotype Ontology (HPO) Köhler et al. [2017]. HPO est un travail récent qui fournit une terminologie standardisée des anomalies phénotypiques retrouvées dans les pathologies humaines. Nous nous en sommes servis pour harmoniser la terminologie.

Nous avons ensuite mis cette liste de symptômes par maladie en correspondance avec OrphaData. Cela nous a permi d'imputer les valeurs manquantes de notre base initiale pour les prévalences des symptômes dans les maladies. En effet, OrphaData fourni l'information pour chaque signe sachant la maladie s'il est obligatoire (100% des cas), très fréquent (entre 80 et 90% des cas), fréquent (entre 30 et 79%) ou occasionnel (entre 5 et 29%). Lorsque notre base initiale avait une valeur manquante nous avons imputé la valeur médiane de la fourchette proposée par OrphaData. Lorsque notre base initiale comportait déjà une valeur pour la prévalence du symptôme dans la maladie nous l'avons conservée sans inclure la prévalence d'OrphaData.

Il nous a également été possible d'extraire de HPO l'information sur l'ontologie des symptômes de notre base. Par ontologie nous entendons ici le fait qu'un symptôme donné puisse être décrit à divers niveau de granularité: par exemple "*anomalie cardiaque*" est un ascendant de "*Tétralogie de Fallot*" (qui est une anomalie cardiaque spécifique). Notre outil d'aide à la décision doit être capable d'intégrer ce type de raisonnement classique en médecine (voir le chapitre 5 pour plus de détail).

Actuellement la base de données OrphaData est utilisée pour ajouter de nouveaux syndromes et leurs symptômes associés afin d'enrichir notre base initiale. Une telle opération nécessite une expertise médicale afin de déterminer les maladies et symptômes à conserver de la base OrphaData, i.e les symptômes observables à l'échographie.

Mentionnons enfin l'existence d'une autre base de données: EUROCAT. Celle-ci est

3. Disponible sur http://www.orpha.net. Consulté le [02/10/2018].

le résultat d'un projet à l'échelle européenne visant depuis 1979 à assurer la surveillance épidémiologique des anomalies congénitales. Le taux de couverture de cette base est relativement élevée puisque pas moins de 29% des naissances en Europe sont recensées. Cette base nous sera utile pour obtenir les taux de diagnostic des différentes anomalies durant le diagnostic prénatal : certaines malformations sont en effet plus difficiles à observer que d'autres à l'échographie et il peut être intéressant d'incorporer cette information à notre outil.

1.2.3 Données cliniques

Comme dans de nombreuses applications médicales l'accès à des données cliniques est difficile. Notre outil d'aide à la décision a cependant vocation à collecter des données au fur et à mesure de son utilisation. Ces données devront nous permettre de mieux inférer la loi de notre environement initialisé par les données expertes.

De manière générale nous espérons que ce travail constituera une opportunité de bâtir une base de données globalisée permettant de mieux connaître les maladies rares diagnosticables par échographie foetale.

1.3 Le besoin médical

L'effort visant à la construction d'un outil d'aide à la décision pour le diagnostic prénatal apparaît pertinent pour plusieurs raisons que nous évoquons ici.

1.3.1 Amélioration du taux de détection des maladies rares

Notre base initiale comporte 81 maladies rares, mais l'objectif final est de pouvoir fournir une aide au diagnostic pour environ 1000 pathologies rares différentes. Le nombre de signes observables à l'échographie est encore beaucoup plus grand. Il est bien évident qu'aucun médecin ne peut avoir la connaissance encyclopédique d'envisager tous les diagnostics.

La figure 1.3 recense les taux de diagnostic par le dépistage anténatal d'un certains nombres de malformations congénitales. Si certaines anomalies sont presque systématiquement diagnostiquées avant la naissance (par exemple une anencéphalie ou encore une agénésie rénale bilatérale) d'autres sont en situation de sous-diagnostic. Ainsi le taux de diagnostic des anomalies non détectables par test génétique lors du diagnostic prénatal est seulement de 39% d'après cette étude d'EUROCAT. Il ne faut bien sûr pas conclure que les 61% des anomalies congénitales non diagnostiquées auraient toutes pu être observées en anténatal. En effet, certaines anomalies peuvent ne pas être détectables du tout ou très difficilement durant l'échographie. Cela est dû aux limitations même de cette technique ou encore à la nécessaire inégalité des niveaux d'expertises des différents praticiens. La figure 1.4 montre bien l'hétérogénéité des taux de diagnostics d'anomalies congénitales selon les différentes régions de l'Europe.

Prenatal diagnosis of 18 selected congenital anomaly subgroups for the following registries: Antwerp (Belgium), Hainaut (Belgium), Zagreb (Croatia), Brittany (France), French West Indies (France), Isle de la Reunion (France), Paris (France), Saxony-Anhalt (Germany), Cork and Kerry (Ireland), Emilia Romagna (Italy), Tuscany (Italy), Malta, N Netherlands (NL), Norway, S Portugal, Valencia Region (Spain), Vaud (Switzerland), Northern England (UK), Thames Valley (UK), Wales (UK), Wessex (UK), Ukraine, from 2012 - 2016

Malformation	Total Cases	Cases Prenatally Diagnosed (% of Total Cases)
Non-chromosomal		
All Anomalies (Excluding genetic conditions)	55057	21452 (39.0%)
Anencephalus and similar (Excluding genetic conditions)	975	964 (98.9%)
Spina Bifida (Excluding genetic conditions)	1151	1019 (88.5%)
Hydrocephalus (Excluding genetic conditions)	1185	965 (81.4%)
Transposition of great vessels (Excluding genetic conditions)	834	551 (66.1%)
Hypoplastic left heart (Excluding genetic conditions)	618	539 (87.2%)
Cleft lip with or without palate (Excluding genetic conditions)	1970	1312 (66.6%)
Diaphragmatic hernia (Excluding genetic conditions)	658	478 (72.6%)
Gastroschisis (Excluding genetic conditions)	611	555 (90.8%)
Omphalocele (Excluding genetic conditions)	547	487 (89.0%)
Bilateral renal agenesis including Potter syndrome (Excluding genetic conditions)	324	306 (94.4%)
Posterior urethral valve and/or prune belly (Excluding genetic conditions)	291	242 (83.2%)
Limb reduction defects (Excluding genetic conditions)	1187	682 (57.5%)
Club foot - talipes equinovarus (Excluding genetic conditions)	2823	1604 (56.8%)
Chromosomal		
Chromosomal	11180	8394 (75.1%)
Down Syndrome	6253	4415 (70.6%)
Patau syndrome/trisomy 13	567	539 (95.1%)
Edward syndrome/trisomy 18	1572	1468 (93.4%)

Figure 1.3 – Taux de diagnostic par le diagnostic prénatal de plusieurs malformations congénitales. Source: données EUROCAT se basant sur des registres européens de l'année 2012 à l'année 2016 (voir le détail des registres concernés dans la figure).

Un outil d'aide à la décision tel que nous cherchons à le concevoir ici, doit permettre au praticien de se focaliser plus rapidement sur les régions anatomiques à risque et ainsi d'accompagner les médecins dans l'effort de minimisation du risque de faux négatif. L'outil final pourra également avoir une fonction de formation, notamment auprès des jeunes praticiens, en décrivant de manière succinte les anomalies à aller observer.

1.3.2 Rassurer les patients (et les médecins) en cas d'examen normal

Un certains nombres d'anomalies sont très courantes et bénignes si elles apparaissent de manière isolée. Citons par exemple les pieds bots, une anomalie congénitale orthopédique touchant une à deux naissances sur 1000 en France qui peut-être corrigée sans trop de complications à la naissance. Cependant si cette malformation est associée à d'autres anomalies phénotypiques alors il est à craindre que le foetus soit affecté d'un syndrome rare d'origine génétique. Si le signe d'appel est "pied bot", notre outil d'aide à la décision doit permettre aux praticiens de rapidement se focaliser sur les zones à risque. Si le praticien n'observe pas d'autres anomalies il pourra rassurer les patients et éviter des examens supplémentaires inutiles.

Ce cas de figure avec un signe d'appel courant qui se révèle après l'examen être isolé et donc bénin est heureusement le plus fréquent en pratique.

Proportion of All Anomalies (Excluding genetic conditions) cases prenatally diagnosed, 2012-2016

Registry	Total number of cases	Number prenatally diagnosed	Percentage of all cases (95% CI)
Wessex (UK)	2,147	1,506	70.1 (68.2 - 72.0)
French West Indies (France)	758	461	60.8 (57.3 - 64.2)
Paris (France)	3,152	1,836	58.2 (56.5 - 60.0)
Northern England (UK)	2,719	1,402	51.6 (49.7 - 53.4)
Thames Valley (UK)	2,468	1,185	48.0 (46.0 - 50.0)
Hainaut (Belgium)	1,057	486	46.0 (43.0 - 49.0)
Vaud (Switzerland)	1,093	499	45.7 (42.7 - 48.6)
Wales (UK)	4,188	1,868	44.6 (43.1 - 46.1)
S Portugal	750	332	44.3 (40.7 - 47.8)
Tuscany (Italy)	2,421	1,011	41.8 (39.8 - 43.7)
Isle de la Reunion (France)	1,975	822	41.6 (39.5 - 43.8)
Brittany (France)	5,179	2,104	40.6 (39.3 - 42.0)
N Netherlands (NL)	1,933	757	39.2 (37.0 - 41.4)
Valencia Region (Spain)	4,438	1,591	35.8 (34.5 - 37.3)
Ukraine	3,531	1,231	34.9 (33.3 - 36.5)
Antwerp (Belgium)	1,957	640	32.7 (30.7 - 34.8)
Saxony-Anhalt (Germany)	2,379	725	30.5 (28.7 - 32.4)
Emilia Romagna (Italy)	4,266	1,171	27.4 (26.1 - 28.8)
Zagreb (Croatia)	505	126	25.0 (21.4 - 28.9)
Malta	594	136	22.9 (19.7 - 26.4)
Norway	6,353	1,329	20.9 (19.9 - 21.9)
Cork and Kerry (Ireland)	1,194	211	17.7 (15.6 - 19.9)
Total	**55,057**	**21,429**	**38.9 (38.5 - 39.3)**

Figure 1.4 – Taux de diagnostic par le diagnostic prénatal des malformations congénitales selon les pays. Source: données EUROCAT se basant sur des registres européens de l'année 2012 à l'année 2016.

1.3.3 Mieux orienter les patients vers des tests génétiques si nécessaire

Si, par malheur, le praticien observe une conjonction d'anomalies il pourra plus rapidement et mieux orienter la patiente vers des examens complémentaires.

Un diagnostic précoce est toujours préférable, qu'il soit possible de corriger la malformation par chirurgie in foeto (sans attendre de dégradation plus importante) ou non et qu'une IMG soit à envisager. L'intensité de la détresse psychologique des patients est en effet intimement liée au terme de la grossesse. La figure 1.5 montre qu'une partie non négligeable des malformations congénitales diagnostiquées le sont en fin de grossesse: 23% pour le registre EUROCAT de Paris.

1.4 Contributions et organisation du manuscrit

Dans cette thèse nous avons fait le choix de présenter les sujets abordés par groupe thématique mélant ainsi tout au long du texte nos contributions à celles de la littérature existante. Afin de faciliter la tâche des rapporteurs nous avons ajouté le symbole ♦ devant les sections présentant une de nos principales contributions.

	<14 weeks		14-23 weeks		>23 weeks		Missing/Not known	
Registry	N	% of all cases (95% CI)	N	% of all cases (95% CI)	N	% of all cases (95% CI)	N	% of all cases (95% CI)
Wessex (UK)	204	9.5 (8.3 - 10.8)	966	45.0 (42.9 - 47.1)	173	8.1 (7.0 - 9.3)	163	7.6 (6.5 - 8.8)
French West Indies (France)	*	*	210	27.7 (24.6 - 31.0)	195	25.7 (22.7 - 29.0)	*	*
Paris (France)	*	*	810	25.7 (24.2 - 27.3)	722	22.9 (21.5 - 24.4)	*	*
Northern England (UK)	172	6.3 (5.5 - 7.3)	784	28.8 (27.2 - 30.6)	238	8.8 (7.7 - 9.9)	208	7.6 (6.7 - 8.7)
Thames Valley (UK)	127	5.1 (4.3 - 6.1)	616	25.0 (23.3 - 26.7)	191	7.7 (6.7 - 8.9)	251	10.2 (9.0 - 11.4)
Hainaut (Belgium)	33	3.1 (2.2 - 4.4)	131	12.4 (10.5 - 14.5)	126	11.9 (10.1 - 14.0)	196	18.5 (16.3 - 21.0)
Vaud (Switzerland)	60	5.5 (4.3 - 7.0)	258	23.6 (21.2 - 26.2)	156	14.3 (12.3 - 16.5)	25	2.3 (1.6 - 3.4)
Wales (UK)	182	4.3 (3.8 - 5.0)	1,409	33.6 (32.2 - 35.1)	249	5.9 (5.3 - 6.7)	28	0.7 (0.5 - 1.0)
S Portugal	42	5.6 (4.2 - 7.5)	136	18.1 (15.5 - 21.1)	104	13.9 (11.6 - 16.5)	50	6.7 (5.1 - 8.7)
Tuscany (Italy)	94	3.9 (3.2 - 4.7)	303	12.5 (11.3 - 13.9)	219	9.0 (8.0 - 10.3)	395	16.3 (14.9 - 17.8)
Isle de la Reunion (France)	95	4.8 (4.0 - 5.8)	349	17.7 (16.1 - 19.4)	378	19.1 (17.5 - 20.9)	0	0.0
Brittany (France)	221	4.3 (3.7 - 4.9)	899	17.4 (16.4 - 18.4)	497	9.6 (8.8 - 10.4)	487	9.4 (8.6 - 10.2)
N Netherlands (NL)	54	2.8 (2.1 - 3.6)	544	28.1 (26.2 - 30.2)	75	3.9 (3.1 - 4.8)	84	4.3 (3.5 - 5.3)
Valencia Region (Spain)	94	2.1 (1.7 - 2.6)	863	19.4 (18.3 - 20.6)	342	7.7 (7.0 - 8.5)	292	6.6 (5.9 - 7.3)
Ukraine	139	3.9 (3.3 - 4.6)	728	20.6 (19.3 - 22.0)	350	9.9 (9.0 - 10.9)	14	0.4 (0.2 - 0.7)
Antwerp (Belgium)	27	1.4 (0.9 - 2.0)	190	9.7 (8.5 - 11.1)	189	9.7 (8.4 - 11.0)	234	12.0 (10.6 - 13.5)
Saxony-Anhalt (Germany)	65	2.7 (2.1 - 3.5)	357	15.0 (13.6 - 16.5)	132	5.5 (4.7 - 6.5)	171	7.2 (6.2 - 8.3)
Emilia Romagna (Italy)	56	1.3 (1.0 - 1.7)	635	14.9 (13.8 - 16.0)	271	6.4 (5.7 - 7.1)	209	4.9 (4.3 - 5.6)
Zagreb (Croatia)	*	*	49	9.7 (7.4 - 12.6)	55	10.9 (8.5 - 13.9)	*	*
Malta	5	0.8 (0.4 - 2.0)	29	4.9 (3.4 - 6.9)	79	13.3 (10.8 - 16.3)	23	3.9 (2.6 - 5.7)
Norway	0	0.0	0	0.0	0	0.0	1,329	20.9 (19.9 - 21.9)
Cork and Kerry (Ireland)	25	2.1 (1.4 - 3.1)	84	7.0 (5.7 - 8.6)	83	7.0 (5.6 - 8.5)	19	1.6 (1.0 - 2.5)
Total	**2,068**	**3.8 (3.6 - 3.9)**	**10,350**	**18.8 (18.5 - 19.1)**	**4,824**	**8.8 (8.5 - 9.0)**	**4,187**	**7.6 (7.4 - 7.8)**

* = Data suppressed to protect the confidentiality of individuals

Figure 1.5 – Taux de diagnostic par le diagnostic prénatal des malformations congénitales selon les pays et les termes de la grossesse. Source: données EUROCAT se basant sur des registres européens de l'année 2012 à l'année 2016.

1.4.1 Livrable

Le but de cette thèse est de construire un outil d'aide au diagnostic pour la recherche de maladies rares par échographie foetale. Il s'agit en direct durant l'examen échographique d'ordonner les différents symptômes possibles dans un ordre décroissant de pertinence, dans un sens que nous préciserons bientôt. Il s'agit également de fournir en temps réel la probabilité de chaque maladie étant donné les signes renseignés par l'utilisateur.

La figure 1.6 donne un aperçu de l'outil d'aide à la décision dans la version beta de son interface. Au départ la population "autre" est de loin la plus probable, il s'agit du cas de figure où le patient n'a aucune des maladies référencées dans notre base de données : il peut donc être sain ou affecté d'une maladie qui n'appartient pas à la base.

Nous commençons avec l'absence du signe "hypertélorisme" qui en tant que telle n'est pas très informative et l'éventualité d'appartenir à la population "autre" reste donc largement la plus probable (probabilité de $99,8/100$). L'information que les oreilles sont situées anormalement bas fait suspecter un syndrome de Noonan (probabilité $73,4/100$). La majorité des signes suivant proposés sont typiques de Noonan mais l'absence d'un signe fréquent chez ce syndrome, spécifiquement le signe "hydrothorax", va réequilibrer les probabilités, rendant plus plausibles les autres syndromes possibles tels que "Apert", Crouzon" etc qui ont une prévalence plus faible dans la population générale. C'est finalement la présence d'une cryptorchidie qui permet de soupçonner fortement un syndrome de Noo-

nan.

Nous présentons ici les différentes problématiques mathématiques qui ont émergé au cours de ce projet et qui constituent les thématiques abordées dans cette thèse.

1.4.2 Chapitre 2: L'optimisation de prise de décision séquentielle

La première question porte à s'interroger sur la manière de quantifier la pertinence des différents symptômes à aller consulter. Que cherchons-nous à optimiser ?

La tentative d'optimiser les stratégies de diagnostic en médecine n'est pas nouvelle. Le problème d'optimisation est généralement formulé sous la forme d'une minimisation d'un compromis entre le coût des tests médicaux et le coût d'un éventuel mauvais diagnostic.

Dans cette thèse nous proposons plutôt de minimiser le nombre de tests médicaux nécessaires pour atteindre un état où l'incertitude concernant la maladie du patient est inférieure à un seuil prédéterminé. Ce faisant, nous tenons compte de la nécessité dans de nombreuses applications médicales, d'éviter autant que possible, tout diagnostic erroné.

Pour résoudre cette tâche d'optimisation, nous étudions plusieurs algorithmes d'apprentissage par renforcement et les rendons opérationnels pour notre problème de très grande dimension: les stratégies apprises se révèlent bien plus performantes que des stratégies gloutonnes classiques.

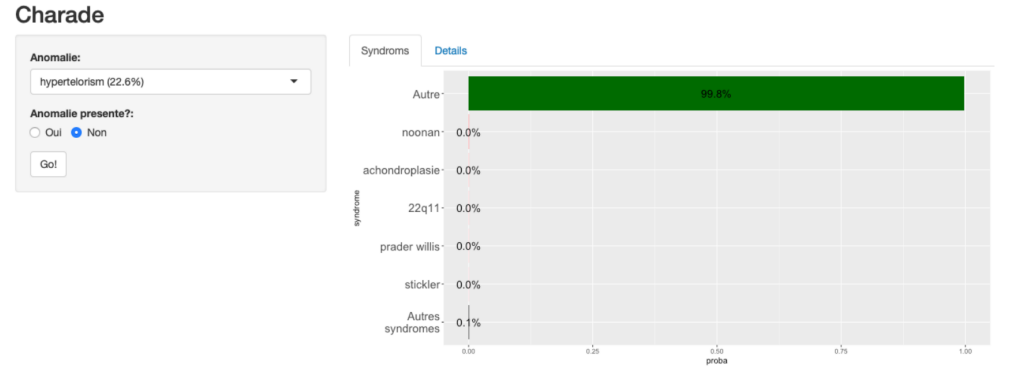
Pour ce faire, nous avons divisé la tâche initiale en plusieurs sous-tâches et nous avons appris une stratégie pour chaque sous-tâche. Nous avons prouvé qu'une utilisation appropriée des intersections entre les sous-tâches peut significativement accélérer la procédure d'apprentissage.

1.4.3 Chapitre 3 : Mélanger experts et données

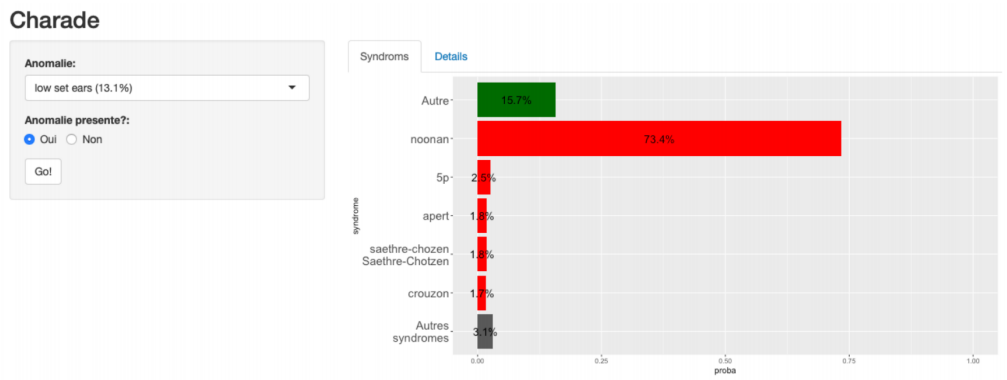
Pour résoudre notre problème d'optimisation et ainsi apprendre notre stratégie de diagnostic il nous faut construire un modèle de l'environnement. Il n'est en effet pas envisageable de déployer dans la vie réelle un algorithme non encore optimisé. De plus le nombre de données cliniques pour les maladies rares est nécessairement limité, c'est pourquoi nous ne pouvons construire notre modèle en se passant des connaissances expertes disponibles.

La deuxième question mathématiques soulevée dans cette thèse est donc celle de l'estimation de la loi jointe des symptômes sachant les maladies. Pour ce faire nous disposons de données expertes (sous la forme de marginales et de règles) ainsi que de données cliniques recueillies au fur et à mesure de l'utilisation de notre algorithme.

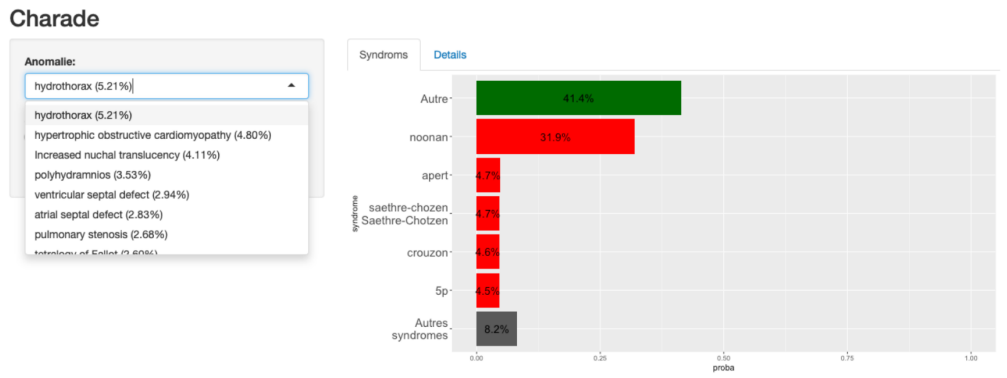
Nous étudions plusieurs manières de combiner modèle expert et modèle empirique. L'idée générale est que si nous avions suffisemment de données nous n'aurions plus besoin des experts et inversement en l'absence de données notre modèle doit reposer entièrement sur les experts. Nous nous intéressons dans cette thèse au régime intermédiaire où nous n'avons pas assez de données pour nous passer des données expertes mais suffisemment



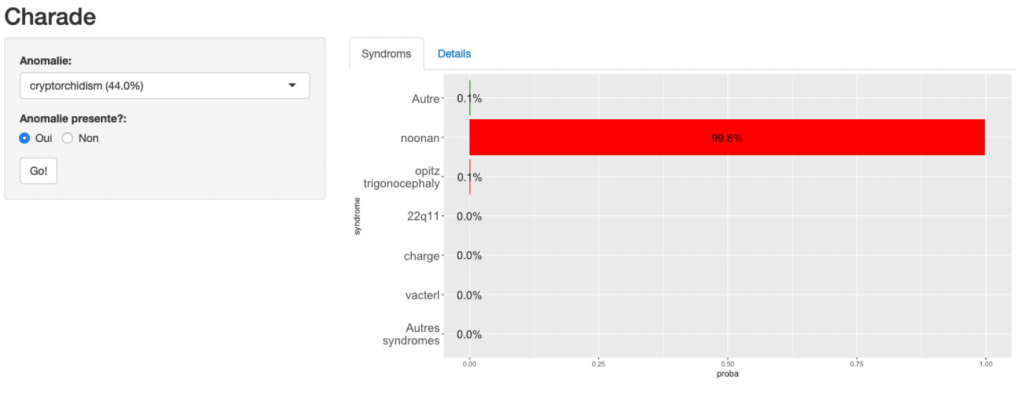
(a)



(b)



(c)



(d)

Figure 1.6 – A demonstration of the current decision support tool diagnosing a Noonan syndrom

pour les corriger si nécessaire, i.e si nous observons une distorsion entre les experts et l'expérience.

Nous proposons d'estimer la loi jointe des symptômes sachant la maladie en choisissant la distribution la plus proche des experts, au sens d'une mesure de dissimilarité adéquate, qui soit consistante avec les données empiriques.

Nous prouvons qu'un estimateur définit de la sorte peut, pour une mesure de dissimilarité reposant sur une norme ou sur une divergence de Kullback-Leibler, s'écrire comme une combinaison linéaire des deux modèles : expert et empirique. Cette propriété nous fournit un algorithme simple et efficace pour calculer notre estimateur.

Nous montrons finalement, tant théoriquement qu'empiriquement, que notre estimateur est toujours plus performant que le meilleur des deux modèles (expert ou données) à une constante près.

1.4.4 Chapitre 4 : Intégrer des raisonnements ontologiques tout en restant dans le cadre probabiliste

La dernière thématique abordée dans cette thèse n'est pas la moins importante pour la réussite pratique du projet. Il s'agit de complexifier quelque peu le tableau présenté plus tôt en intégrant à notre outil d'aide au diagnostic la possibilité pour le praticien de fournir des réponses à différents niveaux de précision pour les questions posées sur les symptômes.

En effet, en médecine un symptôme peut-être décrit à divers niveau de précision. Ainsi le terme "hypoplasie du ventricule gauche" est une dénomination plus précise contenue dans le terme "anomalie du ventricule gauche" qui lui-même est contenu dans le terme "anomalie cardiaque".

La connaissance des relations entre les différentes terminologies des anomalies morphologiques est représentée sous forme d'une arborescence appellée ontologie. Dans cette arborescence nous distinguons deux types de relation entre anomalies : ascendant et descendant. Ainsi "anomalie cardiaque" est un ascendant d' "anomalie du ventricule gauche" alors que la relation inverse est une relation de descendant.

Nous proposons une méthode pour faire de manière efficace les mêmes opérations de calcul des probabilités des maladies étant donné une liste de symptômes et de propositions des prochains symptômes à aller consulter, lorsque les informations sur les symptômes observés sont à différents niveaux de granularité dans l'ontologie.

Chapter references

Jan A Deprest, Alan Flake, Eduard Gratacos, Yves Ville, Kurt Hecher, Kypros Nicolaides, Mark P Johnson, Francois Luks, N Scott Adzick, and Michael Harrison. The making of fetal surgery. *Prenatal diagnosis*, 30:653–67, 07 2010. doi: 10.1002/pd.2571.

N. Scott Adzick, Elizabeth A. Thom, Catherine Y. Spong, John W. Brock, Pamela K. Burrows, Mark P. Johnson, Lori J. Howell, Jody A. Farrell, Mary E. Dabrowiak, Leslie N. Sutton, Nalin Gupta, Noel B. Tulipan, Mary E. D'Alton, and Diana L. Farmer. A randomized trial of prenatal versus postnatal repair of myelomeningocele. *New England Journal of Medicine*, 364(11):993–1004, 2011. doi: 10.1056/NEJMoa1014379. URL https://doi.org/10.1056/NEJMoa1014379. PMID: 21306277.

F. Charlot. *Imaginer et structurer la prise en charge palliative d'un nouveau-né dès la salle de naissance*, pages 99–118. Springer Paris, Paris, 2011. ISBN 978-2-8178-0136-0. doi: 10.1007/978-2-8178-0136-0_12. URL https://doi.org/10.1007/978-2-8178-0136-0_12.

Anne Fagot-Largeault. Le concept de maladie sous-jacent aux tentatives d'informatisation du diagnostic médical. In *History and Philosophy of the Life Sciences, Vol. 10, Supplement: Medicine and Epistemology. Health, Disease and Transformation of Knowledge Perugia, Italy*, pages 89–110. Stazione Zoologica Anton Dohrn - Napoli., 1985. URL http://www.jstor.org/stable/23328898.

Caroline Guibet Lafaye. Pourquoi accepter des refus d'img en cas de pronostic vital néonatal très péjoratif ? *Ethical perspectives / Catholic University of Leuven ; European Centre for Christian Ethics*, 16(4):p. 485–508, 2009. URL https://hal.archives-ouvertes.fr/hal-00357767.

Lecourbe A Ballon M Goffinet F Khoshnood B, Lelong N. Surveillance épidémiologique et diagnostic prénatal des malformations. evolution sur 35 ans, 1981-2014. 2016.

Sebastian Köhler, Nicole A Vasilevsky, Mark Engelstad, Erin Foster, Julie McMurry, Ségolène Aymé, Gareth Baynam, Susan M. Bello, Cornelius F. Boerkoel, Kym M. Boycott, Michael Brudno, Orion J. Buske, Patrick F. Chinnery, Valentina Cipriani, Laureen E. Connell, Hugh J. S. Dawkins, Laura E. DeMare, Andrew Devereau, Bert B. A. de Vries, Helen V. Firth, Kathleen Freson, Daniel Greene, Ada Hamosh, Ingo Helbig, Courtney Hum, Johanna Jähn, Roger James, Roland Krause, Stanley J. F. Laulederkind, Hanns Lochmüller, Gholson J. Lyon, Soichi Ogishima, Annie Olry, Willem H Ouwehand, Nikolas Pontikos, Ana Rath, Franz Schaefer, Richard H. Scott, Michael Segal, Panagiotis I. Sergouniotis, Richard Sever, Cynthia L. Smith, Volker Straub, Rachel Thompson, Catherine Turner, Ernest Turro, Marijcke W. M. Veltman, Tom Vulliamy, Jing Yu, Julie von Ziegenweidt, Andreas Zankl, Stephan Züchner, Tomasz Zemojtel,

Julius O. B. Jacobsen, Tudor Groza, Damian Smedley, Christopher J Mungall, Melissa Haendel, and Peter N. Robinson. The human phenotype ontology in 2017. In *Nucleic Acids Research*, 2017.

Valérie Malan, Laurence Bussières, Norbert Winer, Jean-Philippe Jais, Amandine Baptiste, Marc Le Lorc'h, Caroline Elie, Neil O'Gorman, Nicolas Fries, Véronique Houfflin-Debarge, Loic Sentilhes, Michel Vekemans, Yves Ville, Laurent J. Salomon, and for the SAFE 21 Study Group. Effect of Cell-Free DNA Screening vs Direct Invasive Diagnosis on Miscarriage Rates in Women With Pregnancies at High Risk of Trisomy 21: A Randomized Clinical TrialEffect of Cell-Free DNA Screening on Miscarriage in Women With Pregnancies at High Risk of Trisomy 21Effect of Cell-Free DNA Screening on Miscarriage in Women With Pregnancies at High Risk of Trisomy 21. *JAMA*, 320(6):557–565, 08 2018. ISSN 0098-7484. doi: 10.1001/jama.2018.9396. URL https://dx.doi.org/10.1001/jama.2018.9396.

Laurent Salomon and Yves Ville. Syndrome transfuseur-transfusé : physiopathologie, diagnostic et traitement chirurgical. *Académie nationale de médecin*, 2008.

Chapter 2

INTRODUCTION

Abstract: *In order to set the scene, we begin with a brief overview of the world of prenatal diagnosis by discussing the issues, the actors and the techniques involved. We then discuss the data available and the information we have been able to obtain from the various existing open access databases for rare diseases. We continue by explaining the current need of ultrasound users for a decision support tool and at the same time specifies the desirable properties that our final algorithm must have. We end by trying to step back and formulate the more general mathematical questions that are addressed in this thesis.*

Contents

The work presented in this thesis is the result of my three years of research under the supervision of my two thesis directors, Erwan Le Pennec and Stéphanie Allassonnière, on a subject at the mathematical/medicine interface.

The objective of this project is to provide prenatal diagnostic practitioners with a decision support tool to assist them in the search for rare diseases by ultrasound.

My work during these three years was part of a multidisciplinary team that included mathematicians, Erwan, Stéphanie and myself, obstetricians, Julien Stirnemann, Emmanuel Spaggiari and Yves Ville, as well as a public health specialist, Antoine Neuraz.

2.1 Prenatal screening and diagnosis

2.1.1 Issues in prenatal diagnosis

France registers approximately 780,000 births each year, including 39,000 in the Paris region alone, and it is estimated that 3 to 4 percent of these pregnancies are affected by at least one congenital anomaly (see Khoshnood B [2016]).

Congenital anomalies are a major cause of infant mortality: 21% of deaths in the first year of life are attributable to such anomalies according to the Institut de veille sanitaire [1]. They also represent a significant part of disabilities.

It is therefore clear that the congenital malformations are a public health issue. However, if the existence of a congenital anomaly is a relatively frequent event, the large number of possible different pathologies, many of which are rare diseases, makes the obstetrician's mission difficult.

2.1.2 Tools that can be used for prenatal diagnosis

Ultrasound, because it is a non-invasive and inexpensive technique, remains the preferred tool for prenatal screening. In France, the pregnant woman is required to undergo three ultrasound examinations for control at 12, 21, then 33 weeks of amenorrhea (absence of menstruation). If an abnormality is detected or suspected, the patient is referred to a specialist doctor for a second-line ultrasound, also known as a "diagnostic ultrasound".

If a significant risk of chromosomal abnormality (e. g. trisomy 21) is identified, an amniocentesis should be offered to the mother. Amniocentesis is the removal of amniotic fluid by abdominal puncture. The study of fetal cell karyotype should provide a reliable

1. http://invs.santepubliquefrance.fr//Dossiers-thematiques/Maladies-chroniques-et-traumatismes/Malformations-congenitales-et-anomalies-chromosomiques. Accessed[04/03/2019]

diagnosis of the suspected chromosomal abnormality. It is performed around the 17th week of amenorrhea. This intrusive technique has long been considered to be risky for the fetus, the 1% figure of increased risk of miscarriage attributed to amniocentesis has long been advanced. However, a recent study by Necker doctors showed, in randomized trials on 2051 women, that using current techniques there is no statistically significant increase in the risk of miscarriage Malan et al. [2018]. However, this examination is more expensive and complex than an ultrasound and is therefore not routinely used for screening.

In addition, it should be stressed here that amniocentesis can only detect syndromes with well known karyotype abnormalities, which is far from being the case for all rare diseases. These considerations explain the choice of syndromes selected for our database (see section 2.2): they must be syndromes for which there is not already a reliable and risk-free way to identify them.

Finally, there are magnetic resonance imaging (MRI) and computed tomography (CT) examinations. These examinations are performed following a diagnostic ultrasound when it reveals an abnormality. They must allow "the study of certain organs or structures of the foetus or of its annexes in a different and complementary way to ultrasound. The most common anatomical regions explored on MRI are the central nervous system, thorax and abdomen of the fetus. CT explores the fetal skeleton." [2]

It should therefore be noted that ultrasound is particularly important in prenatal screening since it is systematically used routinely. It is also the best solution for diagnosing a certain number of malformations. Indeed, as recalled in the decree of 20 April 2018 on good practices in prenatal diagnosis: "fetal ultrasound is in some cases the only examination that can identify a risk for the fetus or the current pregnancy, to diagnose a fetal or obstetrical pathology and monitor it".

2.1.3 The actors involved

As specified in article 1 of the decree of 2 May 2017 (2017-702), in France only "midwives and doctors with skills recognised by a diploma" are authorised to carry out obstetrical and foetal ultrasound examinations. However, the prerogatives of these two trades are not identical.

Midwives hold a state diploma enabling them to carry out screening (control) ultrasound examinations. Diagnostic ultrasounds, on the other hand, are not part of their field of competence. Their profession is governed by articles L4151 of the Public Health Code.

The doctors qualified to monitor a woman's pregnancy are obstetricians. When a serious congenital anomaly is suspected, patients are referred to the Pluridisciplinary Centres for Prenatal Diagnosis (CPDPN), which numbered 49 on French territory in 2015. The final diagnosis specifying the nature and severity of the malformation is made by at

2. Extract from Annex 1 of the decree of 20 April 2018 setting out the recommendations for good practice relating to the procedures to carry out imaging examinations contributing to prenatal diagnosis.

least two prenatal diagnostic specialists in a collegial manner.

2.1.4 Possible therapeutic solutions

When a congenital malformation is diagnosed, the question of management is raised. Decision-making is always done in a collegial manner by doctors from different specialties and in collaboration with the Multidisciplinary Centres for Prenatal Diagnosis. The final decision always rests, of course, with the parents.

If the condition of the unborn child is identified as particularly serious, i.e. involving a severe disability without an existing therapeutic solution, a medical termination of pregnancy (IMG) is possible up to the last day of pregnancy. This is Article L2213-1 of the French Public Health Code:

> Voluntary termination of pregnancy may, at any time, be practised if two doctors who are members of a multidisciplinary team certify, after the team has given its advisory opinion, that either the continuation of the pregnancy poses a serious threat to the woman's health or that there is a high probability that the unborn child will suffer from a particularly serious condition recognised as incurable at the time of diagnosis.

It should be noted that it is not uncommon for parents to refuse the IMG when the neonatal vital prognosis is very pessimistic Guibet Lafaye [2009]. In such cases, prenatal diagnosis will have made it possible to consider in advance the possible palliative care to be provided to the newborn from birth Charlot [2011].

It is appropriate here to embellish this bleak picture somewhat and to highlight recent advances in fetal surgery that offer hope for the future. For example, the vital prognosis of fetuses with transfusion-transfused transfusion syndrome for multiple pregnancies has been greatly improved with recent advances in fetal surgery. Without intervention, perinatal mortality is 90 percent and severe neurological sequelae are found in 20 to 40 percent of survivors. Recent techniques allow the survival of one of the two twins in nearly 80% of cases and a risk of sequelae in survivors of less than 10% Salomon and Ville [2008]. There are also possible in-fetal operations for spina bifida malformations: a neural tube closure defect with a prevalence of one in 2000. It has been proven that this operation significantly improves the prognosis of the unborn child's disability compared to the alternative solution of operating in neonatal care (see Adzick et al. [2011]). In general, it is only when such a demonstration has been made that fetal surgery is considered a fully justified alternative since, of course, it is not without risk for the mother and the child it carries. However, this technique is now safe enough to be offered even in cases of non-lethal malformations (A Deprest et al. [2010]).

2.1.5 Criminal liability issues

We do not, of course, claim to provide an exhaustive overview of the criminal liability issues raised by prenatal diagnosis. The reader will easily imagine how complex and ever-changing these issues can be.

In France, two separate state agencies have been created to settle disputes between patients and caregivers in order to offer a quick and free alternative to litigation. These are the Conciliation and Medical Accident Compensation Commission (CCI) and the National Office for Medical Accident Compensation (ONIAM). The ICC's function is to provide expertise for the establishment of potential liability of the caregiver or health facility. More specifically, ONIAM's mission is to organise the compensation system.

If liability is established by the ICC, the health professional's insurer must make a proposal for compensation that will be accepted or refused by the victim, in the event of refusal the victim must turn to the competent courts.

The number of complaints sent to the ICC has been steadily increasing since the 2000s (see figure 1.1). Obstetrical gynaecology is one of the specialities for which the costs of legal compensation are the highest. The Mutuelle d'assurances du corps de santé français (MACSF) publishes its annual report on compensation procedures and, in 2017, obstetrics and gynaecology came in third place with almost 6 million euros in compensation (see figure 1.2).

2.2 The available data

2.2.1 An initial expert database

We have a list of diseases and for each disease a list of generally associated symptoms (which we will call typical symptoms). We write :

$$B_i = \begin{cases} 1 & \text{if the patient has the symptom of kind i} \\ 0 & \text{otherwise}\,. \end{cases}$$

We note D the random variable associated with the patient's disease: $D \in \{d_1, ..., d_k\}$. We assume that the patient has only one disease at a time, which is a reasonable assumption for rare diseases. $\{D = d_k\}$ represents the event for which the patient is healthy or has a disease that is not in the database.

We have an estimate of the prevalence of each disease, as well as the probability of presenting each typical symptom given the disease.

We therefore know $\mathbb{P}[D = d_j]$ abbreviated in $\mathbb{P}[D_j]$ and also $\mathbb{P}[B_i = 1 \mid D = d_j]$ abbreviated in $\mathbb{P}[B_i \mid D_j]$ when B_i is a typical symptom of D_j. Let us add that it is possible to have a disease and to present an atypical symptom but we do not have an estimate for such events. We will assume that atypical symptoms occur with a low

probability (set at 10^{-5}) and independently of other symptoms (conditionally related to the disease).

An important element is that we do not have the joint distributions of symptoms given the disease but only the marginals. We discuss this difficulty in the chapter 4.

Currently, our database references 81 diseases and 220 different symptoms. The disease with the highest number of associated symptoms is VACTERL syndrome, with 19 possible symptoms.

All this information, which we refer to as expert data in the following, was provided to us by doctors at Necker-Enfants Malades Hospital based on the available literature.

In the following we will use the terms syndrome, pathology or disease indistinctly. Similarly, the terms symptoms, anomalies and signs will also be used in an equivalent way in this work. In accordance with the Larousse definition, we define a syndrome as a set of several symptoms related to a given pathological condition and allowing, through their grouping, to guide the diagnosis.

This is the classic way of representing the concept of disease for a computer, although this approach is not without its difficulties, especially when the boundary between healthy and pathological is blurred Fagot-Largeault [1985].

2.2.2 Confrontation à des bases plus générales et extraction de l'arborescence des signes

To our knowledge, there is no such basis as ours in free access, i.e. a basis composed of a list of rare syndromes with the list of signs associated with ultrasound.

The closest open access base to what we want is probably OrphaData [3]. This database on rare and orphan diseases is made available by Orphanet, an organisation created in France by INSERM in 1997 and now supported by the European Union with some forty partner countries.

However, this database is not directly exploitable as such for our problem because it is not limited to obstetrics. Thus many signs of the base are not observable on ultrasound and many syndromes are only observable in neonatal.

We started by mapping the symptoms from our initial database with those from the Human Phenotype Ontology (HPO) database Köhler et al. [2017]. HPO is a recent work that provides a standardized terminology of phenotypic abnormalities found in human pathologies. We used it to harmonize terminology.

We then mapped this list of symptoms by disease with OrphaData. This allowed us to impute missing values from our initial database for symptom prevalence in diseases. Indeed, OrphaData provides information for each sign given the disease if it is mandatory (100% of cases), very frequent (between 80 and 90% of cases), frequent (between 30 and 79%) or occasional (between 5 and 29%). When our initial base had a missing value we

3. Available on http://www.orpha.net. Accessed on [02/10/2018].

imputed the median value of the range proposed by OrphaData. When our initial base already included a value for the prevalence of the symptom in the disease, we retained it without including the prevalence of OrphaData.

It was also possible for us to extract from HPO the information on the ontology of symptoms of our database. By ontology we mean here that a given symptom can be described at various levels of granularity: for example, "*cardiac anomaly*" is an ascendant of "*Tetralogy of Fallot*" (which is a specific cardiac anomaly). Our decision support tool must be able to integrate this kind of classical medical reasoning (see chapter 5 for more details).

Currently we use the OrphaData database to add new syndromes and their associated symptoms to enrich our initial database. Such an operation requires medical expertise to determine the diseases and symptoms to select from the OrphaData database, i.e. the symptoms that can be observed on ultrasound.

Finally, we should mention the existence of another database: EUROCAT. This is the result of a Europe-wide project aimed at epidemiological surveillance of congenital anomalies since 1979. The coverage rate of this base is relatively high since no less than 29% of births in Europe are recorded. This database will be useful for us to obtain the diagnosis rates of the different anomalies during prenatal diagnosis: some malformations are indeed more difficult to observe than others on ultrasound and it may be interesting to incorporate this information into our tool.

2.2.3 Clinical data

As in many medical applications, access to clinical data is difficult. However, our decision support tool is intended to collect data as it is used. These data should allow us to better infer the law of our environment initialized by the expert data.

In general, we hope that this work will provide an opportunity to build a globalized database to better understand rare diseases that can be diagnosed by fetal ultrasound.

2.3 The medical need

The effort to build a decision support tool for prenatal diagnosis seems relevant for several reasons that we mention here.

2.3.1 Improving the detection rate of rare diseases

Our initial base includes 81 rare diseases, but the final objective is to be able to provide diagnostic support for about 1000 different rare diseases. The number of signs that can be observed on ultrasound is even greater. It is obvious that no doctor can have the encyclopedic knowledge to consider all diagnoses.

Figure 1.3 shows the rates of diagnosis by antenatal screening for a number of congenital malformations. While some anomalies are almost systematically diagnosed before birth (e. g. anencephaly or bilateral renal agenesis), others are under-diagnosed. Thus the rate of diagnosis of anomalies not detectable by genetic testing at prenatal diagnosis is only 39% according to this EUROCAT study. Of course, it should not be concluded that all the 61% of undiagnosed congenital anomalies could have been observed antenatally. Indeed, some anomalies may not be detectable at all or very difficult to detect during ultrasound. This is due to the limitations of this technique itself or to the necessary inequality in the levels of expertise of the different practitioners. The figure 1.4 clearly shows the heterogeneity of congenital anomaly diagnosis rates across the different regions of Europe.

A decision support tool such as we are trying to design here should enable the practitioner to focus more quickly on anatomical regions at risk and thus assist physicians in minimizing the risk of false negatives. The final tool may also have a training function, particularly for young practitioners, by briefly describing the anomalies to be observed.

2.3.2 Reassure patients (and doctors) in case of normal examinations

A number of anomalies are very common and benign if they occur in isolation. For example, clubfoot, an orthopedic congenital anomaly affecting between one and two births in 1000 in France can be corrected without too many complications at birth. However, if this malformation is associated with other phenotypic abnormalities, then there is a risk that the fetus may have a rare syndrome of genetic origin. If the call sign is "clubfoot", our decision support tool should allow practitioners to quickly focus on areas at risk. If the practitioner does not observe other abnormalities, he or she can reassure patients and avoid unnecessary additional examinations.

This case, with a frequent call sign that appears after the examination to be isolated and therefore benign, is fortunately the most frequent in practice.

2.3.3 Better refer patients to genetic tests if necessary

If, unfortunately, the practitioner observes a conjunction of anomalies, he will be able to refer the patient more quickly and better to additional examinations.

Early diagnosis is always preferable, whether it is possible to correct the malformation by in-fetal surgery (without waiting for further degradation) or not and whether an IMG is to be considered. The intensity of patients' psychological distress is closely related to the stage of pregnancy. The figure 1.5 shows that a significant part of the congenital malformations diagnosed are diagnosed at the end of pregnancy: 23% for the EUROCAT registry in Paris.

2.4 Contributions and organization of the manuscript

In this thesis we have chosen to present the topics covered by thematic groups, thus mixing our contributions with those of the existing literature throughout the text. In order to facilitate the task of the rapporteurs we have added the symbol ♦ in front of the sections presenting one of our main contributions.

2.4.1 Deliverable

The purpose of this thesis is to build a diagnostic tool for the search for rare diseases by fetal ultrasound. During the ultrasound examination, we aim at ordering the various possible symptoms in a decreasing order of relevance, in a sense that we will soon specify. It is also a question of providing in real time the probability of each disease given the signs given by the user.

The figure 1.6 gives an overview of the decision support tool in the beta version of its interface. Initially the "other" population is by far the most likely, it is the case where the patient has none of the diseases listed in our database: he can therefore be healthy or affected by a disease that does not belong to the base.

We begin with the absence of the "hypertelorism" sign, which as such is not very informative and the possibility of belonging to the "other" population therefore remains largely the most likely (probability of 99.8/100). The information that the ears are abnormally low suggests Noonan syndrome (probability 73.4/100). The majority of the following proposed signs are typical of Noonan but the absence of a frequent sign in this syndrome, specifically the sign "hydrothorax", will rebalance the probabilities, making more plausible the other possible syndromes such as "Apert", Crouzon" etc. that have a lower prevalence in the general population. It is finally the presence of cryptorchidism that makes it possible to strongly suspect Noonan syndrome.

We present here the different mathematical problems that have emerged during this project and which constitute the themes addressed in this thesis.

2.4.2 Chapter 2: Optimizing Sequential Decision Making

The first question is concerned by how to quantify the relevance of the different symptoms to be consulted. What are we looking to optimize?

The attempt to optimize diagnostic strategies in medicine is not new. The optimization problem is usually formulated as a minimization of a trade-off between the cost of medical tests and the cost of a possible misdiagnosis.

In this thesis we propose instead to minimize the number of medical tests necessary to reach a state where the uncertainty regarding the patient's disease is below a predetermined threshold. In doing so, we take into account the need in many medical applications to avoid misdiagnosis as much as possible. This led us to formulate the task of optimizing

symptom checkers as a stochastic shortest path problem.

To solve this optimization task, we study several reinforcement learning algorithms and make them operable in our high-dimensional setting: the strategies learned are much more effective than traditional greedy strategies.

To do this, we divided the initial task into several subtasks and learned a strategy for each subtask. We have proven that appropriate use of intersections between subtasks can significantly accelerate the learning process.

2.4.3 Chapter 3: Mixing experts and data

To solve our optimization problem and learn our diagnostic strategy, we need to build an environmental model. It is not possible to deploy an algorithm in real life that has not yet been optimized. Moreover, the number of clinical data for rare diseases is necessarily limited, which is why we cannot build our model without the available expert knowledge.

The second mathematical question raised in this thesis is therefore that of the estimation of the joint distribution of symptoms given the disease. To do this, we have expert data (in the form of marginals and rules) as well as clinical data collected as we use our algorithm.

We study several ways to combine expert and empirical models. The general idea is that if we had enough data we would no longer need the experts and conversely in the absence of data our model must rely entirely on the experts. In this thesis, we are interested in the intermediate regime where we do not have enough data to dispense with expert data but enough to correct them if necessary, i.e. if we observe a distortion between experts and experience.

We propose to estimate the joint distribution of symptoms given the disease by choosing the distribution closest to the experts, in the sense of an adequate measure of dissimilarity, which is consistent with the empirical data.

We prove that an estimator defined in this way can, for a measure of dissimilarity based on a Kullback-Leibler divergence or a norm, be written as a linear combination of the two models: expert and empirical. This property provides us with a simple and efficient algorithm to calculate our estimator.

Finally, we show, both theoretically and empirically, that our estimator is always more efficient than the best of the two models (expert or data) within one constant.

2.4.4 Chapter 4: Integrating ontological reasoning while remaining within the probabilistic framework

The last theme addressed in this thesis is not the least important for the practical success of the project. The purpose is to make the modelling presented so far more complex by integrating into our diagnostic tool the possibility for the practitioner to provide answers at different levels of granularity for the questions asked about the symptoms.

Indeed, in medicine a symptom can be described at various levels of precision. Thus, the term "hypoplasia of the left ventricle" is a more precise term contained in the term "anomaly of the left ventricle" which itself is contained in the term "cardiac anomaly".

The knowledge of the relationships between the different terminologies of morphological anomalies is represented in the form of a tree structure called ontology. In this tree structure we distinguish two types of relationship between anomalies: ascending and descending. Thus "cardiac anomaly" is an ascendant of "left ventricle anomaly" while the inverse relationship is a descendant relationship.

We propose a method to efficiently perform the same operations for calculating the probabilities of diseases given a list of symptoms and proposals for the next symptoms to consult, when the information on the observed symptoms is at different levels of granularity in the ontology.

Chapter references

Jan A Deprest, Alan Flake, Eduard Gratacos, Yves Ville, Kurt Hecher, Kypros Nicolaides, Mark P Johnson, Francois Luks, N Scott Adzick, and Michael Harrison. The making of fetal surgery. *Prenatal diagnosis*, 30:653–67, 07 2010. doi: 10.1002/pd.2571.

N. Scott Adzick, Elizabeth A. Thom, Catherine Y. Spong, John W. Brock, Pamela K. Burrows, Mark P. Johnson, Lori J. Howell, Jody A. Farrell, Mary E. Dabrowiak, Leslie N. Sutton, Nalin Gupta, Noel B. Tulipan, Mary E. D'Alton, and Diana L. Farmer. A randomized trial of prenatal versus postnatal repair of myelomeningocele. *New England Journal of Medicine*, 364(11):993–1004, 2011. doi: 10.1056/NEJMoa1014379. URL https://doi.org/10.1056/NEJMoa1014379. PMID: 21306277.

F. Charlot. *Imaginer et structurer la prise en charge palliative d'un nouveau-né dès la salle de naissance*, pages 99–118. Springer Paris, Paris, 2011. ISBN 978-2-8178-0136-0. doi: 10.1007/978-2-8178-0136-0_12. URL https://doi.org/10.1007/978-2-8178-0136-0_12.

Anne Fagot-Largeault. Le concept de maladie sous-jacent aux tentatives d'informatisation du diagnostic médical. In *History and Philosophy of the Life Sciences, Vol. 10, Supplement: Medicine and Epistemology. Health, Disease and Transformation of Knowledge Perugia, Italy*, pages 89–110. Stazione Zoologica Anton Dohrn - Napoli., 1985. URL http://www.jstor.org/stable/23328898.

Caroline Guibet Lafaye. Pourquoi accepter des refus d'img en cas de pronostic vital néonatal très péjoratif ? *Ethical perspectives / Catholic University of Leuven ; European Centre for Christian Ethics*, 16(4):p. 485–508, 2009. URL https://hal.archives-ouvertes.fr/hal-00357767.

Lecourbe A Ballon M Goffinet F Khoshnood B, Lelong N. Surveillance épidémiologique et diagnostic prénatal des malformations. evolution sur 35 ans, 1981-2014. 2016.

Sebastian Köhler, Nicole A Vasilevsky, Mark Engelstad, Erin Foster, Julie McMurry, Ségolène Aymé, Gareth Baynam, Susan M. Bello, Cornelius F. Boerkoel, Kym M. Boycott, Michael Brudno, Orion J. Buske, Patrick F. Chinnery, Valentina Cipriani, Laureen E. Connell, Hugh J. S. Dawkins, Laura E. DeMare, Andrew Devereau, Bert B. A. de Vries, Helen V. Firth, Kathleen Freson, Daniel Greene, Ada Hamosh, Ingo Helbig, Courtney Hum, Johanna Jähn, Roger James, Roland Krause, Stanley J. F. Laulederkind, Hanns Lochmüller, Gholson J. Lyon, Soichi Ogishima, Annie Olry, Willem H Ouwehand, Nikolas Pontikos, Ana Rath, Franz Schaefer, Richard H. Scott, Michael Segal, Panagiotis I. Sergouniotis, Richard Sever, Cynthia L. Smith, Volker Straub, Rachel Thompson, Catherine Turner, Ernest Turro, Marijcke W. M. Veltman, Tom Vulliamy, Jing Yu, Julie von Ziegenweidt, Andreas Zankl, Stephan Züchner, Tomasz Zemojtel,

Julius O. B. Jacobsen, Tudor Groza, Damian Smedley, Christopher J Mungall, Melissa Haendel, and Peter N. Robinson. The human phenotype ontology in 2017. In *Nucleic Acids Research*, 2017.

Valérie Malan, Laurence Bussières, Norbert Winer, Jean-Philippe Jais, Amandine Baptiste, Marc Le Lorc'h, Caroline Elie, Neil O'Gorman, Nicolas Fries, Véronique Houfflin-Debarge, Loic Sentilhes, Michel Vekemans, Yves Ville, Laurent J. Salomon, and for the SAFE 21 Study Group. Effect of Cell-Free DNA Screening vs Direct Invasive Diagnosis on Miscarriage Rates in Women With Pregnancies at High Risk of Trisomy 21: A Randomized Clinical TrialEffect of Cell-Free DNA Screening on Miscarriage in Women With Pregnancies at High Risk of Trisomy 21Effect of Cell-Free DNA Screening on Miscarriage in Women With Pregnancies at High Risk of Trisomy 21. *JAMA*, 320(6):557–565, 08 2018. ISSN 0098-7484. doi: 10.1001/jama.2018.9396. URL https://dx.doi.org/10.1001/jama.2018.9396.

Laurent Salomon and Yves Ville. Syndrome transfuseur-transfusé : physiopathologie, diagnostic et traitement chirurgical. *Académie nationale de médecin*, 2008.

Chapter 3

Planning Task

Abstract: *In this chapter, we propose an optimization formulation for the task of building a decision support tool for the diagnosis of rare diseases. We aim to minimize the number of medical tests necessary to achieve a state where the uncertainty regarding the patient's disease is less than a predetermined threshold. In doing so, we take into account the need in many medical applications, to avoid as much as possible, any misdiagnosis. To solve this optimization task, we investigate several reinforcement learning algorithms and make them operable in our high-dimensional setting: the strategies learned are much more efficient than classic greedy strategies.*

Contents

3.1 A Markov Decision Process framework

3.1.1 Formulation of the optimization problem ♦

General framework We can formulate our sequential decision making problem using the Markov Decision Process (MDP) framework. MDP are widely used for modeling the decision making process of an agent acting in an uncertain environment, see Bertsekas [2005], Sutton and Barto [2018]. By uncertain, we mean here that the agent's actions randomly affect the environment.

An MDP is composed of a 4-tuple $(\mathbb{S}, \mathbb{A}, T, r)$ where $\mathbb{S}$ is the state space, an element s of $\mathbb{S}$ should encode the available information about the environment valuable for the decision maker at a given time. $\mathbb{A}$ is the action space, it is composed of the actions that the agent can takes in its environment. More specifically we denote $\mathbb{A}(s)$ the set of action the agent can take in state s.

$T(s, a, s')$ is the transition function which gives the probability to reach state s' when taking action a in state s. The Markov propriety of the MDP allows us to only consider

the last state s when considering the probability to reach state s' while taking action a and not the whole past path followed.

Finally $r(s, a)$ is the reward the agent received when taking action a in state s. The goal of the agent is to find a way to act in its environment which maximise the expected amount of reward.

Modelization in our particular problem Concerning the state space $\mathbb{S}$, we use the ternary base encoding 1 if the considered symptom is present, 0 if it is absent, 2 if non observed yet. We therefore write

$$\mathbb{S} = \{(2, \dots, 2), (1, 2, \dots, 2), \dots, (0, \dots, 0)\}.$$

An element $s \in \mathbb{S}$ is a vector of length 220 (the number of possible symptoms), it sums up our state of knowledge about the patient's condition: the i-th element of s encode information about the symptom whose identifier is i. In the following we will denote $s_t = (s_t^j)_{j=1}^{220}$ the state visited at time t.

Concerning the action space $\mathbb{A}$, we write :

$$\mathbb{A} = \{a^1, ..., a^{220}\}.$$

An action is a symptom that we suggest to the obstetrician to look for, more specifically a^j is the action to suggest to check symptom j.

Our transition function T directly use our environment model which is composed of the joint distributions of the different combinations of symptoms given the disease (see chapter 4 for more details on the construction of such a model) :

$$T(s_t, a^i, s_{t+1}) = \begin{cases} \mathbb{P}[B_i \mid s_t] & \text{if } s_{t+1}^j = s_t^j \text{ for all } j \neq i \text{ and } s_{t+1}^i = 1 \\ \mathbb{P}[\bar{B}_i \mid s_t] & \text{if } s_{t+1}^j = s_t^j \text{ for all } j \neq i \text{ and } s_{t+1}^i = 0 \\ 0 & \text{otherwise}. \end{cases}$$

This equality states that taking a certain action a in a state s we can only reach two different descendant states depending on the presence or absence of the symptom to be consulted. To compute this probability $\mathbb{P}[B_i \mid s_t]$ we can use the Bayes formula and compute the corresponding probabilities $\mathbb{P}[B_i, s_t]$ and $\mathbb{P}[s_t]$ using the law of total probability summing over all the possible diseases.

When referring to the events whose probabilities we measure, we will sometimes use the notation with the symptoms random variables $(B_i)_i$ and sometimes use the state $s \in \mathbb{S}$. Note that these notations are in fact equivalent, it is possible to easily switch from states s to symptom intersections $(B_i)_i$.

Our environment dynamic is by construction Markovian in the sense that:

$$\mathbb{P}[s_{t+1} \mid a_t, s_t, a_{t-1}, s_{t-1}, ...a_0, s_0] = \mathbb{P}[s_{t+1} \mid a_t, s_t]$$

where a_t is the action taken at time t.

Diagnostic policy We aim to learn a diagnostic policy that associates each state of knowledge (list of presence/absence of symptoms) with an action to take (a symptom to check):

$$\pi : \mathbb{S} \to \mathbb{A}. \tag{3.1.1}$$

What should be a good diagnostic policy? Many medical applications consider a trade-off between the cost of performing more medical tests (measuring it in time or money) and the cost of a mis-diagnosis Zubek and Dietterich [2005], Tang et al. [2016], Kao et al. [2018].

However in our case the cost of performing more medical tests (i.e to check more possible symptoms) is negligible against the potential cost of a mis-diagnosis. In theory, the obstetrician have to check all possible symptoms to ensure the fetus does not present any disease. Therefore we do not take the risk of a mis-diagnosis by trying to ask fewer questions. However if the physician observes a sufficient amount of symptoms he can stop the ultrasound examination and perform additional tests, like an amniocentesis, to confirm his hypotheses.

This is why we can label some states as terminal: they satisfy the condition that the entropy of the random variable disease is so low that we have no doubt on the diagnosis. In this setting, our goal is to minimize the average number of inquiries before reaching a terminal state:

$$\pi^\star = \arg\min_{\pi} \; E_{\mathcal{P}}\big[I \big| s_0, \pi\big], \tag{3.1.2}$$

where $s_0 = (2, ..., 2)$ is the initial state, $\mathcal{P}$ the law of the environment currently used, π the diagnostic policy, and I is the random number of inquiries before reaching a terminal states, i.e:

$$I = \inf\{t \mid H(D \mid S_t) \leq \epsilon\} \tag{3.1.3}$$

where

$$\begin{aligned} H(D \mid S_t) &= \sum_{s_t} \mathbb{P}[S_t = s_t] H(D \mid S_t = s_t) \\ &= -\sum_{s_t} \mathbb{P}[S_t = s_t] \sum_{d} \mathbb{P}[D = d \mid s_t] \log \mathbb{P}[D = d \mid s_t] \end{aligned}$$

is the entropy of the random variable disease D given what we know at time t: S_t. We should think s_t as a realization of S_t, this is nothing more than the state we reached for

one examination on a given patient while S_t is the associated random variable. For a given start state s_0 and a policy π there are many possible states that we can reach since the answers are stochastic.

Note that we are not ensured that for all t we had $H(D \mid s_{t+1}) \leq H(D \mid s_t)$. Nevertheless this inequality holds when taking the average

$$H(D \mid S_{t+1}) \leq H(D \mid S_t),$$

see theorem 2.6.5 of Cover and Thomas [2006], "information can't hurt". In summary, when we consider that entropy is sufficiently low and that we can stop and propose a diagnosis, we know that on average, the uncertainty about the patient's disease would not have increased if we had continued checking symptoms.

Setting a reward function as follow, $\forall s_t, a_t$:

$$r_{t+1} := r(s_t, a_t) = -1,$$

we can rewrite (3.1.2) in the classical form of an episodic reinforcement learning problem Sutton and Barto [2018]:

$$\pi^\star = \arg\max_{\pi} \; E_{\mathcal{P}} \left[\sum_{t=1}^{I} r_t \Big| s_0, \pi \right]. \tag{3.1.4}$$

In the RL community such a reward design is called action-penalty representation, since the agent is penalized for every action that it executes Barto et al. [1995]. Note that even though we refer to (3.1.4) as a RL problem, it would be more appropriate to present it as a planning problem since we assume to know the environment law $\mathcal{P}$ (see chapter 4 and Figure 4.1) and aim to solve the MDP associated. Nevertheless, in practice, since the dimension of our problem is high, we solve the MDP by sampling trajectories and then treating the environment model as a simulator. This subtlety explains the common confusion between the terms RL and planning.

3.1.2 Related works.

Non probabilistic symptom checker A *symptom checker* is an algorithm which takes a list of observed symptoms as input and output a corresponding list of possible diseases. There are numerous relevant symptom checker for the diagnostic of rare diseases (in particular in obstetric) such as for example Orphamizer see Köhler et al. [2009] and Köhler et al. [2017]. Nevertheless we think that an algorithm operable during the medical examination would be more useful than an expert system designed for retrospective use.

This is why we aim to propose at each stage of the medical examination the most interesting symptom to check and to display the probability of each disease. We call such algorithm *dynamic symptom checkers*.

Existing dynamic symptom checker Numerous dynamic symptom checker are freely available online. Let us cite for example SymCat [1] a dynamic symptom checker developed for patient self-use. However SymCat, like most of the symptom checker, does not have any academic communication associated, it is therefore difficult to compare our work with theirs.

The goal of such symptom checkers may be to diagnose the disease or to triage patients determining whether the users should seek care more or less urgently or if self care is possible.

Although these systems are pretty popular with the public, several of them reports ten of millions of use each year, their performance are not obvious. A recent study on 23 popular symptom checker and using a sample of 770 standardized patients found that the correct diagnosis was ranked first in only 34% of the case and that the correct diagnostic was ranked within the top 20 diagnoses in 58% of the cases, see Semigran et al. [2015]. These symptom checker are generally risk averse, advising users to seek care for conditions where self care is reasonable. Their effect on patient anxiety is therefore not negligible Doherty-Torstrick et al. [2016].

The symptom checker that made the most noise in the press these last years is certainly Babylon, a product of the London start-up Babylon Health. DeepMind founders, a start-up acquired by Google in 2014, are among the group of investors of this company and they also benefit from support of the NHS (National Health Service) of Great-Britain. Babylon Health claimed to obtain better results than general practitioner on the triage task and a comparable accuracy for the task to identify the disease of the patient, see Razzaki et al. [2018] for more details on the method used.

Once again there is no much detail on the core of the algorithm tested in Razzaki et al. [2018]. They only mention that it is based on a probabilistic graphical model and also refer to their previous paper Middleton et al. [2016]. In Middleton et al. [2016] they represent the diagnostic procedure with "triggers", i.e. certain combinations of signs trigger a series of pre-learned questions. This is a good way to reduce the dimension of the problem by replacing the task of learning the global decision tree by the task of learning several sub-trees (see figure 2 of Middleton et al. [2016]). They do not provide much more information on how these triggers node are determined and how the diagnostic strategy starting from a certain trigger node is learned, they only referred to the classical work of Quinlan [1986] on optimization of decision trees.

In this thesis, we present a particular way of dividing the global decision tree into several sub-trees and propose algorithms that take advantage of sub-trees already resolved to learn the remaining ones, see sections 3.1.3 and 3.4.4.

Moreover as explained in the introduction, our decision support tool is designed to be used by physicians, it should not increase patient anxiety but rather reduce it by offering a

1. http://www.symcat.com/

second perspective and reducing the number of ultrasounds to be performed. Moreover we believe that a decision support tool for the physicians is especially useful for the diagnosis of rare diseases since common disease are easy to diagnose for any physician. Finally obstetrics is a particular field in the sense that the main person concerned is not able to describe his or her symptoms.

Algorithm $AO^\star$ Very few academic works try to address the issue of building a dynamic symptom checker. We reference Zubek and Dietterich [2005] which proposed an algorithm inspired of $AO^\star$ algorithm.

$AO^\star$ algorithm is a variant of the famous $A^\star$ algorithm for finding minimum cost path between a start node and a final node trough a graph Hart et al. [1968]. The cost of a path is defined as the sum of the costs the edge crossed. The algorithm works by maintaining a tree of paths originating at the start node and extending those paths one edge at a time until its stopping criterion is satisfied (see figure 3.1). At each step it consider all the current node n yet expanded and expand the one which minimize :

$$f(n) = g(n) + h(n) \tag{3.1.5}$$

where $g(n)$ is the cost of the path from the start node to node n and h is an heuristic function which estimates the cost of the cheapest path from n to the goal. As long as h is an admissible heuristic, i.e it does not overestimate the costs to reach the node goal, $A^\star$ is guaranteed to converge toward the least-cost path.

Our problem is related to $A^\star$ algorithm as we aim to find the shortest path of our graph between a given start node and a certain class of terminal nodes. The admissible heuristic proposed by Zubek and Dietterich [2005] consists in making a two-step look-ahead. This approach should allow to save computational resources by not exploring part of the decision trees which appears really bad.

Note that such an approach is not tractable in a high-dimensional graph like ours. Indeed an $A^\star$ algorithm assume the possibility to store the whole tree of the paths investigated.

Some improvements of A* have been proposed as for example in Korf [1985] with the algorithm known as IDA*. Such an algorithm provides the same guarantee as $A^\star$ to find the shortest path while using less memory. It consists in cutting off a branch when its total cost (see equation (3.1.5)) exceeds a given threshold. This threshold is updated over the iterations of the algorithm. IDA* needs less memory than A* because it only remember the nodes of the current path, requiring a linear amount of memory in the length of the solution that it constructs.

However, this algorithm does not exactly match our problem. In fact, in our case, we are not looking for a single shortest path in a deterministic graph, but rather to find for each node of the graph the shortest path to the objective states. Indeed, since we

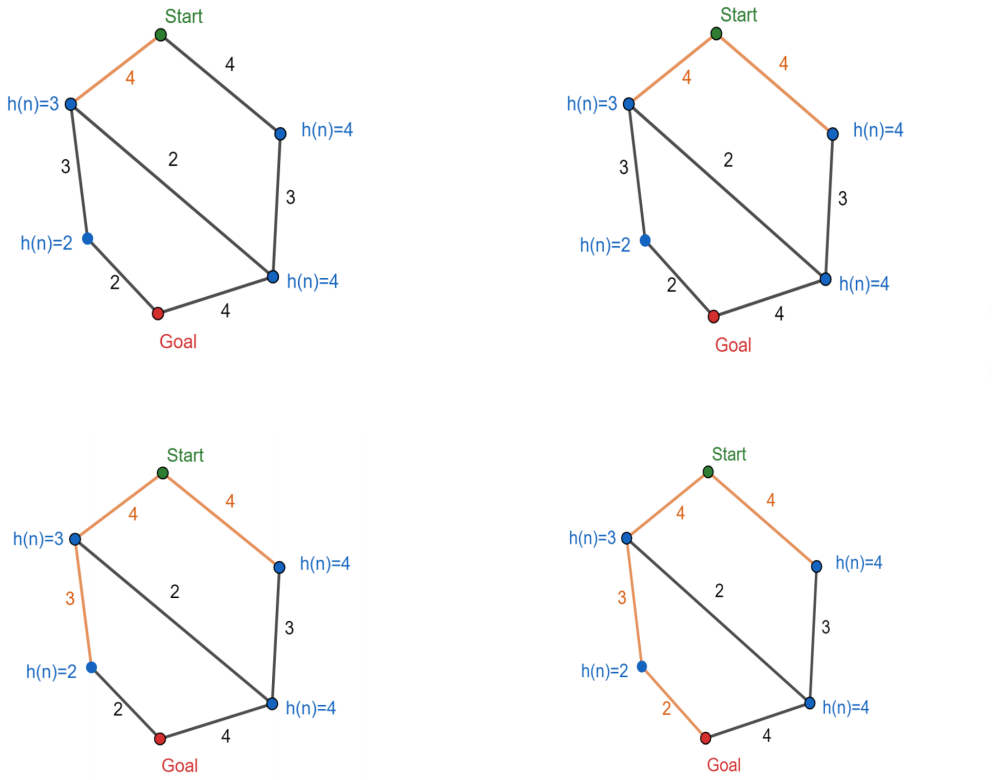


Figure 3.1 – Example of the iteration of an $A^\star$ algorithm on a simple graph.

allow the doctor to answer a different answer than the one we propose, we want to have a good solution even in a part of the tree that is not the optimal path. In this sense we cannot avoid using some form of policy parameterization, so we have the hope that a good solution on the optimized part of the tree will have learned a good enough representation of the state to be good on the invisible part of the tree.

Decision tree optimization Note that in a certain sense our problem can be likened to a decision tree optimization task where the features are the symptoms and the disease is the target. Indeed a policy on a MDP is a generalization of a tree, a policy being less rigid in that sense that it can still propose the next feature to check when the physician made a different choice to the one we proposed. Classic decision tree algorithms, see Breiman et al. [1984] or Quinlan [1986], rely on optimizing an impurity function, the entropy or Gini index of the target random variable, in a greedy way.

Some works as Kortum et al. [2017] proposed to use this algorithm to build a dynamic symptom checker, it chooses to ask the question about the symptom which minimizes the average entropy of the disease. However note that such an algorithm is greedy and is therefore subject to the well-known horizon effect Berliner [1973], an action that appears to be good in horizon 1 may turn out to be bad in a more distant horizon. See section

3.3.1 for more details.

Thus recent works looking for global optimization procedure of decision trees such as Bertsimas and Dunn [2017] can be seen as relevant. However, once again, these algorithms using MIO (Mixed Integer Optimization) solvers cannot cope with our high-dimensional problem. Indeed the complexity of such algorithms is $n \times 2^D$ where n is the number of data and D the maximal depth of the tree. Nevertheless in our case we can not restrict that easily the maximum allowed tree depth since in the worst case, the physician will not observe any symptom and will then have to check them all.

20 questions game Our problem is somewhat related to the old well-known 20 questions game. In this game one of the two players chooses a subject (a famous person, an animal) and does not reveal it to the other. Then the questioner tries to guess the subject by asking 20 questions. Some recent works proposed a reinforcement learning approach to train a questioner algorithm for such kind of game Hu et al. [2018], Zhao and Eskénazi [2016].

In some sense we could see our problem as a 20 questions games, asking question about symptoms to guess the disease. There are nevertheless several notable differences between our problem and a classical 20 question game. First note that in a 20 question game there is a fixed budget of question and that therefore the reward design is different. We do not set ourselves a question budget as it would not be ethical in medicine to do so (to set the examination duration to 10 minutes for example). Then note that the transitions are deterministic in a 20 question games which makes it significantly easier comparing to our stochastic transitions since there is less variance in the returns.

In Hu et al. [2018] the authors propose an interesting alternative approach to modelize the states of the MDP. Instead of representing states as a vector of digit where each number corresponds to a symptom, the state is a vector of the current probabilities of each disease. It is difficult to judge a priori which representation is the most suitable. Note that there is a loss of information when transforming the state from the presence/absence of symptoms to the probabilities of the diseases. Indeed if the list of presence/absence of symptoms directly give the probabilities of the diseases the reverse path is less obvious as different combinations of symptoms can give the same disease distribution. In practice the questions already asked are memorized so as to avoid to propose symptoms which have been checked yet but only the current probabilities of the disease are given to the neural network. Note that this loss of information might have some unfortunate consequence in some specific cases. We can imagine for example two different combinations of presence/absence of symptoms, say (B_1, B_2) and (B_2, B_3), that might give the same disease distribution but that in the first case the most discriminant question to ask would be B_4 when B_4 is actually incompatible with B_3. Thus a policy trained with such a state representation would have to take a sub-optimal action in one of the two cases.

Otherwise these two different alternative do not appear easy to decide between and a

careful empirical comparison is needed. It is not obvious that the approach of Hu et al. [2018] would reduce that much the dimension of the state space but we could argue that the continuing aspect of the disease distribution input would be preferable to the discrete nature of our input (we add an embedding layer to our neural network to face this issue).

This is a common issue in applications of reinforcement learning to be concerned on the feature engineering of the state signal. A good state representation has to satisfy several conditions. First if we want to remain in the MDP framework, the state should only include information that we can observe in real life, otherwise a partially observable Markov decision process (POMDP) must be considered, see Kaelbling et al. [1998]. A second condition is to respect the Markov propriety, for example in the paper of Mnih et al. [2013] the authors had to complexify the first idea for the state representation that would take the raw pixel image to learn Atari with RL algorithms : they took the last 4 images to ensure that the Markov propriety is respected. A last more fuzzy rule is to provide enough information about the state so as to learn good policies without providing too much unnecessary information which will have to force the neural network to remove it itself. Some recent works focus on automatically learning an efficient state representation and to integrate this learning to the training process de Bruin et al. [2018].

It is also interesting to note that the strategies learned in Hu et al. [2018] by training a neural network does not appears to outperform a solution based on entropic minimization ideas (following the ideas of classical algorithm for the decision tree construction). It strengthens our observations presented later that the entropic baseline is often a good solution. Nevertheless we prove empirically in several sub-tasks that a neural network trained with reinforcement learning algorithms can be more effective that this baseline in our problem.

Finally we note that there are some problematic aspects to the work of Hu et al. [2018] that encourages us to take their numerical results with caution. Among these aspects let us cite the fact that they propose to employ a neural network to estimate the immediate rewards of their actions. Nevertheless it is, for obvious reason, problematic to learn simultaneously the rewards and the strategy of the agent. The objective of learning the reward signal is formulated differently in the literature, it can be presented as the objective of learning from expert human demonstrations (the agent strategy is therefore fixed) and is in such case referred as inverse reinforcement learning, see Ng and Russell [2000]. Some recent work as Agarwal et al. [2019] propose to learn simultaneously the parameters of what is called auxiliary reward and the policy. Nevertheless the reward initially designed by the engineer (which is not the auxiliary reward) is not modified and is still used. Auxiliary rewards are optimized so that the policy learned by maximizing it on the training set has also achieved high performance on the test set for the initial reward provided.

From what we have understood, in Hu et al. [2018] it does not seem that there is such

an anchoring of the learned auxiliary reward on the reward initially designed, which seems problematic.

Disease diagnostic task using Reinforcement Learning algorithms Recent works Tang et al. [2016], Kao et al. [2018] and Peng et al. [2018] focus on this problem of building a symptom checker using reinforcement learning algorithms. Nevertheless our approach is fairly different to these previous works, both in our way to formulate the objective (and then in our reward design) than in the solutions that we propose (our ways to deal with high-dimensional issues).

They formulated their optimization problem as a trade-off between asking less questions and making the right diagnosis while we formulate it as the task of reaching as quick as possible, on average, a pre-determined high degree of certainty about the patient disease. In practice, in our case, the only parameter ϵ to be tuned is the degree of certainty we want at the end of the examination: we should stop when the entropy of the disease falls below this threshold. The fewer the ϵ the more symptoms our algorithm will need before considering that the game ends.

Tang et al. [2016] makes use of a discounted factor $\gamma \in [0, 1]$ in their reward signal design. They design the reward associated to each question to be zero until possessing a diagnosis (which is an additional possible action) where the reward is equal to γ^q (if the guess was correct, 0 otherwise), q being the number of questions that have been inquired before possessing the diagnosis. In this context γ makes the compromise between asking fewer question and making the right diagnosis. The smaller γ, the more likely the algorithm is to make a wrong diagnosis by trying to ask fewer questions.

Note that Tang et al. [2016] has to perform its learning algorithm while trying several differents values of γ. On the contrary we can determine which value of ϵ we should take before launching any learning algorithms. We can indeed interact with the physician, in a first step, presenting him a sample of states where our algorithm would possesses a diagnostic. If the physician considers that the algorithm stops too early we should decrease ϵ, otherwise we should increase ϵ. This is an advantage since the main bottleneck in terms of computing time is the learning phase.

Stochastic shortest path Our problem is a particular case of the more general stochastic shortest path problem presented in the book of Bertsekas [2007]. In deterministic shortest path the goal is to find the shortest path from an initial node (state) to a termination node of a graph where the cost of a path is measured as the sum of the cost of the crossed arcs (linking the node).

In the stochastic shortest path the controller assign at each node a probability distribution to reach all the other node. If these probability distributions all assign a probability 1 to reach an unique successor we then return to a problem of deterministic shortest path.

Our problem is a stochastic shortest path where the structure of the graph is particular

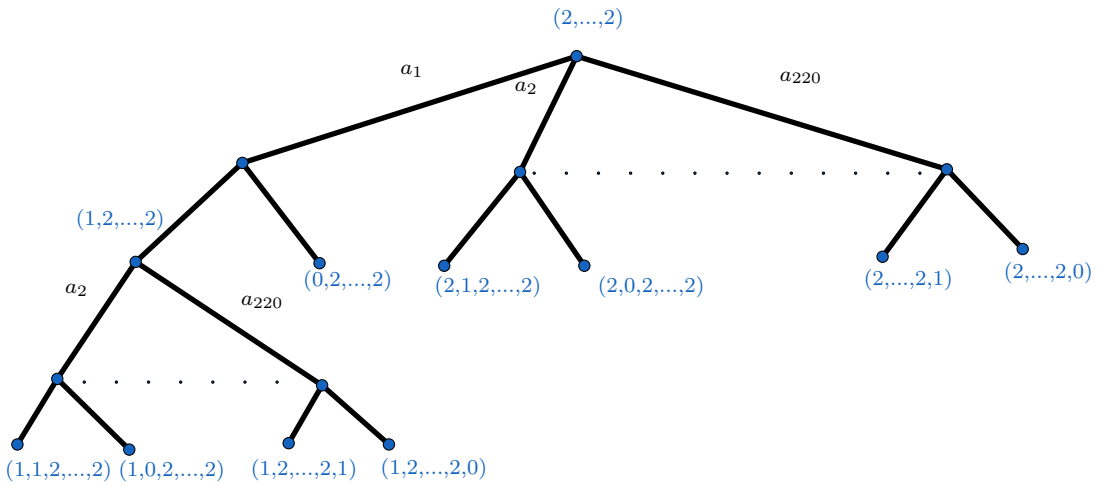


Figure 3.2 – Visualization of a part of the graph of which we are looking for the shortest path.

: there are links only between nodes which are at depth levels diverging by 1 (see figure 3.2). Namely if we denote $s^{(k)} = (s_i^{(k)})_{i=1}^{220}$ and $s^{(l)} = (s_j^{(l)})_{j=1}^{220}$ two different nodes, they are connected, $s^{(l)} \to s^{(k)}$, if an only if it exists h such that $s_i^{(k)} = s_i^{(l)}$ for all $i \neq h$ and $s_h^{(k)} \neq 2 = s_h^{(l)}$. Thus at each node only a few number of descendants nodes are directly reachable. These connections are only in one direction, from the ascendant to the descendant.

The classic formulation of the stochastic shortest path problem considered by Bertsekas [2007] consider an infinite horizon undiscounted problem where transition cycles are possible. Therefore they mainly focus on finding assumptions under which the classic dynamic programming algorithms are ensured to converge to the optimal policy. The first assumption of Bertsekas [2007] is that at least one policy is proper. A policy is said to be proper if there is positive probability that the destination will be reached after at most n step regardless of the initial state (where n is the number of non-terminal state/node). The second assumption is that all improper policies have an infinite cost for at least one state (a reward $-\infty$). The recent work Bertsekas and Yu [2013] prove that a variation of the classic dynamic programming algorithms can be shown to converge even by removing this second assumption.

From this point of view our problem is not theoretically hard as all the policies are proper and we should see in section 3.2 that it is sufficient to ensure convergence of dynamic programming algorithms. What makes our problem difficult is the curse of dimensionality that make many algorithms intractable.

3.1.3 High-dimensional issues ♦

Decomposition of the task using a partition of the state space Our full model is of very high dimension: we have 220 different symptoms and then theoretically $3^{220} \approx 10^{104}$ different states. Thus a classical tabular dynamic programming (DP) approach (see section 3.2) is impossible. According to our experiments, a classical Deep-Q learning is also not numerically tractable (section 3.4).

In order to break the dimension, we capitalize first on the fact that the physicians use our algorithm mainly after seeing a first symptom. In such case, we make the assumption that this initial symptom is typical. It might be possible to have a disease which also presents a non-typical symptom but this happens with a very low probability, sufficiently negligible for the clinicians. Anyway, in this case, we would end up with a high entropy and no disease identification. This leads to switch to another strategy. With such an assumption the dimension drops significantly since we now only consider diseases for which this initial symptom is typical, the only relevant symptoms are the one which are typical of these remaining diseases.

Therefore we created 220 tasks $\mathcal{T}_i$ to solve. We denote $\mathbb{B}_i = (B_{i_1}, ..., B_{i_k})$ the set of symptoms related with the symptom i, i.e this is the set of symptoms which are still relevant to check after observing the presence of symptom i. We denote $\mathbb{S}_i$ the state set associated to the task $\mathcal{T}_i$, the elements of $\mathbb{S}_i$ are of length $|\mathbb{B}_i|$. For an element $s = (s_j)_{j=1}^{|\mathbb{B}_i|}$ of $\mathbb{S}_i$ the m-esima number s_m stands for the symptom $\mathbb{B}_{i_m}$.

For all i, we start from state $s_{(i)}$ which is the state of length $|\mathbb{B}_i|$ with the presence of symptom B_i and no other information on the others symptoms, and we aim to solve :

$$\pi^{\star}_{(i)} = \arg\max_{\pi} E_{\mathcal{P}} \left[\sum_{t=1}^{I} r_t \mid s_{(i)}, \pi \right]. \tag{$\mathcal{T}_i$}$$

The different sub-problems dimensions, i.e the different values of $|\mathbb{B}_i|$, are displayed in Figure 3.8. Fragmenting that much our problem as the advantage of giving us a very good optimized policy on several part of our decision tree that would have been under-optimized otherwise (because these parts of the tree are not often visited). Of course, optimizing the parts of the tree that are not often visited is not very useful to reduce our overall loss function, but it is important to provide, in all cases, a reasonable proposal to the physician if we want him to have confidence in us. This approach will force us to choose a learning algorithm which can handle different sub-problems without needing to tune too many hyper-parameters.

Related works To cope with these high dimensional issues, Tang et al. [2016] proposed in their first paper to learn a different policy for each of the 11 anatomical parts they previously built. As they recognized in their second paper Kao et al. [2018] this approach is problematic. Indeed a symptom may be related to several different anatomical parts.

How to choose which model to use when observing an initial symptom? In their first paper, when a patient give an initial symptom, they choose the model with the best accuracy on their training set and follow this policy until the end of the process. Nevertheless, as they write in Kao et al. [2018], it is possible that the target disease does not belong to the disease set of the chosen anatomical part. This is why they proposed to learn another policy, called master model, which choose at each step the most promising model (among the 11 anatomical parts) to use.

Concerning Peng et al. [2018] the main idea is to use reward-shaping to face the sparse-reward issue typical of the symptom-checker optimization. They add an auxiliary reward when discovering the presence of a symptom. Nevertheless it is well-known that reward-shaping and Q-value initialization are in fact equivalent, see Wiewiora [2011]. Thus initializing our algorithms with a reasonable policy (for example the one which greedily minimize the disease entropy) instead of a random policy would give the same benefits observed in Peng et al. [2018].

In reinforcement learning, there exist several ways to solve a problem like ($\mathcal{T}_i$). If the dimension is small enough it is possible to find the optimal solution explicitly using a dynamic programming algorithm (see section Sutton and Barto [2018], for example using the value iteration algorithm (see section 3.2). If the number of states is too high we have to parameterize the policy (policy-based approach, see section 3.3) or to parameterize the Q-values (value-based approach, see section 3.4). We have investigated both approaches to solve our problem.

3.1.4 The use of the environment model

Theoretically the environment model that we have at our disposal, the joint distribution of the combination of symptoms given the disease learned in chapter 4, provides us with a transition model.

Nevertheless, in practice, we do not use directly the model. It is indeed not possible to store the transition matrix for dimensional reasons. That is why our only alternative to solve the optimization problem (3.1.4) is to simulate games by recalculating the transition probabilities on the fly.

Moreover, the incremental nature of the states implies that the cost for simulating a game from the start s_0 to a certain state s_t is approximately the same than the cost to do an unique transition from state s_{t-1} to s_t. That is why our environment model should be looked in practice as a simulator of games starting from state s_0 to s_I.

Finally, note that a subtlety of our stopping criterion (3.1.3) is that it requires us to compute the probabilities of the diseases given the symptoms combination of the current state at each step of the medical examination. Indeed for dimensional reasons we can not store the set of goal states and have to check at each step if we can stop and possess a diagnostic or not.

Even though we refer to (3.1.4) as a RL problem, it would be more appropriate to present it as a planning problem since we assume to know the environment law $\mathcal{P}$ and aim to solve the MDP associated. Nevertheless, as we explained previously, since the dimension of our problem is high, we solve the MDP by sampling trajectories and then treating the environment model as a simulator which blurs the boundaries between RL and planning.

3.2 Dynamic programming algorithm for solving low dimensional sub-tasks

In this section we assume that we have placed ourselves at the level of a tasks $\mathcal{T}_i$ sufficiently low dimensional to solve it with dynamic programming algorithms. We will always drop the i index for the sake of simplicity even if if we work on a simple sub-task and not on the entire problem of equation (3.1.4).

3.2.1 The different dynamic programming algorithms

Some notations Let us first define the value function as for all state s_t:

$$V_\pi(s_t) = E_{\mathcal{P}}\left[\sum_{t'=t+1}^{I} r_{t'} \mid s_t, \pi\right].$$

This is the expected cumulative reward we get starting from state s and using policy π. Similarly the state-action function is defined as for all state s_t and all action a_t :

$$Q_\pi(s_t, a_t) = E_{\mathcal{P}}\left[\sum_{t'=t+1}^{I} r_{t'} \mid s_t, a_t, \pi\right].$$

The optimal value at state s, $V^\star(s) = V_{\pi^\star}(s)$ gives the highest achievable expected return when starting from state s. Similarly $Q^\star(s,a) = Q_{\pi^\star}(s,a)$. It means that a strategy is optimal if it achieves the highest expected return in all states.

Bellman equations The main idea of dynamic programming algorithms consist in making use of the famous "principle of optimality", namely the tail policy is optimal for the tail subproblem. In our special area of interest this principle takes the form of the following Bellman optimality equations for the value functions :

$$V^\star(s) = \max_{a\in\mathbb{A}}\left\{r(s,a) + \sum_{s'}\mathbb{P}[s' \mid s,a]V^\star(s')\right\} \tag{3.2.1}$$

and for the state-actions values :

$$Q^\star(s,a) = r(s,a) + \sum_{s'} \mathbb{P}[s' \mid s,a] \max_{a' \in \mathbb{A}} Q^\star(s',a').$$

Defining the Bellman optimality operator $T^\star : \mathbb{R}^{\mathbb{S}} \to \mathbb{R}^{\mathbb{S}}$ as :

$$(T^\star V)(s) = \max_{a \in \mathbb{A}} \Big\{ r(s,a) + \sum_{s'} \mathbb{P}[s' \mid s,a] V(s') \Big\}$$

we can rewrite the Bellman equation under the more compact following form :

$$T^\star V^\star = V^\star.$$

Value iteration algorithm The value iteration algorithm is a point-fix algorithm which uses the Bellman optimality equation as a recursive update, namely let V_k refer to the value function estimate at iteration k, the value-iteration algorithm produces the following sequence of value functions, for all k :

$$V_{k+1} = T^\star V_k$$

where V_0 is an arbitrary vector.

Policy Iteration algorithm The policy iteration algorithm starts from an arbitrary policy π_0 and then alternates two steps until convergence :

- A policy evaluation step (E): compute the value function of each state for the policy of the k-th iteration $V_{\pi_k}(s)$.
- A policy improvement step (I): use the value functions of π_k to greedily improve it in a better policy π_{k+1}.

This summarises as:

$$\pi_0 \xrightarrow{E} V_{\pi_0} \xrightarrow{I} \pi_1 \xrightarrow{E} V_{\pi_1} \xrightarrow{I} \pi_2 \xrightarrow{E} ... \xrightarrow{I} \pi^\star \xrightarrow{E} V^\star$$

Concerning the policy evaluation step we use the recursive equality on value functions:

$$V_\pi(s) = r(s,\pi(s)) + \sum_{s'} \mathbb{P}[s' \mid s,\pi(s)] V_\pi(s'). \tag{3.2.2}$$

The Bellman operator underlying π, $T^\pi : \mathbb{R}^{\mathbb{S}} \to \mathbb{R}^{\mathbb{S}}$, is defined by :

$$(T^\pi V)(s) = r(s,\pi(s)) + \sum_{s'} \mathbb{P}[s' \mid \pi(s),a] V(s') \tag{3.2.3}$$

and we can then write (3.2.2) under the more compact form :

$$T^{\pi}V_{\pi} = V_{\pi}.$$

Then the policy evaluation algorithm of a strategy π consists in producing the following sequence of value functions :

$$V_{k+1} = T^{\pi}V_k.$$

Finally the policy improvement step consists in improving greedily the strategy π to obtain a new strategy π' :

$$\begin{aligned}\pi'(s) &= \arg\max_{a\in\mathbb{A}} Q_{\pi}(s,a) \\ &= \arg\max_{a\in\mathbb{A}} \Big\{ r(s,a) + \sum_{s',r} \mathbb{P}[s' \mid s,a] V_{\pi}(s'). \Big\}\end{aligned}$$

3.2.2 Proof of convergence

Most of the books of reinforcement learning, Szepesvari [2010], Sutton and Barto [2018] presents the proofs of convergence of dynamic programming algorithms in the case of infinite horizon discounted problems, namely the reward are discounted by a factor $\gamma \in]0,1[$ at each time step :

$$\pi^{\star}_{\gamma} = \arg\max_{\pi} \mathbb{E}_{\mathcal{P}} \left[\sum_{t=1}^{+\infty} \gamma^t r_t \mid s_0, \pi \right]. \tag{3.2.4}$$

The main idea of the proof is to show that the Bellman optimality operator $T^{\star}_{\gamma}$

$$(T^{\star}_{\gamma}V)(s) = \max_{a\in\mathbb{A}} \Big\{ r(s,a) + \gamma \sum_{s'} \mathbb{P}[s' \mid s,a] V(s) \Big\}$$

is a contraction which is due to the fact that $\gamma \in]0,1[$. The second point is to note that the Bellman optimality equations means that $V^{\star}$ is the unique fixed-point of $T^{\star}_{\gamma}$. Then it suffices to use the Banach's fixed-point theorem to conclude that the value-iteration algorithm converges toward $V^{\star}$.

In an undiscounted horizon setting, $T^{\star}$ may be no longer a contraction but the result still holds, see Bertsekas [2007] (Proposition 1.2), if it verifies the two assumptions that there is at least one proper policy and that all improper policies have an infinite cost for at least one state.

In fact as in our case all the policies are proper we have that $T^{\star}$ is a contraction for a weighted sup-norm. The proof of such a result can be find in Bertsekas and Tsitsiklis [1996] (proposition 2.2) we outline here the main lines. First the idea is to prove that the

Bellman operator T^π (see equation (3.2.3)) is a contraction. We use here $p_{i,j}(\pi(j))$ as a shorthand for $\mathbb{P}[s_{t+1} = j \mid s_t = i, \pi(s_t)]$, thus $\forall i \in \mathbb{S}$, $\forall V, \tilde{V} \in \mathbb{R}^{\mathbb{S}}$ and a vector ψ to be specified later :

$$\begin{aligned}
\left|(T^\pi V)(i) - (T^\pi \tilde{V})(i)\right| &= \left|\sum_j p_{i,j}(\pi(i))(V(j) - \tilde{V}(j))\right| \\
&\leq \sum_j p_{i,j}(\pi(i)) \left|V(j) - \tilde{V}(j)\right| \\
&\leq \left(\sum_j p_{i,j}(\pi(i))|\psi(j)|\right) \max_j \frac{\left|V(j) - \tilde{V}(j)\right|}{|\psi(j)|} \\
&\leq \beta|\psi(i)| \max_j \frac{\left|V(j) - \tilde{V}(j)\right|}{|\psi(j)|} \qquad (3.2.5)
\end{aligned}$$

and then dividing by $|\psi(i)|$ and taking the maximum over i for the left-side of the inequality we have $\|T^\pi V - T^\pi \tilde{V}\|_{\psi,\infty} \leq \beta \|V - \tilde{V}\|_{\psi,\infty}$.

For ψ we set for all i, $\psi(i) = V^\star(i)$, and using the inequality

$$\begin{aligned}
V^\star(i) &= -1 + \max_a \sum_j p_{i,j}(a) V^\star(j) \\
&\geq -1 + \sum_j p_{i,j}(\pi(i)) V^\star(j)
\end{aligned}$$

we have $\sum_j p_{i,j}(\pi(i))\psi(j) \leq \psi(i) + 1 \leq \beta\psi(i)$ where $\beta = \max_i \frac{\psi(i)+1}{\psi(i)}$.

Note that for all i, $\psi(i) \leq -1$ and therefore $0 < \beta < 1$ which completes the demonstration that T^π is a contraction.

Then equation (3.2.5) implies that

$$(T^\pi V)(i) \leq (T^\pi \tilde{V})(i) + \beta\psi(i) \max_j \frac{\left|V(j) - \tilde{V}(j)\right|}{\psi(j)}$$

and taking the maximum over π on both sides we have :

$$(T^\star V)(i) \leq (T^\star \tilde{V})(i) + \beta\psi(i) \max_j \frac{\left|V(j) - \tilde{V}(j)\right|}{\psi(j)}.$$

The same kind of inequality can be obtain interchanging the role of V and $\tilde{V}$ which completes the demonstration that $T^\star$ is a contraction.

3.2.3 Some more details on our implementation

See Annex 1 (confidential).

3.2.4 A qualitative analysis on one of our tasks

We analyze here the policy obtained by using a look-up table value iteration algorithm on a small sub-task (it remains 8 relevant symptoms to check) in order to illustrate some of the dilemmas a medical doctor can face during an examination. We start with the presence of symptom 9. The three diseases which does have symptom 9 in their list of typical symptoms are displayed in table 3.1. We should think, for this one experiment only, that the symptoms are conditionally independent given the disease. Another important information is the prevalence of each disease, we have $\mathbb{P}[D = d_1] = 0.042$, $\mathbb{P}[D = d_2] = 0.0083$ and $\mathbb{P}[D = d_3] = 0.0083$. Finally there is no relation of ascendant/descendant between the 9 symptoms considered in this example. The optimal strategy obtained induces a decision tree which is displayed in Figure 3.3.

Table 3.1 – List of plausible diseases and corresponding list of related symptoms for the sub-task starting with presence of symptom 9.

Disease 1		Disease 2		Disease 3	
Id Symptom	Probability	Id Symptom	Probability	Id Symptom	Probability
1	0.50	6	0.90	2	0.90
2	0.55	7	0.50	4	0.90
3	0.50	9	0.90	6	0.50
5	0.90			9	0.50
8	0.50				
9	0.50				

The first question is comprehensible, it ask about the most plausible symptom of the most plausible disease: the symptom 5. If the answer was positive it continue with a symptom typical of the first disease which is not also typical of other diseases: the symptom 3. The combination of the presence of this two symptoms is sufficient to diagnose the disease 1. The rest of the tree is less obvious. For example when we get a "yes" for symptom 5 and a "no" for symptom 3, should we continue asking symptoms related to disease 1 or should we switch to the symptoms typical of the other diseases? The founded path chooses a symptom related to both disease 1 and disease 3 (symptom 2), probably because it is then easy (and fast) to discard disease 3 by asking a question about symptom 4 (note that the disease 3 has only 3 related symptoms).

Another interesting parts of the tree is when we received a negative answer to our first question about the symptom 5. Then, the initial most plausible disease (the disease 1)

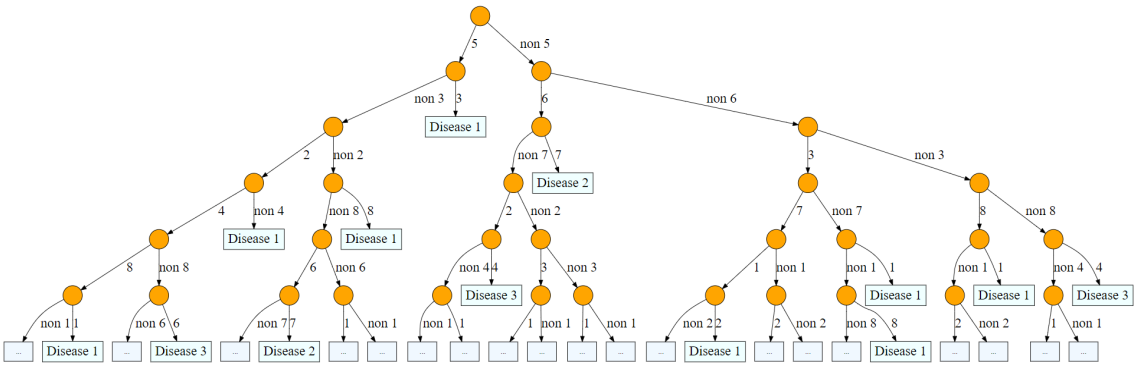


Figure 3.3 – Optimal decision tree for a subtask with 8 possible symptoms.

becomes less likely, but it is not clear if its probability decreases that much that we should check symptoms of other diseases or not. In this case the optimal strategy is to switch to symptoms of disease 3 which has less typical symptoms and must be (in this part of the tree) more plausible than the disease 1.

Due to its size, we do not draw the entire decision tree but only the 6 first symptoms to check, we wrote "..." for the leaves where the obtained diagnostic strategy still propose to check more symptoms.

3.3 A policy-based approach with hand-crafted features as a baseline:

3.3.1 Classic greedy algorithm to optimize decision trees

The classic greedy algorithm for decision tree optimization consists in choosing at each node the features which minimize the average entropy of the target Breiman et al. [1984], Quinlan [1986]. To be more precise we can write it under the form of the following deterministic strategy, $\forall s \in \mathbb{S}$:

$$\pi^{\text{Breiman}}(s) = \arg\max_{a \in \mathbb{A}} H(D \mid s) - \mathbb{E}[H(D \mid s, a)]$$

The quantity $H(D \mid s) - \mathbb{E}[H(D \mid s, a)]$ is sometimes referred as the *information gain* as it measures the average loss of uncertainty about the target outcome that produce action a.

3.3.2 Energy-based policy mixing several reasonable way to play ♦

Stochastic vs deterministic policy We defined in equation (3.1.1) our policy as a deterministic mapping from state to actions but note that we could also define the policy as a mapping associating to each state s the probability to take each possible action a:

$$\pi : \mathbb{S} \times \mathbb{A} \to [0, 1]$$

$\pi(s, a)$ thus stands for the probability to take action a when being in state s.

However, as long as the environment $\mathcal{P}$ is well known and fixed, the optimal policy $\pi^\star$ is deterministic.

This is a direct consequence of the Bellman optimality equation (3.2.1). The max operator induce a deterministic policy which takes at each state the best action, i.e the one which maximize the value function (for all $s \in \mathbb{S}$, and all policy π we have $V_{\pi^\star}(s) \geq V_\pi(s)$). A stochastic policy can therefore be at best as good as the optimal policy and worst if the different action possible at a given state does not give the same outcome.

Nevertheless in our application it can be interesting to propose several symptoms to check at the user,each with its corresponding score (interest to check it), instead of a single one. Indeed physicians might be reluctant to use a decision support tool which do not let them a part of freedom in their choice. This is why we will also consider stochastic policies.

Energy-based policy: We consider here an energy-based formulation, a popular choice as in Heess et al. [2013]:

$$\pi_\theta(s, a) = \frac{e^{\theta^T \phi(s,a)}}{\sum_b e^{\theta^T \phi(s,b)}}$$

where $\pi_\theta(s, a)$ is the probability to take action a in state s, $\phi(s, a)$ is a feature vector: a set of measures linked with the interest of taking action a when we are in state s. To be more precise:

$$\phi(s, a) = \Big(H(D \mid s) - \mathbb{E}[H(D \mid s, a)], P[S_a \mid s], \mathbb{1}_{A_a \in S_{max}(s)}\Big)$$

where $S_{max}(s)$ is the set of typical symptoms of the most likely disease at state s and $H(D \mid s)$ is the entropy of the random variable disease at state s. In words $\phi(s, a)$ summarizes three reasonable way to "play" our game.

- Ask the question that minimizes the expected entropy of the disease random variable. This is exactly the Breiman et al. [1984] way to play.
- Ask the question where the probability of a positive answer is maximum. It is specific to our game where positive answers are much more informative than negative answers (it would not be the case in a classic 20 questions game).

- Inquire symptoms related to the currently most plausible disease.

These features have been identified as relevant measures in rare disease research by physicians we are working with. They represent different way to think and dilemmas faced during medical examination: when I observed a symptom should I think about symptoms usually observed jointly or should I think about the most plausible disease and look for the corresponding symptoms?

Note that this parameterized function π_θ is nothing more than a neural network without hidden layer designed with hand-crafted features. When properly optimized this policy outperforms, by construction, classical decision tree algorithm Breiman et al. [1984].

3.3.3 Learning algorithms for policy gradient methods in RL

Our aim is to learn good parameters θ for each of our 220 sub-problems:

$$\theta^\star_{(i)} = \arg\min_\theta L_i(\theta) := E_{\pi_\theta}[I \mid s_{(i)}].$$

This kind of optimization problem, has been well studied by the reinforcement learning community, see Konda [1999] or Sutton et al. [1999] for the general analysis and Heess et al. [2013] for the energy-based particular case.

Note that we are not able to compute exactly $L_i(\theta)$ neither its gradient. Let us have a closer look to $L(\theta)$ (we drop the i index in the following):

$$L(\theta) = E_{\pi_\theta}[I \mid s_0] = \sum_a \pi_\theta(s_0, a) Q_{\pi_\theta}(s_0, a)$$

Most of the time we use $\nabla_\theta \log(\pi_\theta(s,a))$ instead of $\nabla_\theta \pi_\theta(s,a)$ as in our case of soft-max policy $\nabla_\theta \log(\pi_\theta(s,a))$ can be written in a very simple way:

$$\nabla_\theta \log(\pi_\theta(s,a)) = \phi(s,a) - E_{\pi_\theta}[\phi(s,.)].$$

and because there is a simple relation between both of them:

$$\nabla_\theta \pi_\theta(s,a) = \pi_\theta(s,a) \frac{\nabla_\theta \pi_\theta(s,a)}{\pi_\theta(s,a)} = \pi_\theta(s,a) \nabla_\theta \log(\pi_\theta(s,a)).$$

The policy gradient theorem (see Sutton et al. [1999]) ensure us that:

$$\nabla L(\theta) = E_{\pi_\theta}\left[\nabla_\theta \log(\pi_\theta(s,a)) Q_{\pi_\theta}(s,a)\right].$$

The algorithm REINFORCE (see Williams [1992] and the pseudo-code 1) use directly this theorem to derive a gradient-based algorithm, it uses a return v_t as an unbiased sample of $Q_{\pi_\theta}(s,a)$:

$$\Delta\theta_t = \alpha \nabla_\theta \log(\pi_\theta(s_t, a_t)) v_t.$$

Algorithm 1 REINFORCE algorithm.

Initialization of θ.
while The computational budget has not been exhausted **do**
  Generate a new episode $\{s_1, a_1, r_2, ..., s_{I-1}, a_{I-1}, r_I\}$
  for t=1: I-1 **do**
    $\theta \leftarrow \theta - \alpha \nabla_\theta \log(\pi_\theta(s_t, a_t)) v_t$
  end for
end while

This algorithm suffer from the high variance of the estimate of $Q_{\pi_\theta}(s, a)$. Using a single sample v_t for each gradient descent step generally produces a very unstable learning progress.

In order to reduce the variance of the update, a traditional solution consist in using a baseline. Let B be a function only depending on the state s (and not on the action) then the policy gradient theorem can be rewrite as :

$$\nabla L(\theta) = E_{\pi_\theta} \left[\nabla_\theta \log(\pi_\theta(s, a)) \left(Q_{\pi_\theta}(s, a) - B(s) \right) \right]$$

since

$$\begin{aligned}
\mathbb{E}_{\pi_\theta}[\nabla_\theta \log \pi_\theta(s, a) B(s)] &= \sum_a \pi_\theta(s, a) \nabla_\theta \log \pi_\theta(s, a) B(s) \\
&= \sum_a \nabla_\theta \pi_\theta(s, a) B(s) \\
&= \nabla_\theta \left(B(s) \sum_a \pi_\theta(s, a) \right) \\
&= \nabla_\theta B(s) \\
&= 0
\end{aligned}$$

A classical choice for the baseline is an estimate of the state value function $B(s) = V_w(s)$. The parameters w can also be learned with a Monte-Carlo algorithm see 2.

Algorithm 2 REINFORCE with Baseline.

Initialization of θ and w.
Initialization of learning rates α_w and α_θ.
while The computational budget has not been exhausted **do**
  Generate a new episode $\{s_1, a_1, r_2, ..., s_{I-1}, a_{I-1}, r_I\}$
  for t=1: I-1 **do**
    $\delta \leftarrow G_t - V_w(s_t)$ where G_t is the return at step t.
    $w \leftarrow w - \alpha_w \delta \nabla_w V_w(s_t)$
    $\theta \leftarrow \theta - \alpha_\theta \delta \nabla_\theta \log(\pi_\theta(s_t, a_t))$
  end for
end while

3.4 A value-based approach:

3.4.1 Training Deep Neural Networks

We recall that the Q-values are defined as

$$Q_\pi(s,a) = E\big[\sum_{t'=t}^{I} r_{t'} \mid s_t = s, a_t = a, \pi\big],$$

namely this is the expecting amount of reward when starting from state s, taking action a and then following the policy π. The optimal Q-values, are defined as $Q^\star(s,a) = \max_\pi Q_\pi(s,a)$ and satisfy the following Bellman equation:

$$Q^\star(s,a) = E_{s'\sim\mathcal{P}}[r + \max_{a'} Q^\star(s',a')].$$

The optimal policy $\pi^\star$, is directly derived from $Q^\star$: $\pi^\star(s) = \arg\max_a Q^\star(s,a)$. Therefore we "only" need to evaluate $Q^\star(s,a)$, $\forall s,a$. This can be done by a value-iteration algorithm which uses the Bellman equation as an iterative update:

$$Q_{i+1}(s,a) = E[r + \max_{a'} Q_i(s',a')|s,a].$$

It is known, see Sutton and Barto [2018], that $Q_i \to Q^\star$ when $i \to \infty$.

As the dimension of the problem is too high to store/evaluate all the Q-values, we parameterized it by a neural network: $Q(s,a) \approx Q_w(s,a)$.

The famous Deep Q-Network (DQN) algorithm proposed by Mnih et al. [2013] made possible the use of neural networks to parameterize the Q-values (then called Q-network) in the value iteration algorithm with function approximation. The Q-network, at iteration i, is trained by minimizing the loss function

$$L_i(w_i) = E_{s,a}\big[(y_i - Q_{w_i}(s,a))^2\big]$$

where $y_i = E_{s'\sim\mathcal{P}}[r + \max_{a'} Q_{w_{i-1}}(s',a')|s,a]$ is the target. This can be done by a standard back-propagation algorithm. In practice to successfully combine deep learning with reinforcement learning, the main idea is to use experience replay to break correlation between data: build a batch of experiences (transitions s, a, r, s') from which one samples afterwards. Another trick is to freeze the target network during some iterations to overcome instability while learning.

By doing so acting and learning are dissociated, the policy used to act (called behavior policy) is different from the one learned from the transitions sampled in replay memory (the target policy). In RL, this type of algorithm are called off-policy methods. It is a desirable propriety for a RL algorithm to be off-policy as the behavior policy is designed to enforce exploration.

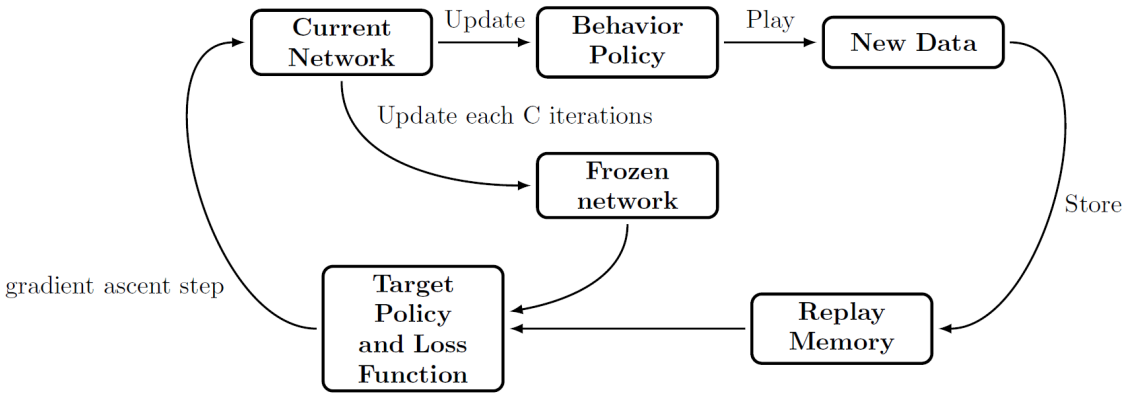


Figure 3.4 – General schema of Deep Reinforcement Learning algorithms.

Figure 3.4 shows the general simplified scheme of the algorithms used in deep reinforcement learning: an agent interact with its environment and collects data (transitions s_t, a_t, r_t, s_{t+1}) which are incorporated to the replay memory from which we sample to form the target policy. Periodically the behavior policy is updated with the current learned policy. In our case we update the behavior policy as soon as we made a gradient ascent step.

3.4.2 Some remarks on the behavior policy

Our behavior policy is an ϵ-greedy version of the current learned policy in order to enforce exploration. We use this policy to simulate games, or more precisely transitions (s_t, a_t, s_{t+1}, r_t).

For this purpose, we need a model of the environment, a transition model which told us the probability to reach a state s_{t+1} when taking action a_t in s_t. Our environment model is composed of the symptoms combination distribution of each disease (see chapter 4). Namely we store $(P[B_1, ..., B_{K_D} \mid D])_{B_1,...,B_{K_D}}$ the probability of all the possible combinations of typical symptoms given the disease, K_D is the number of typical symptoms of D. We add the assumption that a patient can also presents non-typical symptoms but with small probability and independently of the others symptoms (see section 4.6.2).

Then to simulate transitions we need to determine for each disease which are the symptoms of the current list which are typical and which are not. It allows us to find the right combination we have to extract from $(P[B_1, ..., B_{K_D} \mid D])_{B_1,...,B_{K_D}}$. This computation is not that cheap especially when we add ontological considerations (see chapter 5). We can speed it when we play games from the start s_0 to a terminal state s_I: we remember which symptoms are typical for each disease and thus only have to determine if the last symptom is typical or not.

Note that, at each stage of a game, we have to compute the probability of the symptoms combination given each disease so as to determine whether we should stop or not. Another

important observation is that trying to compute directly $\mathbb{P}[s_I \mid s_0]$ is as costly as playing an entire game incrementally (as previously described) from s_0 to s_I.

These two observations should convince the reader that an asynchronous learning approach, as in Mnih et al. [2016], would not be suitable for our problem. From a computational perspective, it is reasonable to play games from the start to a terminal state.

3.4.3 The update target: Temporal-difference and Monte-Carlo algorithm

A remaining question concerns the definition of the update target, should we use Monte Carlo returns or bootstrap with an existing Q-function ?

We recall (following Sutton and Barto [2018]) that an algorithm is a bootstraping method if it bases its update in part on an existing estimate. This is the case of the Temporal-Difference (TD) algorithm defined as:

$$Q_{k+1}(s,a) \leftarrow (1-\alpha) \underbrace{Q_k(s,a)}_{\text{old estimate}} + \alpha \underbrace{\left(r(s,a,s') + \max_{a'} Q_k(s',a') \right)}_{update}$$

where we sampled s, a, s' using the current policy (in a ϵ-greedy way in order to enforce exploration) and the environment model $\mathcal{P}$. Q_k is the estimate at iteration k, α the learning rate. On the contrary a Monte-Carlo method does not bootstrap:

$$Q_{k+1}(s,a) \leftarrow (1-\alpha) Q_k(s,a) + \alpha G$$

where G is the reward we received from a simulated game.

It is not clear at first sight whether we should use a TD method or a MC method to compute the target y_i. This question is the subject of a recent work Amiranashvili et al. [2018] which show that MC approaches can be a viable alternative to TD in the modern reinforcement learning era. Usually TD method is seen as a better alternative than MC method which is often discarded because of the high variance of the return.

Nevertheless our case study is specific: we face a finite-horizon task with a final reward: the reward signal is not very informative before reaching a terminal state. In addition, for the subproblems of intermediate dimensions, we are ensured that games do not last too much time and then that there is a small variance in the return of the Monte-Carlo episodes.

We implemented both solutions referred as DQN-TD and DQN-MC. At each step of DQN-MC, we sample, following the behavior policy, 100 games starting from the initial state $s_{(i)}$ and stopping when they reach a terminal state. All the transitions s, a, s' of all these games are annotated with the reward they received (the number of questions that have been necessary to reach a terminal state during the game concerned) and incorporated in the replay memory. We then sample transitions from this replay memory (one twentieth)

and perform a gradient ascent step with a back-propagation algorithm (we used the Keras library Chollet et al. [2015]).

Concerning the DQN algorithm with TD method, we kept the main features of DQN-MC in order to facilitate their comparison. We play 100 games, still with the behavior policy, and all the transitions s, a, s' of all these games receive a -1 reward when s' is not terminal, 0 otherwise. The learning rate is initialized with a lower value than in the DQN-MC algorithm but it is decreased in exactly the same way in both cases: divided by two each 300 iterations. Another difference is the frozen network we use as target in DQN-TD which is not needed in DQN-MC. We update the frozen network each 2 iterations (we have also tried to update it less frequently but have not observed any major differences with the results presented here).

We compared these two algorithms, DQN-MC and DQN-TD, on severals of our sub-tasks (see figures 3.9 and 3.10). We did not observed much difference on small and intermediate sub-problems: both algorithms converge at the same speed towards solutions of the same quality. Nevertheless DQN-TD appears much more sensitive to the learning rate. Indeed as it can be seen in Figure 3.9, DQN-TD converge on this problem, where it remains 29 relevant symptoms to check and 8 possible diseases, when the learning rate is initialized at 0.0001. Nevertheless if the learning rate is chosen a little bit higher, at 0.001, DQN-TD diverge. On the contrary, DQN-MC converge when the learning rate is initialized to 0.001 and also when initialized to 0.01 even if the returns of the algorithm are less stable in this latter case. These observations have to be combined with the one of Figure 3.10 where it remains 104 relevant symptoms to check and 18 possible diseases. We can see that in this case DQN-TD with an initial learning rate of 0.0001 diverge. Reducing the learning rate to 0.00001 does not change this fact. On the contrary we do not need to reduce the initial learning rate of DQN-MC (we take it equal to 0.001) to make it converge to a good solution. Since we have to train as many neural networks as the number of sub-tasks, we need a robust algorithm able to deal with different task complexity without changing all the hyper-parameters.

This is why we chose to use DQN-MC instead of DQN-TD. It is, indeed, a well-known issue sometimes referred as "deadly triad" Sutton and Barto [2018] that combining function approximation, off-policy learning and bootstrap to compute the target (what the DQN-TD algorithm does) is not safe. We show that DQN-MC performs well on small and intermediate sub-tasks of our problem. The higher dimensional tasks are harder to solve because the games are expected to last longer which is a challenge both in term of computing time that in terms of learning stability (higher variance of the return). To scale up on such problems, we break down the state space into a partition and leverage already solved sub-tasks as bootstrapping methods.

Algorithm 3 DQN-MC with Bootstrapping on already solved sub-tasks.

Start with low dimensional tasks.
for i such that the task $\mathcal{T}_i$ has not been yet optimized **do**
  if $|\mathbb{B}_i| \leq 30$ **then**
    while the budget for the optimization of this task has not been reached **do**
      Play 100 games (ϵ-greedy) from the start $s_{(i)}$ to a terminal state.
      Integrate all the obtained transitions to the Replay-Memory
      Throws part of the Replay-Memory away (the oldest transitions of the replay)
      Sample 1/20 of the Replay-Memory
      Perform a gradient ascent step (backpropagation algorithm) on the sample
    end while
  end if
end for

Continue with higher-dimension tasks.
while there are still tasks to be optimized **do**
  Choose the easiest task to optimize: the one with the highest proportion
  of already solved sub-tasks (weighted by their probability to be faced)
  while the budget for the optimization of this task has not been reached **do**
    Play 100 games (ϵ-greedy) from the start $s_{(i)}$ to a terminal state (condition (j))
    or to a state that was yet encountered in an already solved task (condition (jj))
    if we stopped a game because of condition (jj) **then**
      Bootstrap i.e use the network of the sub-tasks to predict
      the average number of question to reach a terminal state
    end if
    Integrate all the obtained transitions to the Replay-Memory
    Throws part of the Replay-Memory away (the oldest transitions of the replay)
    Sample 1/20 of the Replay-Memory
    Perform a gradient ascent step (backpropagation algorithm)
  end while
end while

3.4.4 Solving higher dimension tasks by bootstrapping with already solved sub-tasks ♦

We denote $\mathbb{B}_i = (B_{i_1}, ..., B_{i_k})$ the set of symptoms related with the symptom i, i.e this is the set of symptoms which are still relevant to check after observing the presence of symptom i. When $|\mathbb{B}_i|$ is small enough (say $|\mathbb{B}_i| < 11$), we can learn the optimal policy $\pi^\star$ by a simple Q-learning lookup table algorithm, see Sutton and Barto [2018].

Considering intermediate dimension problems (say $11 < |\mathbb{B}_i| < 31$) we can use the DQN-MC algorithm which performs pretty well on these problems (see experiences in section 3.5.2). For high-dimensional problems ($|\mathbb{B}_i| > 30$) using directly the DQN algorithm would be time-consuming. An easy way to accelerate the learning phase of these big networks is to make use of the smaller networks previously trained. Indeed if S_i is a symptom for which $|\mathbb{B}_i|$ is high, there must have some $S_j \in \mathbb{B}_i$ such as $|\mathbb{B}_j|$ is small enough and therefore such as the Q-values of $\pi^\star_{(j)}$ have been yet computed or at least approached. Put in another way, when we try to learn the optimal Q-network of a given problem, we yet know, for some inputs, the Q-values that should output a quasi-optimal Q-network.

There are several ways to take advantage of these already optimized subtasks to optimize networks on larger tasks. A first idea would be to incorporate to the replay-memory of the larger task, the replay-memories of the already solved sub-tasks by having previously properly resized the states. Remind that at each iteration, i.e each gradient ascent step, we sample transitions from the replay-memory (s, a, s' and the reward received at the end of the game R) to form the target and train set used to perform the back-propagation algorithm step. We can add to these sets some immovable transitions, the one we already know (because they appear in sub-problems already solved).

However, by doing so we will face several issues. First, when we train our neural network using the replay memory constituted by playing on the concerned task, we are ensured that the transitions that populate our replay-memory will be present in a proportion equivalent to their probability of being encountered in the task. On the contrary, when we add some immovable transitions from already solved sub-tasks to our replay memory, we might over-optimize our network on these sub-tasks. Put it another way, the network will be over-optimized on parts of the decision tree which are not that frequently faced in practice.

Secondly, although the length of the episodes will have been reduced since using the subtasks replay-memories allows us to learn more quickly how to play at the end of the games, it will still be time consuming to play from the beginning until the end of the episodes for tasks of high dimension. The length of the episodes will also be an issue considering the variance of the MC returns.

Therefore a second idea would be to learn a policy on the higher dimension task by bootstrapping on already solved subtasks. Namely we play games starting from the initial state $s^{(i)}$ and bootstrap when reaching a state that belongs to a state set of the partition

where there yet exist an optimized network. In practice we have a function which is called each time we received a positive answer which checks if there already exists a network optimized for such a starting symptom. If this is indeed the case, the current game is stopped and the corresponding optimized network is called to predict the average number of question to ask to reach a terminal state. The main lines of the whole procedure are summarized in the algorithm 3

Note that in doing so, we do not optimize the network for the entire task. It is therefore necessary to change the neural network used for the recommendation during the examination when we change the space of the partition. The advantage is that we do not need to use a more complex architecture for this higher dimension task.

3.4.5 Some remarks on the complexity of a task

When we described the several tasks $\mathcal{T}_i$, we focused mainly on the number of remaining relevant symptoms to check denoted $|\mathbb{B}_i|$. This is the most important parameter since it is the input length of our network and then determines the number of network parameters that we have to optimize.

Nevertheless there are more parameters which influence the complexity of a task. Let us mention the number of possible diseases and especially their probabilities. Indeed, if there are many possible symptoms to check and many possible diseases but there is a disease that is much more plausible than the others, then the task is not so difficult. Another feature that can influence a task complexity is the amount of symptoms which are typical of several of the possible diseases.

Thus as it seems difficult to quantify the difficulty of a task we should avoid to judge the performance of our algorithms in an absolute way but should always compare them to more classical methods.

Finally note that even what we call "task complexity" is not that easy to define. An idea would be to define the complexity of a task as the difference between the average number of question that have to ask a random policy and the average number of question that have to ask the optimal policy.

3.4.6 Considerations on the possibilities of parallelizing computations

One of the other advantage of fragmenting the general task as several sub-tasks more easily optimizable is the possibility to optimize the different sub-tasks in parallel.

As we have seen yet the different sub-tasks $(\mathcal{T}_i)_{i=1}^{220}$ are connected. When we observe the presence of a symptom B_i, there is a whole set of abnormalities $\mathbb{B}_i$ that can be related to this abnormality because they are typical of disease for which symptom B_i is typical.

We display in the graphs of figures 3.5 and 3.6 the links between the different sub-tasks for our database. Each node of the graph represents a task, i.e an initial symptom from which we start the medical examination. The links between the nodes are all the darker

as the probability of observing both symptoms together is high. Note that there no reason that $\mathbb{P}[B_i \mid B_j] = \mathbb{P}[B_j \mid B_i]$ for $i \neq j$ this is why the thickness of a link between node B_i and node B_j is computed by the following formula:

$$\frac{1}{2}\left(\mathbb{P}[B_i \mid B_j] + \mathbb{P}[B_j \mid B_i]\right)$$

There is no links between two nodes when the two corresponding symptoms are not associated to any common disease in the database. In the figure 3.6 the only difference in comparison with figure 3.5 is the fact that we removed, for visual reasons, the symptoms that were connected to less than 8 symptoms (the B_i for which $|\mathbb{B}_i| \leq 8$).

The graph obtained appears deeply connected, only a few symptoms are totally disconnected from the others.

An interesting question would be to determine how to allocate the computational resources among the remaining tasks to optimize so as to minimize the computation time required.

It nevertheless seems hard to face this problem with more than a heuristic : to choose at each step to optimize the Q-networks which has the highest rate of sub-problems already solved (where each sub-tasks is weighted by its probability to be faced).

3.5 Numerical Results

For all the experiments involving neural networks, we used the same architecture detailed in table 3.2. We first use an embedding layer since the inputs processed by our neural network should not be treated as numerical values. We then use two hidden layer with ReLu activation and a final layer with linear activation which outputs the Q-values of the possible actions. The ϵ parameter of our stopping criterion is set to 10^{-6} for all the experiments.

Table 3.2 – Neural network architecture for task $\mathcal{T}_i$. $|\mathbb{B}_i|$ the number of remaining relevant symptoms to check.

Name	Type	Input Size	Output Size
L1	Embedding Layer	$\lvert\mathbb{B}_i\rvert$	$3 \times \lvert\mathbb{B}_i\rvert$
L2	ReLu	$3 \times \lvert\mathbb{B}_i\rvert$	$2 \times \lvert\mathbb{B}_i\rvert$
L3	ReLu	$2 \times \lvert\mathbb{B}_i\rvert$	$\lvert\mathbb{B}_i\rvert$
L4	Linear	$\lvert\mathbb{B}_i\rvert$	$\lvert\mathbb{B}_i\rvert$

3.5.1 Our baseline has quasi-optimal performances on small subproblems ♦

We can compare the performance of our policies optimized by a REINFORCE algorithm (the ones of section 3.3), with a policy derived from the classic decision tree algorithm that

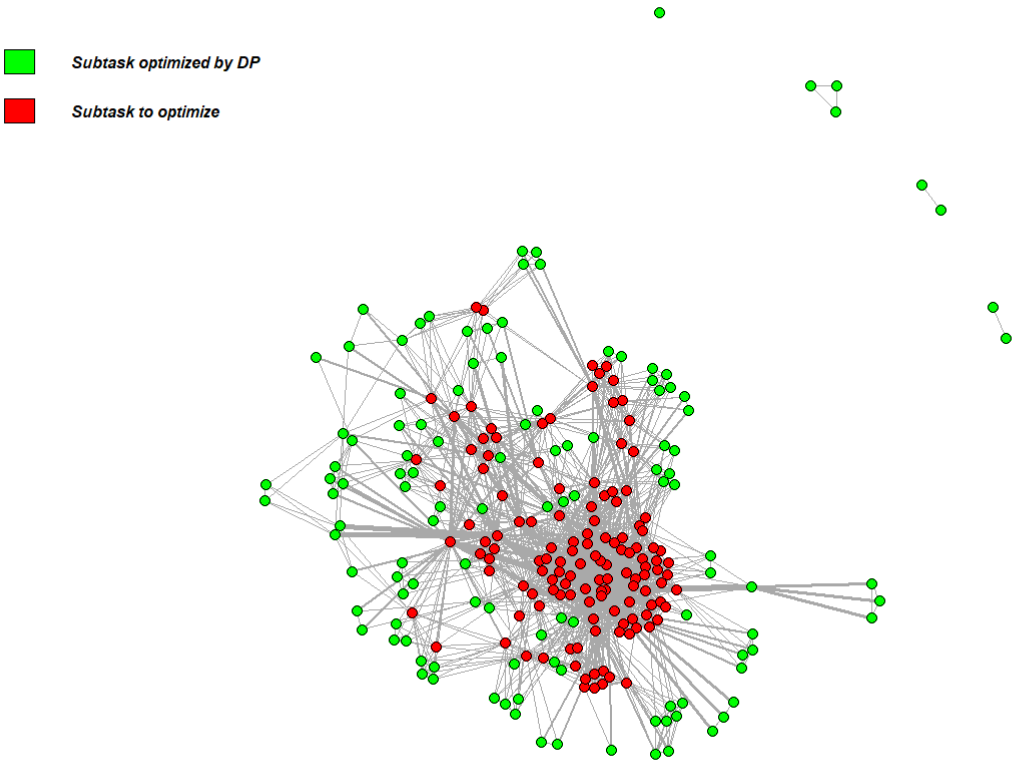


Figure 3.5 – Graph of the links between the sub-tasks.

we call Breiman policy: choosing the action that minimize the target entropy at horizon 1 (see Breiman et al. [1984]). We also compare these policies with the true optimal policies when it was possible to compute the latter, i.e. when the dimension was small enough.

Results on some of our subtasks are presented in Figure 3.7. Our energy-based policy appears to clearly outperform a classic Breiman algorithm and all the more so as the dimension increases: the average number of questions to ask may be divided by two in some cases. On small subproblems where we have been able to compute the optimal policy by a dynamic programing algorithm, our energy-based policy appears to be very close to the optimal policy.

3.5.2 DQN-MC algorithm vs our baseline ♦

We have performed a DQN-MC algorithm on our subtasks. We expect this algorithm to find a better path than the energy-based policy of section 3.3 since a neural network has many more parameters and can therefore handle many more different situations than our baseline. Nevertheless to train such a high dimensional function instead of the three

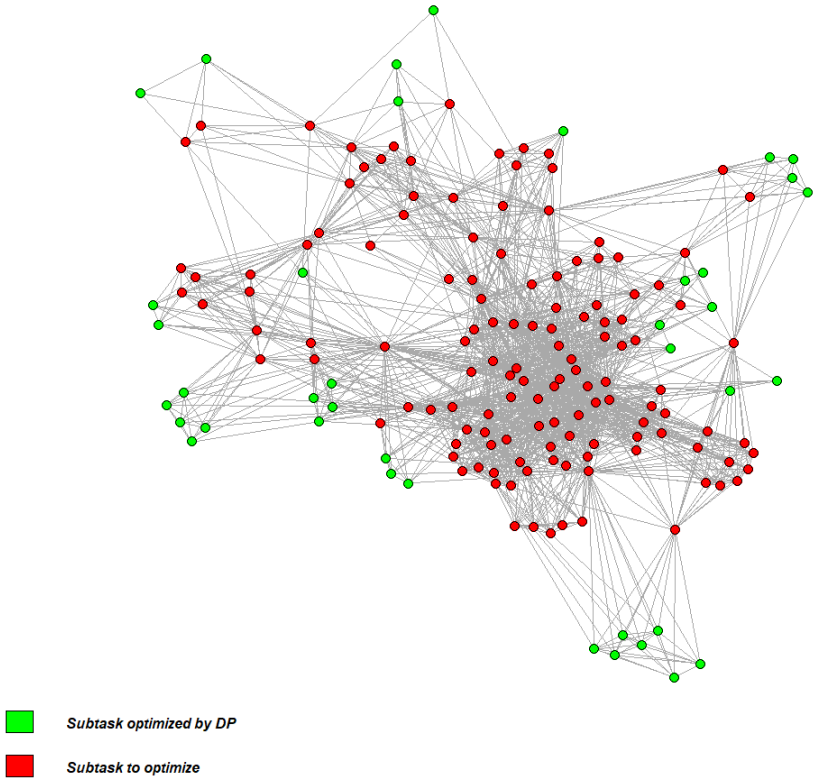


Figure 3.6 – Graph of the links between the sub-tasks.

parameters of our baseline has a cost. How much iterations does need a DQN-MC to outperform our baseline?

We recall here that an iteration of the DQN-MC algorithm consist in playing 100 games that are added to the replay memory, then we sample one twentieth of this replay memory and perform a back-propagation algorithm. For comparison, our baseline has been trained with a REINFORCE algorithm, each iteration consist in playing one game and performing a gradient ascent step accordingly, we stop the training phase when reaching 1000 iterations.

Figures 3.11 and 3.12 show, as expected, that the DQN-MC algorithm needs more simulations of games than our baseline. Indeed in these two sub-tasks, DQN-MC needed respectively 40 and 200 iterations to reach our baseline, so $40 \times 100 = 4000$ and $200 \times 100 = 20000$ games instead of the 1000 which trained our baseline. In Figure 3.11, for a sub-task of dimension 10, we can see that the DQN-MC algorithm needs a reasonable amounts of games to outperform our baseline. In that case, the DQN algorithm found a very good diagnostic policy but did not reach the optimal policy, it is probably stuck in a local extrema (although we do use an exploration parameter).

In Figure 3.12, the DQN algorithm seems to converge toward the baseline. This might

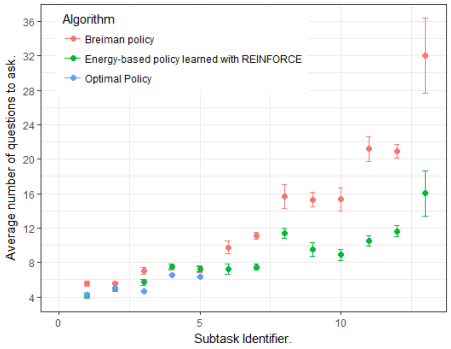


Figure 3.7 – Average number of questions to ask on several subtasks.

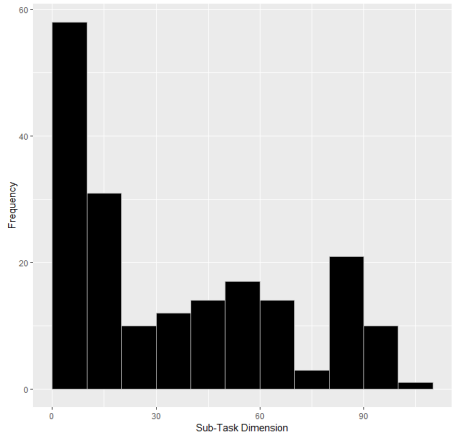


Figure 3.8 – Dimensions of the different sub-tasks.

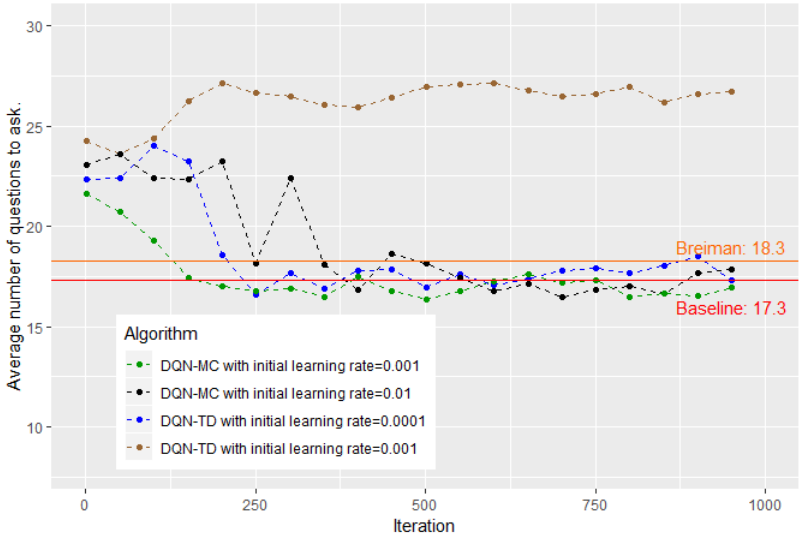


Figure 3.9 – Comparison of DQN-TD and DQN-MC. Task dimension: 29.

be due to the fact that, in these tasks of intermediate dimension (it remains 26 relevant symptoms and 8 diseases), our baseline is yet a good solution close to the optimal policy. Thus the DQN algorithm which is not ensured to converge to the optimal policy might get stuck in a local extrema at the level of the baseline.

These experiments can be conducted in a laptop without use of GPU and should be then easily reproducible using our environment simulator or a similar one.

Finally, as one might expect considering the difference of needed iterations to converge between Figure 3.11 and Figure 3.12, the idea of using previously resolved subtasks will be important to deal with high-dimensional tasks.

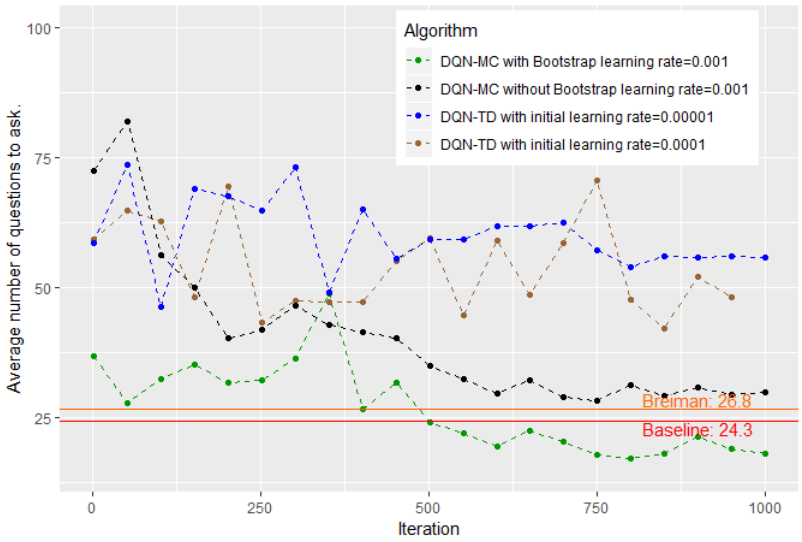


Figure 3.10 – Comparison of DQN-TD, DQN-MC and DQN-MC-Bootstrap. Task dimension: 104.

3.5.3 Bootstraping on already solved sub-tasks helps (a lot) for high-dimensional tasks ♦

In these experiments, we compare the performance of a simple DQN-MC algorithm against a DQN-MC-Bootstrap on some of our tasks. We used the same neural network architecture for both algorithms (see table 3.2). The two algorithms use exactly the same hyper-parameters, the only difference being the bootstrap trick of DQN-MC-Bootstrap.

Figures 3.13 and 3.10 show the benefits of using the solved sub-tasks as bootstraping methods. In both cases a simple DQN-MC is unable to find a good solution while a DQN-MC-Bootstrap outperforms pretty quickly our baseline. Note that the neural network trained with DQN-MC-Bootstrap starts with a policy that is not that bad. It is appreciable as it reduces, since the beginning of the training phase, the length of the episodes and then the computing cost associated.

For the experiment of Figure 3.13 it remains 70 relevant symptoms to check, 9 possible diseases including the disease "other", and 20 sub-tasks have been already solved. Finally the probabilities of presence of each of the subtasks initial symptom given the initial symptom of the main task were (0.01; 0.44; 0.01; 0.15; 0.15; 0.01; 0.03; 0.02; 0.11; 0.01; 0.26; 0.01; 0.03; 0.01; 0.15; 0.01; 0.15; 0.24; 0.16; 0.06).

For the experiment of Figure 3.10 it remains 104 relevant symptoms to check, 18 possible diseases including the disease "other", and 103 sub-tasks have been already solved.

Finally we have been able to learn a good policy for the main task (3.1.2) where it remains 220 relevant symptoms to check, 82 possible diseases including the disease "other" and all the possible sub-tasks have been already solved. Our DQN-MC-Bootstrap algorithm starts with a good policy which only needs 45 questions on average to reach a

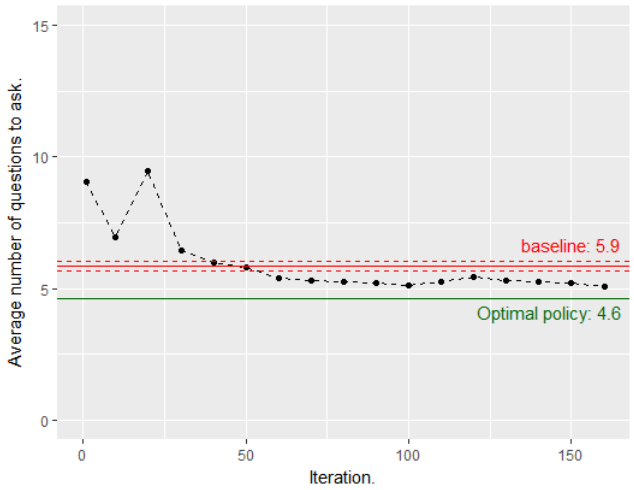


Figure 3.11 – Evolution of the performance of the neural network during the training phase with DQN-MC. Task dimension: 10.

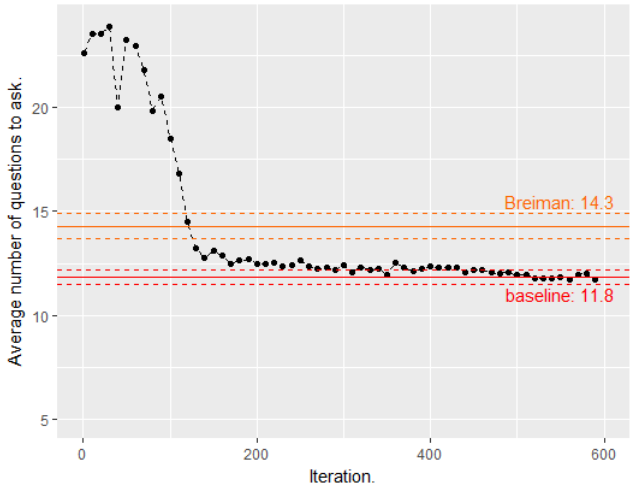


Figure 3.12 – Evolution of the performance of the neural network during the training phase with DQN-MC. Task dimension: 26.

terminal state. Some training iterations allows it to improve until needing 40 questions to reach a terminal state. On the contrary the experiment we made on a DQN-MC which tries to solve from scratch this task has to ask 117 questions, on average, to reach a terminal state and does not improve significantly during the 1000 iterations. We have evaluated also the performance of the Breiman policy on the global task, it needs 89 questions on average to reach a terminal states (with a variance of 10 questions).

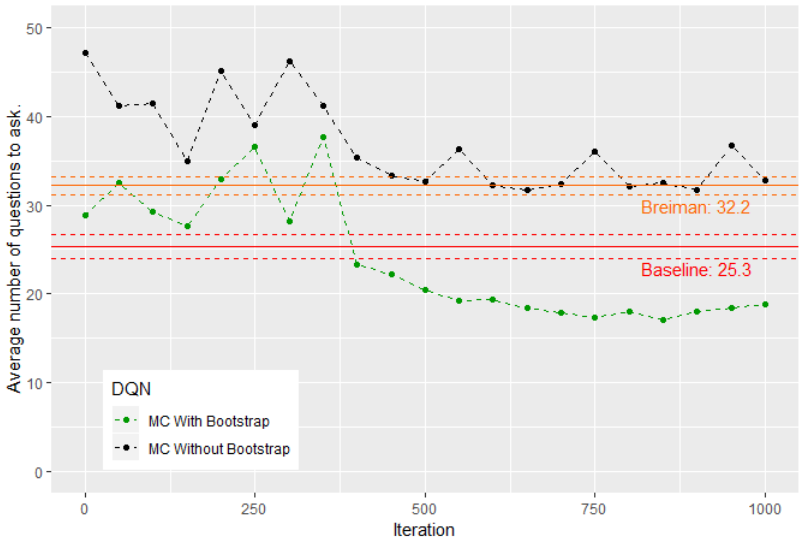


Figure 3.13 – Evolution of the performance of the neural network during the training phase. Task dimension: 70.

3.6 Theoretical and empirical analysis of some difficulties linked to the partitioning of the state space

3.6.1 Partitioned MDP in the look-up table case

Let us assume to simplify the analysis, and without loss of generality, that we split the state space $\mathbb{S}$ into a partition of two sub-spaces $\mathbb{S}_1$ and $\mathbb{S}_2$: $\mathbb{S} = \mathbb{S}_1 \sqcup \mathbb{S}_2$.

We assume that we always start from state $s_0 \in \mathbb{S}_1$ and that all the descendant of the states of $\mathbb{S}_2$ belong to $\mathbb{S}_2$. Namely when we play a game we start in $\mathbb{S}_1$ and when we reach a state of space $\mathbb{S}_2$ we stay in this part of the state partition until the end of the game (we can not move back to $\mathbb{S}_1$).

We aim to learn a strategy $\pi^\star = \arg\max_\pi \mathbb{E}[R \mid s_0, \pi]$ and for that aim to solve the two following optimization problems on each element of the partition :

$$\pi_1^\star = \underset{\pi:\mathbb{S}_1 \to \mathbb{A}}{\arg\max}\, \mathbb{E}[R \mid s_0, \pi]$$

$$\pi_2^\star = \underset{\pi:\mathbb{S}_2 \to \mathbb{A}}{\arg\max}\, \mathbb{E}[R \mid \mathcal{F}(\mathbb{S}_1, \mathbb{S}_2), \pi]$$

where $\mathcal{F}(\mathbb{S}_1, \mathbb{S}_2)$ is the set of states at the frontier between $\mathbb{S}_1$ and $\mathbb{S}_2$, more precisely this is the set of states of $\mathbb{S}_2$ that can be reached from $\mathbb{S}_1$ by taking an unique action. To be exact in the definition of $\pi_2^\star$ we could think that it is necessary to precise the distribution on the initial state $\mathcal{F}(\mathbb{S}_1, \mathbb{S}_2)$. In fact note that as long as all the elements $\tilde{s} \in \mathcal{F}(\mathbb{S}_1, \mathbb{S}_2)$ have a positive probability to be chosen as initial state, the policy $\pi_2^\star$ remains the same. This is a direct consequence of the Bellman principal of optimality.

Moreover let us denote by $\tilde{\pi}$ the aggregation of $\pi_1^\star$ and $\pi_2^\star$, i.e $\tilde{\pi}(s) = \pi_1^\star(s)$ if $s \in \mathbb{S}_1$ and $\tilde{\pi}(s) = \pi_2^\star(s)$ if $s \in \mathbb{S}_2$. Then we have that $\pi^\star = \tilde{\pi}$.

In other words we can learn independently the policy on $\mathbb{S}_1$ and the policy on $\mathbb{S}_2$ when we aim to learn $\pi^\star$.

3.6.2 Partitioned MDP and parameterization of the policy

Things are starting to get more complicated when we use a parameterization of the policy. This is due to the fact that the optimal parameterized value function does not verify the Bellman optimality equation. Intuitively the fact that we presented in the last section that the policy $\pi_2^\star$ remains the same whatever is the distribution put on the initial states set $\mathcal{F}(\mathbb{S}_1, \mathbb{S}_2)$ is no longer verified when we use a parameterized policy $\pi_{\theta_2^\star}$.

Let us denote $V_\pi^{(j)}(s)$ the amount of reward collected by the policy π on the space $\mathbb{S}_j$ starting from the state s.

We denote

$$\pi_{\theta_1^\star} = \underset{\pi_\theta \in \Pi_\theta, \pi_\theta : \mathbb{S}_1 \to \mathbb{A}}{\arg\max} \mathbb{E}[R \mid s_0, \pi]$$

and

$$\pi_{\theta_2^\star} = \underset{\pi_\theta \in \Pi_\theta, \pi_\theta : \mathbb{S}_2 \to \mathbb{A}}{\arg\max} \mathbb{E}[R \mid \mathcal{F}_{\pi_{\theta_1^\star}}(\mathbb{S}_1, \mathbb{S}_2), \pi]$$

where $\mathcal{F}_{\pi_{\theta_1^\star}}(\mathbb{S}_1, \mathbb{S}_2)$ stands for the set of states at the frontier between $\mathbb{S}_1$ and $\mathbb{S}_2$ where each state s of this set has a probability $p^{\pi_{\theta_1^\star}}(s \mid s_0)$ to be the initial state.

We can thus write, for $s \in \mathbb{S}_1$:

$$\begin{aligned}
V_{\pi^\star}(s) - V_{\pi_{\theta^\star}}(s) &= V_{\pi_1^\star}^{(1)}(s) + \sum_{s_T} p^{\pi_1^\star}(s_T \mid s) V_{\pi_2^\star}^{(2)}(s_T) \\
&\quad - \left(V_{\pi_{\theta_1^\star}^\star}^{(1)}(s) + \sum_{s_T} p^{\pi_{\theta_1}^\star}(s_T \mid s) V_{\pi_{\theta_2^\star}^\star}^{(2)}(s_T) \right) \\
&= \Delta_{\pi^\star, \pi_{\theta_1^\star}} V^{(1)}(s) + \sum_{s_T} p^{\pi_{\theta_1^\star}}(s_T \mid s) \Delta_{\pi^\star, \pi_{\theta_2^\star}} V^{(2)}(s_T) \\
&\quad + \sum_{s_T} \left(p^{\pi_1^\star}(s_T \mid s) - p^{\pi_{\theta_1^\star}}(s_T \mid s) \right) V_{\pi_2^\star}^{(2)}(s_T)
\end{aligned}$$

The last equality is composed of three terms. Two of them only depend on the fact that there is no reason why $\pi^\star$ would belong to Π_θ, this is the sense of $\Delta_{\pi^\star, \pi_{\theta_1^\star}} V^{(1)}(s_T)$ and $\Delta_{\pi^\star, \pi_{\theta_2^\star}} V^{(2)}(s_T)$.

The third term of this equality expresses the idea that the discrepancy of the states reached at the frontier following $\pi_{\theta_1^\star}$ instead of $\pi_1^\star$ has a negative impact on the quality of the solution found: $\pi_{\theta^\star}$.

3.6.3 Experiment on a subtask

We have evaluated the difference $V_{\pi^\star}(s) - V_{\pi_{\theta^\star}}(s)$ on a particular sub-task where we were able to learn the optimal policy $\pi^\star$ using the value iteration algorithm.

In the sub-task chosen it remains 12 relevant symptoms to check and 4 suspected diseases, including the disease "other". We have approximated $\pi_{\theta^\star}$ by training a neural network with a DQN-MC algorithm, see algorithm 3, starting from the state with the presence of the call symptom in question and no other information on the other symptoms, let us denote such a state as s_0 in this section. The computational budget is fixed to 500 iterations. We denote $\pi_{\tilde{\theta}}$ the obtained policy.

We computed the difference $V_{\pi^\star}(s) - V_{\pi_{\tilde{\theta}}}(s)$ for all the states s where we add all the possible combinations of 0 to s_0. There are thus $2^{12} = 4096$ possible states.

Each $V_{\pi_{\tilde{\theta}}}(s)$ is estimated by simulating 1000 games starting from state s until the end of the game and we took the average result.

We have obtained $V_{\pi^\star}(s_0) - V_{\pi_{\tilde{\theta}}}(s_0) = 0.15$ where $V_{\pi^\star}(s_0) = 2.44$ so the obtained policy is pretty closed to the performance of the optimal policy when starting from s_0.

We can see in figure 3.14 that most of the time the fact to have trained the neural network by starting the game in s_0 does not have a too bad influence on the performance of $\pi_{\tilde{\theta}}$ compared to $\pi^\star$. In the majority of cases we have to ask less than one additional question comparing with the optimal policy.

Nevertheless this difference can be quite high enough in several cases with up to 4 additional questions. As we are not sure to exactly start in practice from the state s_0 we should try to find solutions to such an issue.

3.6.4 Possible solutions

The discrepancy between the performance of the optimal policy and the performance of the optimal parameterized policy observed in figure 3.14 could be attributed to various factors.

A first hypothesis would be to blame the parameterization itself, i.e to consider that Π_θ is not rich enough to cope with the different situations and a solution would be to use deeper neural network.

Another reason, more plausible, would be that the solution obtained $\pi_{\tilde{\theta}}$ remains far from the performance of $\pi_{\theta^\star}$ especially for the states which are very unlikely to be reached from s_0. Indeed we learn $\pi_{\tilde{\theta}}$ by generating games starting from s_0 because our goal is to maximise the amount of reward collected from s_0 but if we want to be more efficient on less visited part of the tree we will have to balanced differently the exploration/exploitation dilemma.

We recall that so far we tackled the exploration issue by using an exploration factor for the behavior policy, i.e we choose with probability 0.1 a random action instead of the current best action. One could think that the solution is to start with a higher exploration

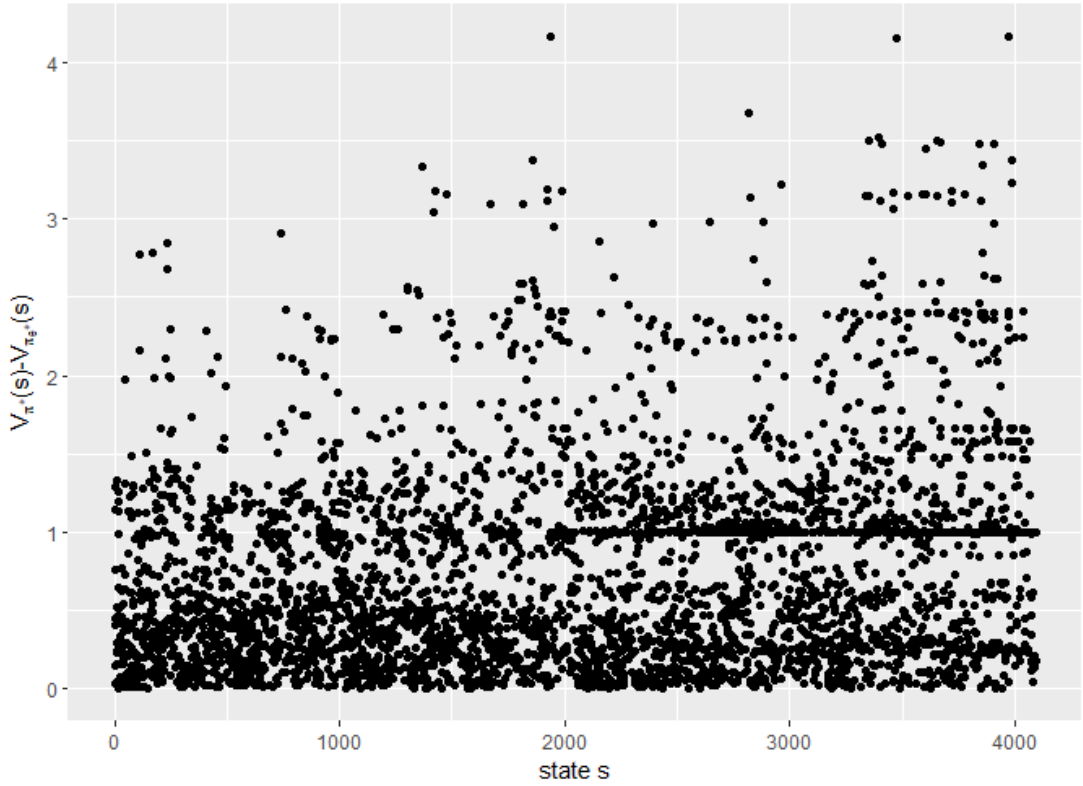


Figure 3.14 – Difference between the value of the optimal policy and the optimal parameterized policy on a particular subtask for several states.

factor and to progressively decrease it. Nevertheless we did not observe any significant improvement in this case.

Another idea consist in training $\tilde{\pi}$ using a DQN-MC algorithm with a random start state. We displayed the results in figure 3.16 where all the states has the same probability to be chosen as the start of a new episode when we simulate games.

We can see that the difference $V_{\pi^\star}(s) - V_{\pi_{\theta^\star}}(s)$ appears much lower in this case, see also the boxplot of figure 3.17. Such a more exploratory learning does not have a too bad influence on the overall quality of the solution found in this particularly case, we got: $V_{\pi^\star}(s_0) - V_{\pi_{\theta^\star}}(s_0) = 0.14$ instead of 0.15 in the less exploratory case.

It then confirms that the exploration/exploitation dilemna is indeed the heart of the matter. Moreover the states s where the difference $V_{\pi^\star}(s) - V_{\pi_{\tilde{\theta}}}(s)$ is the highest in the figure 3.14 were quite predictable. These are the states where we add the absence of symptoms typical of the most plausible disease. Indeed the policy $\pi_{\tilde{\theta}}$ will have done little to explore this parts of the tree since they are relatively unlikely to be reached when starting from s_0. The figure 3.15 shows that the states where the difference $V_{\pi^\star}(s) - V_{\pi_{\tilde{\theta}}}(s)$ is the highest are the states which have been relatively under-visited during the training phase. The probability of reaching each state is calculated as the frequency of such a state in the replay memory divided by the length of the replay memory..

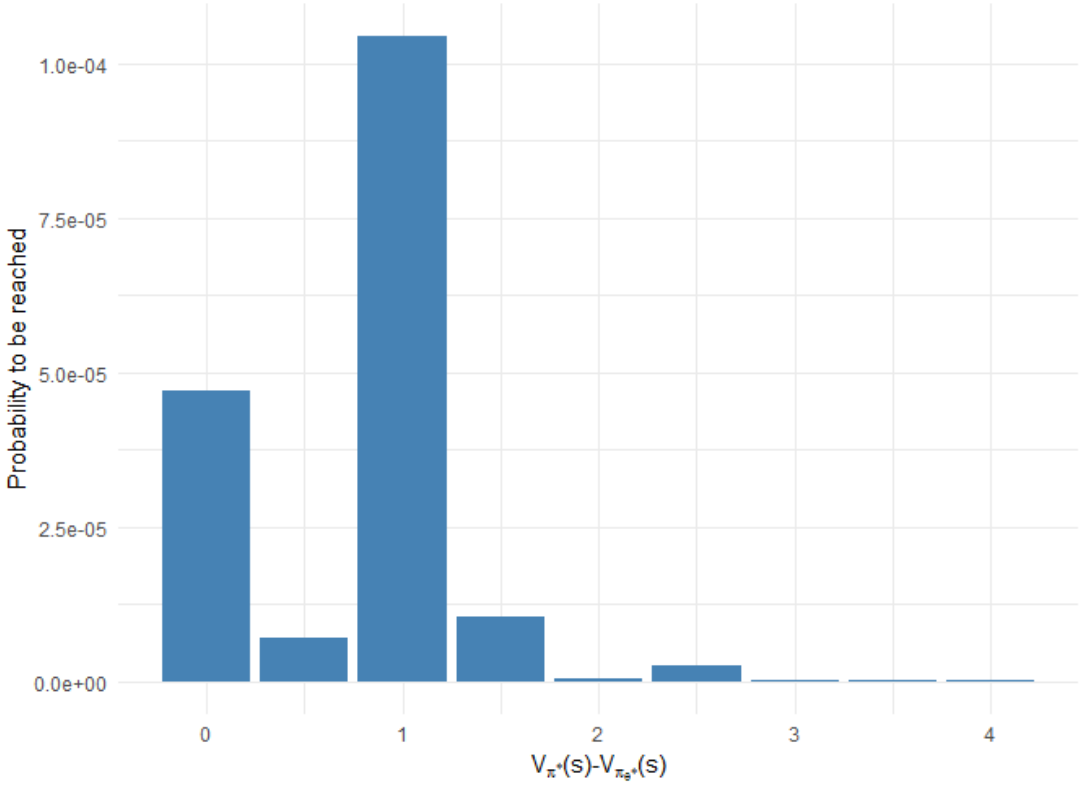


Figure 3.15 – Difference between the value of the optimal policy and the optimal parameterized policy on a particular subtask for several states. Strategy learned with random start.

This is a well known issue in games where Monte-Carlo Tree Search (MCTS) algorithms are designed to face such an exploration/exploitation dilemna, see Browne et al. [2012] for an overview of this field. The standard algorithm used in MCTS is Upper Confidence Tree (UCT) which is inspired of the Upper Confidence Bound (UCB) bandits algorithms, see Kocsis and Szepesvári [2006].

Several amelioration have been proposed since such as Rapid Value Evaluation (RAVE) Gelly and Silver [2007] Gelly and Silver [2011]. A generalization of RAVE (GRAVE) has been developed by Cazenave [2015]. The main idea is to use the statistics of frequently visited states to estimate the values of less visited states.

Indeed the deeper we go in the tree, the less simulations each node receives and the more uncertain our estimates of the values are. The idea of GRAVE is that there is a trade-off between the accuracy of an estimate and compliance with the current state of the estimate. To estimate the value of a state that had been visited only a few number of times (less than a predetermined threshold) Cazenave [2015] use the statistics of a reference state which is the closest ancestor state that has more playouts than the predetermined threshold.

Note that in our case, we have a natural measure of the discrepancy between two

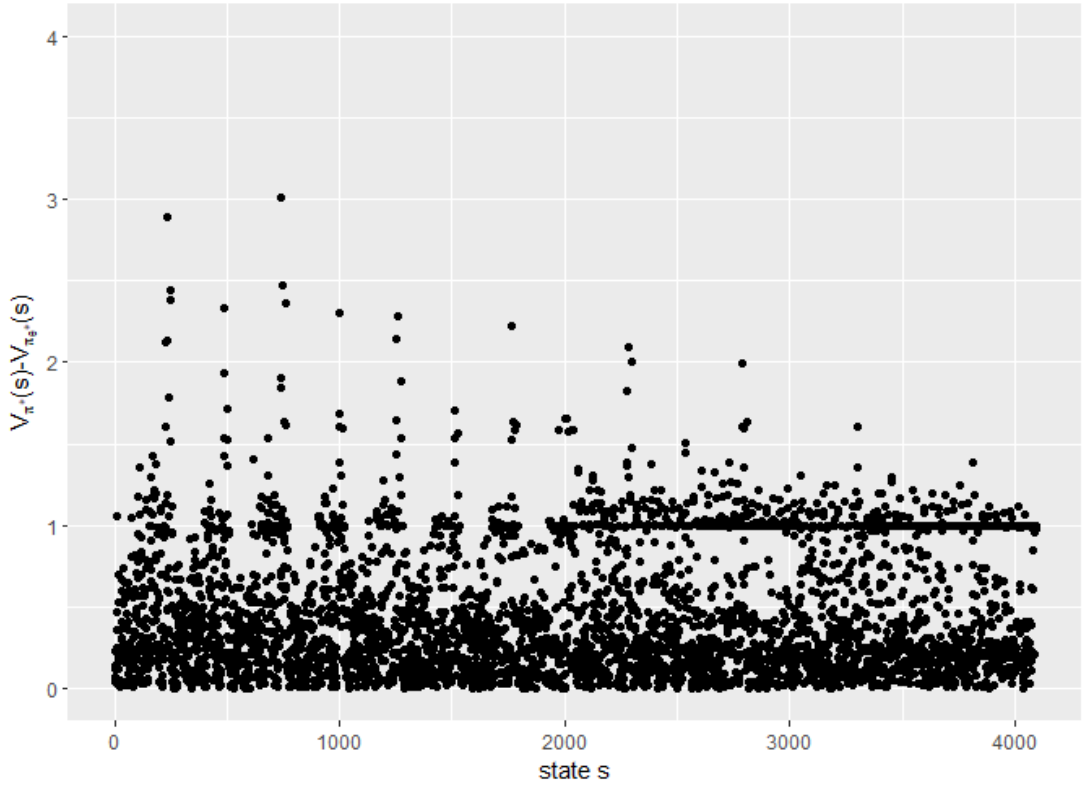


Figure 3.16 – Difference between the value of the optimal policy and the optimal parameterized policy on a particular subtask for several states. Strategy learned with random start.

different states : the Hamming distance. One could then imagine to use not only the statistics of the closest ancestor but the statistics of all the others states that would be weighted in function of their distance. One would then biased the estimate of less visited states by adding a kernel structure. This idea has been studied several years ago for example in Srinivasan et al. [2015] and exhibited good empirical results. More recently Xiao et al. [2018] show that this idea of utilizing information of similar states with a kernel based estimation improves the performance of MCTS in both theory and practice.

3.6.5 Related works

There are many works which aims to find a systematic way to decompose the state space associated to a dynamic programming problem into a family of smaller DP problems that can be independently solved and which together would allow to solve the higher dimensional problem. Let us cite the work of Tsakiris and Tarraf [2012], Tsakiris and Tarraf [2014] for example. Nevertheless such works have to make too strong assumptions about the state transition dynamic. They assume that the state transition dynamic is linear which is the case of a deterministic shortest path but not for a stochastic shortest path as ours.

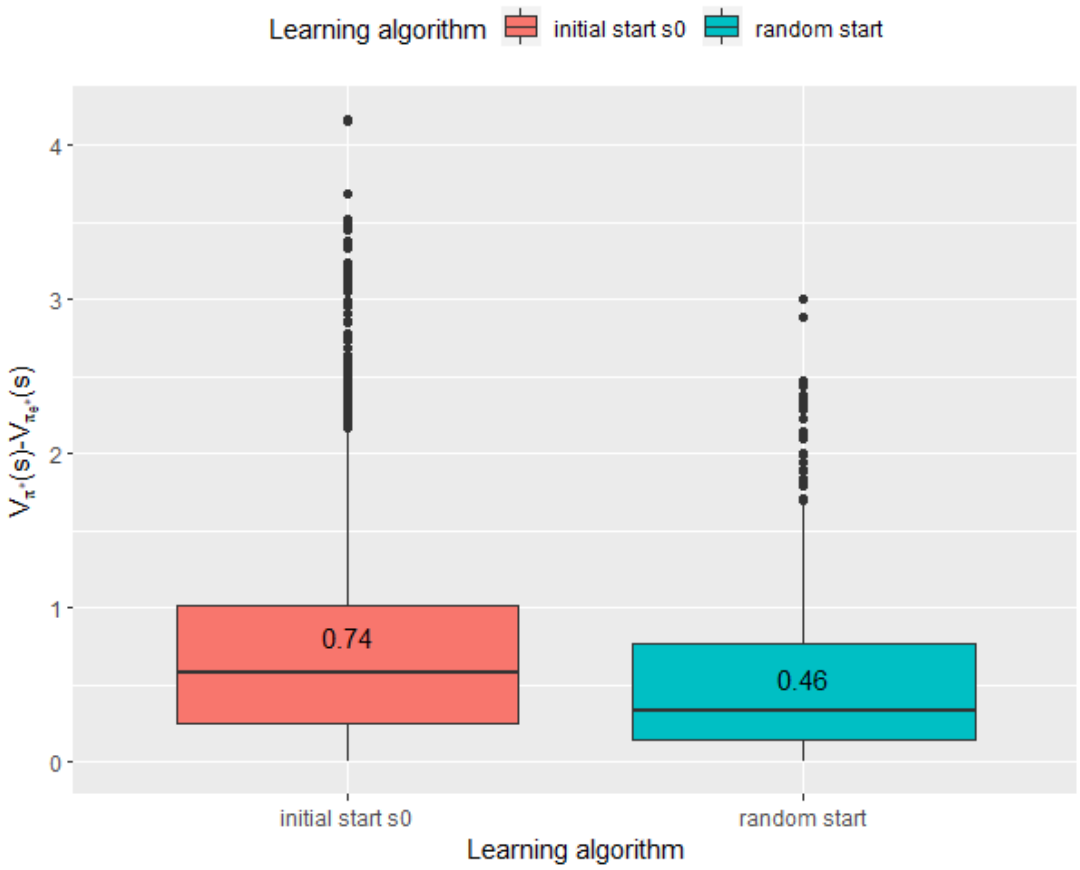


Figure 3.17 – Difference between the value of the optimal policy and the optimal set policy on a particular subtask for several states according to the different way of simulating games.

This notion of partitioned MDP, also referred as decomposed MDP, has also been studied in Dean and Lin [1995] where the authors investigate this "divide and conquer" approach for dimensionality reduction purposes in MDP resolutions. They proposed an "iterative approximation approach" which learn a policy for each sub-task with a given law on the frontier states, glue these local policies in a global policy and thus obtain a new law of the frontier states and iterate this procedure until convergence. However this algorithm is interesting only if the number of states at the frontiers is limited.

This is why several works as Parr [1998], Laroche [2001] or Lozenguez et al. [2012] focused on the particular case of "weakly" coupled MDP, where the number of states connecting two communicating subproblems is assumed to be small.

Nevertheless in our case it does not seem to exist an easy way to decompose the global MDP in several weakly coupled sub-MDP. This is why we proposed to focus on the exploration/exploitation dilemna which reduces the bias induced by the fact that we have an uncertainty about the states in which we start the subtasks.

3.7 Conclusion

In this chapter we have proposed an original formulation of the sequential decision making problem associated to the objective of building a symptom checker for the diagnostic of rare diseases. We have taken into account the need in medicine to reach a high level of certainty when making a diagnosis. Our aim is to minimize the average number of medical tests to be performed before reaching this level of certainty.

We have studied several reinforcement learning algorithms and made them operational in our very high dimensional environment. To do this, we have divided the initial task into several subtasks and learned a strategy for each subtask. We have proven that appropriate use of intersections between subtasks can significantly accelerate the learning process. The strategies learned have proven to be much better than classic greedy strategies.

Chapter references

Rishabh Agarwal, Chen Liang, Dale Schuurmans, and Mohammad Norouzi. Learning to generalize from sparse and underspecified rewards. In *ICML*, 2019.

Artemij Amiranashvili, Alexey Dosovitskiy, Vladlen Koltun, and Thomas Brox. Td or not td: Analyzing the role of temporal differencing in deep reinforcement learning, 2018.

Andrew G. Barto, Steven J. Bradtke, and Satinder P. Singh. Learning to act using real-time dynamic programming. *Artif. Intell.*, 72:81–138, 1995.

Hans J. Berliner. Some necessary conditions for a master chess program. In *Proceedings of the 3rd International Joint Conference on Artificial Intelligence*, IJCAI'73, pages 77–85, San Francisco, CA, USA, 1973. Morgan Kaufmann Publishers Inc. URL http://dl.acm.org/citation.cfm?id=1624775.1624786.

Dimitri P. Bertsekas. *Dynamic Programming and Optimal Control, Vol. I.* Athena Scientific, 2005. ISBN 1886529264. URL http://www.worldcat.org/oclc/314894080.

Dimitri P. Bertsekas. *Dynamic Programming and Optimal Control, Vol. II.* Athena Scientific, 3rd edition, 2007. ISBN 1886529302, 9781886529304.

Dimitri P. Bertsekas and John N. Tsitsiklis. *Neuro-dynamic programming*, volume 3 of *Optimization and neural computation series*. Athena Scientific, 1996. ISBN 1886529108. URL http://www.worldcat.org/oclc/35983505.

Dimitri P. Bertsekas and Huizhen Yu. Stochastic shortest path problems under weak conditions, 2013.

Dimitris Bertsimas and Jack Dunn. Optimal classification trees. *Mach. Learn.*, 106(7): 1039–1082, July 2017. ISSN 0885-6125. doi: 10.1007/s10994-017-5633-9. URL https://doi.org/10.1007/s10994-017-5633-9.

Leo Breiman, Joseph H Friedman, R. A. Olshen, and C. J. Stone. Classification and regression trees. 1984.

Cameron Browne, Edward Jack Powley, Daniel Whitehouse, Simon M. Lucas, Peter I. Cowling, Philipp Rohlfshagen, Stephen Tavener, Diego Perez Liebana, Spyridon Samothrakis, and Simon Colton. A survey of monte carlo tree search methods. *IEEE Trans. Comput. Intellig. and AI in Games*, 4(1):1–43, 2012. URL http://dblp.uni-trier.de/db/journals/tciaig/tciaig4.html#BrownePWLCRTPSC12.

Tristan Cazenave. Generalized rapid action value estimation. In *Proceedings of the 24th International Conference on Artificial Intelligence*, IJCAI'15, pages 754–760. AAAI Press, 2015. ISBN 978-1-57735-738-4. URL http://dl.acm.org/citation.cfm?id=2832249.2832354.

François Chollet et al. Keras. https://github.com/fchollet/keras, 2015.

Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*. Wiley-Interscience, New York, NY, USA, 2006. ISBN 0471241954.

T. de Bruin, J. Kober, K. Tuyls, and R. Babuška. Integrating state representation learning into deep reinforcement learning. *IEEE Robotics and Automation Letters*, 3(3):1394–1401, July 2018. ISSN 2377-3766. doi: 10.1109/LRA.2018.2800101.

Thomas Dean and Shieu-Hong Lin. Decomposition techniques for planning in stochastic domains. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence - Volume 2*, IJCAI'95, pages 1121–1127, San Francisco, CA, USA, 1995. Morgan Kaufmann Publishers Inc. ISBN 1-55860-363-8. URL http://dl.acm.org/citation.cfm?id=1643031.1643045.

ER Doherty-Torstrick, Walton KE, and Fallon BA. Cyberchondria: Parsing health anxiety from online behavior. pages 390–400. Psychosomatics, 2016. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5952212/.

Sylvain Gelly and David Silver. Combining online and offline knowledge in uct. In *Proceedings of the 24th International Conference on Machine Learning*, ICML '07, pages 273–280, New York, NY, USA, 2007. ACM. ISBN 978-1-59593-793-3. doi: 10.1145/1273496.1273531. URL http://doi.acm.org/10.1145/1273496.1273531.

Sylvain Gelly and David Silver. Monte-carlo tree search and rapid action value estimation in computer go. *Artificial Intelligence*, 175:1856–1875, 2011. URL http://www.sciencedirect.com/science/article/pii/S000437021100052X.

P. E. Hart, N. J. Nilsson, and B. Raphael. A formal basis for the heuristic determination of minimum cost paths. *IEEE Transactions on Systems Science and Cybernetics*, 4(2):100–107, July 1968. ISSN 0536-1567. doi: 10.1109/TSSC.1968.300136.

Nicolas Heess, David Silver, and Yee Whye Teh. Actor-critic reinforcement learning with energy-based policies. In Marc Peter Deisenroth, Csaba Szepesvári, and Jan Peters, editors, *Proceedings of the Tenth European Workshop on Reinforcement Learning*, volume 24 of *Proceedings of Machine Learning Research*, pages 45–58, Edinburgh, Scotland, 30 Jun–01 Jul 2013. PMLR. URL http://proceedings.mlr.press/v24/heess12a.html.

Huang Hu, Xianchao Wu, Bingfeng Luo, Chongyang Tao, Can Xu, Wei Wu, and Zhan Chen. Playing 20 question game with policy-based reinforcement learning. *CoRR*, abs/1808.07645, 2018. URL http://arxiv.org/abs/1808.07645.

Leslie Pack Kaelbling, Michael L. Littman, and Anthony R. Cassandra. Planning and acting in partially observable stochastic domains. *Artificial Intelligence*, 101(1):99 – 134, 1998. ISSN 0004-3702. doi: https://doi.org/10.1016/S0004-3702(98)00023-X. URL http://www.sciencedirect.com/science/article/pii/S000437029800023X.

Hao-Cheng Kao, Kai-Fu Tang, and Edward Y. Chang. Context-aware symptom checking for disease diagnosis using hierarchical reinforcement learning. In *AAAI*, 2018.

Levente Kocsis and Csaba Szepesvári. Bandit based monte-carlo planning. In *Proceedings of the 17th European Conference on Machine Learning*, ECML'06, pages 282–293, Berlin, Heidelberg, 2006. Springer-Verlag. ISBN 3-540-45375-X, 978-3-540-45375-8. doi: 10.1007/11871842_29. URL http://dx.doi.org/10.1007/11871842_29.

Sebastian Köhler, Nicole A Vasilevsky, Mark Engelstad, Erin Foster, Julie McMurry, Ségolène Aymé, Gareth Baynam, Susan M. Bello, Cornelius F. Boerkoel, Kym M. Boycott, Michael Brudno, Orion J. Buske, Patrick F. Chinnery, Valentina Cipriani, Laureen E. Connell, Hugh J. S. Dawkins, Laura E. DeMare, Andrew Devereau, Bert B. A. de Vries, Helen V. Firth, Kathleen Freson, Daniel Greene, Ada Hamosh, Ingo Helbig, Courtney Hum, Johanna Jähn, Roger James, Roland Krause, Stanley J. F. Laulederkind, Hanns Lochmüller, Gholson J. Lyon, Soichi Ogishima, Annie Olry, Willem H Ouwehand, Nikolas Pontikos, Ana Rath, Franz Schaefer, Richard H. Scott, Michael Segal, Panagiotis I. Sergouniotis, Richard Sever, Cynthia L. Smith, Volker Straub, Rachel Thompson, Catherine Turner, Ernest Turro, Marijcke W. M. Veltman, Tom Vulliamy, Jing Yu, Julie von Ziegenweidt, Andreas Zankl, Stephan Züchner, Tomasz Zemojtel, Julius O. B. Jacobsen, Tudor Groza, Damian Smedley, Christopher J Mungall, Melissa Haendel, and Peter N. Robinson. The human phenotype ontology in 2017. In *Nucleic Acids Research*, 2017.

Vijaymohan Konda. Actor-critic algorithms. In *NIPS*, 1999.

Richard E. Korf. Depth-first iterative-deepening: An optimal admissible tree search. *Artif. Intell.*, 27(1):97–109, September 1985. ISSN 0004-3702. doi: 10.1016/0004-3702(85)90084-0. URL http://dx.doi.org/10.1016/0004-3702(85)90084-0.

Xiaowei Kortum, Lorenz Grigull, Werner Lechner, and Frank Klawonn. A dynamic adaptive questionnaire for improved disease diagnostics. In Niall Adams, Allan Tucker, and David Weston, editors, *Advances in Intelligent Data Analysis XVI*, pages 162–172, Cham, 2017. Springer International Publishing. ISBN 978-3-319-68765-0.

Sebastian Köhler, Marcel H. Schulz, Peter Krawitz, Sebastian Bauer, Sandra Dölken, Claus E. Ott, Christine Mundlos, Denise Horn, Stefan Mundlos, and Peter N. Robinson. Clinical diagnostics in human genetics with semantic similarity searches in ontologies. *The American Journal of Human Genetics*, 85(4):457 – 464, 2009. ISSN 0002-9297.

doi: https://doi.org/10.1016/j.ajhg.2009.09.003. URL http://www.sciencedirect.com/science/article/pii/S0002929709003991.

Pierre Laroche. GraphMDP: A New Decomposition Tool for Solving Markov Decision Processes. *International Journal on Artificial Intelligence Tools*, 10(3):325–343, 2001. URL https://hal.inria.fr/inria-00100821. Article dans revue scientifique avec comité de lecture.

Guillaume Lozenguez, Lounis Adouane, Aurélie Beynier, Abdel-Illah Mouaddib, and Philippe Martinet. Map Partitioning to Approximate an Exploration Strategy in Mobile Robotics. *Multiagent and Grid Systems - An International Journal of Cloud Computing* , 8(3):275–288, 2012. URL https://hal.archives-ouvertes.fr/hal-00968881.

Katherine Middleton, Mobasher Butt, Nils Y. Hammerla, Steven Hamblin, Karan Mehta, and Ali Parsa. Sorting out symptoms: design and evaluation of the 'babylon check' automated triage system. *CoRR*, abs/1606.02041, 2016. URL http://arxiv.org/abs/1606.02041.

Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Alex Graves, Ioannis Antonoglou, Daan Wierstra, and Martin A. Riedmiller. Playing atari with deep reinforcement learning. *CoRR*, abs/1312.5602, 2013.

Volodymyr Mnih, Adrià Puigdomènech Badia, Mehdi Mirza, Alex Graves, Timothy P. Lillicrap, Tim Harley, David Silver, and Koray Kavukcuoglu. Asynchronous methods for deep reinforcement learning. In *ICML*, 2016.

Andrew Y. Ng and Stuart J. Russell. Algorithms for inverse reinforcement learning. In *Proceedings of the Seventeenth International Conference on Machine Learning*, ICML '00, pages 663–670, San Francisco, CA, USA, 2000. Morgan Kaufmann Publishers Inc. ISBN 1-55860-707-2. URL http://dl.acm.org/citation.cfm?id=645529.657801.

Ronald Parr. Flexible decomposition algorithms for weakly coupled markov decision problems. In *Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence*, UAI'98, pages 422–430, San Francisco, CA, USA, 1998. Morgan Kaufmann Publishers Inc. ISBN 1-55860-555-X. URL http://dl.acm.org/citation.cfm?id=2074094.2074144.

Yu-Shao Peng, Kai-Fu Tang, Hsuan-Tien Lin, and Edward Chang. Refuel: Exploring sparse features in deep reinforcement learning for fast disease diagnosis. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, editors, *Advances in Neural Information Processing Systems 31*, pages 7333–7342. Curran Associates, Inc., 2018. URL http://papers.nips.cc/paper/7962-refuel-exploring-sparse-features-in-deep-reinforcement-learning-for-fast-disease-pdf.

J. R. Quinlan. Induction of decision trees. *Mach. Learn.*, 1(1):81–106, March 1986. ISSN 0885-6125. doi: 10.1023/A:1022643204877. URL http://dx.doi.org/10.1023/A:1022643204877.

Salman Razzaki, Adam Baker, Yura Perov, Katherine Middleton, Janie Baxter, Daniel Mullarkey, Davinder Sangar, Michael Taliercio, Mobasher Butt, Azeem Majeed, Arnold DoRosario, Megan Mahoney, and Saurabh Johri. A comparative study of artificial intelligence and human doctors for the purpose of triage and diagnosis. *CoRR*, abs/1806.10698, 2018. URL http://arxiv.org/abs/1806.10698.

HL Semigran, Linder JA, Gidengil C, and al. Evaluation of symptom checkers for self diagnosis and triage: audit study. page 351. British Medical Journal, 2015. URL https://www.bmj.com/content/351/bmj.h3480.

Sriram Srinivasan, Erik Talvitie, and Michael H. Bowling. Improving exploration in uct using local manifolds. In *AAAI*, 2015.

Richard S. Sutton and Andrew G. Barto. *Introduction to Reinforcement Learning*. MIT Press, Cambridge, MA, USA, 2nd edition, 2018. ISBN 0262193981.

Richard S. Sutton, David A. McAllester, Satinder P. Singh, and Yishay Mansour. Policy gradient methods for reinforcement learning with function approximation. In *NIPS*, 1999.

Csaba Szepesvari. *Algorithms for Reinforcement Learning*. Morgan and Claypool Publishers, 2010. ISBN 1608454924, 9781608454921.

Kai-Fu Tang, Hao-Cheng Kao, Chun-Nan Chou, and Edward Y. Chang. Inquire and diagnose : Neural symptom checking ensemble using deep reinforcement learning. 2016.

M. C. Tsakiris and D. C. Tarraf. On subspace decompositions of finite horizon dynamic programming problems. In *2012 IEEE 51st IEEE Conference on Decision and Control (CDC)*, pages 1890–1895, Dec 2012. doi: 10.1109/CDC.2012.6427001.

Manolis C. Tsakiris and Danielle C. Tarraf. Algebraic Decompositions of DP Problems with Linear Dynamics. *arXiv e-prints*, art. arXiv:1404.5086, Apr 2014.

Eric Wiewiora. Potential-based shaping and q-value initialization are equivalent. *CoRR*, abs/1106.5267, 2011. URL http://arxiv.org/abs/1106.5267.

Ronald J. Williams. Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine Learning*, 8:229–256, 1992.

Chenjun Xiao, Jincheng Mei, and Martin Müller. Memory-augmented monte carlo tree search. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence,*

(AAAI-18), the 30th innovative Applications of Artificial Intelligence (IAAI-18), and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI-18), New Orleans, Louisiana, USA, February 2-7, 2018, pages 1455–1462, 2018. URL https://www.aaai.org/ocs/index.php/AAAI/AAAI18/paper/view/17139.

Tiancheng Zhao and Maxine Eskénazi. Towards end-to-end learning for dialog state tracking and management using deep reinforcement learning. *CoRR*, abs/1606.02560, 2016. URL http://arxiv.org/abs/1606.02560.

Valentina Bayer Zubek and Thomas G. Dietterich. Integrating learning from examples into the search for diagnostic policies. *CoRR*, abs/1109.2127, 2005. URL http://arxiv.org/abs/1109.2127.

Chapter 4

Learning a model of the environment

Abstract: *In this chapter we study the problem of building a model of the environment on which our algorithms will be trained. It is a crucial aspect for many applications, since it is often impossible to deploy in real life an algorithm that has not yet been optimized due to the costs generated by bad decisions. In our case we start without any clinical data and therefore have to rely on expert knowledge entirely. However, we plan to collect empirical data as the decision support tool is used. We present here two main ways to combine the initial experts knowledge, expressed as conditional probabilities, with clinical data: a penalized maximum likelihood approach and a second one consisting in a barycenter between experts and data. We are particularly interested in the intermediate regime where we do not have enough data to do without our initial a priori, but enough to correct it if necessary. We show, empirically and theoretically, that our barycenter estimator is always more efficient than the best of the two models (expert or data) within a constant.*

Contents

4.1 Introduction

4.1.1 The need to learn a model of the environment

As described in chapter 3 our agent will be trained by interacting with its environment. The transitions (initial state, action, reached state, reward)$= (s_t, a_t, s_{t+1}, r_t)$ used to improve the diagnostic strategy can be obtained by interacting directly with the environment in the real world or by simulation using an approximate model of the environment. These two different possible approaches are known as:

- **Model-based RL**: We first build a model of the environment in order to know how our environment will react to our actions. Then our agent is trained using experiences simulated from this model (planning task).
- **Model-free RL**: We do not try to infer the environment dynamic, we just train our agent using trial-and-error directly obtained by the interaction with the environment.

In our problem, as in most cases of practical applications of reinforcement learning, a model-free approach is not an option. Indeed, it is often impossible in a medical or industrial context to deploy in real life an algorithm that has not yet been optimized due to the costs generated by bad decisions. A model-free architecture would need a very

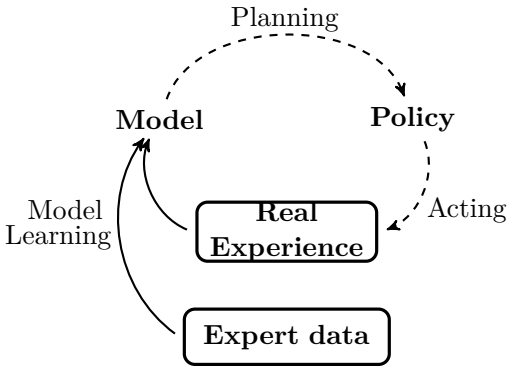


Figure 4.1 – Our global architecture.

large amount of data/time to learn a good policy especially considering the diversity of situations it will face. This is a time that domain knowledge can save us.

Note that in a number of popular applications of reinforcement learning such an model of the environment is not necessary. This is the case in recent advances of computer Go Silver et al. [2017] which shows that it is possible to achieve a superhuman level in a challenging domain as Go without any domain knowledge, using only reinforcement learning with self-play. However, we are not in an adversarial game where we could learn from self play. Another approach is to use expert demonstrations in order to estimate both rewards and environment dynamics Herman et al. [2016] or to learn directly a policy Tossou and Dimitrakakis [2013]. Expert demonstrations are often integrated as a supervised learning initialization step in the AI architecture as in Silver et al. [2016]. We do not have such expert demonstrations and in any case, since we are interested in rare diseases, we would need a very large amount of demonstrations to learn a good policy.

Generally speaking our application area is specific by its lack of data, making the environment dynamic very uncertain. This prevent us from designing our architecture without any domain knowledge.

We detail in this chapter the model learning phase of our architecture (see figure 4.1) where we integrate expert data to the data collected by the experience of the algorithm in order to build a sufficiently accurate model of the environment.

4.1.2 The problem and some notations

Let us denote :

$$p^\star = (p_1^\star, ..., p_K^\star) = \begin{pmatrix} \mathbb{P}[\bar{B}_1, ..., \bar{B}_{J-1}, \bar{B}_J \mid D] \\ \mathbb{P}[\bar{B}_1, ..., \bar{B}_{J-1}, B_J \mid D] \\ \vdots \\ \mathbb{P}[B_1, ..., B_{J-1}, B_J \mid D] \end{pmatrix} \tag{4.1.1}$$

the distribution we aim to estimate. In our specific case, we focus on a particular

disease D and try to estimate the probabilities of the different combinations of symptoms typical of the disease in question, given the event that the disease of the patient is D.

Let B_1,..., B_J be the typical symptoms of the disease D. We aim to estimate the $2^J = K$ different combinations (as $\mathbb{P}[B_1, ..., B_L \mid D]$ for example) when we only have data on the marginals $\mathbb{P}[B_i \mid D]$, for all $i \in [1, J]$.

Of course a first idea would be to assume that the symptoms are conditionally independent given the disease. However, we expect complex correlations between the symptoms typicals of a given disease. Indeed we can imagine two symptoms very plausible individually but which rarely occur together (or even never in the case of incompatible symptoms such as microcephaly and macrocephaly).

Note that the assumption of conditional independence would make it possible to present a disease without having any of the symptoms related to this disease in the database (when there is no B_i such that $\mathbb{P}[B_i \mid D] = 1$), which should be impossible.

We also have empirical data collected as the decision support tool is used. We denote $x^{(1)}$,..., $x^{(n)}$ an i.i.d sample from $p^\star$.

Generally speaking if we had enough empirical data, we would no longer need the experts. Conversely, without empirical data, our model must be based entirely on experts. We detail here the ideas of the two main approaches investigated in this thesis to deal with intermediate regime where we do not have enough data to do without the a priori given by the experts but where we have enough data to correct and specify this initial a priori.

We first recall in section 4.2.1 the principle of maximum entropy which is a common denominator to our two different approaches. We then briefly introduce the ideas of both approaches: the maximum-likelihood with entropic penalization in section 4.2.2 and the barycenter between expert and data in section 4.2.3. We make in section 4.3 a review of the literature. In section 4.4 we present some experiments which confirm that the penalized formulation is very sensitive to the choice of the penalization parameter. On the contrary our barycenter estimator has the advantage to propose an objective way to choose the weight to be given to experts compared to data. We finally show in section 4.5, both empirically and theoretically, that our barycenter estimator is always more efficient than the best of the two models (expert or data) within a constant.

4.2 Mixing expert and empirical data

4.2.1 A common denominator: the maximum entropy principle

The aim to benefit simultaneously from expert data and empirical data has of course a very old history. This is the very essence of Bayesian statistics Gelman et al. [2004] which aims to integrate expert data, in the form of an a priori, which is updated with empirical data using the Bayes' theorem to obtain what will be called the posterior.

Note that in our case we do not have a classical a priori modeling the model parameters with probability distributions. We have an a priori on the marginals as such as a number of constraints on the distribution to be estimated. The absence of an obvious a priori to model the distribution of the parameters naturally leads us to the idea of maximum entropy theorized by Jaynes [1957]. Indeed, if no model seems more plausible to us than another, then we will choose the least informative. This is a generalization of the principle of indifference often attributed to Laplace:

"We consider two events as equally probable, when we see no reason that makes one more probable than the other, because, even if there is an unequal possibility between them, since we don't know which is the biggest, this uncertainty makes us look at one as as likely as the other" de Laplace [1774].

This principle therefore takes the form of an axiom allowing us to construct a method to choose an a priori: the least informative possible consistent with what we know.

We then define the distribution of maximum entropy as follow:

$$p^{\text{maxent}} = \arg\max_{p/p\in\tilde{\mathcal{C}}} H(p) \tag{4.2.1}$$

where $\tilde{\mathcal{C}} = \mathcal{C} \bigcap \mathcal{C}^{\text{expert}}$. $\mathcal{C} = \{p/\sum_i p_i = 1, p_i \geq 0\}$ is the probability simplex and $\mathcal{C}^{\text{expert}}$ is the set of constraints fixed by experts.

Note that p^{maxent} is well-defined, namely it exists and is unique, as long as $\mathcal{C}^{\text{expert}}$ is a convex set. Indeed the function $p \mapsto H(p)$ is strictly concave and it is well-known that a strictly concave function under convex constraints admit an unique maximum.

If $\mathcal{C}^{\text{expert}}$ only contained the constraints for the marginals then p^{maxent} is nothing more that the independent distribution:

Proposition 4.2.1 *Let p^{maxent} defined by (4.2.1) such that $\mathcal{C}^{expert}$ is only composed of the constraints on the marginals $\mathbb{P}[B_1 \mid D]$,..., $\mathbb{P}[B_J \mid D]$ then p^{maxent} is the independent distribution.*

Proof Without loss of generality we can drop the conditioning on the event that the patient has the disease D. Our proposition is then a direct consequence of the following classical inequality for the mutual entropy:

$$H(B_1, ..., B_J) \leq H(B_1) + ... + H(B_J) \tag{4.2.2}$$

with equality if and only if the $(B_i)_{i=1}^J$ are mutually independent.

As the marginals are fixed, so are the $H(B_i)$ for all i. Then $H(B_1, ..., B_J)$ reach its maximum when the $(B_i)_{i=1}^J$ are mutually independent and is then equal to $H(B_1) + ... + H(B_J)$.

Note that the inequality (4.2.2) is a consequence of the chain rule for entropy :

$$H(B_1, ..., B_J) = H(B_J \mid B_1, ..., B_{J-1}) + H(B_{J-1} \mid B_1, ..., B_{J-2}) + ... + H(B_1)$$

and of the fact that for two r.v X and Y: $H(X \mid Y) \leq H(X)$ with equality if and only if $Y \perp\!\!\!\perp X$ ∎

However, in our case, we can add some information about the structure of the desired distribution as constraints integrated to $\mathcal{C}^{\text{expert}}$. We judge impossible to have a disease without having at least a certain amount of its associated symptoms: one, two or more depending on the disease. Indeed the disease we are interested in manifest themselves in combination of symptoms. The combinations allowing the fact to have simultaneously two exclusive symptoms should also be constraints to be equal to 0. All combinations of constraints are conceivable as long as $\tilde{\mathcal{C}}$ remains a convex closed space, in order to ensure the existence and uniqueness of p^{maxent}.

We therefore construct our a priori by taking the maximum entropy distribution checking the constraints imposed by the experts. Thus among the infinite distributions that verify the constraints imposed by the experts, we choose the least informative distribution p^{maxent}, in other words the one closest to the conditional independence distribution.

We need to add information to move from the information provided by the experts to the final distribution and we want to add as little as possible on what we don't know. This approach is referred to as maxent (maximum entropy) and has been widely studied in the literature Jaynes [1957], Cover and Thomas [2006], Berger et al. [1996].

4.2.2 Maximum likelihood with entropic penalization ♦

The idea here is to integrate the collected empirical data by choosing the estimator which maximize the log-likelihood. However as we do not expect to have sufficient amount of data to infer the symptoms distributions we regularize this maximum likelihood approach by an entropic term and by the fact to respect the initial a priori given by experts.

Let us denote

$$\mathcal{P} = \begin{pmatrix} \mathbb{P}[B_1 \mid D] \\ \vdots \\ \mathbb{P}[B_J \mid D] \\ \mathbb{P}[\bar{B}_1, ..., \bar{B}_J \mid D] \\ \vdots \\ \mathbb{P}[B_1, ..., B_J \mid D] \end{pmatrix} =: \begin{pmatrix} \mathbb{P}[B_1 \mid D] \\ \vdots \\ \mathbb{P}[B_J \mid D] \\ p_1 \\ \vdots \\ p_K \end{pmatrix} \tag{4.2.3}$$

and $p = (p_1, ..., p_K)$. Then our estimator is defined as :

$$\begin{aligned}\mathcal{P}^{\text{new}} &= \underset{\mathcal{P}/\mathcal{P}\in\mathcal{C}}{\arg\max}\ L\left(x^{(1)}, ..., x^{(n)} \mid p\right) \\ &\quad + \epsilon\left(H(p) - \sum_i \lambda_i \mathbb{KL}\left(v_i||w_i\right)\right) \\ &=: \underset{\mathcal{P}/\mathcal{P}\in\mathcal{C}}{\arg\max}\ \mathcal{F}(\mathcal{P})\end{aligned} \tag{4.2.4}$$

where $v_i \sim Be\left(\mathbb{P}^{expert}[B_i \mid D]\right)$ and $w_i \sim Be\left(\mathbb{P}[B_i \mid D]\right)$ for all $i \in [1, J]$. The constraint $\mathcal{P} \in \mathcal{C}$ just states the classical probability measure constraints: respect of marginals and sum equal to one, we also add the constraint to set to 0 symptoms combinations considered impossible. Note here that we remove the constraints on the marginal experts and integrate it as a penalization with the Kullback-Leibler term. We have three terms:

- A log-likelihood term for empirical data: $L\left(x^{(1)}, ..., x^{(n)} \mid p\right)$ where $x^{(i)}$ is the i-th combination of symptoms observed in real life. We aim at maximizing this quantity since we want our model to be coherent which what we observed. Symptoms combinations observed in real life should be considered a little bit more plausible. Note that the log-likelihood of independent observations $x = (x^{(1)}, ..., x^{(n)})$ under model $\mathcal{P}$ has a very simple form:

$$L\left(x^{(1)}, ..., x^{(n)} \mid p\right) = \sum_{i=1}^{K} N_i \log(p_i)$$

  where $N_j(x) = \sum_k \mathbb{1}_{\{x^{(k)}=j\}}$ is the number of times we had observed the j-th symptom combination.

- An entropic term, $H(p)$, in order not to consider impossible a symptom combination that has not been yet observed in real life.

- The last term ensures that the marginals of our new distribution will not stray too far from our initial a priori given by expert data: $\mathbb{P}^{expert}[B_i \mid D]$. We recall that $\mathbb{KL}(P||Q) \geq 0$, $\forall P, Q$ and $\mathbb{KL}(P||Q) = 0 \Leftrightarrow P = Q$ Note that each marginal does not have the same coefficient λ_i as we do not have the same confidence in all the expert data. In particular we can handle missing data, i.e when we do not know $\mathbb{P}[B_i \mid D]$, by setting $\lambda_i = 0$.

As we should see in section 4.4 the main issue of such a penalized formulation is the choice of the regularization parameter ϵ.

4.2.3 Barycenters between experts and data ♦

Recall that $p^\star = (p_1^\star, ..., p_K^\star)$ and that $x^{(1)}$,..., $x^{(n)}$ is an i.i.d sample of $p^\star$. The empirical distribution $p_n^{\text{emp}} = (p_{n,i}^{\text{emp}})_{i=1}^K$ is given by:

$$p_{n,i}^{\text{emp}} = \frac{1}{n} \sum_{j=1}^{n} \mathbb{1}_{\{x^{(j)}=i\}}. \tag{4.2.5}$$

Following the ideas of section 4.2.1 we define the expert distribution as the distribution which maximize entropy while satisfying the constraints fixed by experts :

$$p^{\text{expert}} = \arg\max_{p/p\in\tilde{\mathcal{C}}} H(p) \tag{4.2.6}$$

where $\tilde{\mathcal{C}}$ is the intersection of the simplex probabilities with the set of constraints fixed by experts: in our case it is composed of a list of censured combinations and a list of marginals given by experts. Note that it is possible to give more or less credit to the marginals given by experts by formulating the constraint as an interval (more or less wide) rather than a strict equality. The distribution of expert is then defined as the least informative distribution consistent with what we know.

Let $\mathcal{L}$ be any dissimilarity measure between two probability distributions. Our barycenter estimator mixing expert and empirical data is then defined as:

$$\hat{p}_{\epsilon_n}^{\mathcal{L}} = \underset{p\in\mathcal{C}/\mathcal{L}(p_n^{\text{emp}},p)\leq\epsilon_n}{\arg\min} \mathcal{L}(p^{\text{expert}}, p) \tag{4.2.7}$$

where

$$\epsilon_n := \epsilon_n^\delta = \arg\min_l \mathbb{P}[\mathcal{L}(p_n^{\text{emp}}, p^\star) \leq l] \geq 1 - \delta. \tag{4.2.8}$$

$\hat{p}_n^{\mathcal{L}}$ is then defined as the closest distribution from experts, in the sense of the dissimilarity measure $\mathcal{L}$, which is consistant with the observed data.

For such a construction to be possible, we will therefore have to choose a measure of dissimilarity $\mathcal{L}$ such that we have a concentration of the empirical distribution around the true distribution for $\mathcal{L}$.

Such a formulation have several advantages over the first approach. First, we do not have to choose a regularization parameter ϵ which seems to have a strong impact on the results of $\hat{p}^{\text{pen}}$ (see section 4.4.3). This parameter is replaced by the parameter δ, that it is reasonable not to take more than 0.1 and which appears to have low impact on the result of $\hat{p}_n^{\mathcal{L}}$ (see section 4.5.3). Secondly the solution of (4.2.7) can be, it of course depends on the choice of the dissimilarity measure $\mathcal{L}$, easier to compute than the one of the optimization problem (4.2.4) as we should see in section 4.5.

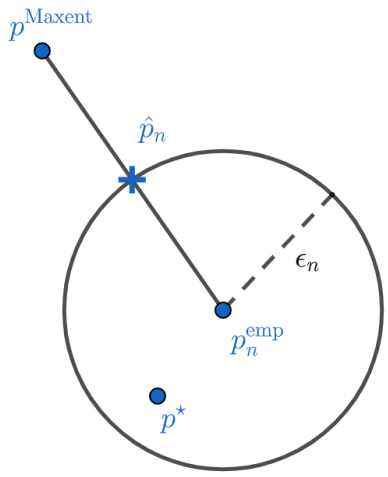


Figure 4.2 – Barycenter between experts and data.

4.3 Related works

4.3.1 Bayesian statistics

The desire to take advantage of expert data and empirical data at the same time has of course a very old history. This is the very essence of Bayesian statistics Gelman et al. [2004] which consists in incorporating expert data, in the form of an a priori, into experimental data by using Bayes' theorem to obtain what will be called the posterior.

Our a priori concerns the marginals and a certain number of constraints on the distribution to be estimated. The absence of an obvious a priori to modelize the parameter's distribution naturally leads us to the idea of maximum entropy theorized by Jaynes [1957].

Indeed, if no model seems more plausible to us than another, then our choice will be the least informative. This is a generalization of the principle of indifference often attributed to Laplace: "We look at two events as equally probable, when we see no reason that makes one more probable than the other, because, even if there is an unequal possibility between them, as we do not know on which side is the greatest, this uncertainty makes us look at one as equally probable as the other" de Laplace [1774]. This principle therefore takes the form of an axiom that allows us to construct a method for choosing an a priori: the least informative possible compatible with what we know.

4.3.2 Expert system with probabilistic reasoning

The creation of a decision support tool for medical diagnosis has been an objective since the beginning of the computer age. Most of the early work proposed a rules-based expert system, but in the 1980s, a significant part of the community studied the possibility of building an expert system using probabilistic reasoning Pearl [1989]. Bayesian probabilities and methods were therefore relatively early considered as good ways to model the

uncertainty inherent in medical diagnosis.

The assumption of conditional independence of symptoms given the disease has been intensively discussed as it is of crucial importance for computational complexity. Some researchers considered this hypothesis harmless Charniak [1983] while others already proposed a maximum entropy approach to face this issue Hunter [1985], Shore [2013] or Miller and Goodman [2013].

However, it seems that none of the work of that time considered the expert vs empirical data trade-off that we face. In the review article Jirousek [1990] presenting the state-of-the-art of the research of that time (1990) about this issue, it is clearly mentioned that these methods only deal with data of probabilistic form. More precisely, they assume that they have an a priori on the marginal but also on some of the combinations of symptoms (in our case we would assume that we have a priori on $\mathbb{P}[B_1, B_2 \mid D]$ for example) and propose a maximum entropy approach where these expert data are treated as constraints in the optimization process. Once again, this is not the case for us since we have only an a priori on the marginal (and a certain number of constraints) as well as experimental data. This field of research was very active in the 1980s and then gradually disappeared, probably due to the computational intractability of the algorithms proposed for the computer resources of the time.

4.3.3 Bayesian Networks

Bayesian networks were then quickly considered as a promising alternative to model probabilistic dependency relationships between symptoms and diseases Pearl [1989]. These are now used in most expert systems, particularly in medicine Koller and Friedman [2009].

A Bayesian network is generally defined as an acyclically oriented graph. The nodes in this graph correspond to the random variables: symptoms or diseases in our case. The edges link two correlated random variables by integrating the information of the conditional law of the son node given the father node. The main advantage of such a model is that it can factorize the joint distribution using the so-called global Markov property. The joint law can indeed be expressed as the product of the conditional distributions of each node given its direct parents in the graph Spiegelhalter et al. [1993].

The construction of a Bayesian network implies first of all to infer its structure, i.e. to determine the nodes that must be linked by an edge of those that can be considered conditionally independent to the rest of the graph (structure learning). Then, learning the network implies learning the parameters, i.e. the probabilities linking the nodes (parameter learning).

It is therefore natural to also find in this area of the literature works that aimed at mixing expert and empirical data. In Zhou et al. [2016] the experts' indications take a particular form since they indicate by hand correlations, positive or negative, between variables. The approach of Constantinou et al. [2016] is also quite distant because it

is preferably based on data. Constantinou et al. [2016] only uses expert indications for additional variables for which there are no data, typically rare events never observed in the database. A work closer to ours is Heckerman et al. [1995] where the authors assume that they have a first Bayesian network built entirely by the experts, to which they associate a degree of trust. The authors then use the available data to correct this expert network. We distinguish ourselves from this work in our effort to find an objective procedure for the weight to be given to experts in relation to the data (and for this weight not to be set by the experts themselves).

Note also that the main interest of Bayesian networks is to take advantage of conditional independence relationships known in advance, as they are pre-filled by experts or inferred from a sufficient amount of data. However, in our case, we do not have such an a priori knowledge about the dependency relationships between symptoms and not enough data to infer them.

4.3.4 Bayesian Reinforcement Learning

We have presented here separately the planning part (optimal policy research) from the transition model learning part. However, some of the literature known as Bayesian Reinforcement Learning proposes to combine these two objectives, see Ghavamzadeh et al. [2016] for an overview of this area. The main aim is to reconsider in traditional reinforcement algorithms the notion of the exploration/exploitation dilemma. It is indeed necessary to grant a bonus to explore further, because the transition probabilities from not often visited states may be poorly estimated. The risk is then to miss out on good solutions because of a poorly estimated transition model.

This exploration, in the sense of using in practice a strategy different from the optimal policy learned, already exists in some way in our system. Indeed, the user can choose to answer any question and not the first one we listed.

Nor have we addressed this issue and assumed that the clinical data received to estimate the transition model are complete. This means that when the fetus is affected by a pathology (which is far from being the most frequent case) it will be necessary to take the time to check the presence/absence of typical symptoms, either at the time of the ultrasound or later via additional examinations.

4.3.5 From the marginals to the joint distribution

Estimating the joint distribution from the marginal is an old problem, which is obviously not necessarily related to expert systems. This problem is sometimes referred to in the literature as the "cell probabilities estimation problem in contingency table with fixed marginals". The book Bishop et al. [1975] gives a good overview of this field. We can trace back to the work of Deming and Stephan [1940] which assumes knowing the marginal and having access to a sample of empirical data and aims to estimate the joint distribution. In

this article, they proposed the "iterative proportional fitting procedure" (IPFP) algorithm, which is still very popular to solve this problem.

An important assumption of Deming and Stephan [1940] is that each cell of the contingency table receives data. In Ireland and Kullback [1968] the authors prove that the asymptotic estimator obtained by an IPFP algorithm is the distribution that minimizes the Kullback-Leibler divergence from the empirical distribution under the constraint to respect the marginal experts.

However, an IPFP algorithm is not suitable for our problem for two main reasons: first, we do not have absolute confidence in the marginals given by experts (we want to allow us to modify them as we collect more data) and second, because since we are interested in rare diseases we do not expect to have a sufficient amount of data. In fact, many of the cells in the contingency table we are trying to estimate will not receive data, but it would be disastrous in our application to assign a zero probability to the corresponding symptom combination.

In a sense, an IPFP algorithm does exactly the opposite of what we are aiming for: it modifies empirical data (as little as possible) to adapt them to experts, while we aim to modify experts (as little as possible) to make them consistent, in a less restrictive sense, with empirical data.

We should also mention the work related to our problem in applications of statistics to the social sciences where researchers aim to construct a synthetic population from marginal coming from several inconsistent sources Barthelemy and Toint [2013]. Their proposed approach also use ideas of maximum entropy but it is still different of our trade-off expert vs empirical data since they build their model without samples.

4.3.6 The Kullback centroid

Our optimization problem (4.2.7) in the particular case where the dissimilarity measure $\mathcal{L}$ is the Kullback Leibler divergence is called moment-projection (M-projection) in the literature. The properties of these projections have been intensely studied Csiszár and Matús [2003].

Note that the Lagrangian associated with such an optimization problem is then nothing more than a Kullback-Leibler centroid. These objects or variations/generalization of them (with Jeffrey's, Bregman's divergences etc...) have been the subject of research since the paper of Veldhuis [2002]. For example, articles Nielsen and Nock [2009] and Nielsen [2013] study cases where an exact formula can be obtained and propose algorithms when this is not the case.

However, we have not found any use of these centroids to find a good trade-off expert vs empirical data as we propose in this paper. Bregman's divergence centroids have been used to mix several potentially contradictory experts, the interested reader may refer to the recent thesis of Adamcik [2014]. We could certainly consider that the empirical

distribution p_n^{emp} is a second expert and that our problem is the same as to mix two experts: literature and data. However, the question of the weight to be given to each expert, which is the question that interests us here, will not be resolved. In Adamcik [2014] the aim is rather to synthesize contradictory opinions of different experts by fixing in advance the weight to be given to each expert. We propose, for our part, an objective procedure to determine the weight to be given to experts comparing to empirical data.

4.4 Proprieties and numerical experiments for the penalized approach

4.4.1 Existence/uniqueness of a solution and numerical considerations ♦

The function $\mathcal{F}$ defined in equation (4.2.4) we aim to optimize is $\mathcal{C}^\infty$ on the constraint space $\mathcal{C}$ which is a compact set (since $\mathcal{C} \subset [0,1]^{2^J+J}$), therefore $\mathcal{F}$ admits a maximum in $\mathcal{C}$. As $\mathcal{F}$ is concave (as a sum of concave functions) this maximum is unique and we can use the Kuhn-Tucker theorem which ensures us that maximizing our function under constraints can be achieved looking for the saddle-point of the Lagrangian.

Deriving the Lagrangian and equating it to 0, we obtain the marginals as function of Lagrangian parameters μ. We write $\mu = (\mu_0, \mu_1, ..., \mu_J)$ the Lagrangian parameters where μ_0 states for the constraint $\sum_j p_j = 1$ and each μ_k states for the marginal constraint respectively to $\mathbb{P}[B_k \mid D]$. If $\lambda_j \neq 0$ we have:

$$\mathbb{P}[B_j \mid D] = \left(1 + \frac{1}{\left(\frac{\mathbb{P}^{expert}[B_j \mid D]}{1 - \mathbb{P}^{expert}[B_j \mid D]} \right) \exp\left(\frac{\mu_j}{\epsilon \lambda_j} \right)} \right)^{-1} ;$$

Note that if $\epsilon\lambda_j \to +\infty$ we indeed recover $\mathbb{P}[B_j \mid D] = \mathbb{P}^{expert}[B_j \mid D]$.

Moreover if $N_j = 0$ we have:

$$p_j = \exp\left(-1 - \frac{\mu_0}{\epsilon} - \sum_k \frac{\mu_k}{\epsilon} \mathbb{1}_{\{B_k=1\}} \right); \tag{4.4.1}$$

If $N_j \neq 0$ we can not obtain a closed form for p_j as function of μ and we have to solve the following equation:

$$-\epsilon \left(\log(p_j) + 1 \right) + \mu_0 + \sum_{j=1}^{K} \mu_j \mathbb{1}_{\{B_j=1\}} + N_j \frac{1}{p_j} = 0;$$

A dichotomy method will be suitable for this task.

Readers familiar with maximum entropy theory should not be surprised by the form of equation (4.4.1) . We recover a classical result, see for example Berger et al. [1996], the

solution of maxent have a nice exponential form: a Gibbs distribution.

We use an Uzawa algorithm to reach the saddle-point of the Lagrangian, see Uzawa [1958]. Since $\mathcal{F}$ is a concave function we are ensured that the saddle-point we converge to by Uzawa iteration is the global maximum of $\mathcal{F}$.

4.4.2 Heuristics for parameters choice ♦

There are two kind of parameters to choose: ϵ and λ_j, $\forall j \in [1, K]$.

We could think that ϵ should decrease with N, but as we have chosen not to renormalize the log-likelihood we have

$$L(x^{(1)}, ..., x^{(N)}) \to \infty$$

when N goes to infinity. A ϵ parameter independent of N seems an easy calibration which provides good results, it should just be chosen large enough to regularize the log-likelihood when N is small (see experiences in section 4.4.3).

However ϵ should depend on the number of unknown parameters of the distribution to be estimated: 2^J (J the number of typical symptoms). Indeed a disease with 12 typical symptoms (i.e. 2^{12} possible symptoms combination) will need far more data than a disease with 4 typical symptoms.

To calibrate ϵ as a function of J, we should look at how the three different terms of (4.2.4) behaves with J. Roughly speaking the entropy term $H(p)$ is of the order of J (the maximal values is $J \times \log(2)$ reached by the uniform distribution). The Kullback-Leibler penalization is linear in J and appears scalable to the entropy (see section 4.4.3).

As it is usually done in log-likelihood regularization we expect the log-likelihood to be of order 2^J: therefore,

$$\epsilon = c \times 2^J$$

where c is a non-negative constant to be determined, seems a reasonable calibration. In practice, our decision support tool will never have a sufficient amount of data and the maxent regularization will allow us to cope with new situations. We will take this into account when choosing c.

Concerning λ_j parameters, the more confident we are in $\mathbb{P}^{expert}[B_j \mid D]$ the higher is λ_j. We simply have to initialize λ_j with sufficiently large values in order to prevent the condition on high entropy to change the marginals too much when N is small as we will see in section 4.4.3. Of course we should not fall into the opposite excess by taking λ_j too large which would have the consequence of staying on the experts' a priori even when the data tell us another reality.

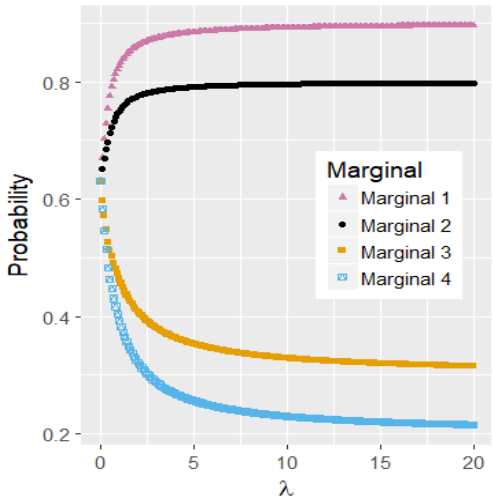


Figure 4.3 – Evolution of marginals' estimates in maxent with marginals regularization.

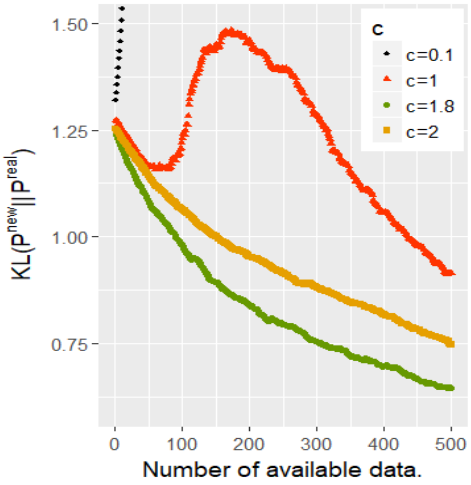


Figure 4.4 – Evolution of the divergence between our estimate and the real symptom combination distribution as function of the amount of available data.

4.4.3 Some experiments ♦

Maxent with Kullback penalization on marginals a priori

Let us start by looking at what happens when we make a maxent with Kullback regularization on marginals, i.e we exclude likelihood for this synthetic experiment. Namely we are interested in a vector $\hat{\mathcal{P}}$ defined as follows:

$$\hat{\mathcal{P}} = \underset{\mathcal{P}/\mathcal{P}\in\mathcal{C}}{\arg\max}\ H(p) - \lambda \sum_{i=1}^{J} \mathbb{KL}\Big(Be\left(\mathbb{P}^{expert}[B_i \mid D]\right) \,||\, Be(\mathbb{P}[B_i \mid D])\Big);$$

where $v_i \sim Be\left(\mathbb{P}^{expert}[B_i \mid D]\right)$ and $w_i \sim Be\left(\mathbb{P}[B_i \mid D]\right)$ for all $i \in [1, J]$.

We set $J = 4$, with the following a priori on marginals $\left(\frac{9}{10}, \frac{8}{10}, \frac{3}{10}, \frac{2}{10}\right) =: (\mathbb{P}^{expert}[B_1 \mid D], \mathbb{P}^{expert}[B_2 \mid D], \mathbb{P}^{expert}[B_3 \mid D], \mathbb{P}^{expert}[B_4 \mid D])$.

For this experiment, we increase λ and look at how it affects the marginals' estimates. We can see in figure 4.3 that all the marginals' estimators start with value 0.63 and then decrease or increase in a monotonous way toward their a priori. This is not surprising since maxent tends to disseminate weight on the entire distribution making parameters of Bernoulli marginals distribution closer to 0.5. In our case we have marginals equal to 0.63 since we enforce combinations with less than one symptom to have zero probability. This gives us an idea of how λ should be initialized.

Adding data

We simulated a symptom combination distribution $\mathcal{P}^{real}$ (with $J = 8$ associated symptoms) using Poisson distribution of parameter 1. The estimate $\mathcal{P}^{new}$ solution of (4.2.4)

given by our Uzawa algorithm has been sequentially updated using data sequentially simulated from $\mathcal{P}^{real}$. For the a priori on marginals, we used the real marginals with an additive Gaussian noise of zero mean and 1/4 variance. The measure of interest is the Kullback-Leibler divergence between the real distribution and our estimate solution of (4.2.4): $\mathbb{KL}(\mathcal{P}^{new}||\mathcal{P}^{real})$ which we would like to minimize. We are interested in how the choice of $\epsilon = c \times 2^J$ affects our estimation of the real distribution. To cope with inherent randomness of this process, an average estimate of the Kullback-Leibler divergence was obtained over 50 repetitions of the same procedure (i.e we simulated 50 Poisson distributions for each different values of c).

In Figure 4.4, the red ($c = 1$) and black ($c = 0.1$) curves clearly show that giving too much weight to the data leads to over-weighing the symptoms combinations observed in real life and keeps us far from the real distribution: we do not sufficiently regularize with the entropy. On the contrary the green ($c = 1.8$) and the orange ($c = 2$) curves performs a good trade-off maxent/maximum likelihood. $c = 2$ is a more cautious choice (we underweight experimental data) than $c = 1.8$ and as a consequence the procedure converge less quickly to the real distribution.

Note that an empirical estimate (solution of a maximum likelihood approach) or an IPFP algorithm would perform very poorly on this task. Indeed many symptoms combinations would be estimated to be 0 when they should not, because of data scarcity: indeed we have $2^8 = 256$ variables and less than 500 data. We have not plotted the Kullback-Leibler divergence of these estimates with respect to the real distribution since it is infinite. In contrast our approach appears robust to data scarcity, provided that we take care of the value of ϵ.

4.5 Numerical experiments and theoretical properties of the barycenter estimator

4.5.1 Barycenter in normed spaces ♦

In this section we work in spaces L^p. Let us recall that the classic norm on the space L^p is given by : $\|x\|_j = \left(\sum_i |x_i|^j\right)^{\frac{1}{j}}$.

Following the ideas presented in section 4.2.3 we define our estimator, $\forall i \geq 1$, $\forall j \geq 1$ as follow :

$$\hat{p}_n^{i,j} = \underset{p \in \mathcal{C} / \|p - p^{\text{expert}}\|_i \leq \epsilon_n}{\arg\min} \|p - p_n^{\text{emp}}\|_j \tag{4.5.1}$$

where

$$\epsilon_n := \epsilon_n^\delta = \arg\min_l \mathbb{P}[\|p_n^{\text{emp}} - p^\star\|_i \leq l] \geq 1 - \delta. \tag{4.5.2}$$

To control ϵ_n we use the concentration inequality obtained in the recent work of Mardia

et al. [2018]. In the literature, most of the concentrations inequalities for the empirical distribution use the L^1 norm. This is why, even if we will present in the following results by trying to generalize for as many couples (i, j) as possible, in practice only the $\hat{p}_n^{1,j}$, for all $j \geq 1$, interest us.

Proposition 4.5.1 *(existence and uniqueness) The estimator $\hat{p}_n^{i,j}$ defined by (4.5.1) exists for all $i \geq 1, j \geq 1$.*

$\hat{p}_n^{i,j}$ is unique if and only if $i \neq 1$.

In the following $\hat{p}_n^{1,1}$ therefore refers to a set of probability measures.

Proof The existence of a solution of (4.5.1) for all $i \geq 1$ and $j \geq 1$ is a consequence of the fact that the projection onto a finite dimension set always exists.

The uniqueness of a solution of (4.5.1) for all $i \neq 1$ is due to the fact that we aim to minimize a strictly convex function under convex constraints. When $j = 1$ the function that we aim to minimize is no longer strictly convex and some counter-examples can be exhibited.

For example if $p^{\text{expert}} = (\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$, $p_n^{\text{emp}} = (\frac{1}{2}, 0, \frac{1}{2}, 0)$ et $\epsilon_n = \frac{9}{10}$.

Note that $\|p^{\text{expert}} - p_n^{\text{emp}}\|_1 = 1 > \frac{9}{10}$. Then using proposition 4.5.2 we know that

$$\begin{aligned}&\frac{\epsilon_n}{\|p_n^{\text{emp}} - p^{\text{expert}}\|_1} p^{\text{expert}} + \left(1 - \frac{\epsilon_n}{\|p_n^{\text{emp}} - p^{\text{expert}}\|_1}\right) p_n^{\text{emp}} \\ &= \left(\frac{11}{40}, \frac{9}{40}, \frac{11}{40}, \frac{9}{40}\right) =: \hat{p}_n^{1,1}\end{aligned}$$

is solution. But $\tilde{p} = (\frac{10}{40}, \frac{9}{40}, \frac{12}{40}, \frac{9}{40})$ is solution too. Indeed :

$$\|\tilde{p} - p^{\text{expert}}\|_1 = \frac{1}{10} = \|\hat{p}_n^{1,1} - p^{\text{expert}}\|_1$$

and:

$$\|\tilde{p} - p_n^{\text{emp}}\|_1 = \frac{36}{40} = \frac{9}{10}.$$

∎

The next proposition shows that one of the solutions of (4.5.1) can always be written as a barycenter between p_n^{emp} and p^{expert} when $i = j$. This property therefore provides us, in these cases, with an explicit expression of a solution of (4.5.1) which was not otherwise trivial to obtain by a direct calculation looking for the saddle points of the Lagrangian (for example in the case $i = j = 1$).

Proposition 4.5.2 *Let $\hat{p}_n^{i,j}$ defined by (4.5.1) then for all $i = j$, it exists $\tilde{p} \in \hat{p}_n^{i,j}$ such that $\exists \alpha_n \in [0, 1]$:*

$$\tilde{p} = \alpha_n p^{expert} + (1 - \alpha_n) p_n^{emp} \tag{4.5.3}$$

where $\alpha_n = \dfrac{\epsilon_n}{\|p_n^{emp} - p^{expert}\|_i}$ if $\epsilon_n \leq \|p_n^{emp} - p^{expert}\|_i$ and $\alpha_n = 1$ otherwise.

Proof Let $\tilde{p} \in \mathcal{C}$ be such that it exists $\alpha \in [0,1]$ where $\tilde{p} = \alpha p^{\text{expert}} + (1-\alpha)p_n^{\text{emp}}$ and such that $\|\tilde{p} - p_n^{\text{emp}}\|_i = \epsilon_n$. We then have :

$$\|\tilde{p} - p_n^{\text{emp}}\|_i = \alpha \|p_n^{\text{emp}} - p^{\text{expert}}\|_i = \epsilon_n$$

and then

$$\alpha = \frac{\epsilon_n}{\|p_n^{\text{emp}} - p^{\text{expert}}\|_i}.$$

Moreover note that we have the following equality since $\tilde{p}$ can be written under the form of a barycenter :

$$\|\tilde{p} - p^{\text{expert}}\|_i + \underbrace{\|\tilde{p} - p_n^{\text{emp}}\|_i}_{=\epsilon_n} = \|p_n^{\text{emp}} - p^{\text{expert}}\|_i$$

Let us make a reasoning by reductio ad absurdum. Let us assume that there exist $p' \in \mathcal{C}$ such that $\|p_n^{\text{emp}} - p'\|_i \leq \epsilon_n$ et $\|p^{\text{expert}} - p'\|_i < \|p^{\text{expert}} - \tilde{p}\|_i$.

We would then have:

$$\begin{aligned}
\|p_n^{\text{emp}} - p^{\text{expert}}\|_i &\leq \|p' - p^{\text{expert}}\|_i + \|p' - p_n^{\text{emp}}\|_i \\
&< \|\tilde{p} - p^{\text{expert}}\|_i + \|p' - p_n^{\text{emp}}\|_i \\
&= \|p_n^{\text{emp}} - p^{\text{expert}}\|_i - \epsilon_n + \|p' - p_n^{\text{emp}}\|_i \\
&\leq \|p_n^{\text{emp}} - p^{\text{expert}}\|_i
\end{aligned}$$

which leads to the desired contradiction. ∎

In particular one of the elements of $\hat{p}_n^{1,1}$ can be written under the form of a barycenter. For the sake of simplicity, we will designate in the following by $\hat{p}_n^{1,1}$ the solution of (4.5.1) for $i = j = 1$ which can be written under the form (4.5.3) and no more the whole set of solutions.

Remark Note that the proposition 4.5.2 is not true when $i = 1$ and $j \neq 1$. This is why we focus on $\hat{p}_n^{1,1}$ for the end of this section.

It is now a question of deriving a result proving that mixing experts and data as we do with $\hat{p}_n^{1,1}$ represents an interest rather than choosing binary one of the two models. For this reason, we show in the following proposition that with a high probability, our estimator $\hat{p}^{1,1}$ is always better than the best of the models within a constant.

Theorem 4.5.3 *Let $\hat{p}_n^{1,1}$ defined by (4.5.1). Then we have with probability at least $1-\delta$:*

$$\|p^\star - \hat{p}_n^{1,1}\|_1 \leq 2 \min\{\epsilon_n, \|p^\star - p^{expert}\|_1\} \tag{4.5.4}$$

Proof A simple application of the triangular inequality gives us :

$$\|p^\star - \hat{p}_n^{1,1}\|_1 \leq \|p^\star - p_n^{\text{emp}}\|_1 + \|p_n^{\text{emp}} - \hat{p}_n^{1,1}\|_1$$

However $\|p_n^{\text{emp}} - \hat{p}_n^{1,1}\|_1 \leq \epsilon_n$ by construction and we have $\|p^\star - p_n^{\text{emp}}\|_1 \leq \epsilon_n$ with probability at least $1 - \delta$.

In addition to that :

$$\|p^\star - \hat{p}_n^{1,1}\|_1 \leq \|p^\star - p^{\text{expert}}\|_1 + \|p^{\text{expert}} - \hat{p}_n^{1,1}\|_1$$

However using the definition of $\hat{p}_n^{1,1}$ and assuming $\|p^\star - p_n^{\text{emp}}\|_1 \leq \epsilon_n$ then:

$$\|p^{\text{expert}} - \hat{p}_n^{1,1}\|_1 \leq \|p^{\text{expert}} - p^\star\|_1.$$

We can conclude that if $\|p^\star - p_n^{\text{emp}}\|_1 \leq \epsilon_n$, which happens with probability at least $1 - \delta$, then :

$$\|p^\star - \hat{p}_n^{1,1}\|_1 \leq 2 \min\{\epsilon_n, \|p^\star - p^{\text{expert}}\|_1\} \tag{4.5.5}$$

∎

4.5.2 Barycenter using the Kullback-Leibler divergence ♦

In this section we study the theoretical properties of the solution of equation (4.2.7) in the particular case where the dissimilarity measure $\mathcal{L}$ is the Kullback-Leibler divergence.

The Kullback-Leibler divergence between two discrete probability measure p and q is defined as:

$$\mathbb{KL}(p||q) = \sum_i p_i \log\left(\frac{p_i}{q_i}\right).$$

Let us recall that the Kullback-Leibler divergence is not a distance since it is not symmetric and does not satisfies the triangular inequality, it is however positive defined Cover and Thomas [2006].

We define our estimator as :

$$\hat{p}_n^L = \underset{p \in \mathcal{C}/\mathbb{KL}(p_n^{\text{emp}}||p) \leq \epsilon_n}{\arg\min} \mathbb{KL}(p^{\text{expert}}||p) \tag{4.5.6}$$

where

$$\epsilon_n := \epsilon_n^\delta = \arg\min_l \mathbb{P}[\mathbb{KL}(p_n^{\text{emp}}||p^\star) \leq l] \geq 1 - \delta. \tag{4.5.7}$$

To calibrate ϵ_n, we can use the concentration inequality obtained in Mardia et al. [2018]. More precisely we have :

$$\epsilon_n = \frac{1}{n}\left(-\log(\delta) + \log \underbrace{\left(3 + 3\sum_{i=1}^{K-2} \left(\sqrt{\frac{e^3 n}{2\pi i}} \right)^i \right)}_{=:G_n} \right). \tag{4.5.8}$$

In the following proposition, we show the existence and uniqueness of our estimator $\hat{p}_n^L$ and the fact that our estimator is a barycenter. However it does not seem possible this time, unlike the case of $\hat{p}_n^{1,1}$ of the equation (4.5.1), to obtain a closed form for $\hat{p}_n^L$.

Proposition 4.5.4 *Let $\hat{p}_n^L$ defined by (4.5.6) then $\hat{p}_n^L$ exists and is unique. Moreover $\hat{p}_n^L$ can be written under the following form :*

$$\hat{p}_n^L = \frac{1}{1+\tilde{\lambda}} p^{expert} + \frac{\tilde{\lambda}}{1+\tilde{\lambda}} p_n^{emp} \tag{4.5.9}$$

where $\tilde{\lambda}$ is a non-negative real such that:

$$\tilde{\lambda} \geq \frac{\mathbb{KL}(p_n^{emp}||p^{expert})}{\epsilon_n} - 1. \tag{4.5.10}$$

Proof The existence and uniqueness of $\hat{p}_n^L$ is a consequence of the fact that $\mathcal{T} = \{p/p \in \mathcal{C} \text{ and } \mathbb{KL}(p_n^{\text{emp}}||p) \leq \epsilon_n\}$ is a convex set. Indeed let $p, q \in \mathcal{T}$, $\alpha \in [0,1]$ then using the classical log-sum inequality we have:

$$\mathbb{KL}(p_n^{\text{emp}}||\alpha p + (1-\alpha)q) \leq \alpha \mathbb{KL}(p_n^{\text{emp}}||p) + (1-\alpha)\mathbb{KL}(p_n^{\text{emp}}||q) \leq \epsilon_n.$$

The Lagrangian associated to the optimization problem (4.5.6) can be written as :

$$\begin{aligned} L(p,\lambda,\mu) = &\sum_i p_i^{\text{expert}} \log\left(\frac{p_i^{\text{expert}}}{p_i}\right) \\ &+ \lambda\left(\sum_i p_i^{\text{emp}} \log\left(\frac{p_i^{\text{emp}}}{p_i}\right) - \epsilon_n\right) \\ &+ \mu\left(\sum_i p_i - 1\right) \end{aligned}$$

Deriving for all $i \in [1, K]$:

$$\frac{\partial L(p,\lambda,\mu)}{\partial p_i} = -\frac{p_i^{\text{expert}}}{p_i} - \lambda\frac{p_i^{\text{emp}}}{p_i} + \mu$$

Equating this last expression to 0 and using the fact that the probability measures sums to 1 we find : $\mu = \lambda + 1$. Then we have for all $i \in [1, K]$:

$$p_i = \frac{1}{1+\lambda} p_i^{\text{expert}} + \frac{\lambda}{1+\lambda} p_i^{\text{emp}}$$

We know that $\hat{p}_n^L$ exists and is unique and using the Kuhn-Tucker theorem (whose assumptions we satisfy since we minimize a convex function under convex inequality constraints) we know that the minimum of the optimization problem (4.5.6) is reached for the saddle-point of the Lagrangian : $(\tilde{\lambda}, \tilde{p}) = (\tilde{\lambda}, \hat{p}_n^L)$. We can then write :

$$\hat{p}_n^L = \frac{1}{1+\tilde{\lambda}} p^{\text{expert}} + \frac{\tilde{\lambda}}{1+\tilde{\lambda}} p_n^{\text{emp}} \tag{4.5.11}$$

We could not obtain a closed form for $\tilde{\lambda}$ unlike the case of $\hat{p}^{i,i}$. however we know that by construction $\mathbb{KL}(p_n^{\text{emp}}||\hat{p}_n) \leq \epsilon_n$.

Moreover using the log-sum inequality and our interpolation formula (4.5.11) we have:

$$\begin{aligned}\mathbb{KL}(p_n^{\text{emp}}||\hat{p}_n) &= \mathbb{KL}\left(p_n^{\text{emp}}||\frac{1}{1+\tilde{\lambda}} p^{\text{expert}} + \frac{\tilde{\lambda}}{1+\tilde{\lambda}} p_n^{\text{emp}}\right) \\ &\leq \frac{1}{1+\tilde{\lambda}} \mathbb{KL}(p_n^{\text{emp}}||p^{\text{expert}}).\end{aligned}$$

We then have the following condition under $\tilde{\lambda}$:

$$\frac{1}{1+\tilde{\lambda}} \mathbb{KL}(p_n^{\text{emp}}||p^{\text{expert}}) \leq \epsilon_n \Leftrightarrow \tilde{\lambda} \geq \frac{\mathbb{KL}(p_n^{\text{emp}}||p^{\text{expert}})}{\epsilon_n} - 1$$

∎

The following proposition is intended to be the analog of the proposition 4.5.3 when $\mathcal{L}$ is the Kullback-Leibler divergence. We prove that the centroid $\hat{p}_n^L$ is better than the experts (with high probability). On the other hand, we obtain that when $\mathbb{KL}(p_n^{\text{emp}}|||p^\star) > \mathbb{KL}(p^{\text{expert}}||p^\star)$, the $\hat{p}_n^L$ barycenter is better than the empirical distribution. To obtain guarantees when $\mathbb{KL}(p_n^{\text{emp}}||p^\star) \leq \mathbb{KL}(p^{\text{expert}}||p^\star)$ seems less obvious and requires control over the quantity $\mathbb{KL}(p_n^{\text{emp}}||p^{\text{expert}})$.

Theorem 4.5.5 *Let $\hat{p}_n^L$ defined by (4.5.6) then we have with probability at least $1-\delta$:*

$$\mathbb{KL}(\hat{p}_n^L||p^\star) \leq \min\left\{\mathbb{KL}(p^{expert}||p^\star), \epsilon_n\,(L_n+1)\right\} \tag{4.5.12}$$

where

$$L_n = \frac{\mathbb{KL}(p^{expert}||p^\star) - \mathbb{KL}(p_n^{emp}||p^\star)}{\mathbb{KL}(p_n^{emp}||p^{expert})}.$$

Proof Using the proposition 4.5.4 we have :

$$\begin{aligned}
\mathbb{KL}(\hat{p}_n^L||p^\star) &= \mathbb{KL}\left(\frac{1}{1+\tilde{\lambda}}p^{\text{expert}} + \frac{\tilde{\lambda}}{1+\tilde{\lambda}}p_n^{\text{emp}}||p^\star\right) \\
&\leq \frac{1}{1+\tilde{\lambda}}\mathbb{KL}(p^{\text{expert}}||p^\star) + \frac{\tilde{\lambda}}{1+\tilde{\lambda}}\mathbb{KL}(p_n^{\text{emp}}||p^\star) \\
&= \frac{1}{1+\tilde{\lambda}}\left(\mathbb{KL}(p^{\text{expert}}||p^\star) - \mathbb{KL}(p_n^{\text{emp}}||p^\star)\right) \\
&+ \mathbb{KL}(p_n^{\text{emp}}||p^\star) \\
&\leq \epsilon_n\left(\frac{\mathbb{KL}(p^{\text{expert}}||p^\star) - \mathbb{KL}(p_n^{\text{emp}}||p^\star)}{\mathbb{KL}(p_n^{\text{emp}}||p^{\text{expert}})}\right) \\
&+ \mathbb{KL}(p_n^{\text{emp}}||p^\star)
\end{aligned}$$

where we used the available inequality to $\tilde{\lambda}$ (the proposition 4.5.4) in the last inequality and the desired result is obtained by assuming that $\mathbb{KL}(p_n^{\text{emp}}||p^\star) \leq \epsilon_n$ which happens with probability at least $1-\delta$.

In addition, note that :

$$\begin{aligned}
\epsilon_n\left(\frac{\mathbb{KL}(p^{\text{expert}}||p^\star) - \mathbb{KL}(p_n^{\text{emp}}||p^\star)}{\mathbb{KL}(p_n^{\text{emp}}||p^{\text{expert}})}\right) + \mathbb{KL}(p_n^{\text{emp}}||p^\star) \\
\leq \mathbb{KL}(p^{\text{expert}}||p^\star) \\
\Leftrightarrow \epsilon_n \leq \mathbb{KL}(p_n^{\text{emp}}||p^{\text{expert}})
\end{aligned}$$

However, if $\epsilon_n \geq \mathbb{KL}(p_n^{\text{emp}}||p^{\text{expert}})$ we have by construction that $\hat{p}_n^L = p^{\text{expert}}$ and therefore $\mathbb{KL}(\hat{p}_n^L||p^\star) = \mathbb{KL}(p^{\text{expert}}||p^\star)$.

We can conclude from all this that :

$$\mathbb{KL}(\hat{p}_n^L||p^\star) \leq \mathbb{KL}(p^{\text{expert}}||p^\star).$$

∎

Remark Note that $\mathbb{KL}(\hat{p}_n^L||p^\star)$ is infinite if p^{expert} does not have the same support that $p^\star$. Nevertheless obtaining a result for $\mathbb{KL}(p^\star||\hat{p}_n^L)$ would require to have a concentration on $\mathbb{KL}(p^\star||\hat{p}_n^{\text{emp}})$ which we do not have. Note that $\mathbb{KL}(p^\star||\hat{p}_n^{\text{emp}})$ is infinite until we have sampled at least one time all the elements of the support of $p^\star$.

4.5.3 Some numerical results ♦

For each experiment in this section, we generate a random distribution $p^\star$ that we try to estimate. To do this, we simulate some realizations of a uniform distribution and renormalize in order to sum up to 1.

We also generate four different distributions that will serve as a priori for the inference: $p^{\text{expert},i}, \forall i \in \{1,2,3,4\}$. The first three priors are obtained by a maximum entropy procedure under constraint to respect marginals of $p^\star$ having undergone a modification. We added to the marginals of $p^\star$ a Gaussian noise of zero expectation and variance equal to $\sigma_1^2 = 0.1$, $\sigma_2^2 = 0.2$ and $\sigma_3^2 = 0.4$ respectively. The last priority $p^{\text{expert},4}$ is chosen equal to the distribution $p^\star$ (the experts provided us with the right distribution).

We then sequentially sample data from $p^\star$, i.e we generate patients, and update for each new data and each different a priori, the left centroid $\hat{p}_n^L$ (using an Uzawa algorithm), the barycenter $\hat{p}_n^{1,1}$, the empirical distribution p_n^{emp} as well as the divergences $\mathbb{KL}(\hat{p}_n^L||p^\star)$ and $\mathbb{KL}(p_n^{\text{emp}}||p^\star)$ and the norms $\|\hat{p}_n^{1,1} - p^\star\|_1$ and $\|p_n^{\text{emp}} - p^\star\|_1$.

The experiments of figures 4.5, 4.6, 4.8 were conducted on a case of a disease with $J = 7$ typical symptoms and where there is therefore $K = 2^7 = 128$ possible combinations. The experiments of figures 4.7 and 4.9 were conducted on a case of a disease with 9 typical symptoms and where there is therefore $K = 2^9 = 512$ possible combinations.

The only parameter we can control is the δ used to construct the confidence interval of the concentration of the empirical distribution around the true distribution. Let us recall, that for the case of the Kullback centroid of the equation (4.5.6) we set :

$$\epsilon_n = \frac{1}{n}\left(-\log(\delta) + \log(G_n)\right) \tag{4.5.13}$$

where G_n is defined in equation (4.5.8).

However, our first numerical experiments show that the choice of ϵ_n defined by the equation (4.5.13) is a little too conservative: see figure 4.5. We need to converge ϵ_n faster towards 0 without abandoning our a priori when it is good.

Our experiments suggest taking a ϵ_n consistent with the proposed concentration in a conjecture of Mardia et al. [2018] for Kullback-Leibler divergence :

$$\epsilon_n = \frac{-\log(\delta) + \frac{n}{2}\log\left(1 + \frac{K-1}{n}\right)}{n}. \tag{4.5.14}$$

Note that we added a constant $\frac{1}{2}$ to the conjecture of Mardia et al. [2018]. As for the choice of δ, this appears important mainly when n is small, taking it sufficiently low avoids an overfitting situation when the number of data is still low without being harmful when n is high. We took it equal to 10^{-6} in all our experiments.

The figure 4.9 shows that our approach is not very sensitive to the choice of δ which is an advantage compared to the penalized approach.

The figures 4.8 and 4.7 show that such a choice for ϵ_n makes a good trade-off between expert and empirical data because we are able to take advantage of these two sources of information when the number of data is small (typically when $n < K$), but also to quickly abandon our a priori when it is bad (see the black curves) or to keep it when it is good (the green curves). Eventually the figures 4.8, and 4.7 were performed on problems of 128

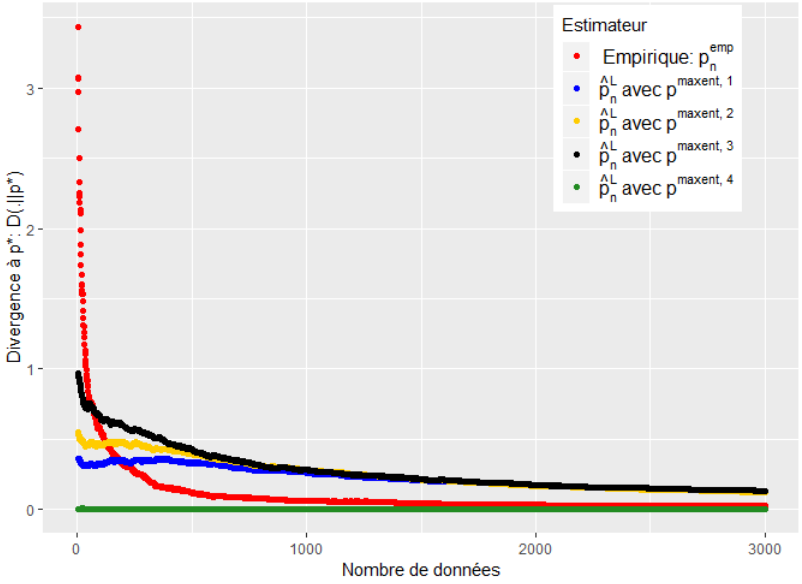


Figure 4.5 – Evolution of the performance of $\hat{p}_n^L$ as a function of the available number of empirical data. ϵ_n defined by equation (4.5.13)

and 512 respectively and this choice of ϵ_n therefore appears relatively robust to changes in size.

Concerning $\hat{p}_n^{1,1}$, we took, still following the conjectures of Mardia et al. [2018]:

$$\epsilon_n = \sqrt{\frac{-\log(\delta) + \frac{n}{2}\log\left(1 + \frac{K-1}{n}\right)}{n}}. \tag{4.5.15}$$

The figure 4.6 shows the error made by our barycenter in norm L^1: $\hat{p}_n^{1,1}$ using such a ϵ_n. We are again able to get rid of a bad a priori relatively quickly to follow the empirical (green curve) while keeping it if it is good (blue curve).

Moreover we show with these experiments that there is an intermediate regime, when we do not have much data, where our estimator is *strictly* better than the two models (experts and data alone). This is particularly visible when we used the ϵ_n of the conjecture of Mardia et al. [2018], see figure 4.7 and 4.8. It is then an empirical evidence that mixing these two heterogeneous sources of information, experts and empirical data, can be useful for statistical inference.

Note that there is some limitations to these experiments. The way we simulate the distributions we are trying to estimate (the $p^\star$) produces quite specific distributions: close to the uniform and dense. If we simulate more sparse distributions, our a priori p^{expert} built from the entropy maximum heuristic will be bad and the experiments will no longer be as interesting.

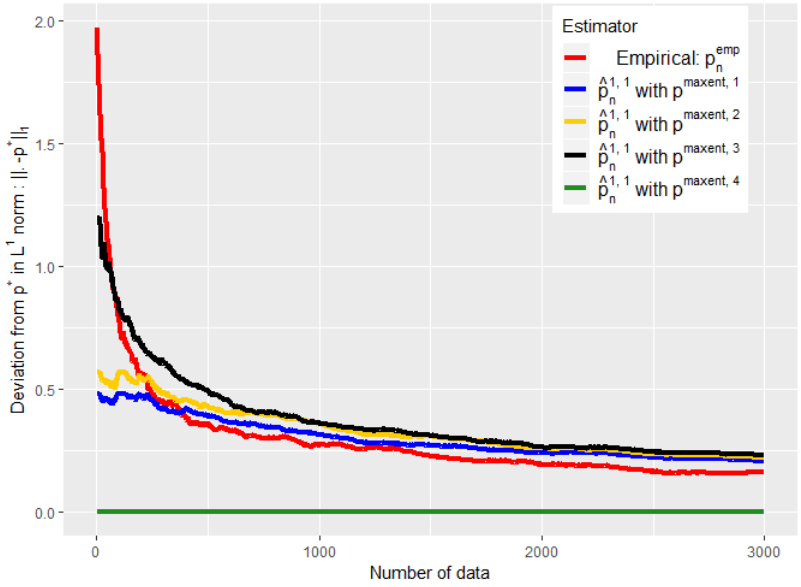


Figure 4.6 – Evolution of the performance of $\hat{p}_n^{1,1}$ as a function of the available number of empirical data. ϵ_n defined by equation (4.5.15).

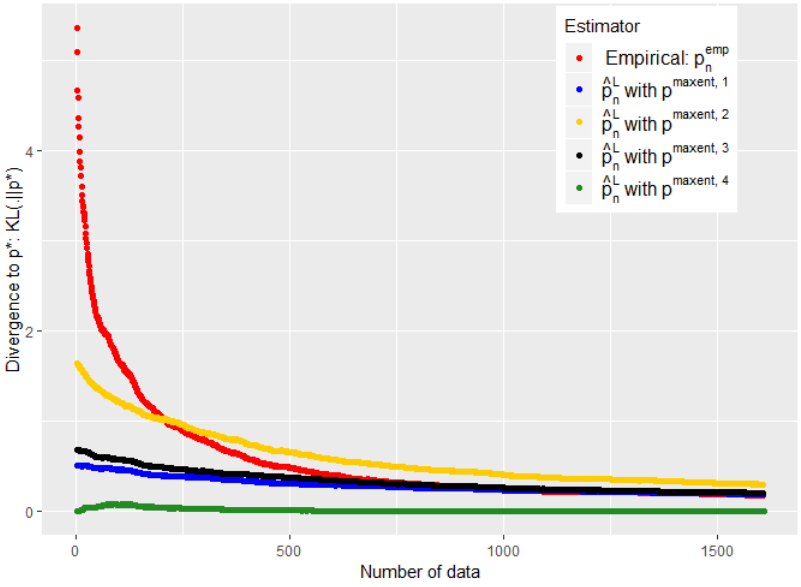


Figure 4.7 – Evolution of the performance of $\hat{p}_n^L$ as a function of the available number of empirical data. ϵ_n defined by (4.5.14). Number of symptom : 9.

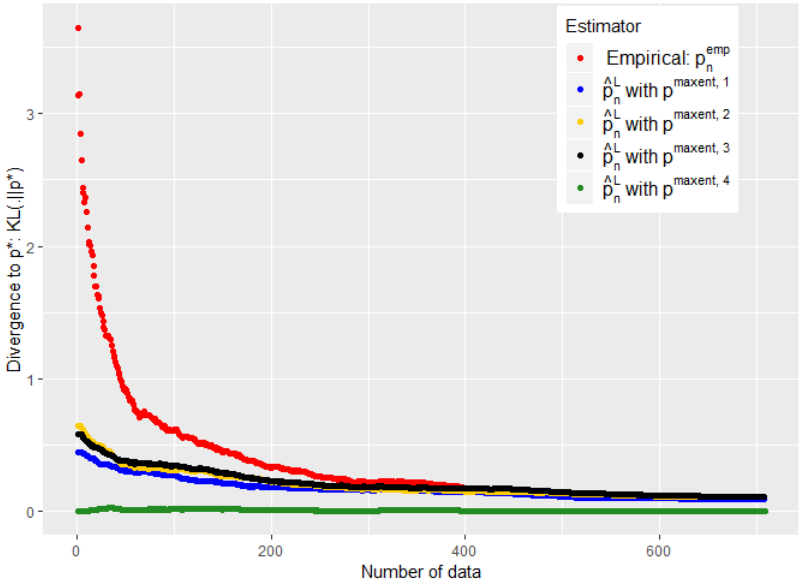


Figure 4.8 – Evolution of the performance of $\hat{p}_n^L$ as a function of the available number of empirical data. ϵ_n defined by (4.5.14). Number of symptom : 7.

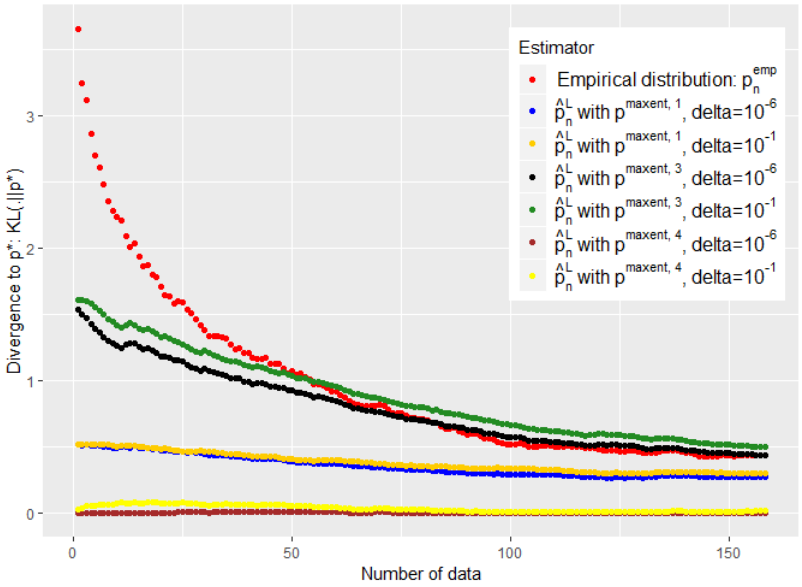


Figure 4.9 – Evolution of the performance of $\hat{p}_n^L$ as a function of the available number of empirical data with different initial a priori and δ. ϵ_n is defined by equation (4.5.13) Number of symptom : 7.

4.6 High-dimensional issues

4.6.1 Explosion of the dimension of symptoms distributions

We are able to estimate the symptom combination distribution ($\mathcal{P}$ of formula (4.2.3)) of each disease D provided that we can store this vector (i.e K is small enough). Note that we actually need a larger vector, as our algorithm processes the information collected by the physician *sequentially*. For example, we will need the $\mathbb{P}[B_1, B_2 \mid D]$ probability, which is not in $\mathcal{P}$ if $K \neq 2$.

There are two possible solution for a disease with K typical symptoms:

- To store the bigger vector of dimension 3^K since we would need to code in ternary to include the information "not seen yet" relative to a given symptom or
- To store the smaller vector of dimension 2^K and compute, on the fly, to recover desired symptoms combination probabilities from available ones.

As we will intensively use our environment model for training our decision support system, we should prefer the first solution, as much as possible.

However, it clearly appears that we will not be able to compute/store the distribution of symptoms combination for all the diseases. Indeed when a diseases has a large number of K typical symptoms, the dimension of the vector $\mathcal{P}^{new}$ we aim to estimate explode: 3^K.

To cope with this issue, we will use the available ontological information about symptoms, i.e the fact that a symptom can be described at different level of precision and make less stringent assumption about the dependence between symptoms (see section 5.3).

4.6.2 Relaxing the model to face potential database default

So far, our model relies heavily on the assumption that expert data gives an exhaustive representation of each diseases. If a symptom has been forgotten for a disease in our expert data list, we would not be able to recover the disease.

That is the reason why we make the assumption that a non-typical symptom (i.e. a symptom that have not been associated to the disease in the expert data) may be observed in a patient with disease D, but with a small (10^{-5}) probability and independently of other symptoms.

4.7 Conclusion

In this chapter we have presented a way to combine expert knowledge, taking the form of marginal probabilities and rules, to real clinical data so as to estimate the joint distribution of the symptoms combinations given the disease.

The particular form of the prior does not allow us to simply adopt a maximum a posteriori (MAP) approach. The absence of a classical prior modeling the model parameters with a particular probability distribution lead us naturally to a maxent approach : if no model seems more plausible to us than another, then we will choose the least informative.

This idea of maximum entropy brings us back to the works of the 80s' and 90 s' where researchers also aimed to build a symptom checker using the marginals. In our work we go further by gradually integrating clinical data as the algorithm is used.

We are interested in the intermediate regime where we do not have enough clinical data to do without experts but have enough to correct them if necessary. Our proposal is to construct our estimator as the distribution closest to the experts' initial a priori, in the sense of a given dissimilarity measure, that is consistent with the empirical data collected. We prove, both theoretically and empirically, that our barycenter estimator mixing the two sources of information is always more efficient than the best of the two models (clinical data or experts alone) within a constant.

We have empirically illustrated the effectiveness of the proposed approach by giving a priori of different quality and incrementally adding clinical data. We have shown that our estimator allows a bad a priori to be abandoned relatively quickly when the inconsistency of the data collected with the initial a priori is observed. At the same time, this same mixture makes it possible to keep the initial a priori if it is good.

Moreover we show with this experiment that, in the intermediate regime, our estimator can be *strictly* better than the best of the two models (experts and data alone). It empirically confirms the idea that mixing these two heterogeneous sources of information can be profitable in statistical inference.

Several refinements are possible such as the addition of a kernel structure for the construction of the empirical distribution. Indeed it is possible that there are omissions of some symptoms in the data collected. Then a kernel approach that would consider states that only differ by some presences as closer would capture such a difficulty and makes a better use of empirical data.

Chapter references

Martin Adamcik. Collective reasoning under uncertainty and inconsistency. *Doctoral thesis, Manchester Institute for Mathematical Sciences*, 2014.

Johan Barthelemy and Philippe L. Toint. Synthetic population generation without a sample. *Transportation Science*, 47:266–279, 2013.

Adam L. Berger, Vincent J. Della Pietra, and Stephen A. Della Pietra. A maximum entropy approach to natural language processing. *Comput. Linguist.*, 22(1):39–71, March 1996. ISSN 0891-2017. URL http://dl.acm.org/citation.cfm?id=234285.234289.

Yvonne M. M. Bishop, Stephen E. Fienberg, Paul W. Holl, Richard J. Light, Frederick Mosteller, and Peter B. Imrey. Discrete multivariate analysis: Theory and practice, 1975.

Eugene Charniak. The bayesian basis of common sense medical diagnosis. In *AAAI*, 1983.

Anthony Costa Constantinou, Norman Fenton, and Martin Neil. Integrating expert knowledge with data in bayesian networks. *Expert Syst. Appl.*, 56(C):197–208, September 2016. ISSN 0957-4174. doi: 10.1016/j.eswa.2016.02.050. URL http://dx.doi.org/10.1016/j.eswa.2016.02.050.

Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*. Wiley-Interscience, New York, NY, USA, 2006. ISBN 0471241954.

Imre Csiszár and Frantisek Matús. Information projections revisited. *IEEE Trans. Information Theory*, 49(6):1474–1490, 2003. doi: 10.1109/TIT.2003.810633. URL https://doi.org/10.1109/TIT.2003.810633.

Pierre-Simon de Laplace. Mémoire sur la probabilité des causes par les évènements. *Mémoires de mathématique et de physique présentés à l'Académie royale des sciences par divers sçavans et lus dans les assemblées*, Tome sixième, 1774.

W. Edwards Deming and Frederick F. Stephan. On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. *Ann. Math. Statist.*, 11 (4):427–444, 12 1940. doi: 10.1214/aoms/1177731829. URL https://doi.org/10.1214/aoms/1177731829.

Andrew Gelman, John B. Carlin, Hal S. Stern, and Donald B. Rubin. *Bayesian Data Analysis*. Chapman and Hall/CRC, 2nd ed. edition, 2004.

Mohammad Ghavamzadeh, Shie Mannor, Joelle Pineau, and Aviv Tamar. Bayesian reinforcement learning: A survey. *CoRR*, abs/1609.04436, 2016. URL http://arxiv.org/abs/1609.04436.

David Heckerman, Dan Geiger, and David M. Chickering. Learning bayesian networks: The combination of knowledge and statistical data. *Machine Learning*, 20(3):197–243, Sep 1995. ISSN 1573-0565. doi: 10.1023/A:1022623210503. URL https://doi.org/10.1023/A:1022623210503.

Michael Herman, Tobias Gindele, Jörg Wagner, Felix Schmitt, and Wolfram Burgard. Inverse reinforcement learning with simultaneous estimation of rewards and dynamics. In Arthur Gretton and Christian C. Robert, editors, *Proceedings of the 19th International Conference on Artificial Intelligence and Statistics*, volume 51 of *Proceedings of Machine Learning Research*, pages 102–110, Cadiz, Spain, 09–11 May 2016. PMLR. URL http://proceedings.mlr.press/v51/herman16.html.

Daniel Hunter. Uncertain reasoning using maximum entropy inference. In *Proceedings of the First Conference on Uncertainty in Artificial Intelligence*, UAI'85, pages 21–27, Arlington, Virginia, United States, 1985. AUAI Press. ISBN 0-444-70058-7. URL http://dl.acm.org/citation.cfm?id=3023810.3023813.

C T Ireland and Solomon Kullback. Contingency tables with given marginals. *Biometrika*, 55 1:179–88, 1968.

E. T. Jaynes. Information theory and statistical mechanics. *Phys. Rev.*, 106(4):620–630, May 1957. doi: 10.1103/PhysRev.106.620. URL http://prola.aps.org/abstract/PR/v106/i4/p620_1.

Radim Jirousek. A survey of methods used in probabilistic expert systems for knowledge integration. *Knowl.-Based Syst.*, 3(1):7–12, 1990. doi: 10.1016/0950-7051(90)90032-D. URL https://doi.org/10.1016/0950-7051(90)90032-D.

Daphne Koller and Nir Friedman. *Probabilistic Graphical Models: Principles and Techniques - Adaptive Computation and Machine Learning*. The MIT Press, 2009. ISBN 0262013193, 9780262013192.

Jay Mardia, Jiantao Jiao, Ervin Tánczos, Robert D. Nowak, and Tsachy Weissman. Concentration inequalities for the empirical distribution, 2018.

John W. Miller and Rodney M. Goodman. A polynomial time algorithm for finding bayesian probabilities from marginal constraints. *CoRR*, abs/1304.1104, 2013. URL http://arxiv.org/abs/1304.1104.

Frank Nielsen. Jeffreys centroids: A closed-form expression for positive histograms and a guaranteed tight approximation for frequency histograms. *IEEE Signal Processing Letters*, 20:657–660, 2013.

Frank Nielsen and Richard Nock. Sided and symmetrized bregman centroids. *IEEE Transactions on Information Theory*, 55:2882–2904, 2009.

Judea Pearl. Probabilistic reasoning in intelligent systems - networks of plausible inference. In *Morgan Kaufmann series in representation and reasoning*, 1989.

John E. Shore. Relative entropy, probabilistic inference and AI. *CoRR*, abs/1304.3423, 2013. URL http://arxiv.org/abs/1304.3423.

David Silver, Aja Huang, Chris J. Maddison, Arthur Guez, Laurent Sifre, George van den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Vedavyas Panneershelvam, Marc Lanctot, Sander Dieleman, Dominik Grewe, John Nham, Nal Kalchbrenner, Ilya Sutskever, Timothy P. Lillicrap, Madeleine Leach, Koray Kavukcuoglu, Thore Graepel, and Demis Hassabis. Mastering the game of go with deep neural networks and tree search. *Nature*, 529:484–489, 2016.

David Silver, Julian Schrittwieser, Karen Simonyan, Ioannis Antonoglou, Aja Huang, Arthur Guez, Thomas Hubert, L Robert Baker, Matthew Lai, Adrian Bolton, Yutian Chen, Timothy P. Lillicrap, Fan Hui, Laurent Sifre, George van den Driessche, Thore Graepel, and Demis Hassabis. Mastering the game of go without human knowledge. *Nature*, 550:354–359, 2017.

David J. Spiegelhalter, A. Philip Dawid, Steffen L. Lauritzen, and Robert G. Cowell. Bayesian analysis in expert systems. *Statist. Sci.*, 8(3):219–247, 08 1993. doi: 10.1214/ss/1177010888. URL https://doi.org/10.1214/ss/1177010888.

A. C. Y. Tossou and C. Dimitrakakis. Probabilistic inverse reinforcement learning in unknown environments. *ArXiv e-prints*, July 2013.

H. Uzawa. Iterative methods for concave programming. *Studies in Linear and Nonlinear Programming*, pages 154–165, 1958.

R Nj Veldhuis. The centroid of the symmetrical kullback-leibler distance. *IEEE Signal Processing Letters*, 9:96–99, 2002.

Yun Zhou, Norman Fenton, and Cheng Zhu. An empirical study of bayesian network parameter learning with monotonic influence constraints. *Decision Support Systems*, 87:69 – 79, 2016. ISSN 0167-9236. doi: https://doi.org/10.1016/j.dss.2016.05.001. URL http://www.sciencedirect.com/science/article/pii/S0167923616300744.

Chapter 5

Probabilistic reasoning on ontologies

Abstract: *In this chapter we give more details concerning the integration of ontological information in our decision support tool for the diagnosis of rare diseases. This is a crucial aspect as we aim to give more freedom to the user allowing him to provide information about the symptoms at different level of granularity. We present a way to integrate this information while remaining in our probabilistic setting and without computational explosion. Finally we make the link between the tree structure of the symptoms and the way to modelize and store the distribution of the symptoms given the disease.*

Contents

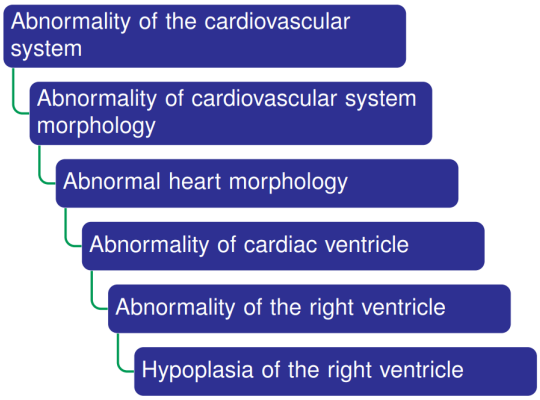


Figure 5.1 – A (very) short extract of the cardiac system ontology.

5.1 Introduction

So far we described the diseases as combinations of symptoms. We designed algorithms inquiring about symptoms so as to find the right disease while minimizing the average number of questions to ask. We have seen that in order to learn a good strategy, we need to learn a model of the environment, i.e to learn the symptom combination distributions given the disease.

Nevertheless a decision support system built in this way will suffer from several issues. Indeed, the symptoms in medicine can be described at several level of granularity. A concrete example is given in Figure 5.1 for the abnormality "hypoplasia of the right ventricle" (the terms range from the least precise to the most precise).

In medicine, these kind of trees, called ontology, are commonly used to represent the knowledge on symptoms hierarchy. These ontologies are built in order to capture the structure and relations between symptoms. Medical ontologies can bear an (almost) infinite level of granularity.

A naive decision support system, i.e which would not include the available ontological information, could ask irrelevant question to a physician. For example, it is perfectly possible that our decision support would advice looking for an hypoplasia of the right ventricle when the physician already mentioned that there is no morphological abnormality of the heart. It is this kind of non-sense that we aim to solve in this chapter.

Furthermore our decision support tool appears, at the moment, too rigid. Indeed, we could ask for an "hypoplasia of the right ventricle" when the physician could not give us such a precise information but rather a more imprecise one like "there is an abnormality of the cardiac ventricle". We should be able to deal with such imprecise answers and then give to the physician more freedom when interacting with our decision support tool while avoiding an explosion in computing time. Once again the use of the ontology allows us

such an improvement.

By doing so we will be able to adapt to less experienced users which are not always able to precise exactly the nature of the abnormality but who can say that something is not going well on a particular region of the patient's body. This is a good way to handle the uncertainty we can have in the users answer. We therefore remain in the setting where the answer are binary ("yes" or "no") but allow the user to provide answers at different level of precision.

5.2 A less rigid decision support tool without computation explosion

Each symptom of our initial database has been mapped to the HPO database. We, then, have been able to extract the underlying tree structure linking the different HPO codes. To be more precise, we know for each HPO code (a given description of a symptom), all its descendants (more precise description of such symptom) as such as all its ascendants (less precise description of such symptom).

5.2.1 The idea ♦

As previously explained, we aim at giving more freedom to the users when describing the symptoms they observed, giving them the possibility to describe the symptoms at different level of granularity.

Then, instead of giving answers at a given level of granularity (our initial list of symptom), one could allow the physician to choose any of the HPO code. It involves an explosion in the number of possible symptoms: our former list of symptoms references some 200 signs when the HPO ontology has around 1300. Both our way to modelize the symptoms combination distributions and our learning algorithms will not be able to cope with such an explosion.

Theoretically each patient could be unique if its symptoms are described to a sufficient level of accuracy. Nevertheless, when we list the typical symptoms of a disease, we try to generalize and find patterns in patient profiles. The idea is too still modelize the symptoms combination distributions with our initial database (the one with 200 symptoms) preserving the ability of generalization our algorithms. We still propose symptoms to check at the level of granularity of the initial database but allow the user to give answers at a different level of granularity (any HPO code can be chosen). By proceeding in this way we obtain a less rigid decision support without computation explosion since all computation are done at the initial level of granularity.

For such an objective, we need a function translating the received imprecise information (the HPO code) into usable information (presence/absence of symptoms at our granularity level). Such a function involves deterministic and stochastic rules.

5.2.2 Deterministic rules

Our function associating to each HPO code the usable information implies some automatic (deterministic) rules. Namely:

- If we received a positive answer for a given HPO code, all its ascendants should be given a positive answer too.
- If we received a negative answer for a given HPO code, all its descendants should be given a negative answer too.

In practice, we store during the medical examination all the information given about the HPO codes selected by the user. In order to compute the probability of each disease, we need to check, for each HPO code and each disease, if this HPO code is in the list of the symptoms related to this disease. If not we have to check whereas ascendants or descendants of this HPO code are in this given list of symptom. Following our two deterministic rules, if the HPO code was declared to be present we have to check if ascendants are in the list, if it was declared to be absent we have to check the descendants.

If the HPO code verifies all the following assertions it can be considered as non typical and treated in consequence (namely its presence is unlikely as in section 4.6.2):

- The HPO code is not in the list of symptoms related to the disease.
- It is present (respectively absent) and its ascendants/descendants (respectively descendants) are not in the list.

Note that the second point involves a relation that we have not studied until then. Indeed, what happens if we observed the presence of an abnormality which HPO code is not in the list of symptoms of the disease but has descendants which are in the list? This issue is studied in the next section.

In practice all these operations can be pre-computed in order to minimize the online computation time required. It is here a matter of adapting the algorithm of section 7 which takes as input a list of presence/absence of symptoms and output the probabilities of all the diseases. An easy way to do it is to store in a list indexed on symptoms, the position the symptom has (or one of its ascendant/descendant has) in the list of typical symptoms of each disease. We should also encode if it was the symptom term, one ascendant or a descendant which was in the list of typical symptoms as it will not involve the same deterministic rule.

5.2.3 Stochastic rules ♦

Let us assume that we have observed the presence of an "abnormal heart morphology" but that the disease we are interested in only has in its list of typical symptom the "Hypoplasia

of the right ventricle". How to take into account such an imprecise information? We need decision stochastic rules for this issue.

When receiving the information of the presence of a HPO code, we have to determine which of its descendants are in the list of symptoms of our first database (the one which we use to build our environment model). All these symptoms have a known probability of apparition (given what we already observed) and we are able to compute them.

Indeed let us denote L a list of symptom for which there is no descendant in our initial database or which are absent. Then let assume that we observed the presence of a symptom which potential descendants are $B_1^{(1)}$, $B_2^{(1)}$, $B_3^{(1)}$ and $B_4^{(1)}$ and the presence of a second symptom which potential descendants are $B_1^{(2)}$ and $B_2^{(2)}$.

There are then $4 \times 2 = 8$ possible combinations. Indeed, without any additional assumption the number of possible combinations could be large. This is why we assume that for each imprecise answer there is only one descendant which is present at a time. Our function first computes $\forall i, j, D$:

$$P[B_i^{(1)}, B_j^{(2)}, L \mid D].$$

It is just the matter of searching for each D which are the typical symptoms in the list $B_i^{(1)}, B_j^{(2)}, L$ and use the deterministic rules if necessary.

We can then compute

$$P[B_i^{(1)}, B_j^{(2)} \mid L] \propto P[B_i^{(1)}, B_j^{(2)}, L] = \sum_D P[B_i^{(1)}, B_j^{(2)}, L \mid D]P[D].$$

We display the probability of each disease by (we denote $\tilde{B}$ for the fuzzy state associated to the 8 possible states $B_i^{(1)}$ and $B_j^{(2)}$):

$$\begin{aligned} P[D \mid \tilde{B}, L] &= P[D \mid L, \cup_{i,j}(B_i^{(1)} \cap B_j^{(2)})] \\ &\propto P[L, \cup_{i,j}(B_i^{(1)} \cap B_j^{(2)}) \mid D]P[D] \\ &= \sum_{i,j} P[L, B_i^{(1)}, B_j^{(2)} \mid D]P[D]. \end{aligned}$$

5.2.4 Optimize the strategy on the leaves of the ontological tree and then go back up ♦

Our stochastic rule can be expensive in terms of computing resources while it is of crucial importance for us to be able to interact quickly with our environment when training our agent. Therefore, the idea is to optimize the subtasks which start from symptoms which does not have any descendants in our database (the leaves). By this way we do not have to use our stochastic rule while training the neural networks.

It is moreover easy to derive the strategy we have to follow when we receive an imprecise

answer during an exam. We denote $\tilde{s}$ the fuzzy state and $(s^{(i)})_{i=1,...,\tilde{d}}$ the associated possible states. We can compute the Q-values in this fuzzy state by averaging on the Q-values on the possible states:

$$Q_\pi(\tilde{s}, a) = \sum_{i=1}^{\tilde{d}} p(s^{(i)} \mid \tilde{s}) \times Q_\pi(s^{(i)}, a). \tag{5.2.1}$$

In practice, when receiving an imprecise answer, our algorithm should ask all the time to the physician if he could provide a more precise answer. If not, a computation as (5.2.1) has to be performed in real time during the examination. This computation should not last more than a second, otherwise we can consider that the provided information was not precise enough and can be overlooked.

To avoid using the stochastic rules while training our agent we need also to remove all the action which has descendants and replace them by their leaves. In a future work it would be interesting to allow different levels of granularity for the action that would suggest the neural network.

5.3 Relations between the ontology and the symptom combination representation ♦

We insisted in section 4.6.1 that there are some cases where we are not able to compute $\mathcal{P}^{new}$ but we still want to be able to compute quickly the probabilities $P[S_1, ..., S_j \mid D]$ without making the assumption of conditional independence. The only solution is to relax our model of dependence between symptoms. We assumed so far that there was dependence between all the symptoms of a disease, we should now consider dependence with a less stringent approach. For the clarity of our presentation, we consider here a two-stage deep ontology with a deeper stage for specific symptoms description and a more vague level for organs.

Let us assume we are interested in a disease with K_1 cardiac typical symptoms (C_1,...,C_{K_1}) and K_2 renal typical symptoms (R_1,...,R_{K_2}). We denote:

$$R = \begin{cases} 1 & \text{if there is at least one renal abnormalities} \\ 0 & \text{otherwise}\,. \end{cases}$$

Then we assume (precise) symptoms from distinct organs are conditionally independent given which organs have abnormalities, so we have the following decomposition:

$$\begin{aligned} P[C_1, ..., C_{K_1}, R_1, ..., R_{K_2} \mid D] = {} & P[C_1, ..., C_{K_1} \mid C, D] \\ & \times P[R_1, ..., R_{K_2} \mid R, D] \\ & \times P[C, R \mid D]. \end{aligned}$$

Note that even if we have lost the possibility to store dependence between precise symptoms from different organs (C_i and R_j), we keep a model of dependence at the higher level in ontology: dependence between organs abnormalities (C and R).

Instead of computing and storing all symptoms combinations, we just store the symptoms combinations inside organs and organs combinations.

The probability of symptom combinations (i.e $P[C_1, ..., C_{K_1} \mid C, D]$ in our example) are computed using the ideas of chapter 4 with the assumption to present at least one symptom (which was yet an assumption before). The organs abnormality combinations $P[R, C \mid D]$ are computed too using the ideas of chapter 4. When marginals $P[R \mid D]$ or $P[C \mid D]$ are not known we can treat them as missing values or try to approximate them using marginals of the lower level, temporarily making some kind of conditional independence assumption.

Each symptom combination can be easily computed using the law of total probability, for example we have the following decomposition:

$$\begin{aligned} P[\bar{R}_1, C_1 \mid D] =& P[\bar{R}_1 \mid \bar{R}, D] \times P[C_1 \mid C, D] \times P[\bar{R}, C \mid D] \\ &+ P[\bar{R}_1 \mid R, D] \times P[C_1 \mid C, D] \times P[R, C \mid D] \end{aligned}$$

where $P[\bar{R}_1 \mid \bar{R}, D] = 1$ and all the other probabilities have been stored making these kind of computations very cheap.

This approach is in fact perfectly adapted for several diseases which manifested themselves in combinations of symptoms coming from specific organs. For example VACTERL syndrome is a rare genetic diseases defined by a combination of at least three abnormalities from three distinct organs Solomon [2011] among vertebral anomalies, anorectal malformation, cardiovascular anomalies, tracheoesophageal fistula, esophageal atresia, renal and/or radial anomalies and limb defects (thus defining the acronym of the disease by their first letter).

Then we are able to cope with any symptom combination distribution even when the number of related symptoms to a disease is high. In such a case we have to find ascendants common to several of these symptoms (which is always possible by definition) that will organize the symptoms in groups (the organs in our example of VACTERL). We then make the conditional independence assumption between symptoms given the ascendants.

Figure 5.2 displays the distribution of symptom combinations of VACTERL syndrom obtained with this group modeling. We plotted the $2^{19} \approx 500000$ points with some in transparency for visibility reasons. Symptom combinations modeled as impossible because there are not sufficient groups turned on, are plotted in red. Comparing this plot to the one obtained making the conditional independence assumption (see Figure 5.3), it appears that group modeling adds much information about the distribution of symptom combinations. The visible symmetries of the distribution are only due to the lexicographic order we use for the different symptom combinations.

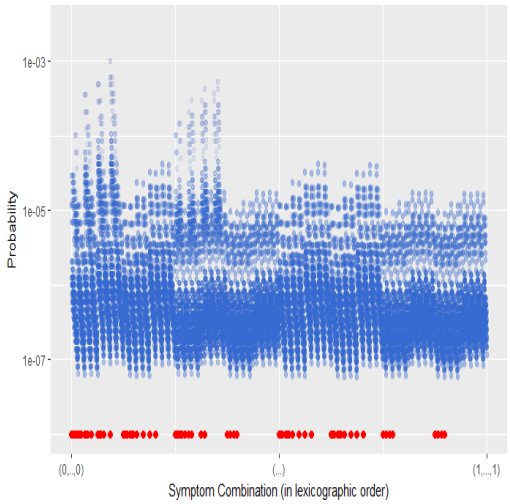


Figure 5.2 – Symptoms combinations distribution for the VACTERL syndrome obtained by maxent with group modeling.

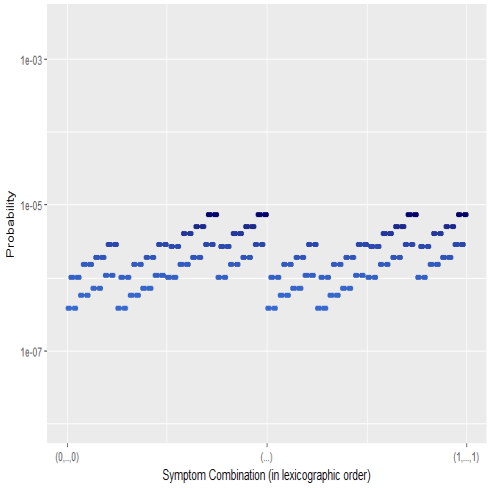


Figure 5.3 – Symptoms combinations distribution for the VACTERL syndrome obtained by making the conditional independence assumption.

5.4 Related works

The objective of integrating the symptom ontology when building a decision support tool for medical diagnosis purpose is an area of research in itself.

The recent thesis of Donfack Guefack [2013] provide a good overview of this research field which is mainly frequented by biostatisticians and computer scientists. Ontologies has been first used for rules-based decision support tool using the Web Ontology Language (OWL), see Valérie Bertaud-Gounot [2012]. OWL is a widely used ontological knowledge model but it does not handle probabilistic reasoning. Thus an effort of the community has been to developed an extension that would handle probabilistic ontologies : Pr-OWL see Costa and Laskey [2006].

A classic way to handle ontology is to use the *information content* (IC) Köhler et al. [2009], Resnik [1995]. The IC of a symptom is defined as the negative natural logarithm of the frequency of this symptom term (and its descendants term) in the database. For example if B_i is a given symptom, following the principle of information theory Cover and Thomas [2006] its information content is given as follow :

$$IC(B_i) = -\log\left(\frac{f_i}{|\mathcal{D}|}\right)$$

where f_i is the number of time the symptom B_i (or one of its descendant) appears in the list of typical symptom of the diseases.

Then a symptom term is all the more informative because it is rare. The symptom term at the root of the ontology (such as "*phenotypical abnormality*" for example) are indeed much less informative than "*hypoplasia of the right ventricle*".

In Köhler et al. [2009] the similarity between a set of signs $Q = (B_1, ..., B_{|Q|})$ and a disease D whose associated symptoms belong to the set $\mathcal{B}(D)$, is defined as the sum of the IC of their most informative common ancestor (MICA):

$$sim(Q \to D) = \frac{1}{|Q|} \left(\sum_{B_i \in Q} \max_{B_j \in \mathcal{B}(D)} IC(\text{MICA}(B_i, B_j)) \right)$$

This similarity score is symmetrized by setting:

$$sim(P, D) = \frac{1}{2} sim(Q \to D) + \frac{1}{2} sim(D \to Q).$$

As we do not know a priori which values of this score of similarity between a set of symptom P and a disease D represents a good match, Köhler et al. [2009] propose to generate random sets of symptom P and to compute the average score of similarity (Monte-Carlo estimation with 100000 simulations). If a score obtained for a given set of symptom P and a given disease D is statistically significantly higher than the score computed for random set it means that it is plausible that the set of symptom P is symptomatic of the presence of disease D.

Note that the score $sim(P, Q)$ depends on the number $|P|$ of symptoms in the set P, then the Monte-Carlo estimation has to be done for every possible value of $|P|$.

Although such an approach is interesting, it does not fit our requirement to remain within a probabilistic framework.

5.5 Conclusion

In this chapter we have presented a way to efficiently deal with the fact that the information about the symptoms can be provided at different level of granularity. The integration of such a feature to our decision support tool is essential as it allows us to be less rigid in our interactions with the users. We can then take into account the user uncertainty while remaining in the binary framework of the abnormalities. We proved that it is possible to integrate this ontological reasoning while remaining in the probabilistic framework presented in the previous chapters.

Finally we connected these reflections to the problem of representing the distribution of the symptoms given the disease. The ontology provide us a natural way to group anomalies together or to simplify their interdependent relationship when there are too many symptoms typical of a disease to assume that they are all connected.

Chapter references

Paulo C. G. Costa and Kathryn B. Laskey. Pr-owl: A framework for probabilistic ontologies. In *Proceedings of the 2006 Conference on Formal Ontology in Information Systems: Proceedings of the Fourth International Conference (FOIS 2006)*, pages 237–249, Amsterdam, The Netherlands, The Netherlands, 2006. IOS Press. ISBN 1-58603-685-8. URL http://dl.acm.org/citation.cfm?id=1566079.1566107.

Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*. Wiley-Interscience, New York, NY, USA, 2006. ISBN 0471241954.

Pierre Sidoine V. Donfack Guefack. *Representation of the signs in the biomedical ontologies for the help to the diagnosis.* Theses, Université Rennes 1, December 2013. URL https://tel.archives-ouvertes.fr/tel-01057310.

Sebastian Köhler, Marcel H. Schulz, Peter Krawitz, Sebastian Bauer, Sandra Dölken, Claus E. Ott, Christine Mundlos, Denise Horn, Stefan Mundlos, and Peter N. Robinson. Clinical diagnostics in human genetics with semantic similarity searches in ontologies. *The American Journal of Human Genetics*, 85(4):457 – 464, 2009. ISSN 0002-9297. doi: https://doi.org/10.1016/j.ajhg.2009.09.003. URL http://www.sciencedirect.com/science/article/pii/S0002929709003991.

Philip Resnik. Using information content to evaluate semantic similarity in a taxonomy. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence - Volume 1*, IJCAI'95, pages 448–453, San Francisco, CA, USA, 1995. Morgan Kaufmann Publishers Inc. ISBN 1-55860-363-8, 978-1-558-60363-9. URL http://dl.acm.org/citation.cfm?id=1625855.1625914.

B. D. Solomon. Vacterl/vater association. *Orphanet J Rare Dis.*, August 2011. doi: doi:10.1186/1750-1172-6-56.

Anita Burgun Valérie Bertaud-Gounot, Régis Duvauferrier. Ontology and medical diagnosis. *Inform Health Soc Care*, 4, 2012. URL https://www.ncbi.nlm.nih.gov/pubmed/22462194.

Chapter 6

CONCLUSION

6.1 Summary

The aim of this thesis was to construct a decision support system for the diagnosis of rare disease in obstetrics.

In a first step, we have tried to define what a good decision support system for the diagnosis of rare diseases should be when the cost of the medical tests is negligible against the potential cost of a misdiagnosis. This led us to an original formulation of the optimization problem related to the task of training a symptom checker. We tried to minimize the average number of medical tests to be performed before reaching a pre-determined high level of certainty about the disease of the patient. We proposed to measure this level of certainty using entropy because its properties ensure us that on average the uncertainty about the patient's disease would not have increased if we had inquired more symptoms.

We investigated several reinforcement learning algorithms and made them operational in our very high dimensional environment. To do this, we have divided the initial task into several subtasks and learned a strategy for each subtask. We have empirically proven that an appropriate use of intersections between subtasks can significantly accelerate the learning process. We have also proven that a traditional greedy strategy (CART) can be largely outperformed by strategies learned with reinforcement learning algorithms.

In addition, we have investigated the task of building a model of the environment that we can use as a simulator to train good diagnostic strategies.

To do so we have reconnected with the first works on expert systems using probabilistic reasoning by proposing a maximum entropy approach. This is because of our intended application, we are interested in rare diseases and we cannot work without the knowledge of experts (or the literature) that is generally expressed as conditional probabilities.

In this thesis, we go further by investigating the possibility to progressively improve the initial a priori given by the experts by integrating the clinical data collected as the algorithm is used. We have investigated several approaches and finally proposed to define the estimator mixing expert and empirical data as the distribution the closest to the initial

expert a priori, in the sense of a certain dissimilarity measure, which is consistent with the collected data. We have proven, both theoretically and empirically, that an estimator defined in this way is always more efficient than the best of the two models (expert or data alone) within a constant. For the intermediate regime that particularly interest us, i.e when we do not have enough data to do without experts but yet enough to correct them if necessary, we empirically observed that our estimator is strictly better than the best of the two models expert or data alone.

In the last chapter we have presented some solutions to integrate the available information on the symptom tree structure to our decision support tool. Most of the work on that question has been done by the biostatistician community. Here we have proven that it is possible to integrate ontological considerations while remaining in our probabilistic setting and without computational explosion.

Finally, this decision support system has been tested internally at Necker Hospital on a fetopathology dataset, which is independent of the dataset that we used to train our algorithms. We get very good results; indeed more than 90% of the scenarios led to a good diagnosis. Most of the errors are due to defects in our dataset, as synonym issues or omissions, that are currently under correction. A larger-scale study is in progress.

6.2 Future research prospects

See annex 2 (confidential).

Chapter 7

Annex 1

Chapter 8

Annex 2

université PARIS-SACLAY | École doctorale de mathématiques Hadamard (EDMH)

Titre : Méthodologie d'aide à la décision pour le dépistage anténatal échographique d'anomalies fœtales par apprentissage statistique

Mots clés : optimisation de prise de décision séquentielle, optimisation d'arbre de décision, aide au diagnostic médical, planification dans des espaces de très grande dimension, mélange experts/données, raisonnement probabiliste dans des ontologies

Résumé : Dans cette thèse, nous proposons une méthode pour construire un outil d'aide à la décision pour le diagnostic de maladie rare. Nous cherchons à minimiser le nombre de tests médicaux nécessaires pour atteindre un état où l'incertitude concernant la maladie du patient est inférieure à un seuil prédéterminé. Ce faisant, nous tenons compte de la nécessité dans de nombreuses applications médicales, d'éviter autant que possible, tout diagnostic erroné. Pour résoudre cette tâche d'optimisation, nous étudions plusieurs algorithmes d'apprentissage par renforcement et les rendons opérationnels pour notre problème de très grande dimension. Pour cela nous décomposons le problème initial sous la forme de plusieurs sous-problèmes et montrons qu'il est possible de tirer partie des intersections entre ces sous-tâches pour accélérer l'apprentissage. Les stratégies apprises se révèlent bien plus performantes que des stratégies gloutonnes classiques. Nous présentons également une façon de combiner les connaissances d'experts, exprimées sous forme de probabilités conditionnelles, avec des données cliniques. Il s'agit d'un aspect crucial car la rareté des données pour les maladies rares empêche toute approche basée uniquement sur des données cliniques. Nous montrons, tant théoriquement qu'empiriquement, que l'estimateur que nous proposons est toujours plus performant que le meilleur des deux modèles (expert ou données) à une constante près. Enfin nous montrons qu'il est possible d'intégrer efficacement des raisonnements tenant compte du niveau de granularité des symptômes renseignés tout en restant dans le cadre probabiliste développé tout au long de ce travail.

Title : Decision support methodology for antenatal ultrasound screening for fetal anomalies by statistical learning

Keywords : sequential decision making, decision tree optimization, medical diagnostic decision support, planning in high-dimensional spaces, mixture experts/data, probabilistic reasoning in ontologies

Abstract: In this thesis, we propose a method to build a decision support tool for the diagnosis of rare diseases. We aim to minimize the number of medical tests necessary to achieve a state where the uncertainty regarding the patient's disease is less than a predetermined threshold. In doing so, we take into account the need in many medical applications, to avoid as much as possible, any misdiagnosis. To solve this optimization task, we investigate several reinforcement learning algorithm and make them operable in our high-dimensional. To do this, we break down the initial problem into several subproblems and show that it is possible to take advantage of the intersections between these sub-tasks to accelerate the learning phase. The strategies learned are much more effective than classic greedy strategies. We also present a way to combine expert knowledge, expressed as conditional probabilities, with clinical data. This is crucial because the scarcity of data in the field of rare diseases prevents any approach based solely on clinical data. We show, both empirically and theoretically, that our proposed estimator is always more efficient than the best of the two models (expert or data) within a constant. Finally, we show that it is possible to effectively integrate reasoning taking into account the level of granularity of the symptoms reported while remaining within the probabilistic framework developed throughout this work.

Université Paris-Saclay
Espace Technologique / Immeuble Discovery
Route de l'Orme aux Merisiers RD 128 / 91190 Saint-Aubin, France