

Rapport de Stage

Débat basé sur la théorie de la décision*

Auteur :
Tony SEGUIN

Encadrants :
Olivier CAILLOUX
Meltem OZTÜRK



Résumé

to do

Table des matières

1	Introduction	2
2	Notations et définitions	2
2.1	Connaissances	2
2.2	Argument et relation	3
3	État de l’art	4
4	Scénario et caractéristique de sélection	4
4.1	Scénario	4
4.2	Critère de sélection	4
5	Approches compatibles	5
5.1	Recherche	5
5.2	Présentations	7
6	Travaux à venir	9

1 Introduction

Un problème d'aide à la décision consiste à formaliser et déterminer les préférences d'un utilisateur dans le but de fournir une alternative recommandable par rapport à un ensemble d'alternative possible. En fonction du contexte du problème fournir juste une recommandation à l'aide de modèle de décision n'est pas forcément suffisant, les utilisateurs peuvent avoir besoin d'explication sur comment la recommandation a été faite et pourquoi elle est la meilleur pour eux. En effet, l'argumentation et la justification d'une recommandation est une part importante de la décision. Construire une explication compréhensible et convaincante est requise pour un utilisateur qui ne possède pas forcément des connaissances mathématiques, par exemple dans le cas des systèmes de recommandation en ligne, il a été montré que les explications augmentent l'acceptabilité d'une alternative recommandée (Pu and Chen, 2007). Dans ce contexte ci, une explication doit être simple et complète pour installer une confiance à l'utilisateur qu'une alternative recommandée est celle qui lui sied le mieux.

L'approche par le biais d'un débat nous semble être un bon compromis entre une explication simple et complète, en effet quoi de mieux qu'un débat quasi naturel pour convaincre une personne que ce que l'on énonce est vrai ou non par contre-argumentation. C'est pour cela que notre travail se penche sur cette problématique d'argumentation pour la recommandation avec cette nuance qu'est le débat.

Ce rapport est découpé (actuellement) en 3 parties. Dans un premier temps, Section 3, un état de l'art sur les systèmes de recommandation et d'aide à la décision, basé sur la review de Nunes et Jannach (2017) afin d'avoir une idée assez large des approches que l'on pourrait utiliser dans le cadre du projet. Dans un deuxième temps nous introduirons notre approche s'inspirant de Cailloux et Meinard (2018), en présentant un scénario et en posant les caractéristiques nécessaires qu'une approche doit avoir pour être compatible avec notre projet. Et finalement, une liste des approches, détaillées, qui nous semble être compatible.

2 Notations et définitions

Avant d'entrer dans les détails de notre approche, il est nécessaire de poser les définitions et les notations que nous allons utiliser tout au long du projet. Nous établissons d'abord les connaissances pour les systèmes de recommandation et ensuite les arguments et les relations entre eux pour le jugement de la décision lors d'un débat.

2.1 Connaissances

- $X = X_1 \times \dots \times X_n$ - l'ensemble des alternatives décrit sur n critères, $x \in X$;
- X_i - le vecteur décrivant l'ensemble des alternatives sur le critère i ;
- x_i - le critère i de l'alternative x ;
- $w = (w_1, \dots, w_n)$ - le vecteur de poids, $\sum_{i=1}^n w_i = 1$;
- w_i le poids du critère i , $w_i \in [0, 1]$.

La sélection d'une alternative par rapport aux autres s'établit par la relation de préférence \succsim_i sur chaque ensemble X_i , représentant ainsi les préférences du décideur sur les éléments de X_i . On note \succ_i et \sim_i comme les parties asymétrique et symétrique de \succsim_i , et nous notons :

- $x_i \succsim_i y_i$ - x est préférée à y sur le critère i .

En combinant les poids des critères et les descriptions des alternatives, notre approche s'appuie naturellement sur les modèles de décision de type *Multi-Attribute Value Theory* (MAVT) afin de d'attribuer un score à chaque alternative permettant ainsi de fournir un résumé de la dominance d'une alternative sur une autre. En général, la fonction de décision associé au

modèle est de la forme, pour $x \in X$, $v(x) = \sum_{i=1}^n w_i \times v_i(x)$, où $v_i(x)$ représente le score de x sur le critère i et $v(x) \rightarrow [0, 1]$. Cette fonction quantifie la relation de préférence \succsim , i.e $\forall x, y \in X, x \succsim_i y \Leftrightarrow v(x_i) \geq v(y_i)$.

2.2 Argument et relation

- Arg - l'ensemble des arguments ;
- $Arg^+(x, y)$ - ensemble des arguments en faveur de x par rapport à y ;
- $Arg^-(x, y)$ - ensemble des contre-arguments ;
- $Arg^=(x, y)$ - ensemble des arguments neutres ;
- $a_{(x,y)} \in Arg$ - un argument de x par rapport à y .

La relation de préférence \succsim nous permet d'avoir une distinction de ces trois types d'argument, nous avons donc :

- $Arg^+(x, y) = \{i \in N, x_i \succ_i y_i\}$;
- $Arg^-(x, y) = \{i \in N, x_i \sim_i y_i\}$;
- $Arg^=(x, y) = \{i \in N, x_i \prec_i y_i\}$.

Ces arguments permettent ainsi de mettre en évidence les phases de compromis pour une alternative qui n'est pas dominante sur l'ensemble des critères. Ces trois ensembles d'arguments sont regroupés dans l'ensemble Arg .

Les relations entre argument suit la notation suivante (Cailloux et Meinard (2018)) :

- T - ensemble des proposition possible ;
- P - ensemble des perspectives.

Dans notre cas, les sujets possibles sont $\{\forall x, y \in X, t_{x \succ y}\}$, représentant ainsi chaque paire possible de domination d'une alternative sur une autre. Une perspective $p \in P$ représente le point de vue de l'utilisateur, qui peut changer au cours du débat s'il a été convaincu d'un argument nouveau. Les ensembles P , T et Arg déterminent la position argumentative de l'utilisateur et est représentée via les relations suivantes :

- $\rightsquigarrow : Arg \times T$, on note $a \rightsquigarrow t$, l'argument a soutien la proposition t ;
- $\triangleright_{\exists} : Arg \times Arg$, on note $a_2 \triangleright_{\exists} a_1$, l'argument a_2 attaque l'argument a_1 , a_1 devient un argument invalide ;
- $\not\triangleright_{\exists} : Arg \times Arg$, on note $a_2 \not\triangleright_{\exists} a_1$, l'argument a_2 n'attaque pas l'argument a_1 , a_1 reste un argument valide.

Un argument s peut soutenir plusieurs proposition comme aucune. On admet $\neg(a_2 \triangleright_{\exists} a_1) \Rightarrow a_2 \not\triangleright_{\exists} a_1$. Si l'utilisateur change d'avis les relations $(\triangleright_{\exists}, \not\triangleright_{\exists})$ suffisent à capturer ce changement.

Definition 1 Une situation de décision est défini par le tuple $(T, Arg, \rightsquigarrow, \triangleright_{\exists}, \not\triangleright_{\exists})$.

- $\triangleright_{\forall} : \text{défini comme } a_2 \triangleright_{\forall} a_1 \Leftrightarrow \neg(a_2 \not\triangleright_{\exists} a_1)$;
- $\not\triangleright_{\forall} : \text{défini comme } a_2 \not\triangleright_{\forall} a_1 \Leftrightarrow \neg(a_2 \triangleright_{\exists} a_1)$.

La relation $a' \triangleright_{\forall} a$ signifie que l'argument a' attaque l'argument a sur toute les perspectives. En revanche la relation $a' \not\triangleright_{\forall} a$ signifie que a' n'attaque jamais a .

Definition 2 étant donné une situation de décision $(T, Arg, \rightsquigarrow, \triangleright_{\exists}, \not\triangleright_{\exists})$, un argument $a \in Arg$ est décisif, si et seulement si, $\forall a' \in Arg : a' \not\triangleright_{\forall} a$.

Definition 3 étant donné une situation de décision $(T, Arg, \rightsquigarrow, \triangleright_{\exists}, \not\triangleright_{\exists})$, une proposition t est :

- acceptable si et seulement si, $\exists a \in Arg \mid a \rightsquigarrow t, \forall a' : a' \not\triangleright_{\forall} a$;
- rejetable si et seulement si, $\forall a \in Arg \mid a \rightsquigarrow t, \exists a_c \mid a_c \triangleright_{\forall} a$ et $\forall a_{cc} : a_{cc} \not\triangleright_{\forall} a_c$.

Definition 4 Une situation de décision $(T, Arg, \rightsquigarrow, \triangleright_{\exists}, \not\triangleright_{\exists})$ est claire, si et seulement si, chaque proposition dans T est acceptable ou rejetable.

Definition 5 Le jugement délibéré de l'utilisateur correspondant à la situation de décision $(T, Arg, \rightsquigarrow, \triangleright_{\exists}, \not\triangleright_{\exists})$ est :

$$T_{utilisateur} = \{ t \in T \mid t \text{ est acceptable} \}.$$

3 État de l'art

En cours de remaniement

4 Scénario et caractéristique de sélection

Dans cette section nous présentons le projet par l'intermédiaire d'un scénario. et nous définissons les caractéristiques que doit valider les approches que nous recherchons.

4.1 Scénario

Si nous étions dans une approche multi-agent, le scénario serait composé de 3 agents (i) un agent U représentant l'utilisateur et (ii) deux agents, R1 et R2, représentant chacun une méthode de recommandation et d'argumentation. Nous prenons pour le moment l'approche avec 2 agents pour la recommandation pour plus de simplicité. Le débat se déroule de la manière suivante :

to do : convenir d'un protocole de dialogue entre les modèles et l'utilisateur, le role de chacun et introduire ici le schéma

4.2 Critère de sélection

Dans le cadre du projet, les approches doivent être cohérente entre elles, i.e elles adoptent, dans les grandes lignes, des modèles de recommandation et donc d'argumentation compatible entre elles. Il est donc nécessaire que les approches que nous allons sélectionner, pour la recommandation basé sur un débat, soient spécifiée à aucun domaine. Ce qui est notre premier critère de sélection.

Le second critère est sur quelles informations nos approches vont s'appuyer pour argumenter ; comme vu Section 3, les approches basées sur les connaissances permettent une argumentation riche et détaillée. En adoptant un Multi-Criteria Decision Model (MCDM), le compromis, les critères décisifs, pertinents ou non, sont des types d'argument qui rentrent dans notre schéma de débat. C'est pourquoi nous nous pencherons seulement sur les approches basées sur les connaissances et utilisant un MCDM.

Nous nous plaçons d'un point de vue, où les modèles possèdent toutes les informations nécessaires pour argumenter, nous allons donc naturellement prendre comme troisième critère les approches utilisant des fonction de décision de type MAVT.

Et comme dernier critère, seules les approches fournissant une argumentation du type «A meilleur que B» ou «A meilleur que tout les autres» seront retenues.

5 Approches compatibles

Nous allons maintenant discuter des approches que nous avons sélectionné. Tout d’abord nous allons préciser notre recherche et ensuite faire un léger résumé de chacune des approches.

5.1 Recherche

La recherche des approches à été faite en deux parties, dans un premier temps une sélection des articles étudiés dans une review (Nunes and Jannach, 2017) à été réalisée et dans un second temps une recherche pour élargir notre vision d’ensemble et obtenir un nombre assez conséquent d’approches à été faite.

Review

Dans la review, les articles étudiés sont classé en quatre catégories : (i) *Technique*, pour les articles apportant une nouvelle forme d’explication (ii) *Tool*, les articles décrivant un outil incluant une explication (iii) *Evaluation*, les articles évaluant ou comparant des approches et (iv) *Foundational*, les articles discutant autour des aspect des explications. Dans un premier temps, seuls les articles *Technique* sont conservé, car ce sont des articles avec des approches non contextualisée (contrairement aux Tools) pour la majorité, soit 101 articles. Ensuite nous avons conservé seulement les articles étiquetés comme étant des approches basée sur les connaissances, plus précisément celles utilisant un MCDM, ce qui nous donne 8 articles. Les résumés et conclusions ont été inspectés, ainsi qu’un lecture transversale de l’ensemble des articles nous à permis de réduire à 6 articles, en effet un article (Bielza C. and al, 2000) est basé sur un modèle de type Multi Attribute Utility Theory (MAUT) et le second article rejeté est (Bohanec and al, 2000) car contextualisé dans le domaine de la médecine. Les articles conservés sont listés en Table 1.

Table 1 Articles retenus de la review

Auteurs	Titres
Klein et Shortliffe	A framework for explaining decision-theoretic advice
Carenini et Moore	Generating and evaluating evaluative arguments
Labreuche	A general framework for explaining the results of a multi-attribute preference model in Automotive Scenarios
Nunes et al	Pattern-based Explanation for Automated Decisions
Belahcene et al	Explaining robust additive utility models by sequences of preference swaps

Élargissement

Ensuite, une recherche supplémentaire à été exécutée pour élargir le nombre d’approches possible, nous avons d’abord établis les mots clés de pour la recherche ainsi que leurs synonymes. En effet dans la littérature tout le monde ne s’accorde pas à utiliser les mêmes termes pour désigner la même chose. Le premier terme est *argumentation*, qui est l’ingrédient essentiel de

notre recherche, ensuite le deuxième terme est *decision support system* incluant ainsi les systèmes de recommandation et le troisième est *multi-criteria* designant toute les variante tel que Multi-Criteria Decision Making par exemple.

Table 2 Termes et synonymes

Termes	Synonymes
argumentation	explanation, justification
decision support system	decision making, recommandation knowledge-based system, knowledge based system
multi-criteria	multi-attribute

La recherche à l’aide de ces mots clés prend forme de la manière suivante sur les plus large base de données de bibliothèques d’articles en ligne, à savoir : ACM Digital Library, IEEE Xplore Digital Library, ScienceDirect et Springer Link.

$$(\text{argumentation} \vee \text{justification} \vee \text{explanation}) \wedge (\text{decision support system} \vee \text{decision making} \vee \text{recommandation} \vee \text{knowledge-based system} \vee \text{knowledge based system}) \wedge (\text{multi-criteria} \vee \text{multi-attribute})$$

Ces termes ont été recherchés dans les titres et les résumés de chaque articles des bases de données, nous avons obtenus 98 articles, voir Table 3. Dans un premiers temps, les titres et résumés ont été étudiés, pour réaliser un premier filtre, ensuite une lecture plus en détails des articles restant à été entrepris.

Table 3 Résultat de recherche par source

Sources	Nombre d’articles
ACM Digital Library	19
IEEE Xplore Digital Library	8
ScienceDirect	19
Springer Link	52
Total	98

Étonnamment, pour la majorité des cas, les articles ont une approche de type machine learning ou les approches sont des outils pour un problème précis. Toute ces approches ont été rejetées. Après ce premier filtre nous sommes parvenu à 8 articles restant et après lecture des textes de ces articles, certains ont été rejeté. Par exemple, [Delle Site et Filippi \(2009\)](#), ne fourni pas d’argumentation, [Yevseyeva et al \(2016\)](#) utilisent des données externes, [Kadziński and al \(2017\)](#) est une analyse expérimentale d’une fonction additive dans plusieurs méthodes de désagrégation des préférences.

Table 4 Articles retenus de la recherche d'élargissement

Auteurs	Titres
Labreuche and al	A Dialogue Game for Recommendation with Adaptive Preference methods : An experimental analysis
Geldermann	Explanation Systems
Papamichail et French	Explaining and justifying the advice of a decision support system : a natural language generation approach

5.2 Présentations

En s'appuyant sur la review de [Nunes et Jannach \(2017\)](#) et l'élargissement présenté précédemment 10 approches respectent les critères imposés. Une présentation par ordre chronologique de chaque articles est établie ci-après.

Klein et Shortliffe

L'approche de [Klein et Shortliffe \(1994\)](#) est l'une des premières approche en MCDA, elle pose les bases de cette branche de la l'aide à la décision. Ils présentent plusieurs stratégies basé sur la MAVT pour expliquer automatiquement les décisions parmi plusieurs objectifs en conflit. Ils décrivent ces stratégies dans un framework prénommé IVA (Interpretive Value Analysis), dans un cadre large d'explication et d'acquisition de pointe dans des systèmes experts qui modélisent des décisions à forte intensité de compromis. Les concepts d'interprétation jouent le rôle de primitives d'explication dans les stratégies d'IVA et sont également utilisés comme fonctions d'évaluation qui guident la composition des explications. Les stratégies génèrent des comparaisons sommaires de paires particulières d'alternatives en limitant la profondeur et la largeur de l'arbre des valeurs. Les stratégies produisent des comparaisons plus détaillées d'alternatives, fournissant des traces pas à pas de calculs des différences de valeurs multi-attributs. Les explications fournies conservent les avantages de l'intelligence artificielle et de la théorie de la décision pour la modélisation des décisions.

Papamichail et French

Cette approche décrit une méthode pour générer une explication dans une décision de contexte analytique. D'après [Papamichail et French \(2003\)](#) le point fort de leurs approche est le développement d'une librairie de texte planifié pour structurer le message soumis à l'utilisateur. L'approche est en revanche générique. Le système fourni deux types de rapport : (i) un rapport de comparaison expliquant le raisonnement derrière le classement des alternatives et (ii) un rapport d'analyse de sensibilité fournissant une évaluation globale du modèle de décision et décrit l'effet de la variation d'un paramètre de décision.

Carenini et Moore

L'article de [Carenini et Moore \(2009\)](#) se veut interdisciplinaire (théorie de l'argumentation, théorie de la décision, linguistique informatique, psychologie sociale et interaction homme-machine). En se concentrant uniquement sur les parties argumentation et théorie de la décision, l'approche se détache des autres car elle forme un graphe d'argumentation pour générer ensuite son explication en langage naturelle adapté aux préférences de l'utilisateur.

Geldermann

L'article de [Geldermann \(2009\)](#) fournit une application pour un système d'explication pour des systèmes d'aide à la décision basé sur MAUT et plus spécifiquement sur MAVT. L'argumentation se fait par le biais d'un rapport comparatif, i.e interprétant les résultats d'évaluations du modèle en comparant deux alternatives. Le rapport discute de la façon dont une alternative évalue l'autre sur chaque critère d'évaluation, en soulignant les arguments pour et contre chaque alternative, sur la base des scores de critères réels. Ainsi, il examine à quel point une alternative est meilleure qu'une autre et souligne les facteurs qui différencient entre deux alternatives. Un autre type de rapport est fourni, l'analyse de sensibilité, qui explique les graphiques d'analyse de sensibilité et illustre l'effet du changement du poids d'un attribut dans le classement des alternatives et discute de la robustesse de la meilleure alternative.

Labreuche

Cette approche se veut être une version plus simple par rapport à l'explication fournie par l'approche de [Klein et Shortliffe \(1994\)](#). En effet l'approche de [Labreuche \(2011\)](#), propose une approche pour sélectionner les arguments utilisables dans une explication faite pour un problème de décision multi-critère pondéré par des poids assignés à ces critères. Il se base sur l'analyse des valeurs de ces poids ainsi que le score des alternatives pour les comparer. Un seul modèle est utilisé sur les trois de l'article nous intéresse, le modèle Expected Utility (EU), qui est un modèle MAVT. Le but de l'approche est de rechercher certains changements dans le vecteur de poids v qui permet une inversion de la décision fournie par le modèle entre deux alternatives. L'explication se focalise alors sur le ou les critères qui ont été modifiés dans v . Les autres critères ne sont pas mentionnés pour une explication se voulant la plus courte possible. Il y a deux stratégies de modification de v : (i) le remplacement de v par un autre vecteur de référence w^F et (ii) la permutation des poids de v . La première stratégie permet de mettre en évidence les critères importants et non importants, et la seconde permet de déterminer les critères décisifs.

Nunes et al

L'approche de [Nunes et al \(2014\)](#) est l'une des plus complètes, elle s'appuie sur une génération d'explication basée sur un groupe d'algorithmes pour identifier les paramètres permettant de remplir le template d'explication qu'ils fournissent dans l'article. En effet une explication possède 7 formes possibles, s'adaptant ainsi aux différents cas particuliers de décision. Si plusieurs formes d'explication sont possibles pour une alternative recommandée, un ordre pré-établi des formes d'explication est installé pour en choisir qu'une seule. Les principales formes sont : (i) critical attribute, l'alternative est choisie car elle a la meilleure valeur sur un critère donné (ii) domination, l'alternative domine sur tous les critères (iii) decisif criteria, l'alternative est sélectionnée en raison d'un ensemble de critères (iv) trade-off resolution, l'alternative a un côté avantageux sur un ensemble de critères qui compense ses défauts. Une étude utilisateur a été réalisée incluant une trentaine de participants, leurs approches ont été comparées à deux autres approches ([Labreuche \(2011\)](#) ; [Klein et Shortliff \(1994\)](#)), les résultats indiquant que leurs approches ont de meilleures performances.

Labreuche and al

L'interaction via un protocole de dialogue entre l'utilisateur et le système en fait sa particularité. En effet, [Labreuche et al \(2015\)](#) proposent une méthode où le système s'adapte en fonction des retours de l'utilisateur et passe d'un modèle de décision à l'autre en fonction des informations disponibles. Dans l'article seul un modèle nous intéresse, car c'est le seul qui est de MAVT.

L'argumentation par contre est sommaire, elle fourni un listing des arguments en faveurs d'une alternative par rapport à une autre.

Belahcene et al

L'approche fournit une explication dite complète, contrairement à d'autre approche qui ne fournissent que les points clés décisifs, dans un contexte de décision multi-critères. Dans le cas de [Belahcene et al \(2016\)](#), les informations initiales prennent une forme de comparaison par paire d'alternatives. La génération d'explication s'inspire d'une méthode «even-swap», une procédure d'élection assumant un modèle additif des valeurs des préférences et basé sur le «trad-off» entre des paires de critère. Leurs version est une généralisation de «even-swaps» à échanges des préférences et simplement montrer une comparaison d'alternatives. L'explication construit un graphe d'arguments positifs et négatifs entre deux alternatives, pour argumenter pourquoi l'une est préférée à l'autre. La visualisation se fait par un graphe biparti entre les deux types d'arguments.

6 Travaux à venir

- Implémentation des approches conservées, obtenir des résultats cohérents avec les articles des approches retenues.
- Adapter les approches pour avoir un langage commun + une communication et interprétation de l'opposant.
- Convenir d'un protocole de dialogue entre les "recommandeurs" et l'utilisateur.
- Mettre en place la convergence des approches dans un seul système.

Références

Cailloux O. & Meinard Y., 2018. A formal framework for deliberated judgment. arXiv:1801.05644v1 [cs.AI]

Nunes, I. & Jannach, D. User Model User-Adap Inter (2017) 27: 393.

Labreuche C., 2011. A general framework for explaining the results of a multi-attribute preference model.

Pu P. Chen L., 2007. Trust-inspiring explanation interfaces for recommender systems.

Klein D.A., Shortliffe E.H., 1994. A framework for explaining decision-theoretic advice

Carenini G, Moore J.D., 2006. Generating and evaluating evaluative arguments

Nunes I., Miles S., Luck M., Barbosa S., and Lucena C., 2014. Pattern-based explanation for automated decisions.

Belahcene, K., Labreuche, C., Maudet, N. et al. Explaining robust additive utility models by sequences of preference. Theory Decis (2017) 82: 151.

Labreuche C., Maudet N., Ouerdane W., and Parsons S., 2015. A Dialogue Game for Recommendation with Adaptive Preference Models.

Geldermann J., 2010. Explanation Systems.

Papamichail K.N., French S., 2003. Explaining and justifying the advice of a decision support system: a natural language generation approach.,