Thesis Proposal : "Towards hybrid and explainable recommender systems mixing content analysis and collaborative filterings"

General context

Over the last twenty years, an increasing attention has been paid to recommender systems, widely popularized by the Netflix Challenge. The main goal of a recommender system is to provide some users, with personalized products, taking into account their profile and preferences.

Recent challenges are about the recommendation of products very complex to describe : jobs, partners... Their characteristics can mix heterogeneous features : quantitative (as ratings) and/or qualitative (as reviews).

Moreover, new questions are emerging about explainability of algorithms. Nowadays, Artificial Intelligence algorithms are democratized in our erveyday life, and consumers want to understand the decision resulting from these algorithms (why this decision and not another one ?) as well as quantify the importance of each factor (element) in the decision process (which element is the most important/sensitive). They require more *explainability* of AI algorithms.

In addition, the new European legislation on data protection foresees to impose more *transparency* to Artificial Intelligence algorithm. The law envisages to make compulsory the agreement of users for using personal data, which will reduce the amount of data that can be collected about users. The customer will also have to be informed about the way their personal data is used. From the algorithms point of view, the decrease of data will impact the quality of the recommmendations.

All these changes, will impact shortly and significantly the design of algorithms. In this thesis, we aim at designing and implementing new explainable and transparent recommender systems for complex products, in the frame of data sparsity.

Scientific challenges and program

The challenges are four fold :

- **Definition**, in a quantitative way, of the concept of transparency, and develop statistical methods to automatically quantify the transparency degree of an algorithm.

- **Classification** of recommender systems from the literature, from the transparency point of view and/or robustness degree with respect to missing data

- **Conception** of new hybrid and explainable recommender systems, robust to sparse data. The products being complex, the heterogeneous descriptions of the products, as well as the multi-sources of information, will be used to construct understandable explanation. Especially, natural language processing, and hybrid (content/social) approaches will be studied. The algorithms will also be able to quantify the weights and the sensitivity of each factor in the final decision.

- **Constitution** of data sets, allowing to evaluate transparency of recommender systems

Application

The application should include a brief description of research interests and past experience, a CV, degrees and grades, a copy of Master thesis (or a draft thereof), motivation letter (short but pertinent to this call), relevant publications (if any), and other relevant documents. Candidates are encouraged to provide letter(s) of recommendation or contact information to reference persons. Please send your application in one single pdf to :

`armelle.brun@univ-lorraine.fr`
`marianne.clausel@univ-lorraine.fr`

The application of the preselected candidates will be reviewed by the Doctoral School IAEM of University of Lorraine in June 2018 for completing the selection process.

Practical informations

Duration : 3 years (full time position)
Starting date : October, 2018
Supervisors :
A. Brun, University of Lorraine/LORIA, France, (`https://members.loria.fr/ABrun/`)
M. Clausel, University of Lorraine/IECL, France, (`https://sites.google.com/site/marianneclausel/`)

Working Environment : The PhD candidate will work between the Probability and Statistic team of the IECL lab and the KIWI Team of the LORIA lab which are two leading institutions, respectively in Mathematics and Computer

Science in France. The two labs are both located at Nancy, France on the same campus.

The Probability and Statistic team of IECL is working on interdisciplinary projects involving probabilistic modeling and inference methods, with a focus on many applications as textual datas, biology, spatial datas...

The KIWI team of LORIA is a dynamic group working on recommender system and connected scientific domains over 20 researchers (including PhD students) and that covers several aspects of the subject from theory to applications, including statistical learning, data-mining, and cognitive science.

Location : Nancy, which is the capital of Lorraine in France, with excellent train connection to Luxembourg (1h30) and Paris (1h30).
Salary after taxes : around 1600 euros.