Chapter 10

An Enhanced Semantic Layer for Hybrid Recommender Systems: Application to News Recommendation

Iván Cantador
Universidad Autónoma de Madrid, Spain

Pablo Castells
Universidad Autónoma de Madrid, Spain

Alejandro Bellogín
Universidad Autónoma de Madrid, Spain

ABSTRACT

Recommender systems have achieved success in a variety of domains, as a means to help users in information overload scenarios by proactively finding items or services on their behalf, taking into account or predicting their tastes, priorities, or goals. Challenging issues in their research agenda include the sparsity of user preference data and the lack of flexibility to incorporate contextual factors in the recommendation methods. To a significant extent, these issues can be related to a limited description and exploitation of the semantics underlying both user and item representations. The authors propose a three-fold knowledge representation, in which an explicit, semantic-rich domain knowledge space is incorporated between user and item spaces. The enhanced semantics support the development of contextualisation capabilities and enable performance improvements in recommendation methods. As a proof of concept and evaluation testbed, the approach is evaluated through its implementation in a news recommender system, in which it is tested with real users. In such scenario, semantic knowledge bases and item annotations are automatically produced from public sources.

DOI: 10.4018/978-1-4666-3610-1.ch010
1. INTRODUCTION

The growing volume and complexity of digital information (news, blogs, music, movies, etc.) of potential value to our daily activities challenge the limits of human processing capabilities in a wide array of information seeking and e-commerce tasks. Users need help to cope with a wealth of readily available information, in order to reach the most interesting items in online retrieval spaces. The problem is not just to find the needle in the haystack, but to select the best among thousands of needles. In such information overload scenarios, recommender systems become particularly appealing as a means to help users make choices, by proactively finding items or services on their behalf, taking into account or predicting their tastes, priorities or goals.

Recommender systems came forth by the early nineties as an emerging area at the confluence of Artificial Intelligence and Information Retrieval, to address the essential research problem of predicting or estimating the relevance of items that a user has not seen or searched for. The way in which the estimation is computed raises the distinction of two main recommendation strategies (Adomavicius & Tuzhilin, 2005): content-based filtering (CBF) strategies, which predict the relevance of an item for a user according to the relevance that other similar items seemed to have for him in the past, and collaborative filtering (CF) strategies, which predict the relevance of an item for a user by considering the relevance that other items had in the past for similar people.

It has been generally observed that combining CBF and CF methods, known as hybrid recommendation, is usually the best approach to mitigate the limitations CBF and CF approaches suffer separately (Adomavicius & Tuzhilin, 2005). Hybrid recommendation approaches are becoming an integral part of a large number of important e-commerce and leisure Web sites like Amazon, Netflix, and many online retailers, where recommendation models have proved successful. Nonetheless, ample room and need for further improvements remains in the current generation of recommender systems to achieve more effective algorithms, in a wider variety of applications. These improvements include, among others:

- Better coping with low data density situations in the available evidence of user preference, e.g., user input scarcity in initial situations (cold-start problem). When a new user enters a recommender system, no personal profile is yet available for him, and no proper recommendations can be made. Similarly, until a new item is rated by a substantial number of users, a CF recommender is not able to recommend it, as observed user preference for specific items is the data CF relies upon. This problem is particularly prevalent in sparse domains such as the News, where there is a constant stream of new items, and each user only rates a few.

- The consideration of contextual information in the recommendation strategies. Traditional recommenders operate on the two-dimensional Users×Items space, i.e., they make recommendations based solely on user and item information, and do not take into consideration additional contextual information that may be crucial in some applications. In many situations, the utility of a certain item to a user may largely depend on the circumstances under which the item would be utilised, or the temporary purpose and changing goals of users with respect to the items. Using multidimensional settings, the inclusion of knowledge about the user’s task, goals, environment, etc., into the recommendation algorithm can lead to better recommendations.

Among other directions, the enhancement of the semantics representation of user preferences and item contents is being identified as a key