Privacy-preserving recommender systems for e-commerce & health services

Fran Casino*

SMART HEALTH Research Group. Department of Computer Engineering and Mathematics, Rovira i Virgili University.
Av. Països Catalans 26. 43007 Tarragona. Catalonia. Spain
Corresponding author e-mail: franciscojose.casino@urv.cat

1 Introduction

Recommender systems [6] evolve from the field of knowledge discovery in databases (KDD) [7]. Systems for KDD are used by companies to discover understandable patterns within large collections of data, which might help, for example, to save money, make better strategic decisions or sell more products. Collaborative Filtering (CF) [4] is a recommender system that comprises a large family of recommendation methods. The aim of CF is to make suggestions on a set of items (I) (*e.g.* books, music, films, monuments or routes), based on the preferences of a set of users (U) that have already acquired and/or rated some of those items.

In order to make recommendations (*i.e.* to predict whether an item would please a given user) CF methods rely on large databases[2] with information regarding the relationships between sets of users and items. These data take the form of matrices composed by n users and m items, and each matrix cell (i, j) stores the evaluation of user i on item j. The recommendations provided by CF methods are based on the assumption that similar users will be interested in the same items. As a result, items well rated by a user u_a could be recommended to another user u_b, if u_a and u_b are similar.

CF methods are classified into three main categories that depend on the data they use as follows: (i) memory-based methods, which use the full matrix with the users' ratings, (ii) model-based methods, which use statistical models and functions of the data matrix but not the data matrix itself, and (iii) hybrid methods, which combine the previous methods with content-based [8] recommendation methods.

* PhD advisor: Agustí Solanas

[2] There are many examples of CF databases [8] referenced in the literature, like Eachmovie, MovieLens, Jester, and Netflix prize data. These databases are frequently used as benchmarks to evaluate the efficiency, quality and robustness of CF methods [5].

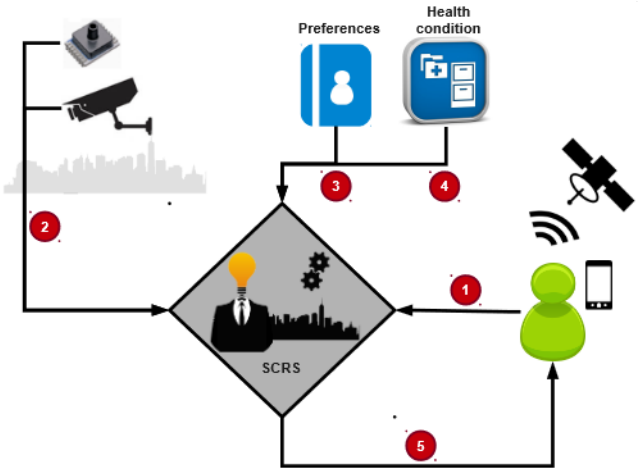

Fig. 1: General scheme and basic operation of a s-health recommender system.

2 Recommender Systems and Smart Health

Nowadays, great efforts are being devoted to build the cities of the future (*i.e.* Smart cities), which will be equipped with full of sensors and actuators (*e.g.* temperature and humidity sensors, pollution and allergens sensors, luminosity sensors or crowds detectors) that would improve the citizens' quality of life. Moreover, the healthcare sector has turned into them to create a powerful symbiosis and create smart health [1] (s-health), which is defined as *the provision of health services by using the context-aware network and sensing infrastructure of smart cities*. Therefore, considering such definition, any application/service that uses the smart city infrastructure to provide healthcare or to promote healthy habits amongst the city population could be considered a smart health application. Hence, great opportunities emerge to consolidate the concept of s-health, for instance, the creation of systems that benefit from this concept and augment it with other powerful well-known information filtering systems such as recommender systems.

It is well-known that many citizens perform physical activities in the city, namely cycling, jogging, running, etc. With the aim to promote these healthy habits, it would be desirable to count with a system that could dynamically adapt to the needs and tastes of the citizens. Within this context, we propose a new way of using the sensing capabilities of smart cities by means of recommender systems that allow citizens/patients to obtain recommendations about the routes that better fit their capacities.

The system would consider real-time constraints and information from several sources: (i) citizens' preferences, (ii) citizens' health conditions and, (iii) real-time information provided by the smart city infrastructure.

An overview of the general scheme of our system architecture and its main actors is shown in Figure 1. Sensors provide real-time environmental information (*e.g.* luminosity, temperature, humidity, pollution) to the Smart City Recommender System (SCRS) through the communication infrastructure of the smart city. Upon the reception of citizen queries, the SCRS checks the health information of citizens and their preferences and cross them with the real-time information of the smart city sensors to finally compute real-time recommendations that are forwarded back to the citizens.

3 Privacy and Collaborative Filtering

The widespread use of CF on the Internet entails great opportunities for both companies and users in multiple contexts. However, the lack of privacy for the contributing users is a major drawback. The relevance of privacy in CF systems is emphasised by the growing pace at which information on each user is collected and stored. Careless management of personal information, apart from being illegal in many countries, has potentially serious consequences for both the users and businesses whose information is disclosed. One of the main problems in CF is that, if customers believe their preferences/profiles may be exposed, they might decide either not to give their assessment on a particular item or to give it incorrectly or inaccurately. Therefore, the feeling of poor privacy protection results in a reduction of the number and quality of evaluations.

Another drawback is that companies can acquire data about the preferences of many users in a given market, getting a big advantage over new competitors if they decide to expand into other markets. Therefore, user profiling through CF promotes in some sense monopolies. Another privacy-related drawback for users comes from the existence of large Internet quasi-monopolies, which massively gather users' preferences and may transfer them within their web of partnered companies in hardly traceable ways, leading to further user profiling.

Whilst privacy preserving CF methods obfuscate and/or hide information on user profiles, sometimes users wish to find other users having similar profiles and form a community. Indeed, communities are very usual in the network, but they can be a double-edged sword. On the one hand, users can conveniently obtain reliable recommendations on items from communities in a particular context. On the other hand, communities can generate a *value homophily* problem in the network, so that recommendations outside the context of the community would give results with little sense, precisely because of the homogeneity of the group.

Therefore, in order to solve the privacy issues raised by the systematic collection of private information on preferences, the Privacy Preserving Collabo-

rative Filtering (PPCF) concept appears [2,3] with the aim to provide quality recommendations without compromising the privacy of users involved.

4 Conclusions and Future Work

Collaborative Filtering is a recommender system used to perform automatic recommendations to users in multiple contexts. Despite the great advantages of using CF, we have highlighted its downside regarding users' privacy, which is probably the most significant challenge to overcome.

Moreover, we have proposed the idea of using recommender systems integrated with the sensing infrastructure of smart cities to provide citizens with routes recommendations that take into account their health conditions and preferences. In addition, the recommendations could be adapted in real-time to the environmental changes of the city.

Future work will focus in the implementation of new PPCF methods that improve the results presented in [3], with the aim to achieve a better privacy/accuracy trade-off.

References

[1] A. Solanas, C. Patsakis, M. Conti, I. Vlachos, V. Ramos, F. Falcone, O. Postolache, P. Pérez-Martínez, R. Di Pietro, D. Perrea, and A. Martínez-Ballesté, Smart health: A context-aware health paradigm within smart cities. *IEEE Communications Magazine*, August. 2014 (In press).

[2] F. Casino, C. Patsakis, D. Puig, and A. Solanas, On privacy preserving collaborative filtering: Current trends, open problems, and new issues. *ICEBE*, pp. 244-249. (2013).

[3] F. Casino, J. Domingo-Ferrer, C. Patsakis, D. Puig, and A. Solanas, Privacy preserving collaborative filtering with k-anonymity through microaggregation. *ICEBE*, pp. 490-497. (2013).

[4] Goldberg, D., Nichols, D., Oki, B. M., Terry, D., *LaTeX User's Guide and Document Reference Manual. Communications of the ACM*, 35(12), 6170. (1992).

[5] Herlocker, J. L., Konstan, J. a., Terveen, L. G., Riedl, J. T., Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems*, 22(1), 553. (2004).

[6] Resnick, P., Varian, H., Recommender systems *Communications of the ACM*, 40(3), 5658. (1997).

[7] Sarwar, B., Karypis, GeorgeKonstan., JRiedl, J., Using collaborative filtering to weave an information tapestry. *ACM WebKDD 2000 Web Mining for ECommerce Workshop*, 1625(1), 2648. (2000).

[8] Su, X., Khoshgoftaar, T. M., A Survey of Collaborative Filtering Techniques. *Advances in Artificial Intelligence*,(Section 3) 119. (2009).