



HAL
open science

Elaboration of personalized projects for users in social and medico-social sector : automatic planning and data objectification

Yinuo Li

► **To cite this version:**

Yinuo Li. Elaboration of personalized projects for users in social and medico-social sector : automatic planning and data objectification. Computation and Language [cs.CL]. Université d'Angers, 2021. English. NNT : 2021ANGE0028 . tel-03921943

HAL Id: tel-03921943

<https://theses.hal.science/tel-03921943>

Submitted on 4 Jan 2023

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

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

THÈSE DE DOCTORAT DE

L'UNIVERSITÉ D'ANGERS
COMUE UNIVERSITÉ BRETAGNE LOIRE

ÉCOLE DOCTORALE N° 601
*Mathématiques et Sciences et Technologies
de l'Information et de la Communication*
Spécialité : *Informatique*

Par

Yinuo LI

**Elaboration of personalized projects for users in social and
medico-social sector : automatic planning and data objectification**

Thèse présentée et soutenue à Angers, le 25 Novembre 2021

Unité de recherche : Laboratoire d'Étude et de Recherche en Informatique d'Angers (LERIA)

Thèse N° :

Rapporteurs avant soutenance :

M. Stéphane BONNEVAY Maître de conférence HDR, Université Claude Bernard Lyon 1
M. Gilles VENTURINI Professeur, Université de Tours

Composition du Jury :

Président :	Mme. Christelle JUSSIEN GUERET	Professeure, Université d'Angers
Examineurs :	M. Stéphane BONNEVAY	Maître de conférence HDR, Université Claude Bernard Lyon 1
	Mme. Christelle JUSSIEN GUERET	Professeure, Université d'Angers
	M. Gilles VENTURINI	Professeur, Université de Tours
Directeur de thèse :	M. Jin-Kao HAO	Professeur, Université d'Angers

Invité(s) :

Mme. Touria AIT EL MEKKI Maître de conférence, Université d'Angers
M. Rachid NAITALI Directeur général, Gépi Conseil

ACKNOWLEDGEMENT

I would first like to express my gratitude and appreciation to my supervisor, Prof. Jin-Kao Hao, whose continued guidance was indispensable in formulating my research directions and methodology.

I would also like to thank Mr. Rachid Naitali for his valuable suggestions and patience.

I thank the jury members: Dr. Stéphane Bonnevey, Prof. Christelle Guéret and Prof. Gilles Venturini, I express my sincere gratitude to all of them for their time and efforts spent on evaluating my work.

I wish to express my gratitude to Dr. Touria Ait El Mekki, for sharing her knowledge and inspiration.

I would like to express thanks to all my colleagues at LERIA, University of Angers and my colleague Dr. Brahim Chabane of Gépi Conseil, for their wonderful collaborations and advices.

Last but not least, I would like to thank my parents: Jinrong Zhang and Weidong Li for their unconditional love and patience during these years, I couldn't have gotten through many difficulties without them.

TABLE OF CONTENTS

General Introduction	9
1 The practical context	15
1.1 Introduction	15
1.2 The social and medico-social sector (SMSS)	16
1.2.1 General description	16
1.2.2 Evolution of the SMSS	16
1.2.3 Missions of the SMSS	18
1.2.4 Social and medico-social structures and services	19
1.2.5 Social and medico-social professions	21
1.3 Decision support in SMSS	23
1.3.1 Literature review	23
1.3.2 MSUsager for the SMSS	24
1.4 Conclusion	25
2 The user project planning problem in SMSS	27
2.1 Introduction	27
2.2 Personalized user project in SMSS	28
2.2.1 Definition	28
2.2.2 Evolution	29
2.2.3 Implementation	30
2.3 Statement and formulation of the user project planning problem	32
2.3.1 Solution representation	32
2.3.2 Constraints	33
2.3.3 Objectives	36
2.4 Related studies	37
2.5 Conclusion	40

3	Mathematical programming and greedy heuristic for the user project planning problem	43
3.1	Introduction	43
3.2	Mathematical programming models	44
3.2.1	Integer programming	44
3.2.2	0/1 programming	46
3.3	Constructive greedy heuristic	49
3.4	Experimentations and results	51
3.4.1	Test instances	51
3.4.2	Evaluation function	54
3.4.3	Experimental setting	55
3.4.4	Computational results	55
3.5	Conclusion	59
4	Multi-neighborhood simulated annealing for the user project planning problem	61
4.1	Introduction	61
4.2	Multi-neighborhood simulated annealing	62
4.2.1	General procedure	62
4.2.2	Neighborhoods based on activity change, swap and exchange	64
4.2.3	Probabilistic exploration of the neighborhoods	68
4.3	Experimentations and results	69
4.3.1	Experimental setting	69
4.3.2	Computational results	70
4.4	Analysis and discussions	72
4.4.1	Analysis of neighborhoods	72
4.4.2	Analysis of parameters	75
4.5	Conclusion	75
5	Adverse event report classification in SMSS	77
5.1	Introduction	77
5.2	Practical context	78
5.2.1	The adverse event in SMSS	78
5.2.2	Problematic and motivation	80
5.3	Literature Review	81

5.4	Adverse event data	83
5.4.1	Data acquisition	84
5.4.2	Annotation	86
5.5	Methods	86
5.5.1	Text preprocessing	88
5.5.2	Keyword lexicon of fact	88
5.5.3	Event-fact similarity	89
5.6	Results	90
5.6.1	Lexicon construction results	90
5.6.2	Classification results	91
5.6.3	Knowledge visualization solution	94
5.7	Analysis and discussion	97
5.7.1	Error analysis	97
5.7.2	Practical contributions	97
5.7.3	Limitations	98
5.8	Conclusion	98
	General Conclusion	101
	List of Figures	105
	List of Tables	108
	List of Algorithms	109
	List of Publications	111
	References	113

GENERAL INTRODUCTION

Context

The work presented in this thesis was carried out under a CIFRE (Industrial Convention of Formation by Research) contract with the laboratory LERIA of the University of Angers and the company GéPI-conseil.

Social and medico-social centers are the main structures in France where different categories of vulnerable populations are hosted. In addition to the daily care, the main mission of the social and medico-social structures is to provide their residents (called “users” hereafter) with a personalized support in terms of a life project that takes into account professional, social, administrative and health needs of the residents. This is in accordance with legal provisions that reinforce the concept of “personalized user project” which aims to improve the person’s life quality with different services.

However, the elaboration and implementation of the personalized user projects are complex. Social and medico-social structures receive heterogeneous populations with different profiles and needs, which forces the structures to adapt their activities to each user profile and, consequently, to create numerous and varied activities requiring different types of materials and human resources. As a result, the decision-makers of these structures have to consider various factors and constraints such as varied user needs, resource availability (staff, budget and materials), activity scheduling, assignment, etc. Moreover, the data and observations collected from users are not always rigorous due to the subjectivity and poor writing quality of professionals, which brings more difficulties to the analysis of user’s needs and desires. Especially, these structures are typically faced with a lack of decision support tools to help them to make the right decisions.

The solutions proposed in this thesis aim to improve the efficiency and quality of the implementation of personalized user project in the social and medico-social sector. On the one hand, we present a general formulation of the user project planning problem and investigate approaches based on mathematical programming and heuristic algorithms to ease the management of large amount of information related to users, activities and resources. On the other hand, we propose a text classification solution to objectify the

information of adverse event reports to improve the user data analysis.

Objectives

This thesis aims to develop decision support solutions for solving the difficulties in the implementation of personalized user project in the social and medico-social sector. This thesis tackles two issues: how to organize information efficiently to construct a high quality project plan and how to collect information reliably during the project implementation for further project improvements. The main objectives of this thesis are summarized as follows.

- To ease the decision-making during project conception phase for structures and then propose the best feasible activity plan for each user, formulate the personalized user project planning problem by identifying the constraints to respect and objectives to maximize with a formal formulation.
- Propose solution approaches based on different techniques (mathematical programming, greedy search, local search) to solve the personalized user project planning problem.
- Generate test instances which could fit as much as possible real-life situations, evaluate the proposed approaches on these instances and perform a comprehensive comparison.
- To ease the information collection task during project execution phase for structures, propose solution to objectify the information of adverse event reports collected by professionals to better follow the project process of users.

Contributions

The main contributions of this thesis are summarized below.

- **Modeling and solution of the personalized user project planning problem.** We formulate this personalized user project planning problem by identifying its constraints and objectives. The personalized user project planning problem is composed of sets of users, activities, resources and timeslots, the aim is to find activity plans for each user and resource during a given period with the satisfaction of a set of practical constraints and the optimization of several objectives. We provide

a formal modeling of this problem with the solution representation and the formulation for each constraint and objective. To solve this problem, we propose two solution approaches: 1) Two mathematical programming models including an integer programming model and a 0/1 programming model. 2) A constructive greedy algorithm which provides feasible solution in a very short time. These approaches are assessed experimentally on 20 realistic test instances, these instances are generated based on real scenarios identified according to the responses received from a large-scale questionnaire survey to more than 3600 structures in the social and medico-social sector in France. This work has been published in *Expert Systems with Applications* [LHC21].

- **Multi-neighborhood simulated annealing algorithm (MNSA).** We propose an effective local search algorithm to improve the results of the personalized user project planning problem. Multi-neighborhood simulated annealing algorithm (MNSA) starts with an initial solution generated by the constructive greedy procedure presented in the first study, then explores three different neighborhood structures in a probabilistic way, based on the simulated annealing framework. We design these neighborhoods according to real planning operations and demonstrate their effectiveness on 20 realistic instances. Within a short time, MNSA can provide feasible solutions of very good quality, even for the largest instances. Comparing with reference approaches, MNSA finds a better solution for 7 instances against 4 cases in favor of the 0/1 programming model by CPLEX and obviously improves on the results of the constructive greedy procedure for the whole set of 20 instances. Finally, we investigate the advantage of the neighborhoods combination of the proposed MNSA algorithm. This work has been submitted to an international journal for publication.
- **Adverse event report classification.** We propose a text classification method which aims to objectify the information of an adverse event (AE) report by proposing the most possible facts according to the analysis of the event description. This work presents the first classification approach to classify French AE reports in the social and medico-social sector by adapting the lexicon-based classification method with lexicons specifically constructed for the sector. The lexicons constructed include professional terms and informal expressions, serve as a guide to improve the writing quality of professionals. We test this method on 388 real AE reports collected from social and medico-social structures and achieve an accuracy up to 88%.

We also develop two visualization charts for adverse event data as an additional decision-support tool to help the structures improve their analysis and complete their knowledge. This work has been published in *Proceedings of the 6th IEEE Conference on ICT Solutions for eHealth (ICTS4eHealth)*, Athens, Greece, September 5-8, 2021.

Organization

The manuscript is organized in the following way:

- In the first chapter, we present the practical context of this thesis. We begin with the main elements of the social and medico-social sector in France including the general definition, evolution and missions defined by the law of January 2, 2002. Then, we introduce representative structures and professions in this sector to show its complexity and importance for society. Next, we review the current informatization level in this sector by presenting related research for decision support and the management software MSUsager.
- In the second chapter, we introduce the personalized user project in the social and medico-social sector with its definition, evolution and implementation process. Based on this context, we present the personalized user project planning problem and provide a formal formulation, including the description and formulation of its constraints and objectives. Finally, we review related studies on project planning problems and representative solution approaches.
- In the third chapter, we develop solution methods for the personalized user project planning problem. We first create two mathematical programming models and run a general tool (CPLEX MIP Solver) to solve them. We create an integer programming model where the decision variables take their values in an integer interval representing the activity and a 0/1 programming model where the decision variables take binary values. We also propose a greedy heuristic approach to compare with mathematical programming models. Then, we introduce 20 simulated test instances of different sizes and tunable characteristics generated according to responses received from a survey. Finally, we provide computational results of these proposed solution approaches tested on 20 instances.
- In the fourth chapter, we propose a multi-neighborhood simulated annealing algorithm (MNSA) for the personalized user project planning problem. We apply the

probabilistic neighborhood search with the simulated annealing framework and describe three neighborhood operators inspired by real planning operations. Afterward, we proceed to the computational studies and comparisons between the proposed MNSA algorithm and solution methods proposed in the third chapter. Finally, we provide analysis of the influence of parameters and the effectiveness of neighborhood combination of MNSA.

- In the fifth chapter, we present a lexicon-based text classification solution for the adverse event report classification. We start by introducing the practical context of adverse event in the social and medico-social sector and the data we used. Then we present the classification method, including text preprocessing, lexicon construction and similarity calculation. Finally, we provide the classification results and the knowledge visualization solution.
- In the last chapter, we give the general conclusion of this thesis as well as some perspectives for future research.

THE PRACTICAL CONTEXT

1.1 Introduction

As in all developed countries, the social and medico-social sector in France is experiencing a fast evolution due to the continuing growth of its aging and disabled population. Since 2002, several laws and decrees have been established to specify the roles and missions of different types of structures. One of the important missions is to ensure the right for the users of the social and medico-social structures to participate in the elaboration and implementation of their “accompanying projects”. As the result, providing users with a personalized support in terms of a life project has become the main mission of all social and medico-social structures. This policy aims to improve the life quality of the aging and disabled population and lighten the financial burden of their family. However, this evolution also brings new challenges to decision-makers of these structures. First, since several years, there has been a significant decrease in staff resources in the social and medico-social sector, on the one hand, and an increase in the societal demands, on the other hand. Second, in order to improve users’ independence and social reintegration capacity as much as possible, structures have to provide multiple and varied services and activities requiring multidisciplinary professionals (doctors, psychologists, educators, etc.) and resources (rooms, vehicles, medical facilities, etc.), making decision-making more complex and difficult. Third, unlike the medical care sector which has well-developed decision support system [Ber+10] and terminology [AF05], structures in the the social and medico-social sector suffer from a lack of decision support and optimization tools to improve their management efficiency.

This thesis aims to help social and medico-social structures meet this challenge by providing them with practical and effective decision support solutions. We first start by introducing important elements of the social and medico-social sector (Section 1.2) and the current research status (Section 1.3) in this sector.

1.2 The social and medico-social sector (SMSS)

In this section, we introduce the social and medico-social sector (SMSS) in France, with its historical evolution, missions and representative structures and professions.

1.2.1 General description

In France, the SMSS assembles all kinds of structures or services whose mission is to host or support people with disabilities, dependents or in a social exclusion situation, for a short or long term.

The social domain: The social field regroups all actions aiming to improve the living conditions of a society. Social activities are often related to exclusion, insertion or reintegration, childhood and family.

The medico-social domain: The medico-social field relates to social medicine, which takes into account the social context and the health conditions of a population.

Target population: The main target population of the SMSS are children, adults in difficulty, the elderly, prisoners and people excluded from society for cultural, financial or medical reasons, etc.

Actors: Social and medico-social actions are implemented by a multitude of actors. The public field includes the state, regions, local authorities, social security, socio-educational penitentiary services and public institutions, etc. The private profit or non-profit field includes medical related centers, private home care services, retire houses, etc.

1.2.2 Evolution of the SMSS

The health and social sector has been undergoing important transformations for several decades. These changes including economic, political and normative changes are the result of an adaptation of the sector to the major evolutions of society. The increase of the aging population and the development of insecurity are leading to the reshaping of the sector so that it will be able to respond to these new demands. However, in the context of the economic crisis and public deficit, resources and funding remain highly constrained especially when the societal demands are increasing. Therefore, the reforms of the sector aim at a better territorial organization, a better economic efficiency and an improvement on the service quality. We list below significant evolutions of the SMSS.

Demand generation. The hospital law of December 31, 1970 [70] refocuses the hospital on its primary function and make a distinction between health domain and social domain. It gave the government three years to draw up a decree to organize the social and medico-social services. As the result, The social and medico-social filed is evolving and being restructured to adapt to intensified and multiplied social needs, for demographic reasons (increase and aging of the population), sociological reasons (weakening of family ties), economic reasons (unemployment, precariousness) and political reasons (social rights). This social demand has also become more complex due to the accumulated difficulties, but also because the expectations of the public have increased.

Official birth. The SMSS has been dedicated since two laws of June 30, 1975:

- Law No.75-534 [75a] about the orientation of disabled people.
- Law No.75-535 [75b] which concerns social and medico-social institutions and affects multiple populations (children and families in difficulty, disabled people, elderly people, people in precarious situations or exclusion). It separates health from social services and organizations. For the first time, the social and medico-social services become a homogeneous and autonomous subject to common rules.

Renovation and standardization. The law of January 2, 2002 [02] standardizes and renovates social actions and specifies the function of regional social and medico-social organizations with the following principles:

- Respect of users' rights.
- Adaptation of structures according to users' needs.
- Establishment of coordination between the different actors.
- Better use of resources.
- Continuous service improvement.
- Implementation of 7 tools (user welcome booklet, charter of user rights and freedoms, residence contract, conciliator or mediator, operating regulations, project of structure or service, council of social life).
- Internal and external evaluation of activities and the quality of services.
- Obligation of training and qualification for professionals in the structure.

Funding agency. The law of June 30, 2004 [04a] about the solidarity and autonomy of the elderly and disabled people creates a new actor for the financing of dependency: the National Solidarity Fund for Autonomy (CNSA), responsible for distributing funds

intended for the operation of medico-social structures and services hosting disabled and elderly people, financed by health insurance and falling under the jurisdiction of the state. In addition, the law of August 13, 2004 [04b] confirms the role of the department as the leader of social and medico-social action.

Quality reform. The most significant recent evolution concerning the SMSS is the law of July 21, 2009 [05] on the reform of the hospitals, patients, health and territories.

- Creation of the National Agency for Performance Support (ANAP) which aims to define criteria for the performance evaluation and develop recommendations and tools to improve the service quality for structures in the SMSS.
- Creation of the Regional Health Agencies (ARS), these agencies aim to implement health policy for regions and simplify health procedures by operating in the health, medico-social and outpatient sectors.

Digital transformation. Announced in September 2018 by the President of the Republic, the “My health 2022”¹ strategy offers an overview and global responses to the challenges facing the French health system. In his speech of September 18, 2018 addressed to the health sector, the President of the Republic gave clear and strong orientations to respond to the difficulties of patients, also to develop the health system of Next 20 years. My health 2022 makes above all the quality and relevance of care the common thread of the transformation. With the establishment of care pathways, patients will not have to assume alone the coordination of the various health professionals, all roles in the sector will coordinate more effectively, thanks to digital technology. This strategy brings new challenge to the SMSS, structures in SMSS are now demanded to adopt informatics tools to manage user data and coordinate with other structures, which promotes the related industrial research and innovation.

1.2.3 Missions of the SMSS

The social utility missions of the SMSS have been defined and clarified by the law of January 2, 2002 renovating social and medico-social actions:

1. Evaluation and prevention of social and medico-social risks, information, investigation, advice, orientation, training, mediation and reparation;
2. Administrative and judicial protection of children and families, teenagers, disabled people, the elderly and people in difficulty;

1. <https://solidarites-sante.gouv.fr/systeme-de-sante-et-medico-social/masante2022/>

3. Educational, medico-educational, medical, therapeutic, pedagogical and training actions adapted to the needs, level of development, potential, evolution of condition and age of the person;
4. Actions of school integration, adaptation, readjustment, social and professional reintegration, support for active life, information and advice on technical aids as well as employment assistance;
5. Actions of assistance in the various acts of life, support, care and company;
6. Actions contributing to social and cultural development, and integration through economic activity.

Before the law of January 2, 2002, structures in the SMSS were not obliged to perform internal or external evaluations. With this law, these structures are obliged to carry out an internal evaluation every 5 years and an external evaluation every 7 years followed by a project and a multi-year action plan. In 2007, the National Agency for the Evaluation and quality of Social and Medico-social structures and services (ANESM) was created to follow the internal and external evaluation of structures in the SMSS. The agency's two main missions are to provide the conditions for structures to evaluate their activities and the quality of services, and to ensure that the authorities receive the results. In 2018, the agency was integrated within the High Authority for Health (HAS).

1.2.4 Social and medico-social structures and services

In France, social and medico-social structures and services are legally defined in article L.312-11 of the Code of Social Action and Families. The field of social and medico-social actors is extremely vast and includes a number of structures that provide supervision and support for a variety of population: children or adults with disabilities, children or adults experiencing social difficulties, independent or dependent elderly people, etc. The public sector is very representative in the housing of the elderly: 44% of structures and 51% of the EHPAD (Accommodation establishments for dependent elderly people) offers are managed by public structures, but private profit organizations are also numerous in this sector where 24% of EHPAD are managed by commercial companies [APE16]. Private non-profit structures account for the majority in the management of structures and services for adults or children with disabilities: they manage 94% of ESAT (Establishments and services for work-related assistance), 78% of MAS (Specialized reception centers), 90% of IME (Medico-educational institutes), they are also predominant in the protection of

children since 95% of MECS (Social children houses) fall within this sector [APE16]. Table 1.1 lists main structures and services of the SMSS with their target population and mission.

Table 1.1 – The main structures and services of the SMSS.

Population	Structures or services name	Mission
Elderly	Accommodation establishments for elderly	Provide non-medical collective accommodation for independent elderly people who need a safe environment to live.
	Accommodation establishments for dependent elderly	Take charge of the deterioration of the health state and the loss of autonomy of dependent elderly people.
	Home care services	Provide household services and personal assistance services at people’s home for regular and essential activities of daily life.
	Home nursing services	Provide nursing care in the form of technical or basic care.
Children and teenagers with disabilities	Medico-psycho-educational centers	Ensure the diagnosis and care of children or teenagers with psychological development disorders and help the patients by keeping them in their family, school or professional and social environment.
	Medico-educational institutes	Provide appropriate education and medico-social support for disabled children.
	Motor education institutes	Provide care for children and teenagers subject to significant motor impairment in order to support them in their family, social and professional integration.
	Special education and home care services	Provide families with advice and support, promote school integration and the acquisition of autonomy by appropriate medical, paramedical, psychosocial, educational and pedagogical measures.
	Early medico-social action centers	Provide outpatient treatment and rehabilitation for children aged from 0 to 6 with sensory, motor or mental deficits

Children and teenagers with social difficulties	Social children houses	Host minors in difficulty such as family violence, psychological or psychiatric difficulties of the parents, serious family conflicts, etc.
	Specialized prevention center	Carry out educational actions aimed at facilitating better social integration for adolescents.
Adults with disabilities	Establishments and services for work-related assistance	Provide support of the social and professional integration of disabled adults. Also medico-social and educational support to promote the personal and social development of the people received.
	Medical-care houses	Host adults slightly affected by disability whose lack of autonomy prevents any professional activity.
	Specialized reception centers	Provide permanent accommodation for adults with multiple disabilities with a strong dependence for daily life.
Adults with social difficulties	Accommodation and social reintegration centers	Ensure the reception, accommodation, support and social integration of people or families experiencing serious difficulties in order to help them gain access or regain the personal and social autonomy.
	Reception centers for asylum seekers	Offer asylum seekers a place of reception for the duration of the study of their application for refugee status.
	Addiction care, support and prevention centers	Ensure prevention and care actions for people with drug or alcohol addiction.

1.2.5 Social and medico-social professions

With around 1.7 million employees, the SMSS become a major sector of the economy, nowadays it brings together around 6.7% of employed workers in France [Dir12]. In addition, it has undergone a very significant development over the last decades, on the strength of an increase of more than a million jobs since the beginning of the 1980s, it has contributed a lot to the growth of tertiary professions in France.

Professionals in the social and medico-social field do not all work within their specific economic activity sectors. Many are employed by public structures and establishments or by individuals, some may be in business (social workers) or in various sectors of activity. But above all, the professional dimension of this field appears to be blurred overall. On the one hand, some of the professions practiced in this sector are not at all specific to it, this is particularly true of the health professions, doctors are mainly involved with the disabled and the elderly, above all the paramedics (nurses, care assistants, occupational therapists, speech therapists, etc.) and psychologists; on the other hand, new “social” activities appear over the last three decades in the context of the development of urban policy such as project managers, local development agents and urban social development workers [Dir12]. Table 1.2 shows some representative professions of the SMSS.

Table 1.2 – Representative professions work for the SMSS.

Field	Profession
Administration and supervision	Manager Social engineer Secretary
Education	Animator Family assistant Special educator Instructor
Help and advice	Social service assistant Auxiliary of social life Economy advisor Family intervention technician
Paramedical	Caregiver Childcare assistant Health executive Nurse manager Occupational therapist Nurse Physiotherapist Psychomotrician Nursery nurse Dietitian Psychologist

This diversity of professions in the SMSS is due to the diversity of the characteristics and situations of users hosted. Different professionals act together in order to succeed

in the elaboration and realization of the user's projects. However, exchanges between multi-professional interveners are often difficult, in this situation, the implementation of the personalized project with the user becomes very complex. Moreover, the available tools currently for the collection and exploitation of observations, notes, multi-disciplinary exchanges, are often either paper file or text editing software, which makes the information exchanges between professionals inefficient and unstable.

1.3 Decision support in SMSS

In this section, we present the progress of the decision support in SMSS, including related research studies and the software MSUsager.

1.3.1 Literature review

Even though the increased societal demands and the popularization of electronic health records [WCY19; ZYS12] enable new research opportunities, little research has been conducted for the SMSS, and the current practice of informatics tools mainly concerns dashboard management software which is only useful for some simple decisions. However, this sector is facing a supply deficit of resources and offers and needs decision support tools urgently, we list below current studies for the decision support in SMSS.

Structure management and assessment:

Chabane *et al.* [CBH17] applied multiobjective optimization to elaborate efficient action plans for social and medico-social structures. They formulated the problem as a multiobjective knapsack model and applied iterated local search to solve it.

Osorio [Mon15] proposed a conception of a system of measurements and monitoring of activities to improve the current uncoordinated situation in SMSS.

Kromm & Ducq [KD12] proposed method enables to analyze existing systems and to implement target models adapted to the different activities in the SMSS which could also integrate all its environmental evolutions.

User reception and assignment:

Varela [And19] proposed an approach to model the user orientation process and presented a mathematical sliding horizon assignment model based on linear programming to make orientation proposals according to the needs-services policy.

Osorio *et al.* [Oso+17] proposed a multi-criteria approach to reduce the waiting time

on the assignment of users to structures. They applied mixed integer linear programming and simulated annealing according to different assignment policies.

As explained above, current studies focus on the structure management and user admission, however, the main mission of all structures of the SMSS is the care and support for their users during their stay period after reception, where many key decisions will be made. To our knowledge, there is no study on the decision support in terms of user's care and life project. In this thesis, we fill the gap by presenting solutions to help structures improve their work efficiency.

1.3.2 MSUsager for the SMSS

To implement the best personalized project for each user, on the one hand, the decision-maker in structures must collect basic data such as the activities that could be carried out in users' projects and the benefit that each activity will bring to the user, also the data during project implementation to follow project progress. On the other hand, they have to choose and schedule from these activities which could achieve at best objectives, with respect of practical constraints, and make necessary adjustment in time.

In this thesis, we propose a solution that aims to integrate into the MSUsager software to help the decision-maker to solve these problems. MSUsager is a software for individual follow-up of users in SMSS from admission to leaving of structure.

1. Data management

The user data collected in SMSS includes all aspects of user from basic civil status and family situation to education state and sensitive medical and health data. The popularization of electronic records has reduced the use of paper documents but there are still many structures manage their users' data with crude recording softwares, which are not secure and difficult for data analysis. MSUsager centralizes the administrative and family information of users in SMSS, as well as the data of their personalized project. The software provides easy and secure access to up-to-date information facilitates exchanges between professionals. MSUsager guarantees compliance with regulatory constraints and it is designed to comply with the General Data Protection Regulation (GDPR) from its conception.

2. Decision making

As explained above, the variety of services and activities brings many decisions to make within structures in SMSS, especially when the available resources are limited.

MSUsager provides several modules for the decision making support during user's stay according to improve the data collection and coordination efficiency. For the personalized user project conception phase, MSUsager provides service proposals to users according to the predefined nomenclature and the optimization of users' projects planning described in Chapter 2, 3 and 4; and for the adverse events during project implementation phase, MSUsager provides the severity qualification, cause tree visualization and auto-classification of users' adverse events described in Chapter 5.

1.4 Conclusion

This chapter presented the social specificities and evolution of the SMSS in France. With the growth of social demand, the SMSS plays a major role to provide all necessary assistance to vulnerable people by offering them a wide variety of adapted services. At the same time, the SMSS is facing more and more challenges, such as the lack of resources and tools, strict quality evaluation, constantly updated policies, etc. These challenges and difficulties have led structures in SMSS to continuously improve the quality of services and management, which brings new opportunities in research and software development.

One of the challenges of the SMSS is the personalized user project implementation for all kinds of structures. We focus on this problem in this thesis, in the next chapters, we present the solutions we proposed to help structures implement personalized project for their users.

THE USER PROJECT PLANNING PROBLEM IN SMSS

2.1 Introduction

As it was described in the previous chapter, the mission of the social and medico-social structures includes the full support for their users adapted to the needs and desires of each. All supports and actions are presented in the form of personalized user project within structures. In this chapter, we present the definition and the modeling of the personalized user project in SMSS.

- We present the definition and elaboration process of the personalized user project in the SMSS. We demonstrate the particularity and complexity of the implementation of this kind of project by comparing with previous studies on project planning related problems.
- We formulate the programming process of the personalized user project as a user project planning problem by defining the components, constraints and objectives of the problem and providing the formal formulation of the problem.

The chapter is organized as follows. Section 2.2 describes the definition, evolution and implementation process of the personalized user project in the SMSS. Section 2.3 presents the problem constraints and the optimization objectives, followed by their mathematical formulation. Section 2.4 presents representative studies and solution methods of project planning related problems and indicates the particularity of the personalized user project in the SMSS. We draw concluding remarks in the last section.

2.2 Personalized user project in SMSS

In this section, we present the personalized user project in the SMSS, including the definition, evolution and implementation process of the personalized user project.

2.2.1 Definition

The personalized user project in the SMSS is defined by a specific document that describes the professional, social and medico-social supports for the user according to his/her needs, expectations and resources. This document also specifies the objectives to achieve, the personalized activities to carry on and the assignment of different roles among various interveners (professionals, family members, etc). The personalized user project is an ongoing process and is revised and modified at regular intervals. It is a comprehensive planning that encompasses all interventions to focus on the development of the user concerned. For each user, a different project will be designed and developed and may implemented by a different team. Specifically, a personalized user project must include following fundamental components:

- Recognition of the rights and freedoms of user deprived of them for reasons related to their age, physical or mental state and social situation.
- The respect of the valued social roles.
- The participation of the user.
- A multidisciplinary team responsible for the design, implementation, supervision and evaluation of the project.
- A comprehensive and relevant evaluation of the user’s abilities in order to identify user’s desires and needs.
- Realistic (long-term) goals based on a series of achievable (short and medium-term) operational objectives.
- A clear assignment of responsibilities of each interveners of the project.
- Documentation of progress towards achieving goals and objectives using assessment tools.
- A mechanism for measuring and revising the progress of the project.

Related official organizations provide useful guides to help the structures to elaborate their user projects, such as the “Guide to Developing a Personalized Support Project”

[CED12] of CEDIS (European committee for the development of social integration) and the “Good professional practice recommendations: User’s expectations and personalized project” [ANE08] of ANESM (National agency for the evaluation and quality of social and medico-social structures and services). In conclusion, the personalized user project aims to:

- Co-construct with the user his/her life project in accordance with his/her resources and the assessment of his needs and expectations.
- Ensure the individualization of each support. This results in the implementation of individual and collective services and activities in line with the resources of the structure and the user’s wishes.
- Take into account several aspects of the user’s life: social, health, professional and/or educational, etc.

2.2.2 Evolution

Depending on the mission of the structure and service, the project related to the user in the SMSS has different appellations: educational project, integration project, personalized support project, individualized care and support project, etc. The concept was first proposed in the decree of April 26, 1999 [99] for aging people which stipulates that a document should be drawn up in which the general characteristics of the “life project” and the care plan are defined; the document is defined by the healthcare team and the coordinating physician, must specify the methods of organization of care within the structure according to the health status of the residents, as well as the arrangements for the coordination of various interveners. Nearly twenty years, a set of laws have obliged and standardized the implementation of users’ projects in social and medico-social structures:

- Law No.2002-2 [02] ensures the right for the users of the social and medico-social structures to participate in elaboration and implementation of their “accompanying projects”.
- Law No.2005-102 [09] provides the users with the right of compensation for the disability consequences, in the form of a “personalized compensation plan” which is implemented according to the “life project” of the person, in the sense of the needs and desires that they express.
- Decree No.2007-975 [07] defines that one of the objectives specific to the evaluation of social and medico-social structures is to assess the activities and the quality of

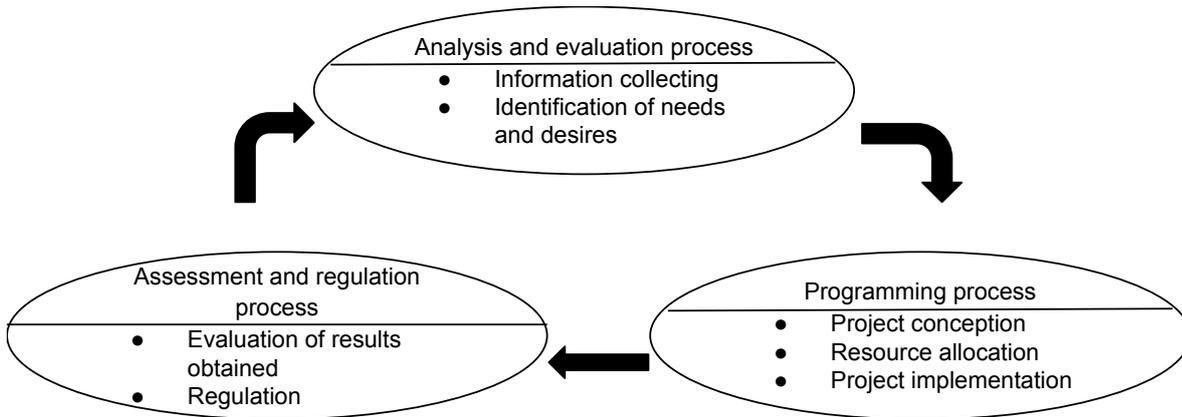


Figure 2.1 – The three main phases of the personalized user project in SMSS.

services with regard to user rights, and the conditions for carrying out the “personalized project”.

The term “personalized project” has been used in the recommendation of CEDIS because it is widely used in practice and suitable for the entire SMSS, we also choose this term in this thesis. Nowadays, the implementation of personalized user project has become an obligation and an important evaluation indicator for all social and medico-social structures.

2.2.3 Implementation

According to the practice guide issued by CEDIS [CED12], there are three main phases in a personalized user project (Figure 2.1).

- *Analysis and evaluation process* This phase aims to identify user’s global desires, abilities and needs. The tasks of this phase are usually implemented by qualified professionals (doctors, psychologists, educators, etc.) to perform special measures such as personal interview, physical and mental test, capacity evaluation, etc. This phase provides necessary information to guide the selection of services and activities for each user.
- *Programming process* This phase is the main phase of project. According to the information collected in the first phase, decision-making team has to draft user’s project by specifying the objectives to achieve and the activities to carry on during a period, while ensuring that the project is feasible by considering resource and

availability constraints of the structure and the user. Once the project is elaborated and validated, the user can start executing his/her project. During the project execution, each user is followed by a professional who notes all observations and events about the user and makes adjustment if necessary.

- *Assessment and regulation process* This phase is carried out at the end of the project. It aims to assess the progress of every activity involved in the project according to the activity completion and observations noted, then make the necessary adjustment and provide information for the next project.

We focus on the second phase, i.e., the programming process. From the first phase, we can obtain a preliminary estimation of activity preference for the user according to his/her needs and personal desires. The activity preference can be understood by the benefits for a user to participate in an activity, which is certainly user specific. This information is the most important indicator for decision-makers because the primary goal of the structure is to help users to improve their daily life. In addition, the differences between activities are reflected in many aspects such as cost, individual activity or group activity, professionals (doctors, educators, drivers, etc.), necessary facilities (bus, room, etc), etc. In general, during the project elaboration, decision-makers are confronted with three questions: i) which activities or services should be implemented for the user, according to his/her needs; ii) how to schedule these activities during the project period; iii) which resources should be allocated to satisfy implementation conditions of each activity. These questions are tightly coupled, and the decision for one of them may influence the others.

This complexity does bring difficulties to the daily work in these structures. In the structure organization, the design of an operational planning is typically realized by a qualified professional or a team who must take account into many factors and unexpected changes. Additionally, the communication cost of this work is very important especially when no digital tools can be used.

Thus, the goal is to propose a model and a tool able to help the decision-maker to elaborate efficient user projects while respecting a set of constraints.

2.3 Statement and formulation of the user project planning problem

In this section, we give a formal description and formulation of the personalized user project planning problem, we present the solution representation, constraints and objectives of the problem.

2.3.1 Solution representation

To introduce the constraints and the objectives of the user project planning problem, we state:

- A set \mathcal{U} of users ($U = |\mathcal{U}|$), each being characterized by his/her budget, availability and preference for each activity. According to our survey mentioned in Section 3.4.1, there are several tens of users in a typical structure.
- A set \mathcal{A} of activities ($A = |\mathcal{A}|$), each being characterized by its price, duration (number of timeslots needed), capacity, availability and features required. Typical activities include medical tests and treatments, entertainment activities (sport, trip, museum visiting), professional training, etc.
- A set \mathcal{R} of resources ($R = |\mathcal{R}|$), including human and facility resources, each being characterized by its availability and features. Typical resources include rooms, vehicles, medical professionals, educators, etc.
- A set \mathcal{T} of timeslots ($T = |\mathcal{T}|$), during which users, activities and resources are to be scheduled. For example, if we take 8 timeslots per day (typically, timeslot equals one hour) and 5 working days per week, then a project of 4 weeks has 160 timeslots.

To formulate the user project planning problem, we represent a candidate solution (i.e., a user project schedule) S by a $U \times T$ matrix \mathbf{X} and a $R \times T$ matrix \mathbf{Y} where

$$x_{ut} = \begin{cases} a & \text{user } u \text{ participates in activity } a \text{ at time } t, \\ -1 & \text{user } u \text{ is not available at time } t \text{ } (ua_{ut} = -1), \\ 0 & \text{user } u \text{ is available at time } t \end{cases}$$

$$y_{rt} = \begin{cases} a & \text{resource } r \text{ used by activity } a \text{ started at time } t, \\ -1 & \text{resource } r \text{ is not available at time } t \text{ } (ra_{rt} = -1), \\ 0 & \text{resource } r \text{ is available at time } t \end{cases}$$

	t1	t2	t3	t4
u1	a1	a1	-1	0
u2	0	a2	a2	a2
u3	a1	a1	0	0

	t1	t2	t3	t4
r1	0	a2	0	0
r2	a1	0	-1	-1
r3	a1	0	0	0

Figure 2.2 – Solution example of the user project planning problem.

The candidate solution of the user project planning problem can be seen as activity schedules for a set of users and resources during a given period, the resources are arranged to satisfy the features required by activities of users. Figure 2.2 shows an example of solution, where user u_1 and u_3 participate in activity a_1 at timeslot t_1 and t_2 with the resources $\{r_2, r_3\}$ and user u_2 participates in activity a_2 from timeslot t_2 to t_4 with the resource $\{r_1\}$.

Below, we describe the problem constraints and the optimization objectives, followed by their mathematical formulation based on Table 2.1.

2.3.2 Constraints

To establish a feasible user project, a number of constraints must be satisfied.

- C_1 . *User availability*. A user cannot be scheduled when he/she is not available. Moreover, a user cannot participate in two activities at the same time.

It is worth noting that the solution representation above ensures that a user can only participate in at most one activity per timeslot, and will not be assigned any activity when the user is not available (preset to -1). In other words, the first constraint *User availability* (C_1) is satisfied by the solution representation.

- C_2 . *Activity availability*. An activity cannot be scheduled for users when it is not available and cannot cross more than one day. Indeed in the social and medico-social structures, it may happen that some activities can be arranged only at certain

Table 2.1 – Settings of important parameters of the user project planning problem.

Symbol	Description
\mathcal{U}	Set of users, $ \mathcal{U} = U$
\mathcal{A}	Set of activities, $ \mathcal{A} = A$
\mathcal{R}	Set of resources, $ \mathcal{R} = R$
\mathcal{T}	Set of timeslots, $ \mathcal{T} = T$
\mathcal{F}	Set of features, $ \mathcal{F} = F$
TS	Number of timeslots per day
ua_{ut}	Whether user u is available at timeslot t , $ua_{ut} = 0$ if available, -1 otherwise
bud_i	Budget of user i
aa_{at}	Whether activity a is available at timeslot t , $aa_{at} = 0$ if available, -1 otherwise
dur_a	Duration of activity a
pri_a	Price of activity a
cap_a	Capacity of activity a
ra_{rt}	Whether resource r is available at timeslot t , $ra_{rt} = 0$ if available, -1 otherwise
pre_{ua}	Preference of user u to activity a
req_{af}	Whether activity a requires feature f , $req_{af} = \{0, 1\}$
fea_{rf}	Whether resource r contains feature f , $fea_{rf} = \{0, 1\}$
$num_{u_{at_k}}$	Number of users starting activity a at timeslot t_k in a candidate solution S ; $num_{u_{at_k}} = \sum_{u \in \mathcal{U}} sta_{uat_k}$, where $sta_{uat_k} = \begin{cases} 1 & \text{if } x_{ut_k} = a \wedge (x_{ut_{k-1}} \neq a \vee k = 1), \\ 0 & \text{otherwise.} \end{cases}$
act_{ua}	Whether user u participates in activity a in a candidate solution S , $act_{ua} = 1$ if $\exists x_{ut} \in S, a = x_{ut}$, $act_{ua} = 0$ otherwise.
res_{at}	Set of resources used by activity a with start-time t in a candidate solution S $res_{at} = \{r \mid r \in \mathcal{R}, y_{rt} = a\}$

timeslots or already prescheduled, especially when external resources are needed.

$$\forall u \in \mathcal{U}; \forall a \in \mathcal{A}; \forall t \in \mathcal{T}; (aa_{at} = 0) \vee (x_{ut} \neq a)$$

- C_3 . *Resource availability*. A resource cannot be scheduled when it is not available and cannot be used by two activities at the same time.

$$\forall r \in \mathcal{R}; \forall t_i \in \mathcal{T}; \forall a \in \mathcal{A}; (y_{rt_i} \neq a) \vee \left(\bigcup_{d=i+1}^{i+dur_a-1} y_{rt_d} = \{0\} \right)$$

- C_4 . *Budget constraint*. The cost of the selected activities in a user's project cannot exceed user's budget.

$$\forall u \in \mathcal{U}; \sum_{a \in \mathcal{A}} act_{ua} \times pri_a \leq bud_u$$

- C_5 . *Activity capacity*. For each activity starting at a given timeslot, the number of users attending this activity cannot bypass the capacity of this activity. This constraint is defined for group activities to ensure the activity quality.

$$\forall a \in \mathcal{A}; \forall t \in \mathcal{T}; num_u_{at} \leq cap_a$$

- C_6 . *Feature constraint*. An activity cannot be scheduled if one or more required resources are unavailable. In social and medico-social structures, coordination between various disciplines is required for most of activities, and an activity cannot be carried out when some required features are not satisfied. For example, if the painting workshop needs a room of at least 20 places and professional(s) with art skills and special education skills, then the resources allocated should cover all these features.

$$\forall a \in \mathcal{A}; \forall t \in \mathcal{T}; \forall f \in \mathcal{F}; (num_u_{at} = 0 \vee req_{af} = 0) \vee \sum_{r \in res_{at}} fea_{rf} > 0$$

This multi-feature constraint is specific to the social and medico-social structures, and to our knowledge, has not been considered in other project planning problems.

- C_7 . *Preference constraints*. A user must indicate a preference for each intended activity. This constraint is defined especially for large structures which propose various services. Unlike classic project planning problems where activities are fixed

or all selectable, in social and medico-social structures, some activities could be identified as unsuitable for some users after a first analysis and evaluation phase. So we add this non-selectable constraint to prevent selecting unsuitable activities for users.

$$\forall u \in \mathcal{U}; \forall a \in \mathcal{A}; (act_{ua} = 0) \vee (pre_{ua} > 0)$$

To keep our model compact, we assume that no relation exists between activities and a user can only participate in an activity at most once during the given period.

2.3.3 Objectives

The user project planning problem has one main objective and two secondary objectives.

Main objective: suitability

The main objective of a personalized user project is to provide the best possible response for each need or expectation of the user. The quality of each response is assessed by the professional and the user (or his/her family), according to the impacts the response could have on the life of the user. However, due to the constraints above, it would be impossible to satisfy all the preferable activities for every user. Consequently, the first goal is to find a solution that covers as many as possible preferable activities for the users.

Suitability maximization: Maximize user satisfaction in terms of the preferences of activities implemented for them.

$$f_1(S) = \sum_{u \in \mathcal{U}} \sum_{a \in \mathcal{A}} act_{ua} \times pre_{ua}$$

Secondary objectives: costs

It is desirable to reduce the costs during project implementation for evident economic reasons and unexpected variation. For the structure, the cost corresponds to the utilization of its resources, while for each user, the cost corresponds mainly to the economic expenditure for the chosen activities.

Structure resource minimization: Minimize the use of resources to save costs for the structure, which is equivalent to maximize the timeslots of the resources that are not

utilized.

$$f_2(S) = R \times T - \sum_{r \in \mathcal{R}} \sum_{t \in \mathcal{T}} dur_{y_{rt}} + \sum_{r \in \mathcal{R}} \sum_{t \in \mathcal{T}} ra_{rt}$$

User cost minimization: Minimize the cost of the planned activities to save expenses for the users, which is equivalent to maximize the unconsumed budget of each user.

$$f_3(S) = \sum_{u \in \mathcal{U}} (bud_u - \sum_{a \in \mathcal{A}} act_{ua} \times pri_a)$$

Even if different types of structures have distinctive operational modes, the above constraints and objectives are shared by all structures and constitute the basis for building our user project planning decision-making system.

2.4 Related studies

From a very general perspective, the personalized user project planning problem studied in this work is related to general scheduling problems, which have various applications such as hospital patient scheduling [DL98], transport scheduling [LH07], course timetabling [LH10], and nurse rostering [LH12]. For project management, several studies have been reported in different fields including IT-projects scheduling and assigning [HK10], assignment of projects optimizing the utilization of employees' time and expertise in chemical industry [Bas00] and software maintenance projects selection and scheduling [BT96]. In particular, a classical model of project scheduling called the resource constrained project scheduling problem (RCPSP) [Bru+99] has been widely investigated. Extended RCPSP models have also been proposed to cope with special needs [HB10] such as stochastic RCPSP (SCPSP) [Sto01], proactive and reactive RCPSP [DD19], and scheduling and staffing multiple projects with a multi-skilled workforce [HK10]. The models in this area are complex, highly constrained and notoriously difficult to solve, but still very attractive because of their wide range of applications in the industry and public services (see the comprehensive surveys on this topic in [HB10], [HL05] and [HBS18]).

Within the context of project planning, there are several planning levels such as project (task) selection, project scheduling and project staffing [HK10]. [EP04] and [Chi+08] treated the problem of project selection with several critical factors based on the Analytic Hierarchy Process. [AIH01] considered the contractor pre-qualification problem in project management. Moreover, some studies only focused on project scheduling [DD19; SV19],

while other works considered both project (task) selection and project scheduling [Kos15; Kos20; KS20] and even a mix of portfolio selection, scheduling as well as staff assignment [Gut+08]. Generally, problems with multiple planning levels are more complex and computationally challenging than problems with only one planning level. To cope with the problem complexity, decomposition-based approaches were typically adopted to handle the subproblems in an independent manner.

Representative solution methods for project planning related problems can be classified into exact approaches and heuristic (metaheuristic) approaches.

1. Exact algorithms:

- Kosztyán proposed a generic matrix-based project planning method to solve the single project scheduling problem [Kos15] and later extended it to schedule the flexible multilevel project with a matrix-based multilevel multimode project scheduling (M4PSP) algorithm [Kos20].
- Heimerl & Kolisch [HK10] modeled the multiple projects scheduling and assigning as a mixed-integer linear program (MIP) model with sharp LP-bound and solved it with CPLEX solver, their results showed the benefit of using the MIP for central planning compared to decentral planning.
- Bellenguez-Morineau, & Néron [BN07] proposed a branch-and-bound method to solve the multi-skill project scheduling problem with experimentation of different branching strategies. Their solution worked well on small and average size instances but failed on most large size instances.
- Davari & Demeulemeester [DD19] proposed four different Markov decision process models which can solve the proactive and reactive RCPSP with duration uncertainty over four different classes of policies; they analyzed the performances of these four models with different parameters.

2. Heuristic and metaheuristic algorithms

- Servranckx & Vanhoucke [SV19] developed a tabu search algorithm using information from the proposed classification matrix to guide the search process towards high-quality solutions of RCPSP.
- Liu *et al.* [LCW14] proposed a memetic algorithm (MA) with a new fitness function and two effective local search procedures for a multi-mode resource-constrained project scheduling problem (MRCPSp).

- Tritschler *et al.* [TNK17] proposed a hybrid metaheuristic algorithm for the RCPSP with flexible resource profiles, it contains the flexible resource profile parallel schedule generation scheme which employs the concepts of delayed scheduling and non-greedy resource allocation, embedded in a genetic algorithm.
- Zhu *et al.* [Zhu+17] proposed a multi-start iterative search heuristic named MSIS for the project scheduling problem with resource availability cost (RACP). They divided the problem into two sub-problems: the sequencing problem and the resource decision problem, then solved these two sub-problems sequentially.
- Khoshjahan *et al.* [KNA13] tried to minimize the present value of earliness-tardiness penalty costs through defining a deadline for each activity. In their research, the mathematical programming model was first proposed, followed by the application of genetic algorithm and simulated annealing.

Compared to the above problems and models, the user project planning problem studied in this work has several particularities, making the problem original in its definition and difficult in terms of problem solving. First, the SMSS is characterized by its diversity of activities. Since the admitted persons are disabled and to reduce the burden of their family, the structures usually provide a range of services and activities from daily life care to education even medical treatment. This high degree of diversity makes the underlying project planning problem a real challenge for the structures. Second, unlike other project planning problems, the main target of the user project planning problem is not really the projects but the users. Therefore, the objective here is not to arrange as many as possible activities nor to minimize the project duration, but to improve the overall quality of the projects of all users. This requires the decision-maker to consider the benefits that each activity can bring to the users when deciding which activities should be selected. Third, the user project planning problem needs to jointly consider activity selection, project scheduling and resource allocation. Indeed, due to the resource shortages in the SMSS, the decision results of scheduling and resource allocation could influence the selection of activities for each user. Finally, user projects need to be planned usually for a group of users at the same time, instead of independent users, adding another dimension of planning difficulty. All these features make user project planning a complex and challenging problem.

2.5 Conclusion

This chapter presented the process to implement the personalized user project in the SMSS. Among the three phases of the personalized user project, the project conception phase is the most complex part where many decisions have to be made for the organization of data and coordination of different interveners. To deal with this issue, we modeled the personalized user project as a constraint optimization problem and presented a formal formulation. We defined the main elements of the project, the constraints to respect and the objectives to optimize. The solution of the problem is two matrices which represent the plans of users and resources during the project period with the satisfaction of all constraints.

- We started with the description of the process to implement a personalized user project in the SMSS. Among the three phases of the personalized user project, the programming process is the most important and complicated one, which includes the project conception and project execution. The project conception is critical for the project quality and execution but it is very complex to take account into variant factors and interveners.
- To cope with this problem, we modeled the personalized user project conception task as a constraint optimization problem with detailed description and formal formulation. We stated a set \mathcal{U} of users, a set \mathcal{A} of activities, a set \mathcal{R} of resources and a set \mathcal{T} of timeslots. The candidate solution is a set of plans of several activities for each resource and user during a period. There are a number of constraints must be satisfied such as the availabilities of users, activities and resources; the cost of user's project cannot exceed his/her budget; the number of users attending an activity cannot be greater than its capacity and an activity cannot be scheduled when there are not enough resources to use. There are three objectives to be optimized: maximize the satisfaction in terms of the preferences of activities of the users, minimize the use of resources to save resources for the structure and minimize the cost of the plan to save expenses for the users.
- We ended the chapter with a state of the art on the project planning related problems, which have been widely investigated to fit different practical needs. The project planning problems are often highly constrained and difficult to solve, integer programming and heuristic(metaheuristic) approaches have been widely applied to solve this kind of problem. Unlike other planning and scheduling problems such as

nurse rostering [LH12], hospital patient scheduling [DL98], university course planning [LH10] and project scheduling [DD19; HB10; HK10], a personalized project in the SMSS implies that the services and resources allocated to a particular user may be very different from the other users and depends on his/her vulnerability situation, family situation, personal desires, finance situation, etc. This particularity makes the elaboration of the personalized project of a user very complex, which becomes even more difficult when a group of users must be considered simultaneously.

In the next two chapters, we describe in detail the solution methods that we propose for the personalized user project planning problem.

MATHEMATICAL PROGRAMMING AND GREEDY HEURISTIC FOR THE USER PROJECT PLANNING PROBLEM

3.1 Introduction

In this chapter, we aim to propose solution methods to solve the user project planning problem defined in Chapter 2. We propose two mathematical programming models as well as a greedy algorithm to solve 20 simulated test instances corresponding to real scenarios identified according to a large-scale survey. We summarize the main contributions as follows.

- We send a large-scale survey to more than 3600 structures in the social and medico-social sector in France. From the responses received (more than 200), we create an instance generator that is parametrized to generate problem instances covering various real scenarios in different structures.
- We implement two mathematical programming models including an integer programming model and a 0/1 programming model and run a general tool (CPLEX MIP Solver) to solve these models. The 0/1 programming model (solved by the CPLEX software) shows a high competitive performance compared to the integer programming model and obtains close-to-optimality solutions for some small instances.
- Since that the mathematical programming approach is too computationally expensive, we additionally experiment a greedy heuristic and compare it with the mathematical programming approach. The greedy algorithm is more successful and obtains more stable and better results for the middle and large instances.

The chapter is organized as follows. Section 3.2 presents the integer programming

model and 0/1 programming model of the user project planning problem. Section 3.3 describes the greedy heuristic algorithm. Section 3.4 reports the parameters of test instances and the computational results of the two mathematical programming models and the greedy heuristic algorithm. We draw the conclusion in the last section.

3.2 Mathematical programming models

Based on the problem description and formulation described in Chapter 2, we propose two mathematical programming models including an integer programming model and a 0/1 programming model and run a general tool (CPLEX MIP Solver) to solve these models.

3.2.1 Integer programming

We firstly create an integer programming model where we use decision variables similar to those described in Section 2.3.1. The decision variables take their values in an integer interval representing the activity. This model has the advantage of requiring a reduced number of decision variables and easing the expression of the constraints.

Decision variables

$$x_{ij} = \begin{cases} k \in \{1, \dots, A\} & \text{activity } a_k \text{ started by user } i \text{ at timeslot } j, \\ -1 & \text{user } i \text{ is not available at time } j \text{ (} ua_{ij} = -1 \text{)}, \\ 0 & \text{otherwise} \end{cases}$$

$$y_{ij} = \begin{cases} k \in \{1, \dots, A\} & \text{activity } a_k \text{ with resource } i \text{ started at timeslot } j, \\ -1 & \text{resource } i \text{ is not available at time } j \text{ (} ra_{ij} = -1 \text{)}, \\ 0 & \text{otherwise} \end{cases}$$

$$\text{Maximize } f(S) = w_1 f_1(S) + w_2 f_2(S) + w_3 f_3(S) \quad (3.1)$$

where

Suitability maximization objective:

$$f_1(S) = \sum_{u \in \mathcal{U}} \sum_{t \in \mathcal{T}} \sum_{a \in \mathcal{A}} (x_{ut} == a) \times pre_{ua} \quad (3.2)$$

User cost minimization objective:

$$f_2(S) = \sum_{u \in \mathcal{U}} bud_u - \sum_{u \in \mathcal{U}} \sum_{t \in \mathcal{T}} \sum_{a \in \mathcal{A}} (x_{ut} == a) \times pri_a \quad (3.3)$$

Structure resource minimization objective:

$$f_3(S) = R \times T - \sum_{r \in \mathcal{R}} \sum_{t \in \mathcal{T}} \sum_{a \in \mathcal{A}} (y_{rt} == a) \times dur_a + \sum_{r \in \mathcal{R}} \sum_{t \in \mathcal{T}} ra_{rt} \quad (3.4)$$

Subject to

C_1 . User availability:

$$\forall u \in \mathcal{U}, \forall t \in \mathcal{T}, x_{ut} = -1, \text{ if } ua_{ut} = -1 \quad (3.5)$$

$$\forall u \in \mathcal{U}, \forall t \in \mathcal{T}, \forall a \in \mathcal{A}, \forall d \in \{t + 1, \dots, t + dur_a - 1\}, \quad (3.6)$$

$$x_{ud} = 0, \text{ if } x_{ut} = a$$

A user can only participate in an activity at most once:

$$\forall u \in \mathcal{U}, \forall a \in \mathcal{A}, \sum_{t \in \mathcal{T}} (x_{ut} == a) \leq 1 \quad (3.7)$$

C_2 . Activity availability:

$$\forall a \in \mathcal{A}, \forall u \in \mathcal{U}, \forall t \in \mathcal{T}, \forall d \in \{t - dur_a + 1, \dots, t\}, \quad (3.8)$$

$$x_{ut} \neq a, \text{ if } aa_{at} = -1$$

$$\forall a \in \mathcal{A}, \forall r \in \mathcal{R}, \forall t \in \mathcal{T}, \forall d \in \{t - dur_a + 1, \dots, t\}, \quad (3.9)$$

$$y_{rt} \neq a, \text{ if } aa_{at} = -1$$

$$\forall u \in \mathcal{U}, \forall t \in \mathcal{T}, \forall a \in \mathcal{A}, \forall d \in \{t - dur_a + 1, \dots, t\}, \quad (3.10)$$

$$((x_{ut} == a) \times ((t \div TS) == ((t + dur_a - 1) \div TS)) > 0) \vee (x_{ut} \neq a)$$

$$\forall r \in \mathcal{R}, \forall t \in \mathcal{T}, \forall a \in \mathcal{A}, \forall d \in \{t - dur_a + 1, \dots, t\}, \quad (3.11)$$

$$((y_{rt} == a) \times ((t \div TS) == ((t + dur_a - 1) \div TS)) > 0) \vee (y_{rt} \neq a)$$

C_3 . *Resource availability:*

$$\forall r \in \mathcal{R}, \forall t \in \mathcal{T}, y_{rt} = -1, \text{ if } ra_{rt} = -1 \quad (3.12)$$

$$\forall r \in \mathcal{R}, \forall t \in \mathcal{T}, \forall a \in \mathcal{A}, \forall d \in \{t + 1, \dots, t + dur_a - 1\}, \quad (3.13)$$

$$y_{rd} = 0, \text{ if } y_{rt} = a$$

C_4 . *Budget constraint:*

$$\forall u \in \mathcal{U}, \forall a \in \mathcal{A}, \sum_{t \in \mathcal{T}} (x_{ut} == a) \times pri_a \leq bud_u \quad (3.14)$$

C_5 . *Activity capacity:*

$$\forall t \in \mathcal{T}, \forall a \in \mathcal{A}, \sum_{u \in \mathcal{U}} (x_{ut} == a) \leq cap_a \quad (3.15)$$

C_6 . *Feature constraint:*

$$\forall u \in \mathcal{U}, \forall t \in \mathcal{T}, \forall a \in \mathcal{A}, \forall f \in \mathcal{F}, \quad (3.16)$$

$$(x_{ut} == a) \times req_{af} - \sum_{r \in \mathcal{R}} (y_{rt} == a) \times fea_{rf} \leq 0$$

C_7 . *Preference constraints.*

$$\forall u \in \mathcal{U}, \forall t \in \mathcal{T}, \forall a \in \mathcal{A}, x_{ut} \neq a \vee (x_{ut} == a) \times pre_{ua} > 0 \quad (3.17)$$

where the “ == ” term is the equivalence linear constraint which behaves like a binary condition such that it returns 1 or 0 according to whether the given term is true and false.

3.2.2 0/1 programming

Defining decision variables with binary values is very popular in mathematical modeling. We also experiment this modeling method for our user project planning problem. Compared to the integer programming model, the 0/1 programming model requires more decision variables and needs more mathematical expressions to describe the problem constraints.

Decision variables

- x_{ijk} Binary variable taking value 1 if user i starts activity k at timeslot j , and 0

otherwise.

- y_{ijk} Binary variable taking value 1 if resource i starts to be used by activity k at timeslot j , and 0 otherwise.

$$\text{Maximize } f(S) = w_1 f_1(S) + w_2 f_2(S) + w_3 f_3(S) \quad (3.18)$$

where

$$f_1(S) = \sum_{u \in \mathcal{U}} \sum_{t \in \mathcal{T}} \sum_{a \in \mathcal{A}} x_{uta} \times pre_{ua} \quad (3.19)$$

$$f_2(S) = \sum_{u \in \mathcal{U}} bud_u - \sum_{u \in \mathcal{U}} \sum_{t \in \mathcal{T}} \sum_{a \in \mathcal{A}} x_{uta} \times pri_a \quad (3.20)$$

$$f_3(S) = R \times T - \sum_{r \in \mathcal{R}} \sum_{t \in \mathcal{T}} \sum_{a \in \mathcal{A}} y_{rta} \times dur_a + \sum_{r \in \mathcal{R}} \sum_{t \in \mathcal{T}} ra_{rt} \quad (3.21)$$

Subject to

C_1 . *User availability:*

$$\forall u \in \mathcal{U}, \forall t \in \mathcal{T}, \forall a \in \mathcal{A}, \forall d \in \{t - dur_a + 1, \dots, t\}, \quad (3.22)$$

$$x_{uda} = 0, \text{ if } ua_{ut} = -1$$

$$\forall u \in \mathcal{U}, \forall t \in \mathcal{T}, \sum_{a \in \mathcal{A}} x_{uta} \leq 1 \quad (3.23)$$

$$\forall u \in \mathcal{U}, \forall t \in \mathcal{T}, \forall a \in \mathcal{A}, \forall d \in \{t + 1, \dots, t + dur_a - 1\}, \quad (3.24)$$

$$x_{uda} = 0, \text{ if } x_{uta} = 1$$

A user can only participate in an activity at most once:

$$\forall u \in \mathcal{U}, \forall a \in \mathcal{A}, \sum_{t \in \mathcal{T}} x_{uta} \leq 1 \quad (3.25)$$

C_2 . Activity availability:

$$\forall u \in \mathcal{U}, \forall t \in \mathcal{T}, \forall a \in \mathcal{A}, \forall d \in \{t - dur_a + 1, \dots, t\}, \quad (3.26)$$

$$x_{uda} = 0, \text{ if } aa_{at} = -1$$

$$\forall r \in \mathcal{R}, \forall t \in \mathcal{T}, \forall a \in \mathcal{A}, \forall d \in \{t - dur_a + 1, \dots, t\}, \quad (3.27)$$

$$y_{rda} = 0, \text{ if } aa_{at} = -1$$

$$\forall u \in \mathcal{U}, \forall t \in \mathcal{T}, \forall a \in \mathcal{A}, \forall d \in \{t - dur_a + 1, \dots, t\}, \quad (3.28)$$

$$(x_{uta} \times ((t \div TS) == ((t + dur_a - 1) \div TS)) > 0) \vee x_{uta} = 0$$

$$\forall r \in \mathcal{R}, \forall t \in \mathcal{T}, \forall a \in \mathcal{A}, \forall d \in \{t - dur_a + 1, \dots, t\}, \quad (3.29)$$

$$(y_{rta} \times ((t \div TS) == ((t + dur_a - 1) \div TS)) > 0) \vee y_{rta} = 0$$

C_3 . Resource availability:

$$\forall r \in \mathcal{R}, \forall t \in \mathcal{T}, \forall a \in \mathcal{A}, \forall d \in \{t - dur_a + 1, \dots, t\}, \quad (3.30)$$

$$y_{rda} = 0, \text{ if } ua_{ut} = -1$$

$$\forall r \in \mathcal{R}, \forall t \in \mathcal{T}, \forall a \in \mathcal{A}, \forall d \in \{t + 1, \dots, t + dur_a - 1\}, \quad (3.31)$$

$$y_{rda} = 0, \text{ if } y_{rta} = 1$$

$$\forall r \in \mathcal{R}, \forall t \in \mathcal{T}, \sum_{a \in \mathcal{A}} y_{rta} \leq 1 \quad (3.32)$$

C_4 . Budget constraint:

$$\forall u \in \mathcal{U}, \sum_{t \in \mathcal{T}} \sum_{a \in \mathcal{A}} pri_a \times x_{uta} \leq bud_u \quad (3.33)$$

C_5 . Activity capacity:

$$\forall a \in \mathcal{A}, \forall t \in \mathcal{T}, \sum_{u \in \mathcal{U}} x_{uta} \leq cap_a \quad (3.34)$$

C_6 . Feature constraint:

$$\forall u \in \mathcal{U}, \forall t \in \mathcal{T}, \forall a \in \mathcal{A}, \forall f \in \mathcal{F}, x_{uta} \times req_{af} - \sum_{r \in \mathcal{R}} y_{rta} \times fea_{rf} \leq 0 \quad (3.35)$$

C_7 . Preference constraints.

$$\forall u \in \mathcal{U}, \forall t \in \mathcal{T}, \forall a \in \mathcal{A}, x_{uta} = 0 \vee x_{uta} \times pre_{ua} > 0 \quad (3.36)$$

3.3 Constructive greedy heuristic

The constructive greedy procedure (Algorithm 1) generates always a feasible solution satisfying all constraints. Following the spirit of the constructive greedy initialization in [LH10] and [LH07], activities with a high preference are given higher priorities to be selected. The construction of solution is implemented with several rounds of activity arrangements for users. In every round, users are traversed according to the increasing order of budgets, users with low budgets are more vulnerable to budget constraint so similar to the solidarity consideration in real-life, users with financial difficulties are given priority. For each user under consideration, the followings steps are performed to achieve an arrangement.

1. Select the most preferred eligible activity a of the user.
2. Iterate through timeslots from the first one until a timeslot t is found where a has been started at t and still has available places, if found, add directly the user to a and move to step 5.
3. Iterate through timeslots from the first one until a timeslot t is found where enough available resources can be used, if found, arrange a for the user with resources and move to step 5.
4. Move to the next preferred activity and repeat steps 2-4 until all eligible activities have been considered.
5. Move to the next user (if the current user is the last one, move to the first user) and repeat steps 1-5.

Specifically, the first step of an arrangement is to select an activity to arrange. To response to the main suitability objective, we always take the most preferred eligible activity, which means that this activity has not been arranged and satisfies the budget constraint and the preference constraint. Then, we consider scheduling and resource allocation for this activity. To arrange this activity for the user, we consider two cases. First, to ensure an efficient use of the resources, we verify whether the same activity has been already arranged for other users and still has available places. If the answer is affirmative, the user will be added directly to the activity in the first available timeslot during the period (step2, lines 12-19 in algorithm). In this case, no new resources need to be allocated. If the first case is impossible, we allocate the required resources to this activity, and arrange the activity at the first timeslot having enough available resources (step3, lines 20-32 in algorithm). The selection of resources is achieved by iterating through the available resources

Algorithm 1 Constructive greedy heuristic algorithm.

```

1: Input: Problem instance  $I$ 
2: Output: Initial solution
3: /* Sort set  $\mathcal{U}$  by increasing order of budget */
4:  $sort\_increasing(\mathcal{U})$ 
5: repeat
6:   for  $u \in \mathcal{U}$  do
7:     /* Sort set  $\mathcal{A}$  by decreasing order of preference of  $u$  */
8:      $sort\_decreasing(\mathcal{A})$ 
9:     for  $a \in \mathcal{A}$  do
10:       $actArranged = 0$ 
11:      if  $a$  is not arranged for  $u$  and  $(u, a)$  satisfies  $C_4, C_7$  then
12:        for  $t \in \mathcal{T}$  do
13:          if  $a$  is already arranged at start-time  $t$  and  $(u, a, t)$  satisfies  $C_1, C_5$  then
14:            /* Arrange  $a$  for  $u$  at  $t$  */
15:             $arrange(u, a, t)$ 
16:             $actArranged = 1$ 
17:            break
18:          end if
19:        end for
20:        if  $actArranged = 0$  then
21:          for  $t \in \mathcal{T}$  do
22:            if  $(u, a, t)$  satisfy  $C_1, C_2, C_5$  then
23:              /* Check if  $C_6$  can be satisfied for  $(u, a, t)$ , if yes, return resources selected as  $\mathcal{R}^*$  */
24:              if  $select\_resources(\mathcal{R}^*)$  then
25:                /* Arrange  $a$  for  $u$  at  $t$  */
26:                 $arrange(u, a, t)$ 
27:                 $actArranged = 1$ 
28:                break
29:              end if
30:            end if
31:          end for
32:        end if
33:      end if
34:      if  $actArranged = 1$  then
35:        break
36:      end if
37:    end for
38:  end for
39: until no more activity can be arranged

```

and taking the resources with the required features until all features are satisfied. If this activity cannot be arranged in either way, the algorithm skips this activity and moves to the next preferred activity (step4). Finally, once an activity is successfully arranged for this user or all activities have been verified as unschedulable, the algorithm moves to the next user (step5). We only arrange at most one activity for each user in a round in order to allocate resources more fairly and prevent users who are first considered from occupying too many resources. When the algorithm finishes the activity arrangement for the last user, it returns to the first user to start a new round of arrangements. The algorithm stops when all the activities of all users have been considered and no more activity can be arranged without violating constraints.

During the constructive greedy procedure, the feasibility of each decision is first verified and then executed. If the decision cannot satisfy all constraints, it will not be performed. As such, the greedy procedure always provides a feasible solution satisfying all the constraints of the problem. On the other hand, as explained above, the algorithm considers the users and assigns activities sequentially and according to some specific orders while ignoring alternative decisions. Consequently, the decisions made during the solution construction are not necessarily optimal, leading typically to sub-optimal solutions.

Finally, for each user and activity, since the decision of activity arrangement is verified by checking each timeslot and resource, the time complexity of this constructive greedy procedure is bounded by $\mathcal{O}(U * A * T * R)$ where U , A , T and R represent the number of users, activities, timeslots and resources, respectively.

3.4 Experimentations and results

In this section, we assess the proposed mathematical programming models and the greedy algorithm for tackling the user project planning problem based on 20 realistic instances. We also give an analysis and comparison of these approaches.

3.4.1 Test instances

To our knowledge, there are no publicly available data for the user project planning problem studied in this work. For our experimental study, we created a set of benchmark instances as follows. We sent a large-scale questionnaire survey to more than 3600 structures in the social and medico-social sector in France. We received more than 200

questionnaire responses and achieved several interviews with professionals who are interested in this research. From the responses received, we created an instance generator that is parametrized to generate problem instances covering various real scenarios in different structures. For our experimental study, we generated 20 test instances with different sizes and characteristics to fit as much as possible typical real situations. Table 3.1 shows the main features of these instances with their parameter values where

- Ins: Instance number;
- U : Number of users;
- R : Number of resources;
- A : Number of activities;
- F : Number of features;
- cap: Maximum capacity of activities.

Table 3.1 – Parameters of the test instances.

Ins	U	R	A	TS	W	F	act_cho%	ua%	ra%	aa%	cap	Structure size	Scheduling length
0	10	5	10	2	2	5	100	100	100	100	5	Small	Short
1	10	8	15	8	1	5	100	100	100	100	5	Small	Short
2	10	8	15	4	2	5	100	100	100	100	5	Small	Short
3	30	10	20	4	4	5	100	100	100	100	5	Medium	Medium
4	30	20	30	2	4	5	100	100	100	100	5	Medium	Medium
5	30	20	40	2	4	5	100	100	100	100	5	Medium	Medium
6	50	20	15	8	1	10	100	100	100	100	10	Medium	Short
7	50	30	20	4	2	10	100	100	100	100	10	Medium	Short
8	50	30	30	2	4	10	100	100	100	100	10	Medium	Medium
9	50	20	30	2	4	10	100	100	100	100	10	Medium	Medium
10	50	30	50	2	2	10	100	100	100	100	10	Medium	Short
11	50	30	40	4	4	10	100	100	100	100	10	Medium	Medium
12	50	30	60	2	8	10	100	100	100	100	10	Medium	Long
13	50	30	80	4	8	10	80	100	100	100	10	Medium	Long
14	50	30	100	4	12	10	80	95	95	95	10	Medium	Long
15	70	30	30	2	2	10	100	100	100	100	10	Large	Short
16	70	40	40	2	4	20	100	100	100	100	10	Large	Medium
17	70	40	50	4	4	20	80	95	95	95	10	Large	Medium
18	100	50	30	2	2	20	100	100	100	100	10	Large	Short
19	100	70	50	2	4	30	800	95	95	95	10	Large	Medium

The values of the above parameters characterize the structure size. And larger structures have more users and of course more resources and activity types. We classify the

instances into three categories according to the structure size: 10 users (small structure), 30/50 (medium structure) and 70/100 users (large structure). These instances cover different possibilities of resource adequacy levels and the distribution matches the realistic results according to the survey that most of structures are of medium size.

- TS : Number of timeslots per day;
- W : Number of weeks, hence $T = TS \times W \times 5$;

These two parameters determine the period length required for scheduling, which is different according to the structure type. The instances cover period lengths from one week to 12 weeks.

- $act_cho\%$: Percentage of activities that can be selected;
- $ua\%$: Percentage of user's available timeslots;
- $ra\%$: Percentage of resource's available timeslots;
- $aa\%$: Percentage of activities' available timeslots;

These four parameters indicate that some activities and timeslots are not selectable. According to our survey, this may happen especially for large structures or long period projects. Even though in most cases, all activities and timeslots are considered selectable during the conception phase, to ensure our model can handle this real situation, we decide to add some unselectable cases for some instances with large structure size or long project period. The constraints C_1 - C_3 ensure that users, resources and activities will not be arranged when they are unavailable and the constraint C_7 ensures that the non-selectable activities are never selected into users' projects.

The other parameters:

- Number of features per resource has: $1 \sim 3$ (uniform distribution)
- Number of features per activity needs: $1 \sim 5$ (uniform distribution)
- Activity price: $0 \sim 10$ (truncated normal distribution $\mu = maxPrice/2, \sigma = (maxPrice/2 - minPrice)/3$)
- User budget: $maxPrice/2 * activityNumber/2 \sim maxPrice * activityNumber$ (truncated normal distribution $\mu = maxBudget/2, \sigma = (maxBudget/2 - minBudget)/3$)
- Preference Level: $0 \sim 10$ (uniform distribution)

To cover as many situations as possible, no two instances have the same parameter settings.

3.4.2 Evaluation function

As mentioned above, the quality of a candidate solution S is evaluated by the weighted sum method as follows.

$$f(S) = \sum_{o \in \{1,2,3\}} w_o \times f_o(S)$$

The weight w_o associated to the o th objective function is computed by

$$w_o = \alpha_o \times \frac{1}{f_o^{max}}$$

with

- α_o the importance coefficient to the o th objective function specified by the decision-maker. To ease the use of the decision-making tool, we predefine five default levels for the decision-maker to choose (i.e., $\alpha \in [1, 5]$).
- $f_o^{max} = \max\{f_o(S) : S \in \mathcal{S}\}$, which represents the maximum value of the o th objective in the whole solution space.

The purpose of this function transformation is to apply a normalization to the objectives which are different in nature and scale. Given that all sub-objective functions are positive, we choose the upper-bound approach [Pro+01] as the transformation method.

$$F_i^{trans} = \frac{F_i(x)}{|F_i^{max}|}$$

This transformation ensures that our transformed sub-objective function has a value between zero and one. To determine F_i^{max} , the most recommended method is to optimize separately each objective function [GR06]. However, this approach is too computationally expensive. In our case, we estimate F_i^{max} as follows:

- $f_1^{max} = \sum_{u \in \mathcal{U}} \sum_{a \in \mathcal{A}} pre_{ua}$;
- $f_2^{max} = R \times T + \sum_{r \in \mathcal{R}} \sum_{t \in \mathcal{T}} ra_{rt}$;
- $f_3^{max} = \sum_{u \in \mathcal{U}} bud_u$.

Other transformation methods are possible as discussed in [MA04; MA05]. However, the adopted transformation has the advantage of being simple and efficient for our application.

3.4.3 Experimental setting

The experiments were carried on a computer with an Intel Core i7-8750H 2.20GHz processor and 32 GB RAM under Linux. The integer programming model and the 0/1 programming model are solved by running the CPLEX MIP solver (Version 12.8) with the MemoryEmphasis parameter as true and other default parameters under a time limit of 2 hours. The proposed greedy algorithm was implemented in C++ and compiled using the g++ compiler with the -O3 option.

The greedy algorithm has no parameter to tune and stops when no more activities can be scheduled. For the importance coefficient α_o , we collected four combinations recommended by professionals in the social and medico-social sector and then distributed them randomly to 20 test instances.

3.4.4 Computational results

Table 3.2 shows computational results on the 20 instances obtained by CPLEX with the two mathematical models and results from the greedy algorithm. Column 2 gives the $\{\alpha\}$ used for an instance. Columns 3 to 8 show the results of the integer model (IP) and the 0/1 model (0/1P) in terms of the lower bound, upper bound and gap achieved by the CPLEX solver. Columns 9 and 10 present the results and run time of the greedy heuristic algorithm (GHA). The best values among the results of the compared approaches are highlighted in boldface. Entries with “-” mean that CPLEX failed to output a result due to one of three reasons: i) out of memory; ii) unknown solution status; iii) the pre-processing phase of CPLEX cannot be finished in two hours.

Table 3.2 shows that the 0/1 programming model led to much better results than the integer programming model. This difference comes from the use of logical constraints in the integer programming model, as described in 3.2.1, to express certain constraints. Indeed, the equivalence relation “==” is used as a term in a numeric expression to behave like a binary value, these nonlinear expressions weakened the performance of the model.

The results of the 0/1 programming model is representative. Even if it does not give the optimal solution for any instance, for the small scale instances (such as instance 0, 1, 2, etc), the model obtains near optimal solutions, as shown by the tiny gaps between the lower bound and the upper bound. But as the instance scale increases, the solution quality is greatly worsen. Indeed, the solutions for the instances 7-10, 15 are feasible, but far from optimality with large gaps between the lower bound and the upper bound. For

Table 3.2 – Experimental results and comparison of the integer model, 0/1 model and greedy heuristic algorithm.

Instance	$\{\alpha\}$	IP			0/1P			GHA	
		LB	UB	GAP	LB	UB	GAP	Obj	Time (s)
0	{5,3,0}	4.1142	15.4908	276.52%	4.8033	4.8068	0.074%	4.0623	0.0003
1	{5,2,1}	2.0186	78.7249	3799.94%	4.9300	5.2124	5.73%	4.2060	0.0008
2	{5,1,2}	4.8469	10.1700	109.82%	5.4966	5.5445	0.87%	5.1804	0.0003
3	{5,1,1}	-	-	-	4.7178	5.4622	15.78%	4.2480	0.0079
4	{5,3,0}	-	-	-	5.2055	7.2272	38.84%	3.9318	0.0059
5	{5,1,2}	-	-	-	4.1436	5.9389	43.33%	4.1916	0.0123
6	{5,2,1}	-	-	-	4.3155	5.6144	30.10%	3.5143	0.0136
7	{5,3,0}	-	-	-	4.3039	130.5689	2933.77%	5.0879	0.0051
8	{5,2,1}	-	-	-	4.1035	119.8726	2821.22%	4.7085	0.0140
9	{5,2,1}	-	-	-	4.2911	123.7569	2783.99%	4.7456	0.0103
10	{5,1,1}	-	-	-	2.7762	4.0878	47.24%	2.8983	0.0110
11	{5,1,1}	-	-	-	-	-	-	4.2377	0.0879
12	{5,1,1}	-	-	-	-	-	-	4.5686	0.1007
13	{5,1,2}	-	-	-	-	-	-	5.4005	0.1621
14	{5,3,0}	-	-	-	-	-	-	4.8303	0.6189
15	{5,1,2}	-	-	-	3.4143	5.3128	55.60%	3.6945	0.0197
16	{5,3,0}	-	-	-	-	-	-	3.8128	0.3053
17	{5,1,2}	-	-	-	-	-	-	4.6198	0.2827
18	{5,2,1}	-	-	-	-	-	-	3.5620	0.0678
19	{5,1,1}	-	-	-	-	-	-	4.2342	0.5341

the remaining large instances, no feasible solution was found by CPLEX with the two mathematical models within 2 hours. We notice that the instance 15 contains 70 users and was solved by CPLEX, but the instances 11-14 with 50 users were not solved. This can be explained by the fact that even if the instance 15 has more users than the instances 11 to 14, it has a shorter scheduling length, fewer timeslots (TS) and fewer weeks (W) (see Table 3.1). This implies that the mathematical model for the instance 15 has much fewer decision variables (60000 against 256000 to 1920000 for the instances 11-14), making it easier to solve than the instances 11 to 14.

Compared to the CPLEX solver with the mathematical programming models, the greedy heuristic algorithm is more successful with a better balance between solution quality and run time. For all instances, the greedy algorithm reached its solutions under one second. Additionally, the solution quality of the greedy algorithm is much better than the mathematical programming models for medium and large scale instances.

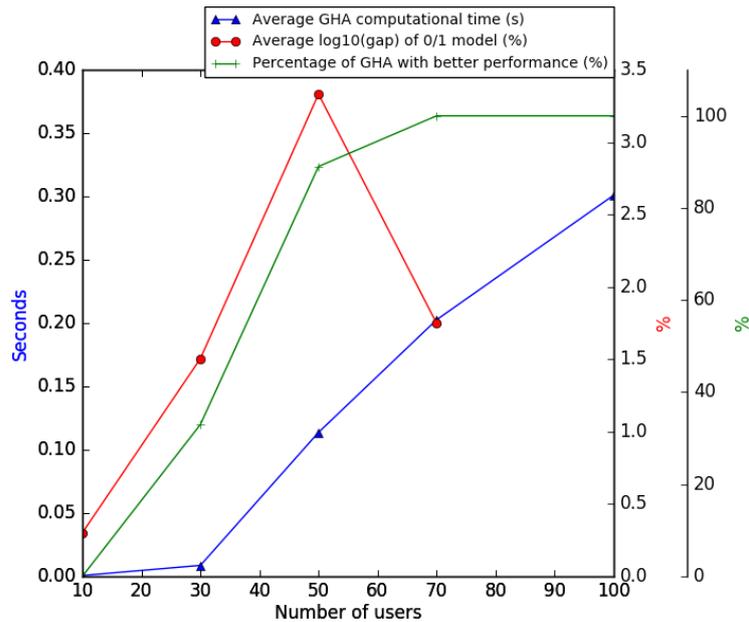


Figure 3.1 – Summary of the computational results by the number of users.

Figure 3.1 shows a visual presentation of the computational results. Given that the integer programming model solved just 3 out of the 20 instances, it is ignored in the figure. The curves in Figure 3.1 correspond to the average \log_{10} *GAP* of the 0/1 programming model (red), the percentage of the instances for which the greedy algorithm gave the best result (green) and the average computational time (in seconds) of the greedy algorithm

(blue). Overall, the results indicate that as the number of users increase, the problem becomes more difficult to solve for both methods. However the greedy algorithm becomes much more computationally efficient compared to the 0/1 programming model.

The plots of Figure 3.2 illustrate the relation between the computational results of the 0/1 programming model and more detailed characteristics of the instances. According to the type of structures, the magnitudes between users, activities and resources are different, which lead to different organizational difficulty levels. To analyze more accurately and avoid the influence of other factors, we tested all instances with a same importance coefficient and compared the results between the instances of similar sizes. The plots show two sets of instances with similar numbers of decision variables of the 0/1 programming model and ordered by increasing order of the gap given by CPLEX (calculated by $|UB - LB|/|LB|$). We observe that instances become harder to solve as the number of users per activity and the number of resources per activity increase. This means that when there are more users and resources to arrange for activities, it is more difficult to make the choice and obtain the best assignment. Indeed, this result corresponds to the decision making difficulties in reality, especially for large structures. For example, when most of the users are interested in several popular activities, the decision-maker has to make the choice to satisfy as many users as possible while guaranteeing the quality of activities at the same time.

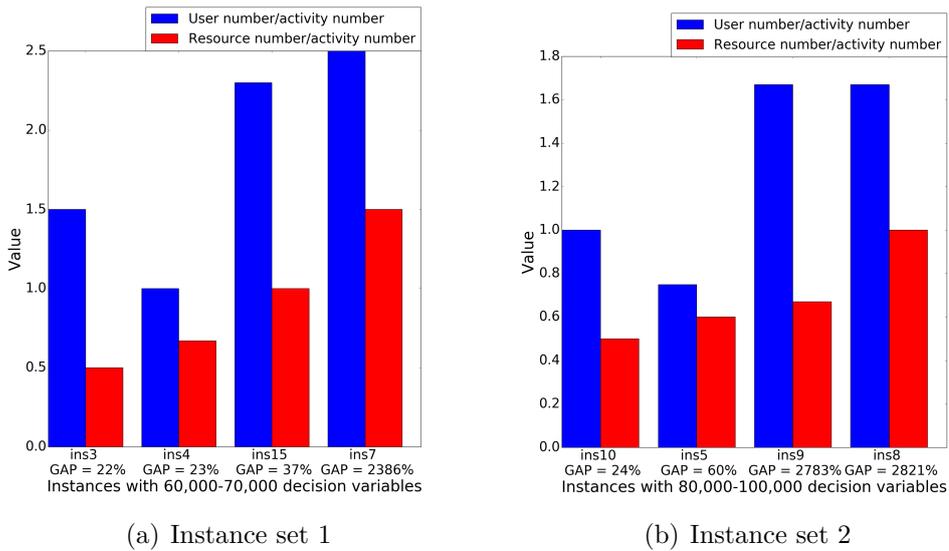


Figure 3.2 – Characteristics of instances and the gap of the 0/1 programming model.

3.5 Conclusion

This chapter presented solution approaches for the personalized user project planning problem introduced in Chapter 2. We created an integer programming model and a 0/1 programming model with the different definition of decision variables. We also studied a greedy heuristic approach which needs less computational cost. These approaches were tested on 20 realistic instances.

- We firstly created an integer programming model where the decision variables take their values in an integer interval representing the activity. This model has the advantage of being easy to understand and less decision variables and constraints. We also experimented a 0/1 programming model where the decision variables take binary values to represent whether this user/resource takes an activity at a timeslot or not. This modeling approach is the most popular and used method but has more decision variables and constraints. We ran these two models on the CPLEX MIP solver.
- Given that the problem is complex and highly constrained, we also proposed a greedy heuristic approach. This constructive greedy procedure generates solution by arranging the preferred activity for each user and provides always a feasible solution satisfying all constraints.
- We created a questionnaire survey about the elaboration of personalized user project to more than 3600 social and medico-social structures and received more than 200 responses. We analyzed the results and achieved several interviews with professionals who are interested in this research. According to responses received from the survey, we generated 20 simulated test instances of different sizes and tunable characteristics to fit as much as possible real situations.
- We tested the two mathematical programming models and the constructive greedy procedure on 20 test instances. The 0/1 programming model showed a better performance compared to the integer programming model but still failed on 8 instances due to the instance size, the constructive greedy procedure generated feasible solutions for all 20 instances in a very short time and performed better than the mathematical models solved by CPLEX on medium and large instances.

MULTI-NEIGHBORHOOD SIMULATED ANNEALING FOR THE USER PROJECT PLANNING PROBLEM

4.1 Introduction

In this chapter, we aim to improve the results of the greedy heuristic algorithm introduced in Chapter 3 by applying the local search techniques. We introduce a multi-neighborhood search with the the general simulated annealing framework called MNSA, which is able to improve significantly the results of the greedy heuristic algorithm and balance the solution quality and computational cost. We summarize the main contributions as follows.

- The MNSA algorithm presented in this chapter combines the efficiency of the SA framework and the effectiveness of multi-neighborhood search. The algorithm has the advantage of finding high quality solution within acceptable computation time.
- The three neighborhoods applied in the algorithm are designed specifically for the user project planning problem, inspired by real-life manual project scheduling operations. We combined these three neighborhoods in a probabilistic exploration way which is able to adjust the importance of each neighborhood.
- We perform experimentation on 20 realistic instances. We demonstrate the high competitiveness of the proposed MNSA algorithm compared to the mathematical programming models and greedy heuristic proposed in Chapter 3.

The chapter is organized as follows. Section 4.2 describes the general scheme of the proposed MNSA algorithm and three neighborhood operators. Section 4.3 reports computational results and comparisons. Section 4.4 investigates the impact of the parameter

settings and the neighborhood combination on the performance of the proposed MNSA algorithm. We draw conclusions in the last section.

4.2 Multi-neighborhood simulated annealing

This section is dedicated to the presentation of the proposed multi-neighborhood simulated annealing algorithm (MNSA) for the user project planning problem.

4.2.1 General procedure

The MNSA algorithm follows the general SA framework [KGV83] and features the combination of three complementary neighborhoods to explore candidate solutions. As shown in Algorithm 2, MNSA starts with an initial solution S and an initial temperature $T = T_0$. Then it performs a number of iterations to find solutions of increasing quality. During each iteration, MNSA first picks a neighborhood N_i ($i = 1, 2, 3$) with a probability p_i where $p_1 + p_2 + p_3 = 1$, then selects randomly a neighbor solution S' from N_i . If S' is feasible, MNSA computes the variation of the objective function between the current solution S and the sampled neighbor solution S' , $\Delta f = f(S') - f(S)$ (Δf is called the move gain). If S' is better than S or as good as S ($\Delta f \geq 0$), then S' replaces the current solution to become the new current solution. This transition is called a move. Otherwise ($\Delta f < 0$), S' is accepted as the new current solution with probability $P(\Delta f, T) = e^{\Delta f/T}$. After performing L moves with the current temperature, where L is a controlling parameter, the parameter T is decreased by a constant cooling factor $\alpha < 1$. Then MNSA starts the next round of its search with the new temperature. MNSA terminates and returns the best solution found during the search when a given stopping condition is met.

It is worth noting that each iteration of MNSA (counter *Iter*) does not necessarily lead to a move (counter *Move*). After each temperature cooling, the counters *Iter* and *Move* will be restarted from zero. As the temperature decreases, the acceptance probability $P(\Delta f, T) = e^{\Delta f/T}$ becomes lower, leading to a lower move acceptance rate $\gamma = Move/Iter$. MNSA uses the move acceptance rate γ as one of the stopping conditions, i.e., the algorithm stops when the move acceptance rate drops to less than a given threshold (empirically fixed to $\gamma = 5\%$) for five consecutive temperature decreases. Another stopping condition is a cutoff time limit. In accordance with the real-life situations in different structures, we fixed the time limit to be 30 minutes, which is an acceptable time for

Algorithm 2 Multi-neighborhood simulated annealing (MNSA).

```

1: Input: Problem instance  $I$ , initial solution  $S$  (Section 3.3), initial temperature  $T_0$ , cooling factor
    $\alpha$ , move counter  $L$ , probabilities  $p_1, p_2, p_3$  for applying neighborhoods  $N_1, N_2, N_3$ 
2: Output: The best solution found  $S^{best}$ 
3:  $T \leftarrow T_0$  /* Set initial temperature */
4:  $S^{best} \leftarrow S$  /* Record the best solution found so far */
5: repeat
6:    $Iter \leftarrow 0$  /* Iteration counter */
7:    $Move \leftarrow 0$  /* Move counter */
8:   repeat
9:     Pick neighborhood  $N_1, N_2$  or  $N_3$  with probabilities  $p_1, p_2, p_3$ 
10:    Pick at random a neighbor solution  $S'$  from the selected neighborhood  $N_i$ 
11:     $Iter \leftarrow Iter + 1$ 
12:    if  $S'$  is feasible then
13:       $\Delta f \leftarrow f(S') - f(S)$ 
14:      if  $\Delta f \geq 0$  then
15:         $S \leftarrow S'$ ;  $Move \leftarrow Move + 1$ 
16:      else if  $e^{\Delta f/T} > random(0, 1)$  then
17:         $S \leftarrow S'$ ;  $Move \leftarrow Move + 1$ 
18:      end if
19:      Update  $S^{best}$  if  $f(S) > f(S^{best})$ 
20:    end if
21:  until  $Move = L$ 
22:   $T \leftarrow T * \alpha$  /* Temperature cooling down */
23: until a stopping condition is met
24: return  $S^{best}$ 

```

decision-makers. During the MNSA search process, only feasible solutions are evaluated and accepted, therefore the best solution found is always feasible. As we explained earlier, the user project planning problem is defined with a set of complex constraints whose feasibility verification is much computationally expensive. As such, the SA framework is quite suitable because only one candidate neighbor solution (instead of all neighbor solutions) is examined at each iteration, which possibly leads to an effective move.

For a given instance, a candidate solution S corresponds to a feasible user project schedule which is represented by a $U \times T$ matrix \mathbf{X} and a $R \times T$ matrix \mathbf{Y} (see Section 2.3.1 and Fig. 2.2). Therefore, the search space Ω explored by our algorithm includes all possible placements of activities in these two matrices that satisfy all constraints ($C_1 - C_7$).

4.2.2 Neighborhoods based on activity change, swap and exchange

It is well known that neighborhood is one key component of any SA algorithm. We present three dedicated neighborhood operators specifically designed for the user project planning problem. These neighborhood operators are inspired by real-life manual project scheduling operations used by human planners.

Activity cancellation and activity arrangement

The cancellation and arrangement of an activity are the most basic operations in user project planning.

The activity cancellation operation is composed of the following steps:

1. Given the attended activity a , user u and a solution S , the start time t and the resources set used res_{at} are known;
2. Set variable(s) $x_{ut} \dots x_{ut+dur_a-1}$ to 0;
3. If no more user participates in activity a from time t , set variable(s) $\{y_{rt} \mid r \in res_{at}\}$ to 0.

The activity arrangement operation is composed of the following steps:

1. Given the unattended activity a , user u and a solution S , verify constraints C_4, C_7 , if feasible, continue the arrangement; else, stop the arrangement;
2. Iterate through timeslots from the first one until a timeslot t is found where a has been started at t and still has available places, if found, set variable(s) $x_{ut} \dots x_{ut+dur_a-1}$

to a , stop the arrangement;

- Pick randomly a timeslot t_i and verify other constraints until tm timeslots have been considered. The verification of resources is achieved by iterating through the available resources from a random start and adding the resources with the required features into resources set res_{at_i} until all features are satisfied or all resources have been verified. If t_i is feasible, set variable(s) $x_{ut_i} \dots x_{ut_i+dur_a-1}$ to a and set variable(s) $\{y_{rt_i} \mid r \in res_{at_i}\}$ to a , stop the arrangement.

Single activity change neighborhood N_1

The single activity change affects one user and one activity, which is a basic operation in user project planning process. Figure 4.1 shows two moves that are eligible for single activity change: a cancellation and an arrangement. For any user and activity, the operation to achieve one move depends on whether this user has participated in this activity.

- Pick a user u from \mathcal{U} and pick an activity a from \mathcal{A} ;
- If u has already a arranged in his/her project, cancel a for u ;
- Otherwise, if u doesn't has a arranged in his/her project, arrange a for u if it doesn't create any conflict with respect to the constraints.

	t1	t2	t3	t4	t5	t6	t7	t8
u1	a1	a1			a3	a3		
u2	a2	a2	a2			a4	a4	a4
u3			a1	a1		a2	a2	a2
u4					a3	a3		

Figure 4.1 – Single activity changes.

As explained above, the cancellation of an activity is always feasible but the arrangement of an activity may fail. The maximum number of timeslot verifications tm is a controlling parameter, the greater it is, the more probably the activity will be successfully arranged, but is more computationally expensive. Hence an appropriate tm is necessary to balance feasibility and efficiency. The single activity change will affect the number of activities in the project of the selected user and may also affect the schedule of resources.

As such, it allows the solution to stay in a stable sparsity level because when the solution matrix is too sparse, which means that users don't have many activities selected in their projects, unattended activities are more probably to be picked and then arranged. Reversely, when the solution is too dense, the cancellation of activities is more probably to be performed. In real-life scheduling, this “single activity change” operation is often used in the conception phase of users' projects to establish feasible projects for every user.

Activity swap neighborhood N_2

An activity swap for a user is performed by a cancellation of an attended activity and an arrangement of an unattended activity. Figure 4.2 shows one eligible move for activity swap.

1. Pick a user u from \mathcal{U} ;
2. Swap the best activity pair (a_1, a_2) : cancel activity a_1 for u , then arrange activity a_2 for u if it doesn't create any conflict.

	t1	t2	t3	t4	t5	t6	t7	t8
u1	a1	a1			a3	a3		
u2	a2	a2	a2			a4	a4	a4
u3			a1	a1		a2	a2	a2
u4			a1	a1	a3	a3		

Figure 4.2 – Activity swap.

We define the best activity pair (a_1, a_2) to swap for user u as follows:

$$(a_1, a_2) = \left(\arg \min_{a \in \mathcal{A}, act_{ua}=1} pre_{ua}, \arg \max_{a \in \mathcal{A}, act_{ua}=0} pre_{ua} \right)$$

In other words, we replace user's least favorite activity in his/her project with his/her favorite unattended activity to maximize the main suitability objective (f_1). An activity swap concerns one user and two activities which can be seen as two particular simultaneous single activity changes. The difference is that the activity swap will never change the number of activities in the project of the selected user. In real-life, the “activity swap” operation is often used during the execution phase of the project when there exists a

better activity for the user or when some selected activities become unfeasible and we want to maintain the same number of the arranged activities.

Activity exchange neighborhood N_3

Unlike the two neighborhoods above, an activity exchange concerns two users. It consists of swapping two different activities with the same start time of two users or simply switching the activity from one user to another. Figure 4.3 shows two eligible moves for activity exchange.

1. Pick a user u_1 from \mathcal{U} ;
2. Take the best user-activity pair (u_2, a_1) to exchange, the start time of a_1 for u_1 is t_1 ;
3. If a user u_2 starts an another activity a_2 at time t_1 , exchange a_1 and a_2 of the two users ;
4. Otherwise, if u_2 is available at t_1 , cancel a_1 for u_1 and arrange a_1 for u_2 at t_1 .

	t1	t2	t3	t4	t5	t6	t7	t8
u1	a1	a1			a3	a3		
u2	a2	a2	a2			a4	a4	a4
u3			a1	a1		a2	a2	a2
u4					a3	a3		

Figure 4.3 – Activity exchanges.

We define the best user-activity pair (u_2, a_1) to exchange for user u_1 as follows:

$$(u_2, a_1) = \arg \max_{u \in \mathcal{U}, a \in \mathcal{A}} pre_{ua} - pre_{u_1a} + pre_{u_1x_{ut_1}} - pre_{ux_{ut_1}}$$

In other words, the best user-activity pair to exchange for u_1 will lead to the largest gain of the evaluation function f . The activity exchange will never change the schedule of resources because there is no activity arrangement or cancellation involved, it only changes the choice of user for an activity already arranged. Hence, it doesn't affect the structure resource objective (f_2) and the user cost objective (f_3). As such, the variation

of the evaluation function f can be calculated quickly by calculating simply the variation of the suitability objective f_1 . The “activity exchange” operation is often applied in real-life during the conception phase and execution phase of users’ projects because many activities need typically to be changed after the evaluation phase to maximize the global user satisfaction and to ensure that every activity will be attended by users who need it the most.

Time complexity

1. *Single activity change*: Given user u and activity a , checking if u has participated in a requires $O(1)$. If the operation is a cancellation, the time complexity is bound by $O(R)$ when the resources are involved. If it’s an arrangement, tm timeslots will be verified where each verification requires $O(R)$. As such, one single activity change takes time $O(tmR)$, which is $O(R)$ since the timeslots verification number tm is a constant parameter.
2. *Activity swap*: Finding the best activity pair (a_1, a_2) to swap for a user requires $O(A)$. Then, the time complexity for a cancellation and an arrangement is $O(R)$. Therefore, one activity swap takes time $O(A + R)$.
3. *Activity exchange*: To find the best user-activity pair (u_2, a_1) to swap for a user, all user-activity pairs have to be evaluated, the variation of the evaluation function f can be calculated in $O(1)$ since only the first objective f_1 is involved. Therefore, the time complexity of one activity exchange is bound by $O(A * U)$.

4.2.3 Probabilistic exploration of the neighborhoods

Given the three neighborhoods presented previously, MNSA explores them in a probabilistic way. Specifically, at each iteration, MNSA selects neighborhood N_i ($i = 1, 2, 3$) to explore with probability p_i such that $p_1 + p_2 + p_3 = 1$. As explained above, these three neighborhoods have very different efficiency and effectiveness. Compared to other neighborhood combination methods such as neighborhood union or token-ring neighborhood exploration [LHG11], the probabilistic combination has the advantage of allowing us to adjust the importance given to each neighborhood by tuning the probability settings (p_1, p_2, p_3) , which in turn influences the behavior of the MNSA algorithm. More information is provided in Section 4.4.1.

Table 4.1 – Settings of MNSA’s parameters.

Parameter	Section	Description	Value
T_0	4.2.1	initial temperature of SA	0.1
L	4.2.1	number of move per temperature of SA	300000
α	4.2.1	cooling ratio of SA	0.97
p_1, p_2, p_3	4.2.1	neighborhood selection probabilities	(0.8,0.1,0.1)
tm	4.2.2	maximum selection number of timeslots	10

4.3 Experimentations and results

In this section, we assess the proposed MNSA on the 20 benchmark instances and give the comparison with several reference methods.

4.3.1 Experimental setting

Computing platform. The proposed MNSA algorithm was implemented in C++ and compiled using the g++ compiler with the -O3 option. The experiments were carried on an Intel Xeon E5-2670 processor with 2.5 GHz and 200 MB RAM under the Linux operating system.

Parameter settings. We used the automatic configuration tool “irace” [Lóp+16] to calibrate the parameters of the MNSA algorithm. This was achieved by feeding irace with six randomly selected instances with a tuning budget of 1000. We adopted the best parameter configuration given by irace and shown in Table 4.1.

Reference methods for comparisons. To evaluate the performance of MNSA, we performed comparisons with three reference methods: the constructive greedy algorithm proposed in 3.3, the 0/1 programming model proposed in 3.2.2 and a classic algorithm for multiple neighborhood structures: the variable neighborhood search (VNS) [HM01]. We adopted the basic VNS described in [HM01] with following steps:

1. Set $k = 1$;
2. Until $k = k_{max}$ ($k_{max} = 3$), repeat steps 3 - 5;
3. Shaking: generate a random neighborhood solution from N_k ;
4. Local search: variable neighborhood descent with first improvement (VND);
5. If this local optimum is better than the incumbent, move there and continue the search with $k = 1$; otherwise, set $k = k + 1$;

4.3.2 Computational results

Given the stochastic nature of the MNSA algorithm, we ran MNSA 20 times independently with 20 random seeds to solve each test instance. Each run was stopped when the cutoff time of 30 minutes was reached or when the move acceptance rate γ drops below the threshold of 5% for five consecutive search rounds (see Section 4.2.1).

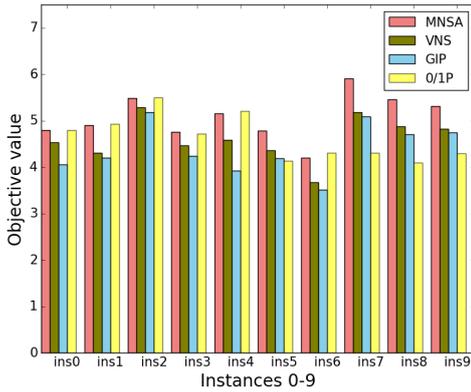
Table 4.2 shows the computational results of the proposed MNSA algorithm and the compared approaches. Column 2 gives the objective function weight used for each instance (see Section 3.4.2). Columns 3 to 6 report the results of MNSA with its best objective value f_{best} found among 20 runs, average objective value f_{avg} , standard deviations std and the average time t_{avg} in seconds to reach the best objective value. Note f_{best} and f_{avg} correspond to lower bounds of the optimal objective value. Columns 7 to 9 (0/1P) give the lower bound (LB), upper bound (UB) and $gap\%$ ($gap\% = |UB - LB|/|LB|$) achieved by the CPLEX solver with the 0/1 programming model. Columns 10 and 11 (GIP) present the results and run time of the greedy procedure. Columns 12 and 13 (VNS) present the best objective value found and the average time to reach the best objective value among 20 runs of the variable neighborhood search. The best values among the results of the compared approaches are highlighted in boldface. Entries with “-” mean that CPLEX failed to output a result due to one of three reasons: i) out of memory; ii) unknown solution status; iii) the pre-processing phase of CPLEX cannot be finished in two hours.

From the results of Table 4.2, we can make the following comments. First, compared to the powerful MIP solver CPLEX, among the 12 instances where CPLEX produced a result, MNSA found a better solution for seven instances against four cases in favor of the CPLEX. The winning cases of CPLEX concerns only some small or medium scale instances. Moreover MNSA has very small standard deviations, indicating its high robustness. Second, MNSA obviously improves on the results of the constructive greedy procedure GIP and dominates the VNS for the whole set of 20 instances. The small p -value ($8.86e - 5 \ll 0.05$) from the non-parametric Wilcoxon signed-rank test confirms the statistical significance between the results of MNSA and those of GIP and VNS. To complement this numerical comparison, Figure 4.4 summarizes the comparison using barchart, which clearly shows the dominance of the MNSA algorithm over the reference approaches.

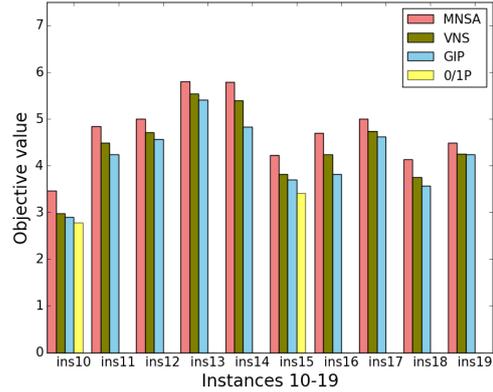
Finally, to further illustrate the the performance difference among the three compared approaches in a global manner, we show in Figure 4.5 their performance profiles [DM02]. Since the performance profile is defined with respect to a minimization criterion and

Table 4.2 – Computational results of the proposed MNSA algorithm and comparison with the reference approaches: 0/1 mathematical programming with CPLEX (0/1P), greedy initialization procedure (GIP) and variable neighborhood search (VNS).

Ins	α	MNSA				0/1P			GIP		VNS	
		f_{best}	f_{avg}	std	t_{avg}	LB	UB	gap%	f	t	f	t_{avg}
0	{5,3,0}	4.8033	4.8027	0.001137	158.87	4.8033	4.8068	0.074%	4.0623	0.0003	4.5333	0.0019
1	{5,2,1}	4.9113	4.8990	0.010970	171.29	4.9300	5.2124	5.73%	4.2060	0.0008	4.3098	0.0031
2	{5,1,2}	5.4940	5.4842	0.007514	100.20	5.4966	5.5445	0.87%	5.1804	0.0003	5.296	0.0042
3	{5,1,1}	4.7599	4.7494	0.005576	385.44	4.7178	5.4622	15.78%	4.2480	0.0079	4.4682	0.0251
4	{5,3,0}	5.1557	5.1267	0.014311	241.09	5.2055	7.2272	38.84%	3.9318	0.0059	4.5893	0.0237
5	{5,1,2}	4.7824	4.7708	0.005865	638.44	4.1436	5.9389	43.33%	4.1916	0.0123	4.3606	0.0862
6	{5,2,1}	4.1999	4.1537	0.030969	203.50	4.3155	5.6144	30.10%	3.5143	0.0136	3.6691	0.0233
7	{5,3,0}	5.9196	5.8732	0.029707	413.04	4.3039	130.5689	2933.77%	5.0879	0.0051	5.1856	0.0095
8	{5,2,1}	5.4681	5.4393	0.015094	443.30	4.1035	119.8726	2821.22%	4.7085	0.0140	4.8762	0.0417
9	{5,2,1}	5.3193	5.2901	0.013284	401.42	4.2911	123.7569	2783.99%	4.7456	0.0103	4.8281	0.0298
10	{5,1,1}	3.4648	3.4508	0.010898	772.11	2.7762	4.0878	47.24%	2.8983	0.0110	2.9788	0.0271
11	{5,1,1}	4.8414	4.8293	0.007239	1011.47	-	-	-	4.2377	0.0879	4.4845	0.3210
12	{5,1,1}	4.9955	4.9865	0.005668	1279.76	-	-	-	4.5686	0.1007	4.7167	0.3445
13	{5,1,2}	5.7975	5.7845	0.005673	1580.86	-	-	-	5.4005	0.1621	5.5392	0.7872
14	{5,3,0}	5.7826	5.7672	0.011426	1415.78	-	-	-	4.8303	0.6189	5.3935	1.5429
15	{5,1,2}	4.2270	4.2042	0.011491	604.31	3.4143	5.3128	55.60%	3.6945	0.0197	3.8136	0.1077
16	{5,3,0}	4.6985	4.6606	0.021251	939.02	-	-	-	3.8128	0.3053	4.2333	0.6114
17	{5,1,2}	5.0051	4.9885	0.007458	1497.09	-	-	-	4.6198	0.2827	4.7399	0.7498
18	{5,2,1}	4.1314	4.1107	0.011794	775.01	-	-	-	3.5620	0.0678	3.7551	0.1456
19	{5,1,1}	4.4820	4.4716	0.005591	1789.225	-	-	-	4.2342	0.5341	4.2499	0.3389
p -value									8.86e-5		8.86e-5	



(a) Instances 0-9



(b) Instances 10-19

Figure 4.4 – Comparison summary of MNSA with 0/1 mathematical programming with CPLEX (0/1P), greedy initialization procedure (GIP) and variable neighborhood search (VNS).

our problem has a maximization objective, we first applied a transformation to convert our objective function into a minimization case. Given a set \mathcal{I} of problem instances and a set \mathcal{A} of algorithms, we convert the objective value $f_{i,a}$ on instance i by algorithm a to $\hat{f}_{i,a} = f_{i,best} + \Delta f_{i,a}$ where $f_{i,best}$ is the best objective value on instance i by all algorithms in \mathcal{A} and $\Delta f_{i,a} = f_{i,best} - f_{i,a} (\Delta f_{i,a} \geq 0)$, which represents the gap of the objective value of approach a relative to the very best objective value. If $\Delta f_{i,a} = 0$, the approach a is the best approach in \mathcal{A} on instance i , and we have $\hat{f}_{i,a} = f_{i,a} = f_{i,best}$. If $\Delta f_{i,a} > 0$, the approach a performs worse than the best approach in \mathcal{A} , and we have $\hat{f}_{i,a} > f_{i,best}$. This transformation method thus ensures that all transformed objective values are greater than zero and maintain the original performance quality order. Based on the transformed objective value $\hat{f}_{i,a}$, we define the performance ratio by $r_{i,a} = \hat{f}_{i,a} / \min\{\hat{f}_{i,a} : a \in \mathcal{A}\}$ to represent the performance on instance i by approach a compared to the best performance by any approach on i . We set $r_{i,a}$ to $+\infty$ if approach a does not solve instance i . Then, to obtain an overall performance assessment of approach a , we define $P_a(\tau) = size\{i \in \mathcal{I} : r_{i,a} \leq \tau\} / |\mathcal{I}|$, which is the probability for approach a that its performance ratio $r_{i,a}$ is within a factor τ . The function P_a is the (cumulative) distribution function for the performance ratio. The value $P_a(1)$ is the number of instances that approach a matches or wins over other approaches. We draw the performance profiles with the software “*perprof-py*”¹. From Figure 4.5, we observe that MNSA dominates the other reference approaches with the best performance profile either in terms of the best objective value or the average objective value.

4.4 Analysis and discussions

In this section, we present an analysis of the neighborhoods and parameters used in the proposed algorithm.

4.4.1 Analysis of neighborhoods

MNSA explores three neighborhoods N_1, N_2, N_3 in a probabilistic way. In order to assess the benefit of this neighborhood combination strategy, we present a comparison of MNSA with three MNSA variants where only one single neighborhood is used. We ran these variants under the same experimental conditions as for MNSA to solve the 20

1. Available at <https://ufpr-opt.github.io/perprof-py/index.html>

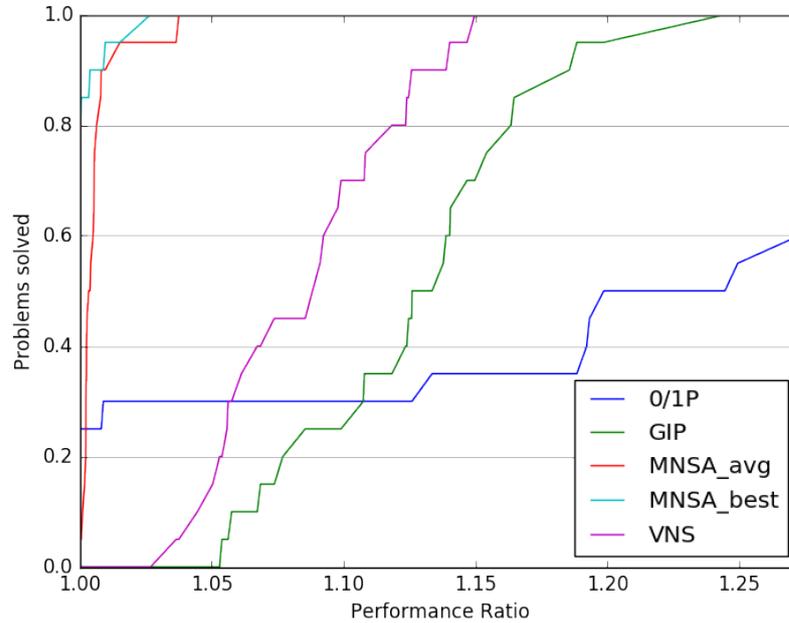


Figure 4.5 – Performance profiles of the 0/1 programming model, greedy initialization procedure, MNSA average results, MNSA best results and variable neighborhood search.

instances. The computational results are reported in Table 4.3 and the visual presentation is shown in Figure 4.6. We observe that the combination of the three neighborhoods perform significantly better than each single neighborhood used alone. Moreover, among the three neighborhoods, the “single activity change” neighborhood N_1 performs much better than the other two neighborhoods. This explains why in MNSA, N_1 is given a much higher application probability ($p_1 = 0.8$ against $p_2 = p_3 = 0.1$, see Table 4.1), allowing thus this dominant neighborhood to play its role.

Finally, we find that the “activity swap” neighborhood N_2 alone can barely improve the results of the constructive greedy initialization. In fact, the greedy procedure typically produces a very dense solution with all possible activities and resources arranged. As the result, the activity swap can rarely find a feasible move from such a solution. Interestingly, when combined with other neighborhoods especially with N_1 , the variant state of solution provides N_2 with more opportunities to find a feasible move. In conclusion, the combination of neighborhoods brings highly positive effects for the search thanks to the collective symbiosis of these complementary neighborhoods.

Table 4.3 – Computational results (f_{best}) of MNSA with the combined neighborhoods ($N_1 + N_2 + N_3$) and each individual neighborhood: “single activity change” neighborhood (N_1), “activity swap” neighborhood (N_2), “activity exchange” neighborhood (N_3).

Instance	N_1	N_2	N_3	$N_1 + N_2 + N_3$
0	4.8033	4.0623	4.1682	4.8033
1	4.9050	4.2060	4.2179	4.9113
2	5.4874	5.1804	5.2096	5.4940
3	4.6947	4.2480	4.3150	4.7599
4	5.1088	3.9318	4.0195	5.1557
5	4.7107	4.1916	4.2361	4.7824
6	4.1559	3.5143	3.5731	4.1999
7	5.8859	5.0879	5.1509	5.9196
8	5.4326	4.7091	4.7537	5.4681
9	5.2823	4.7456	4.7936	5.3193
10	3.3632	2.8983	2.9389	3.4648
11	4.7769	4.2377	4.2905	4.8414
12	4.9326	4.5686	4.6048	4.9955
13	5.7463	5.4005	5.4172	5.7975
14	5.7448	4.8303	4.8695	5.7826
15	4.1891	3.6945	3.7509	4.2270
16	4.6488	3.8128	3.8526	4.6985
17	4.9438	4.6198	4.6610	5.0051
18	4.0512	3.5620	3.6170	4.1314
19	4.3904	4.2342	4.2933	4.4820
<i>p</i> -value	8.86e-05	8.86e-05	8.86e-05	

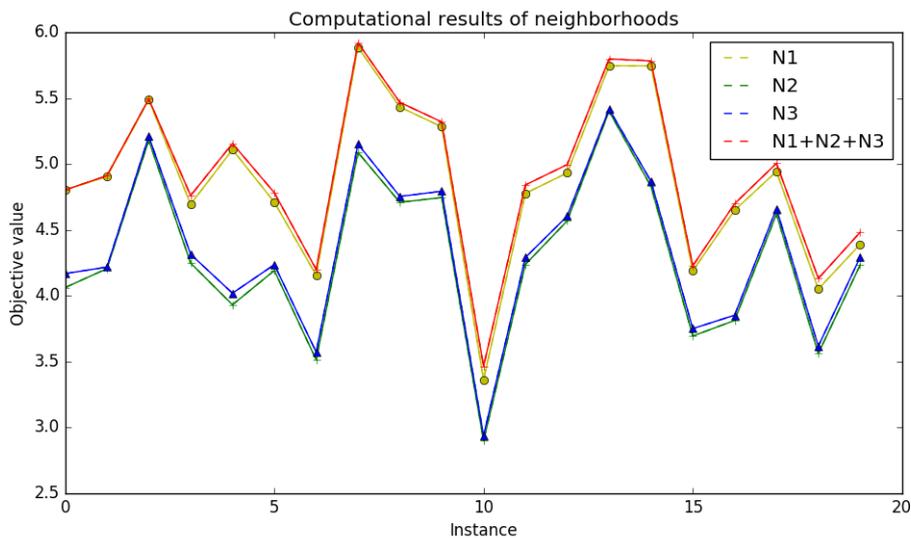


Figure 4.6 – Comparison of four neighborhood combinations.

Table 4.4 – Parameter levels for the 2-level full factorial experiment of MNSA.

Parameters	Low level	High level
T_0	0.05	1.0
L	100000	500000
α	0.95	0.99
p_1, p_2, p_3	(0.4,0.3,0.3)	(0.8,0.1,0.1)
tm	10	50

4.4.2 Analysis of parameters

The proposed MNSA algorithm requires five parameters. The initial temperature T_0 , the number of move per temperature L , and the cooling ratio α are required by the SA framework, while the neighborhood probabilities (p_1, p_2, p_3) and the timeslots maximum selection number tm are used to define the neighborhoods. To investigate the effects of these parameters and their interaction effects, we performed a 2-level full factorial experiment [Mon17] with the levels of these parameters shown in Table 4.4.

We conducted the factorial experiment on six randomly selected instances and each instance was solved 20 times independently. We recorded the average value of the best objective values found on the six instances. We applied the multivariate analysis of variance on the results obtained with the verification of normality distributions and the variance homogeneity. Table 4.5 provides the results of the analysis with the p -values from the F -statistic test for each parameter, all the combinations of two parameters and all five parameters. For a significance level of 0.05, a p -value less than 0.05 indicates a significant relationship between each parameter (or parameter combination) and the results. We observe that each parameter has a significant effect on the algorithm performance according to the very small p -values. This is specially true for the number of moves per temperature L and the timeslots maximum selection number tm . Moreover, the interaction effects between any two parameters are also significant, which explains why we didn't apply the one-at-a-time sensitivity analysis to determine our parameters' values. In conclusion, the MNSA algorithm is sensitive to the settings of its parameters and the interaction effects between them are significant.

4.5 Conclusion

This chapter described an effective local search algorithm called MNSA for the personalized user project planning problem. Based on the simulated annealing framework, the

Table 4.5 – Results of the analysis of variance of the 2-level full factorial experiment of MNSA.

Source	<i>p</i> -value
T_0	2.615913e-297
L	0.000000e+00
α	5.014828e-213
(p_1, p_2, p_3)	3.125439e-231
tm	0.000000e+00
α^*T_0	3.126968e-233
α^*L	2.143320e-263
T_0^*L	9.885850e-284
α^*tm	2.727317e-107
T_0^*tm	6.006472e-246
L^*tm	4.492402e-304
$\alpha^*(p_1, p_2, p_3)$	1.734318e-29
$T_0^*(p_1, p_2, p_3)$	1.622446e-04
$L^*(p_1, p_2, p_3)$	1.310061e-144
$tm^*(p_1, p_2, p_3)$	5.137107e-40
$\alpha^*T_0^*L^*tm^*(p_1, p_2, p_3)$	1.592908e-121

proposed algorithm explores three dedicated neighborhood structures in a probabilistic way.

- Since the problem is highly constrained, the combination of simulated annealing method and probabilistic neighborhood search achieved an efficient exploration of the neighborhood structures, where each feasibility verification could lead to an effective move.
- We showed the effectiveness of the multi-neighborhood local search. These three neighborhoods are inspired by real planning operations, the single activity change neighborhood can explores the search space quickly and the the activity swap and exchange neighborhoods focus on the improvement of evaluation function. We also proved the advantage of combining these three neighborhoods compared with each single one.
- MNSA perform well on the realistic test instances, compared with approaches proposed in Chapter 3, MNSA found a better solution for 7 instances against 4 cases in favor of the CPLEX and obviously improved the result of the constructive greedy procedure.

ADVERSE EVENT REPORT CLASSIFICATION IN SMSS

5.1 Introduction

In the previous chapters, we presented the modeling and solutions of the user project planning problem, which aims to ease the project conception task for decision-makers. Once the project plan is validated, the user will start to participate in activities according to this plan, which is the second part of the project programming process (see Section 2.2.3). To better follow user's project, professionals have to note all related information and one of the major sources is the adverse events (AEs). However, structures are struggling to conduct an efficient analysis of AEs due to the abysmal quality of event reports. In this chapter, we aim to objectify the information of an AE in the SMSS by proposing the most corresponding facts according to the analysis of the event description. Specifically, the proposed classification method is based on the keyword lexicons which are constructed with the real event cases and a terminology of synonyms. In addition, we develop a visualization tool which helps the management and analysis of AEs for structures. We summarize the main contributions of this work as follows.

- First, we collect 500 real AE descriptions from more than 30 social and medico-social structures. For each event, we extract its keywords and classify it with a fact manually. Based on these real AEs, we construct a list of facts that contains almost all possible situations of AE within a social and medico-social structure.
- Second, we propose a method to predict automatically the fact of an event according to its text description to ease the task of reporter. The method contains three phases: a preprocessing phase, a lexicon construction phase and a similarity calculation phase. The knowledge used contains not only the classified events but also a terminology of synonyms specially designed for the SMSS, which can guide the

reporters to employ more technical expressions and terms to improve their writing quality. We test our method on the collected real AEs with a 5-fold cross-validation and obtain appreciable results.

- Third, based on the events classified, we develop a knowledge visualization tool including the keyword cluster graph and the event cluster graph. The visualization tool provides an intuitive view to observe and analyze the distribution and relation between terms and events, which can help structures to correct and improve the writing and service quality.

This chapter is introduced as follows: Section 5.2 presents the practical context about the AE in the SMSS and the motivation of this work. Section 5.3 presents representative studies on AE and text classification. Section 5.4 introduces the data that we work on and the annotation process. Section 5.5 presents the main components of our classification method and Section 5.6 reports the classification results and the visualization tool. We draw conclusions in the last section.

5.2 Practical context

In this section, we introduce the practical background of the adverse event in the SMSS and the motivation of this work.

5.2.1 The adverse event in SMSS

In France, Adverse Events (AEs) in the SMSS include any serious dysfunctions in the management or organization which could affect the care of users, the support or respect for their rights and any event having the effect of threatening the physical or moral health of users or employees, directly or potentially. Different from the AE in the medical sector which relates mostly with drug, care or operation, the AE in the SMSS involves more complex situations since the hosted population is very diverse. Typical AE types include:

- Violent behavior (verbal or physical aggression).
- Accidents or incidents related to an error of care or supervision (mistreatment, medication error).
- Malicious acts within the structure (steal, weapon holding).
- Disasters and exceptional weather events (flood, fire).

- Accidents or incidents related to technical equipment or organization (power breakdown, absence, planning error).
- Suicides and suicide attempts.

The AEs occur frequently in all types of social and medico-social structures, which reflect directly the life quality of users and the management capacity of the structure. An AE may lead to very serious consequences from property damages to casualties. Therefore, to avoid the recurrence and improve the service quality of structures, AEs have received increasing attention from the sector and the government in recent years. Indeed, several policies have been established to standardize the declaration and analysis of AEs in the SMSS:

- The decree No.2016-1813 [16b] defines the nature of serious dysfunctions and events of which the administrative authorities must be informed and obliges the structures to declare to the competent administrative authority, without any delays.
- The decree No.2016-1606 [16a] specifies the modalities of declaration of serious AEs associated with care. This declaration includes two parts: the first part must be reported without delay, which includes the first elements relating to the event, then, after analysis, the second part must be carried out in the following three months, which includes the elements of feedback, as well as the corrective measures taken or envisaged.
- In 2017, The common digital portal¹ is on-line for reporting AEs with various different types of form for structure, professionals and users. The bylaw of February 27, 2017 [17] defines the list of categories of adverse health events for which the declaration or report can be made through the digital portal.

The treatment procedure of an AE is shown in Figure 5.1. When an AE occurs in the structure, the witness(es) (could be user(s), professional(s) or visitor(s)) has(have) to report the detail of event to the structure and the event will be documented. Then, according to the type and severity of the event, the structure has to declare to the administrative authorities or the digital portal using official templates, explaining in detail the event in a chronological order. After the declaration, some analysis and measures could be carried out if necessary. The analysis of an AE is critically important especially for the serious AEs associated with care. An adequate analysis allows the structure to find the root cause of the event and then implement effective measures to prevent the

1. <https://signalement.social-sante.gouv.fr>

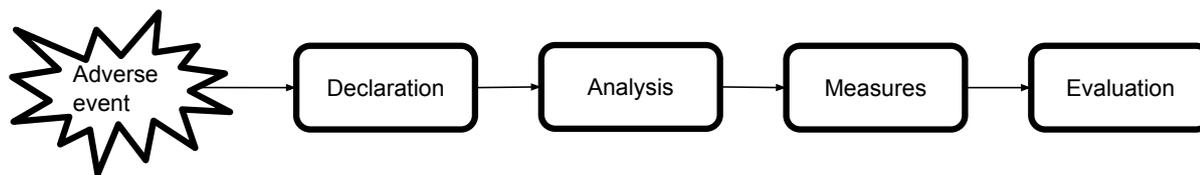


Figure 5.1 – AE treatment procedure in SMSS.

recurrence. The measures to carry out are determined according to analysis, and could be implemented by the structure, such as personnel adjustments, repair, etc.; or by the user, such as canceling or adding certain activities in his/her personalized project. When the measures have been implemented, the structure has to evaluate the effectiveness to make better decision for the next similar AE.

5.2.2 Problematic and motivation

Despite the importance of the declaration and analysis of AE, structures are still confronted with many difficulties in the implementation of these policies. According to the 2019 annual activity report on serious AEs from the French National Authority for Health (HAS) [HAS20], the quality assessment of the declarations received by HAS shows that the analysis is insufficiently performed for 49% of AEs and 53% of AEs could have been avoided. According to our survey and interviews with professionals in the SMSS, there are mainly the following difficulties in practical operations:

- Lack of tools, methods and official guides to record and analyze efficiently the AEs within structures.
- Poor writing quality and subjectivity of the reporter who describes the AE, such as misspellings, colloquialism and imprecision, which often lead to incomprehension or misunderstanding.
- Missing of meta-data, even though most templates provide many additional information choices to detail the event (such as importance, fact, severity), due to the wrong evaluation or lack of concern, these types of data are mostly unavailable.

As a result, extracting objective information from an AE report is indispensable to the analysis and treatment. To represent the AE objectively and accurately, we introduce the concept of *fact* of AE, that can be seen as a very specific AE category. The fact is selected from a definitive fact list defined by the structure with all possible situations that

could happen within the structure (absence of staff, physical aggression, traffic accident, etc). Four examples of AE description and their fact are presented in Table 5.1:

Table 5.1 – AE examples and their fact.

Event description	Fact
The names on two pill organizers prepared by nurses have been reversed. In fact, one in these two persons has received treatment that is not theirs; the second person realized that the treatment was not theirs.	Medication error
Accident between the Abej vehicle (Clio3) and another vehicle during an accompaniment.	Traffic accident
Mr.S (educator) calls to inform that he had an accident on his way to work. He informs that he would be late because he was waiting for the tow truck.	Traffic accident
Fight at the reception between two young people, a third who got involved by giving a punch from behind. The team had to intervene to separate. To calm things down, I had to go out with whoever seemed the most angry.	Physical aggression

As shown in Table 5.1, the fact of an AE represents the most essential information in a concise way which can ensure an objective understanding. Indeed, even some AEs are described in very different way, they may still have the same fact and can be treated with similar measures. Hence, the correct selection of fact can significantly simplify the understanding, statistic, analysis and treatment. However in most AE report cases in the SMSS, the fact of AE is missing or incorrect due to the unprofessional reporting habits of reporters, especially with the lack of efficient decision support reporting tool. To cope with this problem, we explore the possibility of assisting the determination of fact for an AE case according to the analysis of the event description. Next, we present representative studies on AE classification and extraction and our solution method.

5.3 Literature Review

There has been a considerable amount of studies in the filed of AE classification and extraction and table 5.2 summarizes the aim and approaches of some representative related studies.

Table 5.2 – Summary of related studies on AE classification and extraction.

Reference	Aim	Data source	Approach(es)
AE classification			
[Bot+11]	Potential anaphylaxis detection	Unstructured free text	Rule-based classification Supervised machine learning
[Dev+17]	Seriousness determination	Structured adverse event cases	Supervised machine learning Deep Learning
[Lee+17]	Adverse drug event detection	Unstructured tweets	Deep Learning
[Roc+15]	AE detection	Structured data and unstructured description	Supervised machine learning
[Han+17]	Medication-related AE identification	Structured data and unstructured description	Text mining Supervised machine learning
[Mén+19]	AE number prediction	Structured clinical data	Supervised machine learning
[Han+19]	AEs prediction after spine surgery	Structured patient data	Supervised machine learning
AE extraction			
[Pan+19]	AE detection	Structured Product Labels	Event-based text mining
[Ram+14]	AE detection	Structured data and unstructured description	Supervised machine learning
[Lea+10]	Adverse drug reactions extraction	Unstructured user comments to health-related websites	Text mining with a lexicon approach

AE classification task aims to predict an output variable of an AE case. The determination of AE severity has been widely investigated, which is usually formulated as a binary classification task [Bot+11; Dev+17; Han+17; Lee+17; Roc+15]. In addition, Han *et al.* [Han+19] predicted the type of AE from the data of patients which can be seen as a multiclass classification task and Ménard *et al.* [Mén+19] predicted the expected number of AEs to solve the under-reporting issue. Machine learning algorithms have been successfully applied to solve these kinds of problems [Bot+11; Dev+17; Han+17; Han+19; Mén+19; Roc+15] and neural network algorithms [Dev+17; Lee+17] have also been investigated recently. However, these methods are only effective on structured data and rely

on a large set of labeled samples.

AE extraction task aims to extract AE related terms from data or text. Pandey *et al.* [Pan+19] extracted and encoded AE terms from structured product labels and Leaman *et al.* [Lea+10] extracted AE tokens from unstructured comments with a lexicon approach. Ramesh *et al.* [Ram+14] developed an annotated corpus and biomedical named entity tagger to automatically identify medication information and AE entities in text. Text mining tasks require usually qualified experts in the domain to predefine specific rules or lexicon as the basis of mining.

Moreover, all these studies focus on English AE reports, to the best of our knowledge, there is no study on French AE reports. Related work about French medical text includes disease detection from dialogue [Amb+20], automatic classification of doctor-patient questions [Lal+19] and drug non-adherence detection [BG19]. This work presents the first classification approach to classify French AE reports in the SMSS by adapting the lexicon-based classification method with lexicons specifically constructed for the sector including professional terms and informal expressions.

Natural language processing (NLP) refers to automatic computational processing of human languages to computer-understandable format [All88]. Advanced NLP techniques have been successful to solve AE classification problems [MH05], including part-of-speech (POS) tagging [Sch13] to mark up the POS of words in text and named-entity recognition (NER) [NS07] to identify entity (person, organization, etc.) names, etc. In addition, lexicon-based classification approaches have also been applied to text classification [Ban+17; MG16]. However, most lexicon generation methods require a large corpus [Ban+14] or official thesaurus [Lea+10] to extract terms, which are not available for the AE in the SMSS. We propose hence a method adapted to the particular situation for the AE of SMSS in France where corpus resource is still missing.

5.4 Adverse event data

In this section, we present the main features of the text data that we work on and the annotation process.

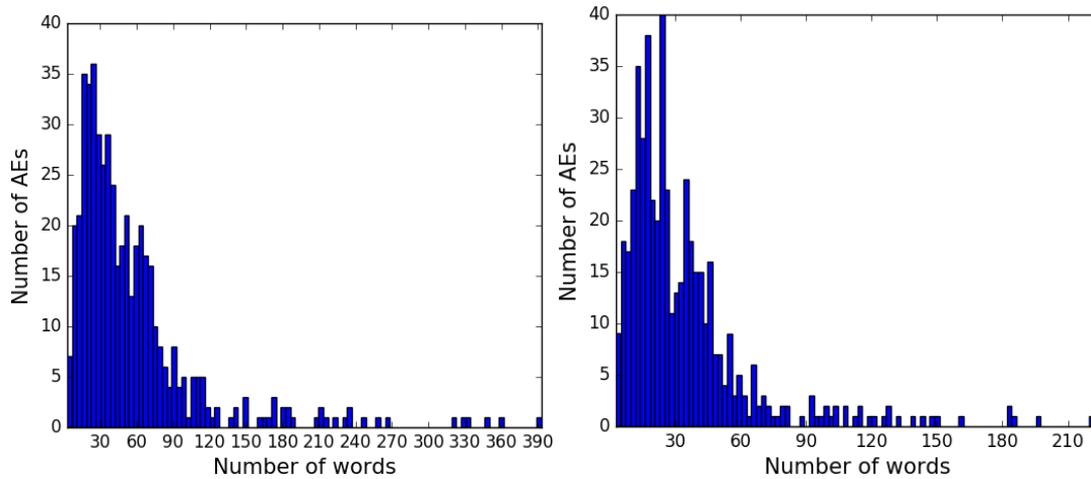


Figure 5.2 – Length distribution of 500 AE descriptions before and after removing stop words.

5.4.1 Data acquisition

We work with 500 AE descriptions collected from the management system of more than 30 social and medico-social structures during 3 years. Each AE case is an unstructured text description in French. An anonymization process was firstly performed for all events to protect users' privacy.

We perform some basic statistics to obtain an immediate insights of these AE data. First, we draw the distribution of description length, Figure 5.2 shows the length distribution of 500 AE descriptions before and after removing stop words where up to 35% of original descriptions have less than 30 words and 57% after removing stop words. This result indicates that most event descriptions are too short to describe the event in detail enough.

Then, we detect many spelling errors and special abbreviations in these texts. These small spelling mistakes will not cause much difficulty for reading but most NLP approaches do suffer from it. Some problematic AE examples are given in Table 5.3.

Finally, we draw the distribution of the top 20 unigrams and bigrams after removing stop words (Figure 5.3) to visualize the most prominent words utilized to describe these events. The top unigrams are mostly very regular and common words that can't present much information, but top bigrams do show some features of events such as "come work" and "absence employee" indicate that there are many cases of absence in these AEs.

Table 5.3 – Problematic AE examples.

Original description	English translation	Problem
Accident, l’infirmière partait au laboratoire pour identifier un bilan sanguin.	Accident, the nurse went to the laboratory to identify a blood test.	Difficult to understand due to the overly brief description.
L’intérimaire AS prévue demain matin ne viendra pas.	The interim AS scheduled for tomorrow morning will not come.	Unclear abbreviation “AS” and “prévue” is misspelled as “prevue”.
Appel de l’infirmière intérimaire pour signaler le non retour de permission de Mr T prévu ce jour.	Call from the temporary nurse to report the non-return of Mr T’s permission scheduled for this day.	“non” is misspelled as “nom”, which is also an existing french word but incorrect here.

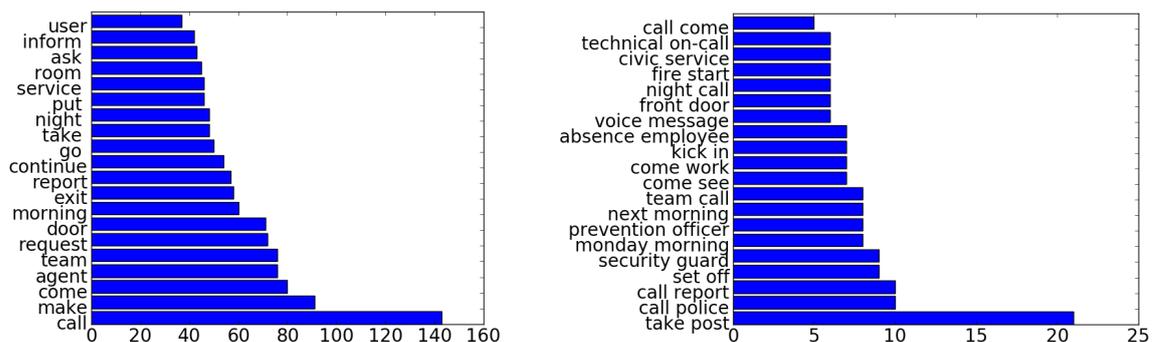


Figure 5.3 – The distribution of top 20 unigrams and bigrams of 500 AE reports after removing stop words.

5.4.2 Annotation

We review manually these 500 event descriptions to determine the adverse level of each case and label each qualified AE with a fact. Then the annotation results have been validated by an expert in the sector. Among these 500 cases, 112 cases with no adverse or potential adverse effect were identified as nonAE. As such, we qualify 388 validated AE cases and a class label (fact) is assigned to each of them. 36 different facts have been extracted among these AEs which can cover almost all situations that could happen in a social and medico-social structure. If an AE case includes several facts, the most serious one was selected (e.g., when there is “verbal aggression” and “physical aggression” within one event, the “physical aggression” will be selected).

Figure 5.4 shows the count distribution of these 388 AEs by 36 facts, facts have less than 5 samples are merged in the class "others". The distribution indicates the extreme class imbalance of events that some types of events happen frequently such as absence of staff, malfunctioning and verbal aggression and some classes have only one sample such as burglary, suicide, death of user, which are much more serious.

Since these AE reports are unstructured text descriptions, the task of selecting the fact of an AE can be formulated as a multiclass text classification problem. Based on these features, the preprocessing and classification methods are presented in following sections.

5.5 Methods

In this section, we present the proposed method for the auto-classification of AE. To classify AE cases, we first preprocess AE descriptions to extract features of AE with NLP techniques. Then we construct lexicon for facts based on the keywords extracted from AE descriptions and a dedicated terminology of synonyms. Finally, we calculate the event-fact similarity to determine the fact with several text similarity approaches. Figure 5.5 shows a general procedure of the classification method.

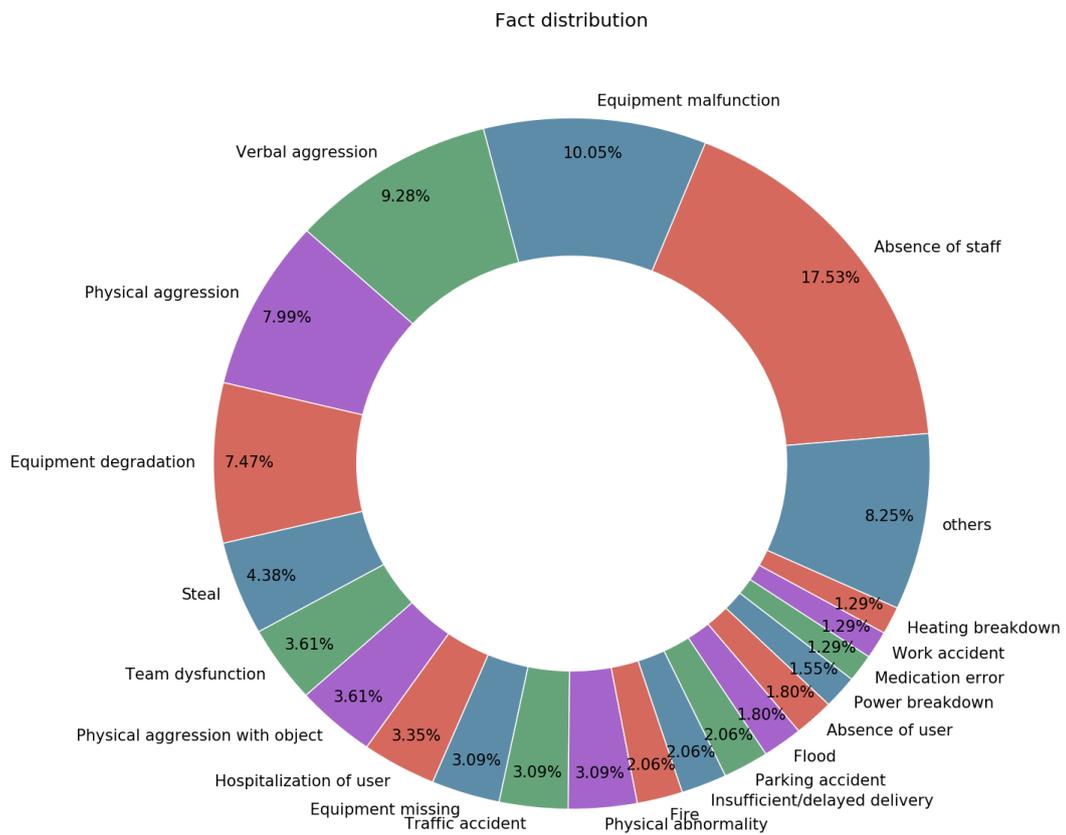


Figure 5.4 – Fact distribution of 388 AEs.

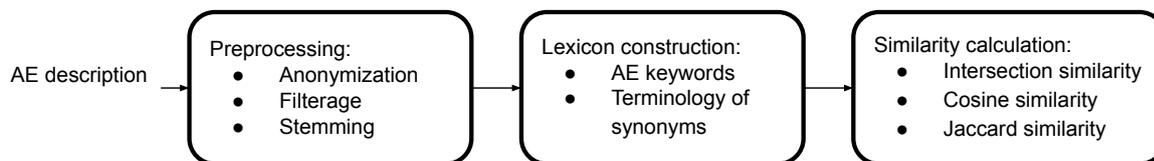


Figure 5.5 – General procedure of the automated AE classification.

5.5.1 Text preprocessing

In preparation for the classification, each unstructured text description is transformed into a word stem set with only the most significant words. We first apply the spelling correction of French on the text to correct spelling errors. Then, to further extract the informative features of the event, we remove all stop-words and keep only verbs, adjectives, adverbs and nouns in the text according to the POS tagging results. Finally, we apply word stemming on the rest of words. Figure 5.6 shows an example of the event description preprocessing. Text preprocessing was conducted in Python 3.6 with packages NLTK, spaCy and pyspellchecker.

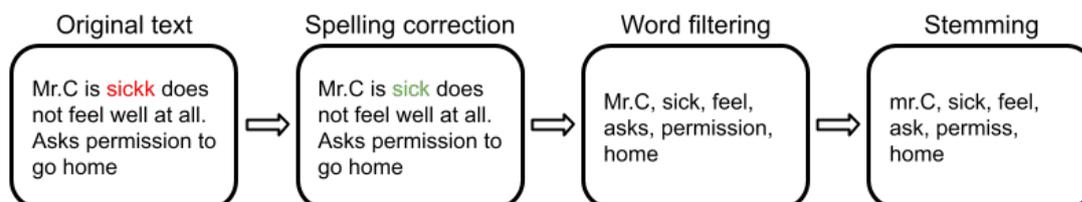


Figure 5.6 – An example of AE description preprocessing.

5.5.2 Keyword lexicon of fact

To the best of our knowledge, there is no public thesaurus or ontology in French in the field of AE in the SMSS. As the result, we establish a keyword lexicon for each fact as the basis of classification. We work in two directions:

1. The keywords collected from AE descriptions. According to the manual annotations validated by the expert (Section 5.4.2), we extract the keywords including informal

expressions and formal terms of each AE based on its fact. We obtain hence a list of keywords for each fact from the 388 classified AEs.

2. A dedicated terminology of synonyms specially constructed for AE in the SMSS. The purpose of this terminology is to 1) enrich the keyword list of facts extracted from real AE data in the first step; 2) provide a lexicon guide for reporters in the structures to improve their writing quality. We co-construct this terminology with an expert in the SMSS by collecting terms and words from three sources: official documents such as annual reports in the SMSS, practical professional writings in structures and the general French dictionary of synonyms, terms collected were later divided into groups of synonyms. Specifically, the terminology is composed with groups of synonyms including two parts: one term and several variants. Table 5.4 shows a part of the terminology. As shown in Table 5.4, words in a same group are considered as synonyms in the context of the AE in SMSS. Moreover, the term in each group is chosen as the representative of the group which means it is recommended to use in practice to describe an AE.

Table 5.4 – A part of the terminology of synonyms for AE in SMSS.

Term	Variants
aggression	assault, attack, violence, provocation, hit, slap
automobile	vehicle, car, tow truck
traffic	road, journey, highway, route
injury	bruise, wound, bleeding

To generate the final keyword lexicon, we combine these two word sets by extending the keyword list of facts collected from real AEs with the terminology of synonyms. Specifically, for each word in the keyword list for a given fact, this word and all its synonyms will be added in the keyword lexicon of the fact.

5.5.3 Event-fact similarity

The determination of fact for an AE is based on the value of the event-fact similarity. Various text similarity measures have been proposed and applied to solve text classification and clustering problems [Hua+08]. Given an AE e and a fact f , we state \mathcal{E} as the

preprocessed word set of event e (Section 5.5.1) and \mathcal{F} as the keyword lexicon of fact f (Section 5.5.2). We test three types of similarity:

- Intersection similarity:

$$S_I(e, f) = \frac{|\mathcal{E} \cap \mathcal{F}|}{|\mathcal{E}|}$$

This measures similarity as the the number of common words between \mathcal{E} and \mathcal{F} divided by the length of \mathcal{E} .

- Cosine similarity:

$$S_C(e, f) = \frac{\sum_{w \in \mathcal{E} \cup \mathcal{F}} Q_e(w) Q_f(w)}{\sqrt{(\sum_{w \in \mathcal{E} \cup \mathcal{F}} Q_e^2(w)) (\sum_{w \in \mathcal{E} \cup \mathcal{F}} Q_f^2(w))}}$$

where $Q_e(w)$ and $Q_f(w)$ is the occurrence of word w in \mathcal{E} and \mathcal{F} , this similarity measures the cosine of the angle between two vectors over \mathcal{E} and \mathcal{F} projected in a multi-dimensional space.

- Jaccard similarity:

$$S_J(e, f) = \frac{|\mathcal{E} \cap \mathcal{F}|}{|\mathcal{E} \cup \mathcal{F}|}$$

The Jaccard metric measures similarity as the size of intersection divided by the size of union of \mathcal{E} and \mathcal{F} .

5.6 Results

5.6.1 Lexicon construction results

A lexicon is constructed for each fact based on 388 AE cases, including informal expressions used by reporters within structures and formal terms extracted from official documents. We develop in addition a management system which allows structures to add, edit or delete terms in lexicons and terminology (see Table 5.5 for some examples).

Table 5.5 – A part of the lexicon for three facts of AE.

Fact	Lexicon
Absence of staff	cannot take, did not show up, illness, time off, not there, absence, individual, absent, not arrived, service, leave, not present, will not come, isolated, suffering...
Traffic accident	traffic, transport, drive, accident, shock, car, hooked up, vehicle, high speed, hit, fuel, road, highway, crash, tow truck...
Equipment malfunction	no longer works, malfunction, no access, alarm, out of service, damaged, blocked, badge forgetting, locked out, impossible to open, forgotten, settled, locked up, fault, degrade, broken down, not working...

We also test WordNet to extend the keywords of AEs. WordNet is a popular lexical database for English where words are grouped into sets of cognitive synonyms called synsets, the French version of WordNet is called the WOLF (Wordnet Libre du Français) [SF08]. However, we find the result constructed with WordNet is not appropriate due to several reasons. First, WordNet is a general lexical database where synonyms are divided by different concepts of the word, which leads to that the synonyms provided by WordNet are too general and not specific enough in the context of the SMSS. Second, WordNet contains only formal vocabulary words, where many particular expressions used in practice are ignored. Third, synonyms in WordNet are indexed with the lemmatisation of the word, which needs the POS tagging of the whole sentence and the lemmatisation of single keyword is very unreliable. As the result, a specific synonymy dictionary is necessary for our study.

5.6.2 Classification results

Evaluation metrics

For each AE to classify, we take up to five facts with the greatest non-zero similarity as the prediction result and consider classification as correct if the true fact is included in these proposals. We chose the multiple output prediction because many AEs include more than one facts, and the selection of the final fact is mostly a subjective choice according

to the actual situation, on the one hand; and it is more preferable that the decision is handled by the reporter instead of replacing them to decide, on the other hand.

In order to examine the performance of classifier, we apply a standard 5-fold cross-validation method following the common 80%-20% split ratio. We randomly divide the AE data into 5 subsets (folds) with same size. For each iteration, the k th fold ($k=1\dots5$) is used to be the testing fold to evaluate the performance of the classifier based on event-fact similarity and the other four training folds provide their keywords as the source to construct the keyword lexicon for facts. Note that due to the extreme class imbalance of data, not all classes (facts) are included in each fold. Considering the imbalanced class distribution, we use the weighted-average recall (weighted-R), which is always equal to the accuracy according to its definition, the weighted-average precision (weighted-P) and the weighted-average specificity (weighted-S) to evaluate the performance of the classifier. Note that the recall (accuracy) is preferable here since we require that the true fact can be found as possible to help reporters make their decision quickly, the other false predicted facts will not affect the decision-making a lot.

Experimental results

Table 5.6 shows the weighted-average recall (weighted-R), weighted-average precision (weighted-P), and weighted-average specificity (weighted-S) of the classifier with three different similarities (Section 5.5.3) on the testing fold and training folds for five iterations of the 5-fold cross-validation. The results show that the three similarities perform well on testing folds (84%–88%) and training folds (88%–92%) in terms of recall but they do struggle with the precision, which is inevitable due to the multiple prediction outputs. The Friedman test (p-value=0.45) indicates that there is no statistically significant difference between the results of three similarities for weighted-R.

Table 5.7 shows the performance of classifier with the intersection similarity of each class (fact) on testing folds with the name of fact, number of events (NE) of the fact, recall, precision and F1-score. First, we attempt to investigate the relation between the performance and the characteristic of fact. According to Table 5.7, we observe that certain facts do have a better result than the others such as hospitalization of user, fire, flood, etc, with the recall of 1.0 and also a good precision. This kind of facts usually have a very accurate and clear definition and distinct keywords which could help the identification. Inversely, facts that contains more various cases such as team dysfunction will bring more uncertainty to the classification which leads to a lower recall and precision. For this kind

Table 5.6 – Top 5 classification performance with three similarity methods on 5-fold cross-validation.

Similarity		Testing fold			Training folds		
		weighted-R	weighted-P	weighted-S	weighted-R	weighted-P	weighted-S
Intersection	Iteration 1	0.77	0.39	0.84	0.9	0.32	0.78
	Iteration 2	0.79	0.34	0.81	0.9	0.30	0.80
	Iteration 3	0.84	0.39	0.82	0.9	0.28	0.78
	Iteration 4	0.91	0.46	0.85	0.89	0.27	0.78
	Iteration 5	0.90	0.33	0.71	0.9	0.29	0.81
	Average	0.84	0.38	0.81	0.90	0.29	0.79
Cosine	Iteration 1	0.93	0.29	0.85	0.92	0.31	0.82
	Iteration 2	0.90	0.35	0.79	0.93	0.29	0.83
	Iteration 3	0.94	0.34	0.85	0.91	0.29	0.81
	Iteration 4	0.92	0.42	0.85	0.92	0.28	0.82
	Iteration 5	0.71	0.29	0.77	0.92	0.30	0.84
	Average	0.88	0.34	0.82	0.92	0.29	0.82
Jaccard	Iteration 1	0.91	0.31	0.85	0.87	0.31	0.83
	Iteration 2	0.91	0.37	0.79	0.87	0.29	0.84
	Iteration 3	0.86	0.33	0.84	0.88	0.29	0.83
	Iteration 4	0.87	0.42	0.86	0.88	0.28	0.83
	Iteration 5	0.85	0.35	0.80	0.89	0.30	0.84
	Average	0.88	0.36	0.83	0.88	0.29	0.83

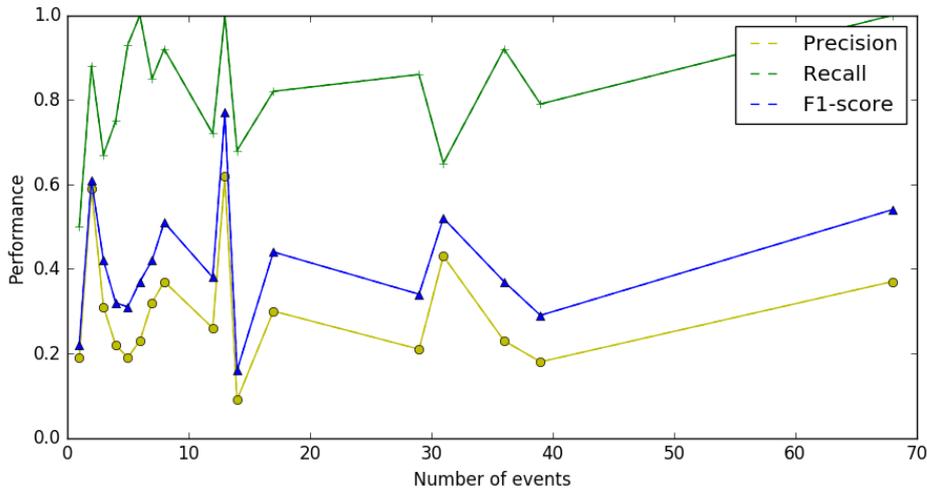


Figure 5.7 – Average performance of classifier of facts grouped by number of events.

of facts, dividing them into more detailed classes could be the solution. Then, we want to figure out whether the number of samples of fact influences the classification performance. For example, the facts with more events may lead to better performances. Figure 5.7 shows the average performance (precision, recall, f1-score) of facts grouped by the number of events. We observe that there is no obvious relation between the classification performance and the number of samples, except that some facts with just one sample have a precision and recall of zero, which is explainable because there is no knowledge existing to predict it. Apart from this, facts with limited samples can still achieve an effective prediction, which indicates the robustness of the classifier.

5.6.3 Knowledge visualization solution

In addition to the fact classifier, we propose two visualization charts for AE data as an additional decision-support tool to help the structures improve their AE analysis and complete their knowledge.

- AE graph (Figure 5.8). Each vertex in the graph represents an AE case and its color represents the fact that it belongs to, while the label of vertex is the keywords of the AE. Edge between two vertexes means that these two AEs have common keyword(s) and the edge width indicates this overlap size. This visualization of AEs provides first, an immediate vision of the distribution of AEs within structure, which allows

Table 5.7 – Classification performance with the intersection similarity of each fact on testing fold.

Fact	Number of events	Recall	Precision	F1-score
Absence of staff	68	1.0	0.37	0.54
Equipment malfunction	38	0.82	0.18	0.30
Verbal aggression	36	0.92	0.23	0.37
Physical aggression	31	0.65	0.43	0.52
Equipment degradation	29	0.86	0.21	0.34
Steal	17	0.82	0.30	0.44
Team dysfunction	15	0.47	0.07	0.12
Physical aggression with object	14	0.93	0.12	0.21
Hospitalization of user	13	1.0	0.62	0.77
Physical abnormality	12	0.42	0.17	0.24
Equipment missing	12	0.83	0.26	0.40
Traffic accident	12	0.92	0.34	0.50
Fire	8	1.0	0.62	0.77
Insufficient/delayed delivery	8	0.88	0.28	0.42
Parking accident	8	0.88	0.22	0.35
Absence of user	7	0.71	0.06	0.11
Flood	7	1.0	0.58	0.73
Power breakdown	6	1.0	0.23	0.37
Work accident	5	1.0	0.24	0.39
Heating breakdown	5	1.0	0.09	0.17
Medication error	5	0.80	0.25	0.38
Suicide attempts	4	0.75	0.38	0.50
Falling of user	4	1.0	0.24	0.39
Sexual aggression	4	0.50	0.03	0.06
Death of user	3	0.67	0.29	0.40
Discovery of weapon	3	0.67	0.33	0.44
Escape of user	2	1.0	0.67	0.80
Accidental exposure to blood	2	1.0	0.50	0.67
Burglary	2	1.0	0.18	0.31
Illegal activity	2	0.50	1.0	0.67
Unauthorized occupancy	1	0.0	0	0
Falling of material	1	1.0	0.11	0.20
Irregular introduction	1	0.0	0	0
Psychological abnormality	1	1.0	0.06	0.11
Harassment	1	1.0	1.0	1.0
Suicide	1	0.0	0	0

the structure to implement targeted measures for frequent facts; second, the possibility to correct and improve the existing knowledge. For example we can observe from Fig. 5.8 the vertex “mistaken, car, tow truck, fuel” has more connections with vertices of other facts instead of its own fact, which means that the annotation of this AE should be reverified.

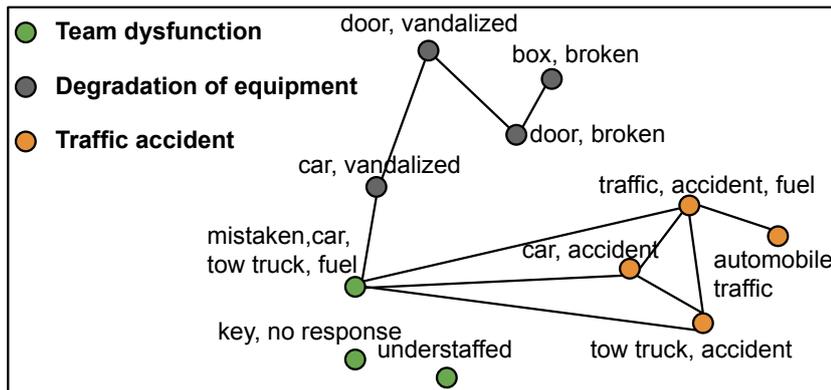


Figure 5.8 – A part of the AE graph.

- Keyword graph (Fig. 5.9). Each vertex in the graph represents a keyword used in AE descriptions, its color indicates the main fact it belongs to and the size of vertex represents its occurrence, the edge between two vertexes means that these two keywords are used together in an AE and the edge width indicates this frequency. This graph gives a direct vision about the vocabulary usage within structure, from that we can observe the evolution of writing quality of reporters in the structure.

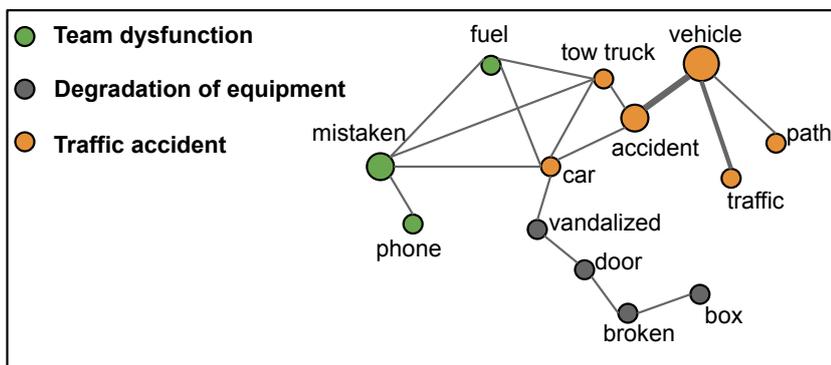


Figure 5.9 – A part of the keyword graph.

5.7 Analysis and discussion

5.7.1 Error analysis

We randomly select 20 prediction failure cases to determine the primary reason of error for the classification. The largest error source (70%) comes from novel AE cases or terms that are not contained in the fact keyword lexicons, especially for the rare facts with very few samples. This problem can be solved by enriching the fact keyword lexicons with more cases and terms gradually, since the data source that we use is really insufficient for most facts. Moreover, this kind of error is also caused by the use of slang words by reporters, which requires more background knowledge to correctly understand them.

The second error source (15%) is the multiple facts case in one AE. These event cases are usually described in more detail with several facts within one description. As the result, the facts with a larger keyword lexicon are more probably to be predicted and the true fact is excluded from the five most possible facts. This problem can be handled by balancing the size of keyword lexicons of facts, or giving a predefined factor to each fact according to their importance.

Even if we apply the spelling correction in the preprocessing phase, there are still errors (10%) caused by spelling errors. The reason is that the misspelled word happens to be another correct word that the corrector can't detect it. This kind of spelling errors are more difficult to identify automatically which requires the improvement of writing quality of reports.

Interestingly, there is one error case (5%) which makes us reconsider the result of our manual classification. Indeed, some AEs are difficult to determine which fact should be selected and in this case, the selection depends fully on the subjective choice of the reporter, considering the actual situation of the structure. Anyhow, the proposals of facts by the classifier bring a positive affect to the decision-making of reporters by providing them more possibilities before making their final decision.

5.7.2 Practical contributions

This work has responded to the urgent need for research in the SMSS by providing decision support solutions for AE report and analysis. The predictive classifier helps reporters select efficiently and correctly the fact of event during documenting by proposing the most related choices which can greatly improve the metadata missing situation. Even if there

are still reporting errors, the visualization tools allow structures reverify and reconsider their past decisions and existing knowledge and then implement corresponding measures to correct. The predictive classifier and the knowledge visualization tool presented in this work serve as a decision support module of the MSUsager software.

5.7.3 Limitations

There are some limitations in the current work. First, we did not consider the multi-words terms in the preprocessing phase. We have tested some term extraction tools however the results given are very inaccurate. In fact, some phrases are used in AE descriptions which also represent important information (e.g., “can’t come” is equal to “absent”). Hence the classification results will be more accurate by adding these informative elements. Second, due to the limited amount of data, we did not experiment the supervised learning methods to solve this classification task like previous studies. In future work we intend to experiment the machine learning algorithms when a considerable amount of data is collected. Finally, the lexicons of facts can be further improved and enriched with the management software which provides interfaces to add and edit terms in lexicons.

5.8 Conclusion

This chapter presented a solution for the adverse event (AE) report classification in the SMSS. AE reports are very important information source to follow the personalized project of users, however, structures are struggling to conduct an efficient analysis of AEs due to the abysmal quality of event reports. This work aims to objectify the information of an AE report by proposing the most related facts according to the analysis of the event description.

- The proposed lexicon-based classification method is based on keywords lexicons which are constructed with real AE reports and a terminology of synonyms. Compared with traditional lexicon construction methods, this method does not need large corpus resources to generate lexicons. The constructed lexicons also serve as a reference to improve the writing quality of reporters.
- We tested our classification method on 388 real AE reports of the SMSS and achieved an accuracy up to 88%, in addition, since the classification method doesn’t need learning process, the predicted facts can be quickly proposed. This predictive classi-

fier helps reporters select efficiently and correctly the fact of event during documenting by proposing the most related choices which can greatly improve the metadata missing situation.

- We developed two graphs as the knowledge visualization solution as an additional decision-support tool. This visualization solution gives a direct vision about the AEs and vocabulary usage within the structure, which allows structures reverify and reconsider their past decisions and existing knowledge and then implement corresponding measures to correct.

GENERAL CONCLUSION

Conclusions

This thesis focuses on a practical problem, the personalized user project implementation in the social and medico-social sector. Since the law of January 2, 2002 which modernizes the operation and evaluation of the sector, the implementation of personalized user project become an obligation and an important evaluation indicator for all social and medico-social structures. The personalized user project includes all the professional, social and medico-social supports for the user according to his/her needs, expectations and resources. However, the elaboration and the implementation of personalized user projects are complex. The characteristic of population received forces the structures to adapt their activities to each user profile and, consequently, to create numerous and varied activities requiring different types of materials and human resources.

In Chapter 2, we presented the process to implement the personalized user project and the current practical difficulties. To deal with these difficulties, we modeled the personalized user project planning as a constraint optimization problem. The main components of the problem include a set of users, a set of activities, a set resources and a set of timeslots. We represented the candidate solution by two matrices which represent the activity plan of users and resources during a given number of timeslots. We also provided the formulation of the constraints to respect and the objectives to maximize.

In Chapter 3, we proposed solution methods to solve the user project planning problem defined in Chapter 2. We first tested the mathematical programming approach with CPLEX, we created an integer programming model where the decision variables take their values in an integer interval representing the activity and a 0/1 programming model where the decision variables take binary values. Another approach that we proposed is a greedy heuristic algorithm, which generates always a feasible solution satisfying all constraints with also good quality. For the experimentation of these approaches, we generated 20 simulated test instances of different sizes and tunable characteristics. We sent a questionnaire survey to more than 3600 social and medico-social structures and received more than 200 responses. These instances were generated according to the survey results to fit as much as

possible real situations. On these 20 test instances, The 0/1 programming model showed a better performance compared to the integer programming model but still failed on 8 instances, the constructive greedy procedure generated feasible solutions in a very short time and performed better than the mathematical programming models on middle and large size instances.

In Chapter 4, we explored an effective multi-neighborhood simulated annealing (MNSA) algorithm for the personalized user project planning problem. Based on the simulated annealing framework, MNSA explores three dedicated neighborhood structures in a probabilistic way. These three neighborhoods are inspired by real planning operations, we also proved the advantage of combining these three neighborhoods compared with each single one. In a given acceptable time, MNSA found a better solution for 7 instances against 4 cases in favor of the CPLEX and obviously improved the result of the constructive greedy procedure.

In Chapter 5, we presented a solution for the social and medico-social adverse event report classification and analysis. Adverse events reports are very important information sources to follow the personalized project of users, however, structures are struggling to conduct an efficient analysis of adverse events due to the abysmal quality of event reports. We proposed a predictive classifier which could help reporters select efficiently and correctly the fact of event during documenting by proposing the most related facts, based on keywords lexicons which are constructed with real adverse event reports and a terminology of synonyms. We tested on 388 real adverse event reports and achieved an accuracy up to 88%. In addition, we developed two graphs as the knowledge visualization solution as an additional decision-support tool, which allows structures reverify and reconsider their past decisions and improve their existing knowledge.

Perspectives

From the work presented in this thesis, several interesting perspectives can be identified for future research.

- To solve the user project planning problem, we used a weighted sum method to combine the three objectives. An interesting possibility would be to consider the multiobjective optimization approach to obtain multiple alternative non-dominated solutions from which the decision-maker can choose the most suitable solutions.
- The definition of the model could be improved or adjusted. The activities in the

proposed model are considered to be independent. It would be interesting to study possible relationships within the activities. Such an extension could be useful in structures where such situations can be encountered, meanwhile, other special constraint and objective can be considered according to the actual needs.

- We provide a automatic planning solution for the user project conception phase. An interesting future study would be to investigate the dynamic user project planning problem for the execution phase. Indeed, even though the project planning is validated before execution, unexpected changes happen frequently such as last-minute unavailability of resources or professionals, activity interruption caused by an adverse event, additional user needs, etc. All these events would lead to the need of adjustments of users' projects. For these situations, a new feasible planning with the smallest changes as possible is needed quickly. As a result, a fast schedule adjustment mechanism could be highly relevant to enrich the decision support system.
- For the adverse event report classification presented in Chapter 5, we applied the lexicon-based classification method due to the lack of data. When a considerable amount of labeled data is collected, the experimentation of more sophisticated machine learning and deep learning methods could be of interest.

LIST OF FIGURES

2.1	The three main phases of the personalized user project in SMSS.	30
2.2	Solution example of the user project planning problem.	33
3.1	Summary of the computational results by the number of users.	57
3.2	Characteristics of instances and the gap of the 0/1 programming model. . .	58
4.1	Single activity changes.	65
4.2	Activity swap.	66
4.3	Activity exchanges.	67
4.4	Comparison summary of MNSA with 0/1 mathematical programming with CPLEX (0/1P), greedy initialization procedure (GIP) and variable neigh- borhood search (VNS).	71
4.5	Performance profiles of the 0/1 programming model, greedy initialization procedure, MNSA average results, MNSA best results and variable neigh- borhood search.	73
4.6	Comparison of four neighborhood combinations.	74
5.1	AE treatment procedure in SMSS.	80
5.2	Length distribution of 500 AE descriptions before and after removing stop words.	84
5.3	The distribution of top 20 unigrams and bigrams of 500 AE reports after removing stop words.	85
5.4	Fact distribution of 388 AEs.	87
5.5	General procedure of the automated AE classification.	88
5.6	An example of AE description preprocessing.	88
5.7	Average performance of classifier of facts grouped by number of events. . .	94
5.8	A part of the AE graph.	96
5.9	A part of the keyword graph.	96

LIST OF TABLES

1.1	The main structures and services of the SMSS.	20
1.2	Representative professions work for the SMSS.	22
2.1	Settings of important parameters of the user project planning problem. . .	34
3.1	Parameters of the test instances.	52
3.2	Experimental results and comparison of the integer model, 0/1 model and greedy heuristic algorithm.	56
4.1	Settings of MNSA's parameters.	69
4.2	Computational results of the proposed MNSA algorithm and comparison with the reference approaches: 0/1 mathematical programming with CPLEX (0/1P), greedy initialization procedure (GIP) and variable neighborhood search (VNS).	71
4.3	Computational results (f_{best}) of MNSA with the combined neighborhoods ($N_1 + N_2 + N_3$) and each individual neighborhood: "single activity change" neighborhood (N_1), "activity swap" neighborhood (N_2), "activity exchange" neighborhood (N_3).	74
4.4	Parameter levels for the 2-level full factorial experiment of MNSA.	75
4.5	Results of the analysis of variance of the 2-level full factorial experiment of MNSA.	76
5.1	AE examples and their fact.	81
5.2	Summary of related studies on AE classification and extraction.	82
5.3	Problematic AE examples.	85
5.4	A part of the terminology of synonyms for AE in SMSS.	89
5.5	A part of the lexicon for three facts of AE.	91
5.6	Top 5 classification performance with three similarity methods on 5-fold cross-validation.	93

5.7	Classification performance with the intersection similarity of each fact on testing fold.	95
-----	---	----

LIST OF ALGORITHMS

1	Constructive greedy heuristic algorithm.	50
2	Multi-neighborhood simulated annealing (MNSA).	63

LIST OF PUBLICATIONS

Published journal papers

- Yinuo Li, Jin-Kao Hao, and Brahim Chabane. User project planning in social and medico-social sector: Models and solution methods. *Expert Systems with Applications* 173: 114684, 2021.

Published conference papers

- Yinuo Li, Touria Ait El Mekki, Jin-Kao Hao. Adverse event report classification in social and medico-social sector. *Proceedings of the 6th IEEE Conference on ICT Solutions for eHealth (ICTS4eHealth)*, Athens, Greece, September 5-8, 2021.

Submitted journal papers

- Yinuo Li, Jin-Kao Hao. Multi-neighborhood simulated annealing for a practical user project planning in social and medico-social sector. Submitted to *Applied Soft Computing*, Under review, May, 2020.

REFERENCES

- [All88] James Allen, *Natural language understanding*, Benjamin-Cummings Publishing Co., Inc., 1988 (cit. on p. 83).
- [Amb+20] Maxime Amblard, Chloé Braud, Chuyuan Li, Caroline Demily, Nicolas Franck, and Michel Musiol, « Investigation par méthodes d'apprentissage des spécificités langagières propres aux personnes avec schizophrénie », *in*: ATALA; AFCP, 2020, pp. 12–26 (cit. on p. 83).
- [And19] Mosquera Varela Anderson Fabian, « Approche territoriale d'une gestion dynamique et multidimensionnelle de l'orientation des personnes en situation de handicap: D'une orientation handicap-établissement à une orientation besoins-prestations », Doctoral Thesis, Lyon, France: Institut National des Sciences Appliquées de Lyon, 2019 (cit. on p. 23).
- [ANE08] ANESM, *Les attentes de la personne et le projet personnalisé*, tech. rep., Agence Nationale de l'Evaluation et de la qualité des Etablissements et Services Sociaux et Médico-sociaux, 2008 (cit. on p. 29).
- [APE16] APEC, *Les métiers du sanitaire, social et médico-social*, tech. rep., Association pour l'emploi des cadres, 2016 (cit. on pp. 19, 20).
- [AF05] Jeffrey K Aronson and Robin E Ferner, « Clarification of terminology in drug safety », *in*: *Drug Safety* 28.10 (2005), pp. 851–870 (cit. on p. 15).
- [17] *Arrêté du 27 février 2017 fixant la liste des catégories d'événements sanitaires indésirables pour lesquels la déclaration ou le signalement peut s'effectuer au moyen du portail de signalement des événements sanitaires indésirables*, <https://www.legifrance.gouv.fr/loda/id/JORFTEXT000034150863/>, February 27, 2017 (cit. on p. 79).
- [BT96] Donald P Ballou and Giri Kumar Tayi, « A decision aid for the selection and scheduling of software maintenance projects », *in*: *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans* 26.2 (1996), pp. 203–212 (cit. on p. 37).

-
- [Ban+14] Anil Bandhakavi, Nirmalie Wiratunga, P Deepak, and Stewart Massie, « Generating a word-emotion lexicon from #emotional tweets », *in: Proceedings of the 3rd Joint Conference on Lexical and Computational Semantics*, 2014, pp. 12–21 (cit. on p. 83).
- [Ban+17] Anil Bandhakavi, Nirmalie Wiratunga, Deepak Padmanabhan, and Stewart Massie, « Lexicon based feature extraction for emotion text classification », *in: Pattern Recognition Letters* 93 (2017), pp. 133–142 (cit. on p. 83).
- [Bas00] Matt Bassett, « Assigning projects to optimize the utilization of employees' time and expertise », *in: Computers & Chemical Engineering* 24.2-7 (2000), pp. 1013–1021 (cit. on p. 37).
- [BN07] Odile Bellenguez-Morineau and Emmanuel Néron, « A branch-and-bound method for solving multi-skill project scheduling problem », *in: RAIRO-Operations Research* 41.2 (2007), pp. 155–170 (cit. on p. 38).
- [Ber+10] Thilo Bertsche, Johannes Pfaff, Petra Schiller, Jens Kaltschmidt, Markus G Pruszydlo, Wolfgang Stremmel, Ingeborg Walter-Sack, Walter E Haefeli, and Jens Encke, « Prevention of adverse drug reactions in intensive care patients by personal intervention based on an electronic clinical decision support system », *in: Intensive Care Medicine* 36.4 (2010), pp. 665–672 (cit. on p. 15).
- [BG19] Elise Bigeard and Natalia Grabar, « Détecter la non-adhérence médicamenteuse dans les forums de discussion avec les méthodes de recherche d'information », *in: TALN 2019*, 2019 (cit. on p. 83).
- [Bot+11] Taxiarchis Botsis, Michael D Nguyen, Emily Jane Woo, Marianthi Markatou, and Robert Ball, « Text mining for the Vaccine Adverse Event Reporting System: medical text classification using informative feature selection », *in: Journal of the American Medical Informatics Association* 18.5 (2011), pp. 631–638 (cit. on p. 82).
- [Bru+99] Peter Brucker, Andreas Drexler, Rolf Möhring, Klaus Neumann, and Erwin Pesch, « Resource-constrained project scheduling: Notation, classification, models, and methods », *in: European Journal of Operational Research* 112.1 (1999), pp. 3–41 (cit. on p. 37).

-
- [CED12] CEDIS, *Guide d'Elaboration du Projet d'Accompagnement Personnalisé*, tech. rep., Le Comité Européen pour le Développement de l'Intégration Sociale, 2012 (cit. on pp. 29, 30).
- [CBH17] Brahim Chabane, Matthieu Basseur, and Jin-Kao Hao, « R2-IBMOLS applied to a practical case of the multiobjective knapsack problem », *in: Expert Systems with Applications* 71 (2017), pp. 457–468 (cit. on p. 23).
- [Chi+08] Kwai-Sang Chin, Dong-ling Xu, Jian-Bo Yang, and James Ping-Kit Lam, « Group-based ER-AHP system for product project screening », *in: Expert Systems with Applications* 35.4 (2008), pp. 1909–1929 (cit. on p. 37).
- [DD19] Morteza Davari and Erik Demeulemeester, « The proactive and reactive resource-constrained project scheduling problem », *in: Journal of Scheduling* 22.2 (2019), pp. 211–237 (cit. on pp. 37, 38, 41).
- [DL98] Keith Decker and Jinjiang Li, « Coordinated hospital patient scheduling », *in: Proceedings of the International Conference on Multi Agent Systems (Cat. No. 98EX160)*, IEEE, 1998, pp. 104–111 (cit. on pp. 37, 41).
- [07] *Décret No.2007-975 du 15 mai 2007 fixant le contenu du cahier des charges pour l'évaluation des activités et de la qualité des prestations des établissements et services sociaux et médico-sociaux*, <https://www.legifrance.gouv.fr/loda/id/JORFTEXT000000822015/>, May 15, 2007 (cit. on p. 29).
- [16a] *Décret No.2016-1606 du 25 novembre 2016 relatif à la déclaration des événements indésirables graves associés à des soins et aux structures régionales d'appui à la qualité des soins et à la sécurité des patients*, <https://www.legifrance.gouv.fr/loda/id/JORFTEXT000033479591/>, November 25, 2016 (cit. on p. 79).
- [16b] *Décret No.2016-1813 du 21 décembre 2016 relatif à l'obligation de signalement des structures sociales et médico-sociales*, <https://www.legifrance.gouv.fr/loda/id/JORFTEXT000033670880/>, December 21, 2016 (cit. on p. 79).
- [99] *Décret No.99-316 du 26 avril 1999 relatif aux modalités de tarification et de financement des établissements hébergeant des personnes âgées dépendantes*,

<https://www.legifrance.gouv.fr/loda/id/JORFTEXT000000211160/>,
April 26, 1999 (cit. on p. 29).

- [Dev+17] Shantanu Dev, Shinan Zhang, Joseph Voyles, and Anand S Rao, « Automated classification of adverse events in pharmacovigilance », *in: 2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, IEEE, 2017, pp. 905–909 (cit. on p. 82).
- [Dir12] statistiques et Prévisions du MASS Direction des études, *Le champ social et médico-social : une activité en forte croissance, des métiers qui se développent et se diversifient*, tech. rep., Ministère des Solidarités et de la Santé, 2012 (cit. on pp. 21, 22).
- [DM02] Elizabeth D Dolan and Jorge J Moré, « Benchmarking optimization software with performance profiles », *in: Mathematical Programming* 91.2 (2002), pp. 201–213 (cit. on p. 70).
- [EP04] Mario Enea and Tommaso Piazza, « Project selection by constrained fuzzy AHP », *in: Fuzzy Optimization and Decision Making* 3.1 (2004), pp. 39–62 (cit. on p. 37).
- [GR06] Oleg Grodzevich and Oleksandr Romanko, « Normalization and other topics in multi-objective optimization », *in: Proceedings of the Fields-MITACS Industrial Problems Workshop*, 2006, pp. 89–101 (cit. on p. 54).
- [Gut+08] Walter J Gutjahr, Stefan Katzensteiner, Peter Reiter, Christian Stummer, and Michaela Denk, « Competence-driven project portfolio selection, scheduling and staff assignment », *in: Central European Journal of Operations Research* 16.3 (2008), pp. 281–306 (cit. on p. 38).
- [HBS18] Farhad Habibi, Farnaz Barzinpour, and Seyed Sadjadi, « Resource-constrained project scheduling problem: review of past and recent developments », *in: Journal of Project Management* 3.2 (2018), pp. 55–88 (cit. on p. 37).
- [Han+17] Lichy Han, Robert Ball, Carol A Pamer, Russ B Altman, and Scott Proestel, « Development of an automated assessment tool for MedWatch reports in the FDA adverse event reporting system », *in: Journal of the American Medical Informatics Association* 24.5 (2017), pp. 913–920 (cit. on p. 82).

-
- [Han+19] Summer S Han, Tej D Azad, Paola A Suarez, and John K Ratliff, « A machine learning approach for predictive models of adverse events following spine surgery », *in: The Spine Journal* 19.11 (2019), pp. 1772–1781 (cit. on p. 82).
- [HM01] Pierre Hansen and Nenad Mladenović, « Variable neighborhood search: Principles and applications », *in: European Journal of Operational Research* 130.3 (2001), pp. 449–467 (cit. on p. 69).
- [AIH01] Kamal M Al-Subhi Al-Harbi, « Application of the AHP in project management », *in: International Journal of Project Management* 19.1 (2001), pp. 19–27 (cit. on p. 37).
- [HB10] Sönke Hartmann and Dirk Briskorn, « A survey of variants and extensions of the resource-constrained project scheduling problem », *in: European Journal of Operational Research* 207.1 (2010), pp. 1–14 (cit. on pp. 37, 41).
- [HAS20] HAS, *Retour d’expérience sur les événements indésirables graves associés a des soins rapport annuel 2019*, tech. rep., Haute Autorité de Santé, 2020 (cit. on p. 80).
- [HK10] Christian Heimerl and Rainer Kolisch, « Scheduling and staffing multiple projects with a multi-skilled workforce », *in: OR Spectrum* 32.2 (2010), pp. 343–368 (cit. on pp. 37, 38, 41).
- [HL05] Willy Herroelen and Roel Leus, « Project scheduling under uncertainty: Survey and research potentials », *in: European Journal of Operational Research* 165.2 (2005), pp. 289–306 (cit. on p. 37).
- [Hua+08] Anna Huang et al., « Similarity measures for text document clustering », *in: Proceedings of the sixth new zealand computer science research student conference (NZCSRSC2008)*, Christchurch, New Zealand, vol. 4, 2008, pp. 9–56 (cit. on p. 89).
- [KNA13] Yazdan Khoshjahan, Amir Abbas Najafi, and Behrouz Afshar-Nadjafi, « Resource constrained project scheduling problem with discounted earliness–tardiness penalties: Mathematical modeling and solving procedure », *in: Computers & Industrial Engineering* 66.2 (2013), pp. 293–300 (cit. on p. 39).
- [KGV83] Scott Kirkpatrick, C Daniel Gelatt, and Mario P Vecchi, « Optimization by simulated annealing », *in: Science* 220.4598 (1983), pp. 671–680 (cit. on p. 62).

-
- [Kos15] Zsolt T Kosztyán, « Exact algorithm for matrix-based project planning problems », *in: Expert Systems with Applications* 42.9 (2015), pp. 4460–4473 (cit. on p. 38).
- [Kos20] Zsolt T Kosztyán, « An exact algorithm for the flexible multilevel project scheduling problem », *in: Expert Systems with Applications* 158 (2020), p. 113485 (cit. on p. 38).
- [KS20] Zsolt T Kosztyán and István Szalkai, « Multimode resource-constrained project scheduling in flexible projects », *in: Journal of Global Optimization* 76.1 (2020), pp. 211–241 (cit. on p. 38).
- [KD12] Henri Kromm and Yves Ducq, « Reference decision models in the medico-social service sector », *in: IFIP International Conference on Advances in Production Management Systems*, Springer, 2012, pp. 422–429 (cit. on p. 23).
- [Lal+19] Fréjus AA Laleye, Gaël de Chalendar, Antoine Brouquet, Antonia Blanié, and Dan Benhamou, « Hybridation d’un agent conversationnel avec des plongements lexicaux pour la formation au diagnostic médical (Hybridization of a conversational agent with word embeddings for medical diagnostic training) », *in: Actes de la Conférence sur le Traitement Automatique des Langues Naturelles (TALN) PFIA 2019. Volume II: Articles courts*, 2019, pp. 313–322 (cit. on p. 83).
- [LH07] Benoît Laurent and Jin-Kao Hao, « Simultaneous vehicle and driver scheduling: A case study in a limousine rental company », *in: Computers & Industrial Engineering* 53.3 (2007), pp. 542–558 (cit. on pp. 37, 49).
- [Lea+10] Robert Leaman, Laura Wojtulewicz, Ryan Sullivan, Annie Skariah, Jian Yang, and Graciela Gonzalez, « Towards internet-age pharmacovigilance: extracting adverse drug reactions from user posts in health-related social networks », *in: Proceedings of the 2010 workshop on biomedical natural language processing*, 2010, pp. 117–125 (cit. on pp. 82, 83).
- [Lee+17] Kathy Lee, Ashequl Qadir, Sadid A Hasan, Vivek Datla, Aaditya Prakash, Joey Liu, and Oladimeji Farri, « Adverse drug event detection in tweets with semi-supervised convolutional neural networks », *in: Proceedings of the 26th international conference on world wide web*, 2017, pp. 705–714 (cit. on p. 82).

-
- [LHC21] Yinuo Li, Jin-Kao Hao, and Brahim Chabane, « User project planning in social and medico-social sector: Models and solution methods », *in: Expert Systems with Applications* 173 (2021), p. 114684 (cit. on p. 11).
- [LCW14] Shixin Liu, Di Chen, and Yifan Wang, « Memetic algorithm for multi-mode resource-constrained project scheduling problems », *in: Journal of Systems Engineering and Electronics* 25.4 (2014), pp. 609–617 (cit. on p. 38).
- [02] *Loi No.2002-2 du 2 janvier 2002 rénovant l'action sociale et médico-sociale*, <https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000000215460/>, January 2, 2002 (cit. on pp. 17, 29).
- [04a] *Loi No.2004-626 du 30 juin 2004 relative à la solidarité pour l'autonomie des personnes âgées et des personnes handicapées*, <https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000000622485/>, June 30, 2004 (cit. on p. 17).
- [04b] *Loi No.2004-810 du 13 août 2004 relative à l'assurance maladie*, <https://www.legifrance.gouv.fr/loda/id/JORFTEXT000000625158/>, August 13, 2004 (cit. on p. 18).
- [09] *Loi No.2005-102 du 11 février 2005 pour l'égalité des droits et des chances, la participation et la citoyenneté des personnes handicapées*, <https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000000809647/>, July 21, 2009 (cit. on p. 29).
- [05] *Loi No.2009-879 du 21 juillet 2009 portant réforme de l'hôpital et relative aux patients, à la santé et aux territoires*, <https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000020879475/>, February 11, 2005 (cit. on p. 18).
- [70] *Loi No.70-1318 du 31 décembre 1970 portant réforme hospitalière*, <https://www.legifrance.gouv.fr/loda/id/JORFTEXT000000874228/>, December 31, 1970 (cit. on p. 17).
- [75a] *Loi No.75-534 du 30 juin 1975 d'orientation en faveur des personnes handicapées*, <https://www.legifrance.gouv.fr/loda/id/JORFTEXT000000333976/>, June 30, 1975 (cit. on p. 17).

-
- [75b] *Loi No.75-535 du 30 juin 1975 relative aux institutions sociales et médico-sociales*,
<https://www.legifrance.gouv.fr/loda/id/JORFTEXT000000699217/>,
June 30, 1975 (cit. on p. 17).
- [Lóp+16] Manuel López-Ibáñez, Jérémie Dubois-Lacoste, Leslie Pérez Cáceres, Mauro Birattari, and Thomas Stützle, « The irace package: Iterated racing for automatic algorithm configuration », in: *Operations Research Perspectives* 3 (2016), pp. 43–58 (cit. on p. 69).
- [LH10] Zhipeng Lü and Jin-Kao Hao, « Adaptive tabu search for course timetabling », in: *European Journal of Operational Research* 200.1 (2010), pp. 235–244 (cit. on pp. 37, 41, 49).
- [LH12] Zhipeng Lü and Jin-Kao Hao, « Adaptive neighborhood search for nurse rostering », in: *European Journal of Operational Research* 218.3 (2012), pp. 865–876 (cit. on pp. 37, 41).
- [LHG11] Zhipeng Lü, Jin-Kao Hao, and Fred W. Glover, « Neighborhood analysis: a case study on curriculum-based course timetabling », in: *Journal of Heuristics* 17.2 (2011), pp. 97–118 (cit. on p. 68).
- [MG16] Santanu Mandal and Sumit Gupta, « A Lexicon-based text classification model to analyse and predict sentiments from online reviews », in: *2016 International Conference on Computer, Electrical & Communication Engineering*, IEEE, 2016, pp. 1–7 (cit. on p. 83).
- [MA04] R Timothy Marler and Jasbir S Arora, « Survey of multi-objective optimization methods for engineering », in: *Structural and Multidisciplinary Optimization* 26.6 (2004), pp. 369–395 (cit. on p. 54).
- [MA05] R Timothy Marler and Jasbir S Arora, « Function-transformation methods for multi-objective optimization », in: *Engineering Optimization* 37.6 (2005), pp. 551–570 (cit. on p. 54).
- [McL07] Saul McLeod, « Maslow’s hierarchy of needs », in: *Simply Psychology* 1 (2007), pp. 1–8.

-
- [MH05] Genevieve B Melton and George Hripcsak, « Automated detection of adverse events using natural language processing of discharge summaries », *in: Journal of the American Medical Informatics Association* 12.4 (2005), pp. 448–457 (cit. on p. 83).
- [Mén+19] Timothé Ménard, Yves Barmaz, Björn Koneswarakantha, Rich Bowling, and Leszek Popko, « Enabling data-driven clinical quality assurance: predicting adverse event reporting in clinical trials using machine learning », *in: Drug safety* 42.9 (2019), pp. 1045–1053 (cit. on p. 82).
- [Mon17] Douglas C Montgomery, *Design and analysis of experiments*, John Wiley & Sons, 2017 (cit. on p. 75).
- [Mon15] Geovanny Osorio Montoya, « Conception des outils pour le suivi des activités et l’aide au pilotage dans le secteur médico-social », Doctoral Thesis, Lyon, France: Institut National des Sciences Appliquées de Lyon, 2015 (cit. on p. 23).
- [NS07] David Nadeau and Satoshi Sekine, « A survey of named entity recognition and classification », *in: Linguisticae Investigationes* 30.1 (2007), pp. 3–26 (cit. on p. 83).
- [Oso+17] Geovanny Osorio, Thibaud Monteiro, Lorraine Trilling, and Frédéric Albert, « Multi-criteria assignment policies to improve global effectiveness of medico-social service sector », *in: Engineering applications of artificial intelligence* 61 (2017), pp. 21–34 (cit. on p. 23).
- [Pan+19] Abhishek Pandey, Kory Kreimeyer, Matthew Foster, Oanh Dang, Thomas Ly, Wei Wang, Richard Forshee, and Taxiarchis Botsis, « Adverse event extraction from structured product labels using the event-based text-mining of health electronic records (ETHER) system », *in: Health informatics journal* 25.4 (2019), pp. 1232–1243 (cit. on pp. 82, 83).
- [PJ11] Sepehr Proon and Mingzhou Jin, « A genetic algorithm with neighborhood search for the resource-constrained project scheduling problem », *in: Naval Research Logistics (NRL)* 58.2 (2011), pp. 73–82.
- [Pro+01] KA Proos, GP Steven, OM Querin, and YM Xie, « Multicriterion evolutionary structural optimization using the weighting and the global criterion methods », *in: AIAA Journal* 39.10 (2001), pp. 2006–2012 (cit. on p. 54).

-
- [Ram+14] Balaji Polepalli Ramesh, Steven M Belknap, Zuofeng Li, Nadya Frid, Dennis P West, and Hong Yu, « Automatically recognizing medication and adverse event information from food and drug administration’s adverse event reporting system narratives », *in: JMIR medical informatics 2.1* (2014), e10 (cit. on pp. 82, 83).
- [Roc+15] Christian M Rochefort, Aman D Verma, Tewodros Eguale, Todd C Lee, and David L Buckeridge, « A novel method of adverse event detection can accurately identify venous thromboembolisms (VTEs) from narrative electronic health record data », *in: Journal of the American Medical Informatics Association 22.1* (2015), pp. 155–165 (cit. on p. 82).
- [SF08] Benoît Sagot and Darja Fišer, « Building a free French wordnet from multilingual resources », *in: OntoLex*, 2008 (cit. on p. 91).
- [Sch13] Helmut Schmid, « Probabilistic part-of-speech tagging using decision trees », *in: New Methods in Language Processing*, 2013, p. 154 (cit. on p. 83).
- [SV19] Tom Servranckx and Mario Vanhoucke, « A tabu search procedure for the resource-constrained project scheduling problem with alternative subgraphs », *in: European Journal of Operational Research 273.3* (2019), pp. 841–860 (cit. on pp. 37, 38).
- [Sto01] Frederik Stork, « Stochastic resource-constrained project scheduling », Doctoral Thesis, Berlin, Germany: Technische Universität Berlin, Fakultät II - Mathematik und Naturwissenschaften, 2001 (cit. on p. 37).
- [TNK17] Martin Tritschler, Anulark Naber, and Rainer Kolisch, « A hybrid meta-heuristic for resource-constrained project scheduling with flexible resource profiles », *in: European Journal of Operational Research 262.1* (2017), pp. 262–273 (cit. on p. 39).
- [WCY19] Yan Wang, Shuang Cang, and Hongnian Yu, « A survey on wearable sensor modality centred human activity recognition in health care », *in: Expert Systems with Applications 137* (2019), pp. 167–190 (cit. on p. 23).
- [ZYS12] Yiting Zhang, Ping Yu, and Jun Shen, « The benefits of introducing electronic health records in residential aged care facilities: a multiple case study », *in: International Journal of Medical Informatics 81.10* (2012), pp. 690–704 (cit. on p. 23).

-
- [Zhu+17] Xia Zhu, Rubén Ruiz, Shiyu Li, and Xiaoping Li, « An effective heuristic for project scheduling with resource availability cost », *in: European journal of operational research* 257.3 (2017), pp. 746–762 (cit. on p. 39).

Titre : Élaboration de projets personnalisés pour les usagers du secteur social et médico-social : planification automatique et objectivation des données

Mot clés : Planification de projets, Optimisation sous contraintes, Programmation mathématique, Recherche de voisinage, Classification de textes.

Résumé : Le secteur social et médico-social en France regroupe l'ensemble des structures où sont accueillies différentes catégories de populations vulnérables. Ces structures doivent élaborer et assurer la mise en œuvre d'un projet personnalisé pour chaque résident, en plus des soins quotidiens. Un projet personnalisé est composé d'un ensemble d'activités et services choisis parmi les activités et services proposés par la structure. Cette thèse étudie le problème principal de l'élaboration de projets personnalisés réalisables et adaptés aux besoins des usagers. Elle vise à améliorer l'efficacité globale de la mise en

œuvre des projets dans les structures sociales et médico-sociales. Concrètement, nous fournissons une modélisation formelle du problème de la planification des projets, plusieurs méthodes de résolution dont deux modèles de programmation mathématique, un algorithme heuristique constructif de type greedy et un algorithme de recuit simulé multi-voisinages. Nous proposons aussi une méthode de classification des événements indésirables et une solution de visualisation des connaissances pour l'objectivation des informations dans le projet personnalisé.

Title : Elaboration of personalized projects for users in social and medico-social sector: automatic planning and data objectification

Keywords : Project planning, Constrained optimization, Mathematical programming, Neighborhood search, Text classification.

Abstract : Social and medico-social sector in France assembles all structures where different categories of vulnerable populations are hosted. In addition to the daily care, these centers have to ensure the implementation of the personalized project for each resident in response to his/her needs in comply with national legal provisions. A personalized project is composed of a set of activities chosen among available activities proposed by a center. This thesis deals with the main issue of elaborating feasible and thoughtful personal-

ized projects, which aims to improve the whole efficiency of project implementation in social and medico-social structures. Specifically, we provide a formal modeling for the user project planning problem and several solution methods including two mathematical programming models, a constructive greedy heuristic algorithm and a multi-neighborhood simulated annealing algorithm. In addition, we propose a adverse event classification method and a knowledge visualization solution for the information objectification in personalized project.