Ph.D. Thesis Proposal

Explainability in reference based Recommender Systems with multiple criteria

From the first expert systems to the recent recommender systems which flourish on commercial websites or are available in the context business-to-business sales applications, decision-aiding has been a central concern in Artificial Intelligence. It has become clear that providing recommendations is only part of the challenge: as decision-aiding tools become everyday more popular and more sophisticated, it is of utmost importance to develop their explanatory capabilities. Consequently, *"Why should I trust you?"* is a question every system should be prepared to answer - especially when the stakes are high. Moreover, in the business-to-business sales scenario, the explanation of how a given recommendation has been reached is often more important than the recommendation itself: it gives the seller appropriate arguments to approach the client. Finally, there is a growing demand of institutions and citizens to make algorithmic decisions transparent and trustworthy. Indeed, the recent General Data Protection Regulation (GDPR) adopted by the European Parliament goes further by adding a "right to explanation".

As a recommender system is targeted to an end-user, it must take into account the user's preferences and specificities. Therefore, such preferences have first to be elicited and then used to constrain the recommendations. In addition, they can also be used to better formulate and reinforce the explanation.

Under such perspectives, our project addresses the problems of recommendations, where the aim for an "artificial agent adviser" is to help a human user (a decision maker) in building and understanding the recommendations of a particular decision problem. Decision aiding is thus a situation involving two parties: a user, with preferences which may be very incompletely defined or difficult to convey, and an agent, which will have the capabilities of representing explicitly and accountably the reasons for which it recommends a solution to a user. *W*e are interested by decision problems involving preference information and data carefully designed and elaborated on *human data sets.* However, such data (e.g. preferences, values, etc.) could be incomplete, imprecise, issued from different potentially conflicting sources. The recommendations are based on Multiple Criteria Decision Aiding models that are well founded from the point of view of Decision Theory [1]*.*

An interesting starting point will be to consider *the Non-Compensatory model (NCS)* [2] and the *Simplified Ranking with Multiple Points (S-RMP) [3].* In short, the first model allows to classify a set of options into an ordered predefined categories, while the second one has the aim to construct a ranking among a set of options. These two models have the advantage to be

interpretable in the sense that the decision rules are simple and easy to grasp by a decision maker [4,5].

We will investigate several use cases using both public and real-world industrial data sets. Also we will demonstrate how explanations improve optimization models generated by Cognitive Optimization.

The Ph.D. student will be hosted by IBM France Lab in Saclay and the thesis is run in collaboration between CentraleSupélec in the context of DATAIA Convergence Institute, IBM France Lab in Saclay and IBM Research Zurich.

Candidates should have a Master Degree in Computer Science from an University or "grande école".He/she should have a good background in artificial intelligence and operational research. Knowledge on multiple criteria decision analysis will be appreciated. Candidates are invited to send their CV, motivation letter, their grades obtained in Master/Grande Ecole, and references. Application deadline : 30/06/2018

Contacts :

- IBM France Lab : Christian de Sainte Marie, <u>csma@fr.ibm.com</u>

- IBM Research : Michalis Vlachos, Paolo Scotton {mvl,psc}@zurich.ibm.com

- CentraleSupélec : Wassila Ouerdane, Vincent Mousseau, {wassila.ouerdane,vincent.mousseau}@centralesupelec.fr

References

[1] D. Bouyssou, T. Marchant, M. Pirlot, A. Tsoukiàs and P. Vincke, Evaluation and Decision models: stepping stones for the analyst, Berlin, Springer Verlag, 2006.

[2] D. Bouyssou and T. Marchant, An axiomatic approach to Non compensatoty sorting methods in mcdm I : the case of two categories, *European Journal of Operation Research*, 2007, 178(1), 217-245.

[3] A. Rolland, Reference-Based preference aggregation procedure in multi-criteria decision making, *European Journal of Operation Research*, 2013, 225(3), 479-486.

[4] V. Ferreti, J. Liu, V. Mousseau and W. Ouerdane, Reference-based ranking procedure for environmental decision making: Insights from an ex-post analysis, *Environmental Modelling & Software*, 2018, 99, 11-24.

[5] O. Sobrie, M. Lazouni, S. Mahmoudi, V. Mousseau and M. Pirlot, A new decision support model for preanesthetic evaluation, *Computer Methods and Programs in Biomedicine*, 2016, 133, 183-196.