MSc Internship – Stage M2 2017-18 : Decentralized Similarity Learning for Peer-to-Peer Recommenders

1 Supervision

Advisers	François Taiani, `francois.tainai@irisa.fr`, 02 99 84 75 04
	Marc Tommasi (Magnet Team, Inria Lille), `marc.tommasi@inria.fr`
	Aurélien Bellet (Magnet Team, Inria Lille), `aurelien.bellet@inria.fr`
Lab	IRISA (UMR 6074)
Team	ASAP (As Scalable As Possible)
	Équipes-Projet Inria / Département D1 IRISA

2 Context

With the growth of on-line content, personalized recommendation has emerged as a key service to help users navigate today's extensive digital universe. To provide personalization, most companies rely today on *centralized* or *tightly coupled machine learning systems* hosted in data centers or in the cloud. This is problematic as this concentration poses strong risks to the privacy of users, and limits the scope of ML applications to tightly integrated datasets under unified learning models.

One promising approach to overcome these limitations consists in exploiting fully decentralised mechanisms [1, 4, 5]. Decentralization is now a growing trend in machine learning research [6], and is attracting considerable attention, including from industrial players.

Decentralized recommenders, however, have so far used a mostly homogeneous design. They typically rely on one homogeneous similarity metric [7] to self-organise large numbers of users in implicit k-nearest-neighbor communities and exploit collaborative filtering. Figuring out the right similarity metric that best fits the needs of a large collection of users is, however, highly challenging.

Our earlier work has shown that allowing individual users to exploit different similarity metrics could greatly improve the quality of recommendations [3]. We have also demonstrated how these similarities could be dynamically selected at run-time on a prototype decentralized algorithm called *Similitude/ [2]. In effect /Similitude* provides in effect a lightweight form a *decentralized similarity learning* applicable to peer-to-peer recommenders.

3 Objective

The objective of this MSc project is to **better understand the convergence properties of *Similitude***. To this aim, **the current algorithm needs to be reformulated** and adapted in order to become more amenable to a state-of-the-art recursive analysis. If time permits, we would also be interested in exploring how privacy preserving mechanisms (such as randomization and obfuscation) could be added to the system while maintaining the quality of its recommendations.

4 Tasks

- Perform a bibliographic study of existing decentralized non-parametric learning approaches, and the convergence guarantees they provide.
- Revisit the Similitude algorithm in the light of recent advances in decentralized learning techniques, and formulate a new version that is more amenable to an analytic investigation.
- Explore which guarantees can be demonstrated for the algorithm, for instance by using a recursive matrix-based formulation and an argument on the matrix's maximum eigenvalue.
- Experiment with both synthetic and real datasets the actual behavior of the algorithm.

References

[1] Marin Bertier, Davide Frey, Rachid Guerraoui, Anne-Marie Kermarrec, and Vincent Leroy. The gossple anonymous social network. In *Middleware'2010*.

[2] Davide Frey, Anne-Marie Kermarrec, Christopher Maddock, Andreas Mauthe, Pierre-Louis Roman, and François Taïani. Similitude: Decentralised adaptation in large-scale p2p recommenders. In *Distributed Applications and Interoperable Systems*, pages 51–65. Springer, 2015.

[3] Anne-Marie Kermarrec and François Taïani. Diverging towards the common good: heterogeneous self-organisation in decentralised recommenders. In *SNS'2012*.

[4] V. Leroy, B. B. Cambazoglu, and F. Bonchi. Cold start link prediction. In *KDD'2010*.

[5] Juan M Tirado, Daniel Higuero, Florin Isaila, Jesús Carretero, and Adriana Iamnitchi. Affinity p2p: A self-organizing content-based locality-aware collaborative peer-to-peer network. *Comp. Net.*, 54, 2010.

[6] Paul Vanhaesebrouck, Aurélien Bellet, and Marc Tommasi. Decentralized collaborative learning of personalized models over networks. In Aarti Singh and Xiaojin (Jerry) Zhu, editors, *20th Int. Cont. on Artificial Intelligence and Statistics, AISTATS*, pages 509–517, 2017.

[7] S. Voulgaris and M. v. Steen. Epidemic-style management of semantic overlays for content-based searching. In *Euro-Par'05*.