LEARNING AND PREDICTING GRAPHS FROM LIMITED INFORMATION.

Doctoral position, HEUDIASYC (UTC)

Thesis supervisor: Sebastien Destercke

Context

Today, many applications, be it in **machine learning, artificial intelligence or optimisation**, require to deal with large combinatorial spaces. Some examples include:

- preferences handling in recommendation and ranking systems [3], where having n objects lead to consider $n!$ different solutions;
- partial orders such as transitive relations ("is a kind of", "is part of") in knowledge bases;
- occupation grids of vehicles that can include hundreds of individual cells [5];

In such large domains, it is often the case that the information we have about the state of the world is scarce, incomplete and possibly noisy. Such severe uncertainties can be modelled in different ways, and imprecise probabilistic approaches offer one of the most general and accurate model to do so [1]. Yet, using such tools means that operations that are already quite complex in a well-known but combinatorial environment become even more complex.

In particular, the task of model learning or of making cautious inferences can be quite challenging, and is an essential pre-requisite to make either safe decision or to gather additional information that could lead to better decisions.

Topic

The thesis topic will primarily focus on how to make inferences over graphs when our knowledge about them is scarce. The reason for choosing graphs as a basic representation tool is that they naturally appear in numerous modern machine learning problems involving complex structures such as relational data or n-ary vectors:

- complete or partial preferences over a set of objects correspond to acyclic directed graphs over this set of objects;
- clusterings or classes of equivalence can be seen as a set of disjoint cliques over the set of object to partition;
- sorting problems such as (graded) multi-label problems or ordinal regression issues can be seen as multi-partite graphs;
- relations in ontologies can be seen as graphs with particular structures: for example taxonomies correspond most of the time to rooted trees, while other relations may have other properties (e.g., "brother of" is symmetric).

The candidate will have to develop efficient methods to both learn models over such structures from scarce or imprecise data, as well as efficient decision rules returning cautious predictions in the form of partially specified graphs. The candidate will benefit from the group expertise on the use of imprecise probabilistic approaches (e.g., convex sets of probabilities [2], evidence theory [4]) to perform her/his research activities.

The proposed line of research participates to a larger program consisting in studying cautious inferences in large combinatorial domains. Beyond inferences over graph structures, we can mention as examples

- making inferences over large binary matrices or highly dimensional Boolean functions from marginal information, such functions or matrices (and their extensions to non-binary ordinal domains) being common in the estimation of occupational grids or in reliability analysis;
- estimating preferences over multiple criteria, where the size of the domain increases exponentially with the number of criteria. Such problems indeed often involve uncertain feedback from the decision maker, or the need to construct new alternatives (e.g., touristic trips) from previously observed preferences;
- extending classical optimisation problems (minimal spanning trees, shortest path problem, ...) over weighted graphs to uncertain weighted graphs, with the goal to characterize sets of potentially optimal solutions.

The candidate will not be expected, *a priori*, to explore such topics, but will be able to participate to research activities concerning them.

Environment

- Place of research: Heudiasyc laboratory, CompiÃÍgne University of Technology (40 minutes by train from Paris)
- Salary: $\sim$ 20K€/year basic salary, possibly complemented by paid teaching activities
- Duration: 3 years
- Starting date: September/October 2017

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.

Applicant 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 explicitly what are the interest of the applicant in the proposed topic;
- 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, or not tailored to this proposal, 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] T. Augustin, F. P. Coolen, G. de Cooman, and M. C. Troffaes. *Introduction to imprecise probabilities*. John Wiley & Sons, 2014.

[2] S. Destercke. Multilabel predictions with sets of probabilities: The hamming and ranking loss cases. *Pattern Recognition*, 48(11):3757–3765, 2015.

[3] J. Fürnkranz and E. Hüllermeier. *Preference learning: An introduction*. Springer, 2010.

[4] M.-H. Masson, S. Destercke, and T. Denoeux. Modelling and predicting partial orders from pairwise belief functions. *Soft Computing*, 20(3):939–950, 2016.

[5] J. Moras, V. Cherfaoui, and P. Bonnifait. Credibilist occupancy grids for vehicle perception in dynamic environments. In *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, pages 84–89. IEEE, 2011.