Internship Project
A biclustering approach to Recommender Systems

Advisors: A. Mucherino and D. Frey; *Teams*: MimeTIC and ASAP

Research Project

Given a set of user profiles U, represented by a list of items J, Recommender Systems (RS) have the very hard task to suggest, to a given user $u \in U$, the item that a group of different users, having a profile very close to u, have already appreciated. Although there exist several algorithms that are currently employed by important web companies , such as Amazon [4], Facebook [5], Netflix [7], and obviously Google [6], there is a growing need to develop both more efficient solutions as well as techniques that can achieve good recommendation while protecting privacy [8].

This internship project focuses on the development of a novel algorithm for RS that is based on the concept of consistent biclustering. Biclustering is a common approach to data analysis in the context of data mining [3]. It was traditionally applied to biological data, in an unsupervised or supervised fashion, and subsequently extended for feature selection [2]. This latest development is particularly interesting for RS, because the associated method is able to select, from a given database U, the list of *features* in J (used for representing the samples $u \in U$) that are mostly relevant for the classification of the samples in the database.

The first step in this internship will therefore consist in adapting and extending the existing tools for feature selection by biclustering to RS. Given an initial database U of user profiles, we will use a clustering techiques for obtaining an initial classification of such profiles. This classification will then be the starting point for constructing a *consistent* biclustering capable of grouping similar profiles, together with the item lists that cause in fact this similarity. The construction of a consistent biclustering will allow us to select only the relevant items: all unknown item values for the users belonging to the same bicluster can potentially be "recommended" in accordance with the values corresponding to the other users. Once this first step is complete, we will evaluate the biclustering approach in terms of recommendation quality, and we will explore whether we can employ it to offer better privacy guarantees than existing systems [8].

Environment

The candidate will work in the joined Inria/IRISA research centre located in Rennes. Inria (`www.inria.fr`) and IRISA (`http://www.irisa.fr`) are amongst the leading research centres in Computer Sciences in France. The internship will be hosted by the ASAP team, and the work will be performed in collaboration with the MimeTIC team.

Requirements for candidacy

- Python, Java, or Matlab programming skills

- Solid mathematical background

- Previous experience in the research topic is a plus

References

[1] A. Mucherino, *Extending the Definition of β-Consistent Biclustering for Feature Selection*, IEEE Conference Proceedings, Federated Conference on Computer Science and Information Systems (Fed-CSIS11), Workshop on Computational Optimization (WCO11), Szczecin, Poland, 269–274, 2011.

[2] A. Mucherino, L. Liberti, *A VNS-based Heuristic for Feature Selection in Data Mining.* In: "Hybrid Meta-Heuristics", Studies in Computational Intelligence **434**, E-G. Talbi (Ed.), 353–368, 2013.

[3] A. Mucherino, P. Papajorgji, P.M. Pardalos, *Data Mining in Agriculture*, 274 pages, Springer, 2009.

[4] G. Linden, B. Smith, and J. York, "Amazon.com recommendations: Item-to-item collaborative filtering," *IEEE Internet Computing*, vol. 7, no. 1, pp. 76–80, 2003.

[5] Recommending items to more than a billion people. `https://code.facebook.com/posts/861999383875667/recommending-items-to-more-than-a-billion-people/`. Accessed on 19/09/2017.

[6] Abhinandan S. Das, Mayur Datar, Ashutosh Garg, and Shyam Rajaram. Google news personalization: Scalable online collaborative filtering. In *Proceedings of the 16th International Conference on World Wide Web*, WWW '07, pages 271–280, New York, NY, USA, 2007. ACM.

[7] Carlos A. Gomez-Uribe and Neil Hunt. The netflix recommender system: Algorithms, business value, and innovation. *ACM Trans. Manage. Inf. Syst.*, 6(4):13:1–13:19, December 2015.

[8] Davide Frey, Rachid Guerraoui, Anne-Marie Kermarrec, Antoine Rault, Francois Taiani, and Jingjing Wang. Hide & Share: Landmark-Based Similarity for Private KNN Computation. In *2015 45th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN)*, pages 263–274, June 2015.

Contacts

We are looking for motivated candidates!
Please send your CV by email, together with a motivation letter and any relevant material:

Antonio Mucherino
Web: `http://www.antoniomucherino.it`
Email: `antonio.mucherino@irisa.fr`

Davide Frey
Web: `https://www.irisa.fr/asap/?page_id=246`
Email: `davide.frey@inria.fr`