User Modeling, Adaptation, and Personalization: 21th International Conference, UMAP 2013, Rome, Italy, June 10-14, 2013 Proceedings. Lecture Notes in Computer Science, Volumen 7899. Springer, 2013. 371-375

DOI: http://dx.doi.org/10.1007/978-3-642-38844-6_42

Mining Semantic Data, User Generated Contents, and Contextual Information for Cross-domain Recommendation

Ignacio Fernández-Tobías

Universidad Autónoma de Madrid, Madrid, Spain
i.fernandez@uam.es

Abstract. Cross-domain recommender systems suggest items in a target domain by exploiting user preferences and/or domain knowledge available in a source domain. In this thesis we aim to develop a framework for cross-domain recommendation capable of mining heterogeneous sources of information such as semantically annotated data, user generated contents, and contextual signals. For this purpose, we investigate a number of approaches to extract, process, and integrate knowledge for linking distinct domains, and various models that exploit such knowledge for making effective recommendations across domains.

Keywords: Cross-domain recommender systems, semantic networks, social tagging, context-aware recommendations.

1 Introduction

To cope with information overload recommender systems aim to find the items that best suit the user's tastes, interests and priorities, without requiring her to launch explicit search queries, as is usually done in information retrieval systems.

Most of current recommender systems target items in a single domain. For instance, Netflix makes personalized recommendations of movies and TV series, and Last.fm suggests music compositions and artists. E-commerce sites like Amazon, however, could take benefit from exploiting the user's preferences on diverse types of items to provide recommendations in different but somehow related domains. Recommendations across domains could mitigate the cold-start problem when little information about the user's preferences is available in a target domain, and are potentially more diverse and serendipitous than single-domain recommendations.

The goal of cross-domain recommender systems is to suggest items in a target domain by exploiting user preferences and/or domain knowledge available in a source domain. Hence, a major challenge for making cross-domain recommendations is how to transfer knowledge across different domains, i.e., how to build *bridges* between domains, in which usually there is little or no overlap between user preferences or item attributes. Motivated by this fact, in this thesis we aim to address the following research question: **How to effectively extract and exploit knowledge from a source domain to improve recommendations in a distinct, target domain?**

For such purpose, we shall investigate a number of methods that exploit different data sources to establish *links* between domains, and shall develop recommendation algorithms that make use of such links. We hypothesize these links will let improve recommendation approaches based only on user preference overlap between domains.

Specifically, we propose to exploit semantically annotated data, user-generated contents (social tags, in particular), and contextual information.

In the so-called Semantic Web, and in its reference implementation Linked Data, there is a large number of inter-connected ontologies that describe and relate a wide array of domains. From these ontologies a system that identifies and extracts semantic concepts and properties (relations) belonging to particular domains of interest, could build and exploit semantic networks linking such domains. Similarly to ontologies, social tags assigned to items in folksonomies represent vocabularies describing distinct domains. A system able to somehow find related tags across folksonomies could establish similarities between social tag-based user and item profiles in source and target domains. Finally, contextual signals, such as time, location, mood, and social companion may be used by a system to establish domain-independent similarities between user preferences and item attributes in different domains.

2 Related work

Cross-domain recommendation has been mostly addressed in the collaborative filtering setting, where there is some user preference (rating) overlap between domains, and where item attributes are not needed. In a seminal paper Winoto and Tang [10] identify three issues to investigate in cross-domain recommendation: the existence of global correlations between user preferences for items in different domains, the method to exploit data on a source domain for predicting preferences on a target domain, and the methodology and metrics to evaluate cross-domain recommendations. In [5] Li surveys works on cross-domain collaborative filtering. He classifies existing approaches according to the type of knowledge transferred: rating patterns, latent features, and user/item inter-domain correlations. Tiroshi and Kuflik [9] evaluate the influence of the involved domains in the recommendation using a kNN approach in which the neighborhood of a user in the target domain is selected among the most similar users from a source domain. To address the problem of little domain overlap, Cremonesi et al. [2] model user and item similarities through graphs, in which all possible paths linking two users or items are used to enhance user- and item-based neighborhood algorithms.

Besides collaborative filtering there have been some attempts to establish semantic relations between items of different domains. In [6] Loizou annotates and links items by means of concepts and relations extracted from Wikipedia. Then, using such relations, users and items are incorporated into a graph upon which a probabilistic recommendation model is built. Social tags have also been used to establish relations between items of different domains. In [4] Kaminskas and Ricci show that emotional tags can be used to effectively select music that fits places of interest. Shi et al. [8] utilize tags to build inter-domain user-to-user and item-to-item similarities. These similarities are proportional to the numbers of tags shared by profiles from different domains, and are incorporated as constraints into a probabilistic matrix factorization model for collaborative filtering. To the best of our knowledge, there is no work in the literature addressing the cross-domain recommendation task by means of contextual features. We believe that a recommender system could exploit relations established through contextual signals, such as the time (e.g. movies and music compositions

usually consumed on Christmas), and the user's mood and emotions (e.g. movies and music compositions that usually yield nostalgic feelings).

3 Proposed solution and future work

The main contribution we expect to achieve in this thesis is the development of a **framework for cross-domain recommendation capable of mining heterogeneous sources of information** such as semantic data, social annotations, and contextual features. For this purpose, we aim to investigate a number of approaches to extract, process, and integrate knowledge for linking domains, and various models that exploit such knowledge for making effective cross-domain recommendations.

3.1 Crossing domains with semantic relations

In [1] and [3] we present an approach that exploits the multi-domain DBpedia ontology (http://dbpedia.org) to build semantic networks linking items from different domains. The approach consists of two main components: a class-level network of DBpedia classes and relations belonging to certain domains of interest, and instance-level networks that are automatically generated instantiations of the class-level network for particular input items, by extracting and filtering information from the above ontology. On the built instance-level networks, a graph-based spreading activation algorithm ranks items in the target domain according to their *semantic relatedness* with the input items.

We have evaluated our approach instantiating the framework for recommending musicians and music compositions semantically related to places of interest. Through two user studies we have shown our approach is able to effectively identify music items relevant for the users and suited to the target places. Current work is focused on using open information extraction tools like ReVerb (http://reverb.cs.washington.edu) to discover complex, arbitrary semantic relations across domains that are not available in DBpedia. Moreover, we shall evaluate the framework in other domains such as movies, music and books, and shall enhance it by incorporating user preferences as prior relevance values to compute personalized semantic similarities.

3.2 Crossing domains with social tags

As shown in [4] and [8] tags can play the role of a common feature space for user/item profiles from distinct domains. In [4] Kaminskas and Ricci show that a limited set of emotional tags assigned to both music and places can reveal latent similarities between both types of items. In [8] Shi et al. use the tags common to two folksonomies for computing similarities between users/items in the folksonomies' domains.

We hypothesize that generic lexicons can be built with social tags, and that these lexicons can be used to generate profiles which would act as *bridges* to compute cross-domain user/item similarities. We have developed an approach that automatically builds a lexicon that describes core emotions (e.g. *happy*, *sad*), by mining synonyms and antonyms extracted from the online Dictionary.com thesaurus. We have proved that our emotion representation is in accordance with the well-known psychological circumplex model of affect [7]. Moreover, using tag co-occurrences, we have linked the lexicon with domain-specific emotion folksonomies extracted from social tagging systems, which let us transform domain-specific tag-based profiles into generic emotion-oriented profiles. For instance, in our approach, a movie annotated with tags associated to the

emotional category *suspenseful* strongly overlaps with the *tense* and *nervous* core emotions of the circumplex model. We have conducted a user study showing our method accurately infers emotions evoked by items in the movie and music domains.

Our current work is aimed to exploit the emotion-oriented profiles for making personalized cross-domain recommendations sensible to the user's mood. We also intend to explore other sources of user generated contents such as textual reviews and comments to build generic lexicons suitable for opinion mining and sentiment analysis.

3.3 Crossing domains with contextual information

Contextual information can be also valuable inter-domain features that characterize user preferences. Context features can be obtained through many signals such as the user's current time, location, and social companion. So far we have explored a context modeling approach, formulating the recommendation task as a classification problem, and investigating if using or not those features in addition to content attributes actually leads to better accuracy. Preliminary empirical results have indicated that time context has preference predictive power in the movies and music domains, and it is improved by including social context in the movies domain. Nonetheless, we still have to evaluate if users find common contexts meaningful enough for cross-domain recommendation. Future work in this direction begins with an analysis of correlations between context features and user preferences in different domains.

References

1. Fernández-Tobías, I., Kaminskas, M., Cantador, I., Ricci, F. A Generic semantic-based framework for cross-domain recommendation. In: 2nd International Workshop on Information Heterogeneity and Fusion in Recommender Systems, pp.25-32 (2011)
2. Cremonesi, P., Tripodi, A., Turrin, R.: Cross-domain recommender systems. In: 2011 IEEE 11th International Conference on Data Mining Workshops, pp. 496–503 (2011)
3. Kaminskas, M., Fernández-Tobías, I., Ricci, F., Cantador, I.: Ontology-based identification of music for places. In: 13th International Conference on Information and Communication Technologies in Tourism, to appear (2013)
4. Kaminskas, M., Ricci, F.: Location-Adapted music recommendation using tags. In: 19th Intl. Conference on User Modeling, Adaptation, and Personalization, pp. 183-194 (2011)
5. Li, B.: Cross-domain collaborative filtering: A brief survey. In: 23rd International Conference on Tools with Artificial Intelligence, pp. 1085–1086 (2011)
6. Loizou, A.: How to recommend music to film buffs: enabling the provision of recommendations from multiple domains. PhD dissertation, Univ. of Southampton (2009)
7. Russell, J. A.: A circumplex model of affect. Journal of Personality and Social Psychology 39(6), pp. 1161–1178 (1980)
8. Shi, Y., Larson, M., Hanjalic, A.: Tags as bridges between domains: Improving recommendation with tag-induced cross-domain collaborative filtering. In: 19th International Conference on User modeling, Adaption, and Personalization, pp. 305-316 (2011)
9. Tiroshi, A., Kuflik, T.: Domain ranking for cross domain collaborative filtering. In: 20th Intl. Conference on User Modeling, Adaptation, and Personalization, pp. 328-333 (2012)
10. Winoto, P., Ya Tang, T.: If you like The devil wears Prada the book, will you also enjoy The devil wears Prada the movie? A study of cross-domain recommendations. New Generation Computing 26(3), pp. 209-225 (2008)