

Titre : Détection et Réparation d'Incohérences de Données via les Techniques de Machine Learning dans un Environnement Incertain

Mots-clés : Décision multicritères, Préférences, Explicabilité, Détection d'incohérence, Fonctions d'agrégation, Dilemme Exploration/Exploitation, Incertitude, Machine learning.

Encadrants : M. Loïc ADAM, MCF-LIAS/ENSMA (loic.adam@ensma.fr) et M. Allel HADJALI, PROF-LIAS/ENSMA (allel.hadjali@ensma.fr)

1. Contexte

L'éllicitation de préférences cherche à **aider** une utilisatrice à prendre une **décision** parmi un large choix d'alternatives (choisir un plan d'investissement, par exemple), tout en étant capable d'expliquer pourquoi une telle décision a été prise (**explicabilité**) [7, 9]. Pour cela, une série de questions est posée à l'utilisatrice jusqu'à avoir suffisamment d'informations sur ses préférences pour lui recommander la meilleure alternative. L'une des stratégies d'éllicitation les plus populaires est l'éllicitation incrémentale robuste [3, 4, 8]. Associée avec l'heuristique de la stratégie de la solution courante (CSS) [5], elle permet de minimiser l'effort cognitif en posant un minimum de questions simples comparant deux alternatives (préférez-vous ce plan, ou celui-là ?). Néanmoins, cette stratégie d'éllicitation ne garantit pas une solution optimale quand l'utilisatrice est incertaine de ses préférences et/ou commet des erreurs ; ou que les interactions entre les différents critères des alternatives ne sont pas modélisées correctement. Pire encore, cette stratégie peut conduire à une mauvaise recommandation, sans signal d'alerte.

Dans des précédents travaux [2], nous avons proposé une stratégie d'éllicitation qui permet, grâce à la modélisation de l'**incertitude** des réponses via la théorie des possibilités [6], de détecter quand des réponses deviennent incohérentes entre-elles (soit car des réponses sont fausses, soit quand le modèle supposé n'est pas suffisamment riche pour prendre en compte des interactions complexes). Ainsi, l'éllicitation est plus robuste face à des incohérences, permettant d'avoir des meilleures recommandations. De plus, nous avons :

- Des stratégies restaurant la cohérence en présence de mauvaises réponses [1], permettant à la fois d'améliorer la qualité de la recommandation finale pour l'utilisatrice, mais aussi d'identifier les réponses erronées.
- De premiers résultats très prometteurs sur le changement de fonction d'agrégation quand celle-ci n'est pas adéquate (conférence nationale LFA 2025). Changer pour le bon modèle n'améliore pas toujours la recommandation finale (voir l'empire), montrant qu'une stratégie de questions pour une fonction d'agrégation n'est pas forcément optimale pour une autre, en lien avec les travaux existants [3].

2. Défis scientifiques

Comme présenté précédemment, les incohérences détectées durant l'éllicitation sont de deux natures différentes : mauvaises réponses de l'utilisatrice (avec un caractère potentiellement aléatoire) et modélisation incorrecte de ses préférences (avec un caractère plus systémique). Du

fait de cette différence de nature, il paraît tout à fait raisonnable de vouloir détecter la nature de l'incohérence afin d'appliquer une stratégie de correction/restauration de la cohérence adaptée.

Dans cette thèse, nous envisageons les contributions scientifiques suivantes :

- 1) L'utilisation des outils du **machine learning** (et de la **science des données**) pour identifier la nature de l'incohérence, afin d'appliquer des stratégies de correction/restauration de la cohérence adaptées, malgré une information limitée.
- 2) Développer une approche d'élicitation permettant de détecter au plus vite de l'incohérence, tout en cherchant à converger rapidement vers une recommandation, en lien avec les travaux en **dilemme exploration/exploitation**.
- 3) En parallèle du premier point, améliorer l'évaluation de la confiance dans les réponses de l'utilisatrice, afin d'enrichir l'**explicabilité** du modèle et potentiellement rendre la détection de la nature de l'incohérence plus aisée.

3. Références

- [1] Adam, L., & Destercke, S. (2024). Handling inconsistency in (numerical) preferences using possibility theory. *Information Fusion*, 103, 102089.
- [2] Adam, L., & Destercke, S. (2021, December). Possibilistic preference elicitation by minimax regret. In *Uncertainty in artificial intelligence* (pp. 718-727). PMLR.
- [3] Benabbou, N., Perny, P., & Viappiani, P. (2017). Incremental elicitation of Choquet capacities for multicriteria choice, ranking and sorting problems. *Artificial Intelligence*, 246, 152-180.
- [4] Bourdache, N., Perny, P., & Spanjaard, O. (2019, August). Incremental elicitation of rank-dependent aggregation functions based on Bayesian linear regression. In *IJCAI-19-twenty-eighth international joint conference on artificial intelligence* (pp. 2023-2029).
- [5] Boutilier, C., Patrascu, R., Poupart, P., & Schuurmans, D. (2006). Constraint-based optimization and utility elicitation using the minimax decision criterion. *Artificial Intelligence*, 170(8-9), 686-713.
- [6] Dubois D., Prade, H. (1988). Possibility Theory: An Approach to Computerized Processing of Uncertainty. *Plenum Press*.
- [7] Ouerdane, W. (2022). *From Preference Elicitation to Explaining Decisions : A Dialectical Perspective* [Habilitation à diriger des recherches, Université Paris - Saclay].
- [8] Wang, T., & Boutilier, C. (2003, August). Incremental utility elicitation with the minimax regret decision criterion. In *IJCAI* (Vol. 3, pp. 309-316).
- [9] Willot, H., Belahcène, K., & Destercke, S. (2024). Principled Explanations for Robust Redistributive Decisions. In *ECAI 2024* (Vol. 392, p. 979-986).