LEARNING OPTIMAL INFORMATION QUERIES UNDER SEVERE UNCERTAINTY

Doctoral position
HEUDIASYC (UTC)

Thesis supervisors: Sébastien Destercke and Mylène Masson

Context

Many modern applications of **machine learning, artificial intelligence and system engineering** involve severe uncertainty: **medical diagnosis**, **sparse preferences** provided by users [8], system design using new and untested technologies, system dimensioning to **face uncertain events such as climate change**. In such cases, determining the optimal queries that will reduce the most our uncertainty and allow us to make optimal decisions is both a very important and challenging task.

Ideal tools to perform such a task are the imprecise probability theories, that are very generic models of uncertainty encompassing both sets (used, e.g., in robust optimisation) and probabilities (used, e.g., in stochastic optimisation). Yet, there are few practical methods or practical studies as to how such models can be used in the task of determining optimal queries. Indeed, the following questions, that have been addressed for a long time in the probabilistic framework [3], have only received little attention in other uncertainty theories.

(1) How to efficiently, coherently and adequately extract valuable information from users or models, by asking relevant questions (possibly in a sequential way)?
(2) How to assess the quality of the (imprecise) answers provided by the user or model when the true value of the quantity is known (possibly uncertainly as well)?

All these questions touch upon both theoretical and practical questions that are essential in many fields such as risk and reliability analysis or machine learning problems, where the elicitation of data labels can be a critical issue.

Topic and environment

Depending on the candidate skills and preferences, the thesis will focus on one of the two following topics

Topic 1: active learning. Active learning techniques consist in acquiring more data, through (optimal) querying, in order to improve the learned model or the predictions made from it [7]. However, most techniques focus on reducing the ambiguity of the posterior predicted probability or on increasing a precisely valued accuracy.

When information is lacking, imprecise probabilistic approaches usually produce sets of probabilities or imprecise predictions (possibly leading to imprecise accuracy). It would therefore be interesting to consider active learning techniques where the primary goal is to reduce the imprecision rather than the ambiguity in the probability, or to make the prediction more precise. The candidate will then explore how active learning can be revisited to use at their advantage the properties of imprecise probability theory [1].

While we expect the candidate to start with standard active learning setting (unlabelled examples, no misclassification cost, negligible cost of labelling), there will be opportunity to explore more complex settings, such as cost-sensitive active learning frameworks [5].

Topic 2: expert elicitation. A critical question when using imprecise probabilistic model, for instance in risk and reliability analysis, in decision aid or in system design, is how the uncertainty can be reduced efficiently to reach a decision, or a satisfying solution (e.g., in terms of design). A similar question has recently gained a lot of attention in preference elicitation [9].

The candidate will explore how such ideas, often issued from the robust optimisation literature, can be adapted to the framework of imprecise probabilities. The goal will then be to recommend optimal questions or queries to ask to the expert. We expect the candidate to first focus on specific imprecise probabilistic models [4], making the elicitation easier, as well as on one-at-a-time optimisation problems [9], that are usually simpler to solve. The candidate will then be free to explore more complex situations, such as imprecise probabilistic models based on generic assessments [6] or sequential optimisation problems [2].

Although the candidate will be able to work independently on one (or both) of these topics if she/he wants, the Heudiasyc environment will allow her/him to collaborate with team-mates of the Decision and Image team, and on applications such as **reliability analysis of railway systems** (within the current national ANR project RECIF), **recognition problems for intelligent vehicles** (a major application included in the LABEX MS2T project) or **learning preferences of museum visitors** (within a regional project). The PhD candidate will also be part of the **Uncertainty In Machine Learning Network (UML-NET)** funded by the French government, providing her/him with opportunities to work with top researchers in Europe on the topic.

The Heudiasyc laboratory has also been recognized as a laboratory of excellence (LABEX) by the French government, providing it with necessary funds to ensure top-quality research as well as an international recognition.

Sought Profile and application requirements

The candidate must demonstrate (through her/his formation, previous projects, recommendations, grades, . . .) excellent skills in either mathematics or computer science. In particular, we are searching for excellent skills in at least one of the following fields:

- Machine Learning
- Probability/statistics
- Optimisation

- Artificial intelligence

Applications and questions can be sent to <sebastien.destercke@hds.utc.fr>. Applications **must** include the following items:

- a letter of motivation detailing what are the interest of the applicant in the proposed topic(s);
- a curriculum vitae clearly showing how the candidate profile matches the above requirements;
- contact information of at least one reference (two or more would be appreciated).

Any application not containing these three items will not be considered further. In addition, the following optional items may be included:

- existing scientific papers or significant project reports;
- any link to significant realisations (e.g., software, ...)
- copy of previously obtained grades.

References

[1] A. Antonucci, G. Corani, and S. Gabaglio. Active learning by the naive credal classifier.

[2] C. Boutilier. A POMDP formulation of preference elicitation problems. In *Proceedings of the Eighteenth National Conference on Artificial Intelligence and Fourteenth Conference on Innovative Applications of Artificial Intelligence, July 28 - August 1, 2002, Edmonton, Alberta, Canada.*, pages 239–246, 2002.

[3] R. M. Cooke. Experts in uncertainty: opinion and subjective probability in science. 1991.

[4] S. Destercke and D. Dubois. *Introduction to Imprecise Probabilities*, chapter Special Cases, pages –. Wiley series in probability and statistics. Wiley, May 2014.

[5] R. Greiner, A. J. Grove, and D. Roth. Learning cost-sensitive active classifiers. *Artificial Intelligence*, 139(2):137–174, 2002.

[6] E. Quaeghebeur, G. de Cooman, and F. Hermans. Accept & reject statement-based uncertainty models. *International Journal of Approximate Reasoning*, 57:69–102, 2015.

[7] B. Settles. Active learning literature survey. *University of Wisconsin, Madison*, 52(55-66):11, 2010.

[8] P. Viappiani. Preference modeling and preference elicitation: An overview. In *Proceedings of the First International Workshop on Decision Making and Recommender Systems (DMRS2014), Bolzano, Italy, September 18-19, 2014.*, pages 19–24, 2014.

[9] P. Viappiani and C. Boutilier. Regret-based optimal recommendation sets in conversational recommender systems. In *RecSys*, pages 101–108, 2009.