Cross-domain recommender systems: A survey of the State of the Art

Article · January 2012

4 authors, including:


Iván Cantador
Universidad Autónoma de Madrid


Marius Kaminskas
Tele 2


Francesco Ricci
Free University of Bozen-Bolzano

Cross-domain recommender systems: A survey of the State of the Art

Ignacio Fernández-Tobías[1], Iván Cantador[1], Marius Kaminskas[2], Francesco Ricci[2]

[1] Escuela Politécnica Superior
Universidad Autónoma de Madrid
28049 Madrid, Spain
{i.fernandez, ivan.cantador}@uam.es

[2] Faculty of Computer Science
Free University of Bozen-Bolzano
39100 Bolzano, Italy
{mkaminskas, fricci}@unibz.it

Abstract. Cross-domain recommendation is an emerging research topic. In the last few years an increasing amount of work has been published in various areas related to the Recommender System field, namely User Modeling, Information Retrieval, Knowledge Management, and Machine Learning. The problem has thus been addressed from distinct perspectives. Hence there are even conflicting definitions of the cross-domain recommendation task, and there is no rigorous comparison of existing approaches. In this paper we provide a formal statement of the problem, and present a review of the state of the art. We also establish a general taxonomy that let us to better characterize, categorize and compare the revised work. Finally, we conclude this review with a survey of interesting research topics on cross-domain recommendation.

Keywords: recommender systems, cross-domain recommender systems, knowledge integration, transfer learning

1 Introduction

The huge and ever increasing amount, complexity and heterogeneity of available digital information overwhelm the human processing capabilities in a wide array of information seeking and e-commerce tasks. To cope with information overload recommender systems have been introduced to filter those items –Web pages, images, videos, audio– that are of low relevance or utility for the user, and present only a small selection better suiting the user's tastes, interests, and priorities. Often these suggestions are presented while the user is browsing an information service, and without requiring her to launch explicit search queries, as is usually done in information retrieval systems.

Recommender systems are an active research field and are being used successfully in numerous e-commerce and leisure Web sites such as Amazon, Netflix, YouTube, iTunes, and Lasf.fm. The vast majority of these systems offer their recommendations only for items belonging to a single domain. Hence, for example, Netflix suggests movies and TV series, and Last.fm makes personalized recommendations of music artists and compositions. In both cases the recommendations are computed using user feedback (ratings) about items in the target domain. In e-commerce sites like Amazon, nonetheless, it would be useful to exploit the user's evaluations about diverse types of items in order to generate a more general model of the user preferences. In fact, there could be dependencies and correlations between preferences in different domains and instead of treating each type of items (e.g. electronics and music) independently, user knowledge acquired in one domain could be transferred and exploited in several other domains. Moreover, although it is not the main goal of cross-domain recommendation, a system could offer joint, personalized recommendations of items in multiple domains, e.g. suggesting not only a particular movie, but also music CDs, books or videogames somehow related with that movie. Analogously, in a touristic application it would be valuable to suggest a cultural event to a customer who has booked a room in a recommended hotel, or in an e-learning system, to present a student with bibliographic references related to a video-lecture that has been recently recommended.

Some systems already offer joint recommendations of items in different domains, but in general, in order to build a recommendation in one domain, they exploit only user preferences on that target domain. Cross-domain recommendation is thus an emerging research topic in the Recommender Systems area. One of the first studies on cross-domain recommendation was that presented by Winoto and Tang in [20]. In that work, the authors identify three important issues to investigate: a) verifying the existence of global correlations of user preferences for items in different domains, b) designing models able to exploit user preferences on a source domain for predicting user preferences on a target domain, and c) developing evaluations appropriate for cross-domain recommendations.

Winoto and Tang speculate that, although cross-domain recommendations may tend to be less precise than single-domain recommendations, the former will be more diverse, which may lead to a higher user satisfaction and engagement [2]. Moreover, cross-domain recommendation techniques may have other advantages, such as addressing the cold-start problem [1], [4] and mitigating the sparsity problem [10], [15]. By identifying relations between items in two different domains, one could suggest a user with items in a novel, unexplored domain, simply exploiting her preferences for items in other known domains.

Despite the above benefits and work, this research topic is quite new and still largely unexplored. Hence, to the best of our knowledge, in the literature an agreed definition of the cross-domain problem has not emerged so far and there is not a general classification of existing approaches. In fact, the issue of how to exploit information from various domains to provide recommendations has been addressed from distinct perspectives in diverse disciplines such as User Modeling [1], [18], Information Retrieval [4], [5], [20], Knowledge Management, and Machine Learning [10], [11], [15], [21].

In this paper we present a survey of the state of the art on cross-domain recommendation, analyzing how diverse domains can be related, and providing a taxonomy that lets characterizing, categorizing and comparing the revised work, and identifying interesting unexplored research issues.

The reminder of the paper is structured as follows. In Section 2 we propose a formal statement of the cross-domain recommendation task, starting from a discussion about the notion of domain and the potential sources of relations between domains. We also provide a classification of existing cross-domain recommendation techniques, and a description of used evaluation methodologies. In Sections 3, 4 and 5 we describe and compare revised approaches in accordance with the identified techniques. Finally, in Section 6 we present potential research lines on the topic.

2 Cross-domain recommendation

Cross-domain recommendation has been addressed recently from different perspectives, and in diverse research areas. Based on such perspectives, in this section we provide a general definition of the cross-domain recommendation task, starting from a discussion about the notion of domain, and a description of the types of relations between domains.

2.1 Notion of domain

In the literature on cross-domain recommendation there is no a consensus on the notion of domain. Authors have utilized the term *domain* sometime for referring to types of items (e.g. movies vs. books) or in other cases to groups of similar items with common characteristics (e.g. movies vs. TV shows). This may be due to the lack of public datasets including users' evaluations on diverse types of items, to be used in the evaluation of the proposed cross-domain recommendation approaches. In fact, several authors have considered artificial data splits to simulate different domains but using data coming from one, single domain, dataset. For instance, [4], [20] and [21] group movies based on their genres (e.g. drama, comedy, thriller), and consider these groups as different movie domains.

In [9] Li presents a review of collaborative filtering approaches for cross-domain recommendation, distinguishing three types of domains: system domains, data domains, and temporal domains. System domains are the different datasets upon which the recommender systems are built, and in which some kind of transfer learning is performed. Data domains are the different representations of user preferences, which can be implicit (e.g. clicks, purchases) or explicit (e.g. ratings). Finally, temporal domains are subsets in which a dataset is split based on timestamps. We consider a notion of domain similar to Li's definition of system domain, which is the one more frequently used in the literature. We define domain as a set of items that share certain characteristics that are exploited by a particular recommender system. As we shall explain, these characteristics can be manifold, e.g. content attributes, ratings, and tags.

2.2 Statement of the cross-domain recommendation task

To date the large majority of the proposed approaches to cross-domain recommendation deals with collaborative filtering (CF). Collaborative filtering strategies exploit user preferences (usually expressed as explicit ratings for items), and ignore any content-based description (attributes) of the items. This feature of CF represents a great advantage when the items belong to heterogeneous sources. In the following we propose a definition of the cross-domain recommendation task that is valid for both content-based and collaborative filtering approaches. Without loss of generality we consider the task of cross-domain recommendation when only two domains $\mathcal{A}$ and $\mathcal{B}$ are involved.

Using the notation of [5], let $\mathcal{U}_A, \mathcal{U}_B$ be the sets of users and $\mathcal{I}_A, \mathcal{I}_B$ be the sets of items with "characteristics" (user preferences and item attributes) in the domains $\mathcal{A}$ and $\mathcal{B}$ respectively. We define two cross-domain recommendation tasks:

- Exploit knowledge about users and items in the source domain $\mathcal{A}$ for **improving the quality of the recommendations** for items in the target domain $\mathcal{B}$.
- **Making joint recommendations** for items belonging to different domains, i.e., suggesting items in $\mathcal{I}_A \cup \mathcal{I}_B$ to users in $\mathcal{U}_A \cup \mathcal{U}_B$.

In this context, as in single-domain recommendation, we assume that the cross-domain recommendation task involves personalization. We thus ignore those recommendation strategies, such as popularity-based, that do not take into account the target user's preferences, and simply suggest items positively evaluated by a large number of users. Moreover, we also exclude those cross-type systems that are initially built with user evaluations about items of different types (e.g. movies and books). In these systems the users' preferences already comprise the domains of interest, and thus any collaborative filtering strategy can be used to provide recommendations of items, which can be considered as belonging to a single domain.

We assume a recommendation scenario in which user and item profiles are distributed in multiple systems (domains), and in which we have to establish a mechanism to link or transfer domain knowledge (e.g. content attribute mappings, semantic similarities, rating patterns) between such systems. In this context, classic nearest neighbor strategies are not valid since it is not possible to directly compute rating-based similarities between pairs of users/items.

2.3 Types of explicit relations between domains

According to the type of overlap between domains, in [5] Cremonesi et al. identify four situations in which a cross-domain recommendation task may be conducted: a) no-overlap, $\mathcal{U}_A \cap \mathcal{U}_B = \emptyset \wedge \mathcal{I}_A \cap \mathcal{I}_B = \emptyset$; b) user overlap, $\mathcal{U}_A \cap \mathcal{U}_B \neq \emptyset$; c) item overlap, $\mathcal{I}_A \cap \mathcal{I}_B \neq \emptyset$; and d) full overlap, $\mathcal{U}_A \cap \mathcal{U}_B \neq \emptyset \wedge \mathcal{I}_A \cap \mathcal{I}_B \neq \emptyset$. In all these situations but the first –no-overlap–, we could obtain effective recommendations with a classic collaborative filtering strategy by considering all the user/item ratings belong to a single, common domain. Nonetheless, assuming a memory-based approach to CF, when the overlap is small, user and item similarities, and consequently generated recommendations, may tend to be inaccurate. To address the above mentioned no-overlap situation, we have to develop approaches that find or build some type of explicit/implicit relations between domains, which would be used as semantic bridges connecting different domains in a recommender system.

In the following we extend the ideas presented in [5] to deal with the domain overlap issue in a more flexible way. Instead of focusing only on users/items with rating information in the considered domains, we define (overlap) relations between domains through "characteristics" that are shared by the user/item profiles in the different domains. In a vector space model users can be represented as vectors in which each component is associated to certain characteristic, that is:

$$\mathcal{U} \cong \underbrace{\mathbb{R} \times \ldots \times \mathbb{R}}_{|\mathcal{X}^{\mathcal{U}}|} = \mathbb{R}^m \tag{1}$$

where $\mathcal{X}^{\mathcal{U}} = \{X_1^{\mathcal{U}}, \ldots, X_m^{\mathcal{U}}\}$ is the set of m characteristics used to describe the user preferences. Analogously, items can be represented as:

$$\mathcal{I} \cong \underbrace{\mathbb{R} \times \ldots \times \mathbb{R}}_{|\mathcal{X}^{\mathcal{I}}|} = \mathbb{R}^n \tag{2}$$

where $\mathcal{X}^{\mathcal{I}} = \{X_1^{\mathcal{I}}, \ldots, X_n^{\mathcal{I}}\}$ is the set of n item characteristics. In both cases, as we shall explain, the nature of these characteristics can be quite diverse, e.g. ratings, content attribute-value pairs, social tags, explicit semantic relations, and implicit latent factors.

Upon the above representation domains $\mathcal{A}$ and $\mathcal{B}$ will be linked if $\mathcal{X}_A^{\mathcal{U}} \cap \mathcal{X}_B^{\mathcal{U}} \neq \emptyset$ or $\mathcal{X}_A^{\mathcal{I}} \cap \mathcal{X}_B^{\mathcal{I}} \neq \emptyset$, i.e., if there are user or item characteristics shared by the two domains. Obviously, the larger the number of shared characteristics, the more robust the relation between domains. In a real situation, due to the heterogeneity of domain representations, which may be provided by various systems, we will have to establish a number of functions mapping characteristics between domains, i.e., $f: \mathcal{X}_A^{\mathcal{I}} \longrightarrow \mathcal{X}_B^{\mathcal{I}}$ and $g: \mathcal{X}_A^{\mathcal{U}} \longrightarrow \mathcal{X}_{\mathrm{B}}^{\mathcal{U}}$. For instance, in a movie and book recommender system, a mapping function could be applied to the genre, $f(\textit{comedy movie}) = \textit{humor book}$, or could identify a user registered in both systems, $g\left(u_i^{\textit{movie recommender}}\right) = u_j^{\textit{book recommender}}$. We note that this mapping is a particular case of a more general approach in which both features are mapped to a new one. We do not enter into this for simplicity purposes.

User and item characteristics together with their relations depend on the implemented recommendation strategy. In the following we describe representative examples.

Content-based relations

An item's content is described as a set of features $\mathcal{F} = \{F_1, \ldots, F_n\}$. Items are represented with vectors of real numbers in which the *j*-th component is the weight or relevance of the corresponding feature F_j. Analogously, a user's preferences are described with the features in $\mathcal{F}$. The user's profile is composed of a vector of real numbers whose components correspond to the degrees in which the user likes or is interested in the features. Hence, in this case $\mathcal{X}^{\mathcal{U}} = \mathcal{X}^{\mathcal{I}} = \mathcal{F}$, and referring to the cross-domain scenario, an overlap between the domains $\mathcal{A}$ and $\mathcal{B}$ occurs when $\mathcal{F}_A \cap \mathcal{F}_B \neq \emptyset$. Figure 1 shows an example of item profiles in a book and movie content-based recommender system. The literary genre is the bridge between such domains, which lets computing similarities between a user profile in the book domain and an item profile in the movie domain. The function f relates pairs of attributes that have the same meaning, but belong to distinct domains.

Relations based on social tags. A particular case of content-based filtering approaches are those that exploit social tags, which constitute a vocabulary T. In these approaches a user is characterized by the tags she assigned to the items she is interested in and, analogously, an item is represented by the set of tags the users have assigned to it. That is, $\mathcal{X}^{\mathcal{U}} = \mathcal{X}^{I} = T$.

Alice

Attribute-Value		Rating
literary genre	Science fiction novel	4.6
literary genre	Fantasy novel	3.8
book	The Name of the Wind	5.0
book	American Gods	4.0

$\xrightarrow{f}$

Star Wars: Episode IV (1977)

Attribute-Value		Weight
genre	Science fiction	1.0
genre	Adventure	0.8
director	George Lucas	0.7
cast	Mark Hamill	1.0
cast	Carrie Fisher	1.0
cast	Harrison Ford	1.0

Figure 1. Content-based user and item profiles in books and movies domains, linked through the attributes *literary genre* and *movie genre*.

Collaborative filtering-based relations

In CF the users' preferences are modeled as a matrix $\mathcal{M}$ in which the element $m_{ij} \in \mathcal{R}$ is the rating assigned by user i to item j. The set $\mathcal{R} \subseteq \mathbb{R}$ of allowed rating values may differ from system to system, but usually are integers in the range [1,5], and the null value $\emptyset$ is assigned to m_{ij} when the item j has not been evaluated by user i. The most popular approach to CF is to exploit $\mathcal{M}$ with nearest neighbor algorithms. They first compute rating-based similarities between either items or users, and then generate a rating prediction for a target item by averaging the ratings for the same item provided by similar users, or ratings of similar items provided by the same target user.

Relations between items. Each item is described with the ratings assigned to it, i.e., the ratings in its corresponding $\mathcal{M}$ column. No additional information is used, only items and users identifiers are needed to relate items and users with each other. Thus, $\mathcal{X}^{\mathcal{I}} = \mathcal{U}$ ($\mathcal{U}$ being the users' ratings) and $\mathcal{X}^{\mathcal{U}} = \{\text{id}\}$, and items in domains $\mathcal{A}$ and $\mathcal{B}$ can be compared if $\mathcal{X}_{\mathcal{A}}^{\mathcal{I}} \cap \mathcal{X}_{\mathcal{B}}^{\mathcal{I}} \neq \emptyset$, i.e., $\mathcal{U}_{\mathcal{A}} \cap \mathcal{U}_{\mathcal{B}} \neq \emptyset$. Let us consider the example shown in Figure 2. In order to recommend movies (domain $\mathcal{B}$) to Alice, who only has evaluations about books (domain $\mathcal{A}$), we compute similarities between her profile and the profiles of users who evaluated items in both domains. In particular, Alice and Bob are interested in the book 'The Name of the Wind', and the latter evaluated positively the movie 'The Lord of the Rings.' For such reason, this movie may be suggested to Alice.

	The Name of the Wind	American Gods	The Lord of the Rings (2001)	The Matrix (1999)	Star Wars (1977)
Alice	5		?		
Bob	4		5		
Carol		4			3
Dave			5	5	

$\mathcal{U}_A$ $\mathcal{U}_B$ $\mathcal{I}_A$ $\mathcal{I}_B$

Figure 2. Example of a joint matrix of ratings for items in books and movies domains.

Relations between users. Users are described and compared through the ratings with which they have evaluated items, i.e., the ratings in their corresponding $\mathcal{M}$ rows. In this case $\mathcal{X}^{\mathcal{U}} = \mathcal{I}$ ($\mathcal{I}$ being the items' ratings) and $\mathcal{X}^{\mathcal{I}} = \{id\}$, and users in domains $\mathcal{A}$ and $\mathcal{B}$ can be compared if $\mathcal{X}_A^{\mathcal{U}} \cap \mathcal{X}_B^{\mathcal{U}} \neq \emptyset$, i.e., $\mathcal{I}_A \cap \mathcal{I}_B \neq \emptyset$.

Other relations

In addition to content- and collaborative filtering-based relations, we believe that a system may exploit other types of relations established through contextual features $\mathcal{C}$, such as the time (e.g. movies and music compositions *usually consumed on Christmas*) and the users' mood and sentiments (e.g. movies and music compositions that *usually yield nostalgic feelings*). To the best of our knowledge, there is not work in the literature addressing the cross-domain recommendation task by means of contextual features. We state that, in general, a hybrid cross-domain recommender system could exploit relations between user preferences and/or item attributes based on content, collaborative filtering, and contextual information, $\mathcal{X} = \mathcal{X}^{\mathcal{U}} \cup \mathcal{X}^{\mathcal{I}} \cup \mathcal{C}$, across multiple domains.

Moreover, as we shall explain, a number of approaches has been proposed that represent user and/or items in low dimension spaces of latent factors, in which the proposed vector space representations are valid. In these cases, if $\mathcal{Z}, \mathcal{W}$ are the sets of user and item latent factors respectively, then $\mathcal{X}^{\mathcal{U}} = \mathcal{Z}$ and $\mathcal{X}^{\mathcal{I}} = \mathcal{W}$. We note that usually there is only one space of latent factors common to users and items, i.e. $\mathcal{Z} = \mathcal{W}$.

2.4 Classification of cross-domain recommendation techniques

According to Loizou [12] cross-domain recommendation has been addressed by following three main types of strategies. There are strategies that consist of integrating and exploiting explicit user preferences distributed in various systems. There are other strategies that are based on recording user behavior and actions aiming to learn user preferences, and use them for generating joint recommendations on multiple domains. Finally, there are strategies that focus on combining recommendations from different domains to build a single system.

Since Loizou's work a number of new recommendation models have been proposed to generate cross-domain recommendations. Some approaches have used social tags and semantic knowledge to establish relations between cross-domain user preferences and/or items attributes. Other approaches have applied Transfer Learning techniques – investigated in Machine Learning – to perform collaborative filtering in cases where there is no explicit user/item overlap between domains. Hence, in [9] Li proposes a categorization of representative approaches based on the type of knowledge transferred to establish links between domains, namely a) collective rating patterns, b) common latent factors, and c) correlations of user/item latent factors.

Based on Li's classification and extending Pan et al.'s [15] beyond CF, we propose a new categorization of existing approaches on cross-domain recommendation. Our classification takes into account the cross-domain recommendation task definition and the domain relation types provided in Sections 2.2 and 2.3. We distinguish adaptive models –which in a directional way exploit information from a source domain to make recommendations in a target domain– and collective models –which are built with data from several domains and potentially can make joint recommendations for such domains. Table 1 shows the proposed classification for the approaches revised in subsequent sections. We can observe that hybrid cross-domain recommendation approaches have been barely investigated. This may be due to the lack of public datasets with both content-based and collaborative filtering information about several domains. In the following we describe datasets utilized in the literature, and other issues related to the evaluation of cross-domain recommenders.

Relations between domains		Recommendation models	
		Adaptive	***Collective***
Content-based	*Attributes*	Azak 2010 [3]	
	Social tags	Kaminskas & Ricci 2011 [8]	Abel et al. 2011 [1] Szomszor et al. 2008 [18]
	Semantic properties	Fernández-Tobías et al. 2011 [6]	Loizou 2009 [12]
	Correlations		Shi et al. 2011 [17]
Collaborative filtering-based	*Ratings*	Azak 2010 [3] Berkovsky et al. 2008 [4] Winoto & Tang 2008 [20]	Loizou 2009 [12]
	Rating patterns	Li et al. 2009a [10]	Li et al. 2009b [11]
	Latent factors	Pan et al. 2010 [15]	Pan et al. 2011 [14]
	Correlations		Cremonesi et al. 2011 [5] Zhang et al. 2010 [21]

Table 1. Proposed taxonomy to classify cross-domain recommendation approaches.

2.5 Evaluation of cross-domain recommendations

To date cross-domain recommendation has been investigated to mainly address the *cold-start problem*, i.e., recommending items to a new user whose preferences are unknown in the target domain, and to mitigate the *sparsity problem*, i.e., recommending items when very few user preferences are known. Hence, many of the existing approaches have been evaluated by varying the sparstity level in the data [1], [4], [15]. Alternatively, recommendations have also been evaluated by varying the degree of rating overlap [5] and social tag overlap [17] between domains. In any of the above cases, cross validation is usually performed to compute average evaluation results [10], [11], [20], [21].

MAE and RMSE are the most widely used metrics for evaluating collaborative filtering approaches, but ranking-based metrics such as Precision, Success, Mean Reciprocal Rank, and F-measure have been used recently as well [1], [5]. According to Winoto and Tang [20], we could expect that cross-domain recommendations, generated with a certain amount of user preferences, are less precise than those based on the same amount of preferences but expressed in the target domain. The advantage of cross-domain recommendation may be not improved accuracy, but added novelty and more diverse recommendations, and thus may offer higher satisfaction and utility to the user. In this context recently proposed novelty and diversity metrics could be taken into consideration [16], [19].

Another important issue related to cross-domain recommendation evaluation is the lack of repositories with data from multiple domains. To cope with this limitation a usual practice has been the splitting of a system's dataset into subsets that are considered as different domains, e.g. movies with distinct genres. The most used datasets have been Movielens, Netflix and EachMovie for the movie domain; BookCrossing and LibraryThing for the book domain; and Last.fm for the music domain.

3 Integrating user preferences about multiple domains

A direct approach to make cross-domain recommendations is exploiting user models that capture preferences about items in each of the considered domains. In this case the main challenge is how to integrate single-domain user models from several systems. Some authors have investigated mechanisms to obtain user preferences belonging to different domains, and to integrate them into a single user model. One of the first works in this line is [7], where González et al. define a domain-independent smart user model composed of a weighted graph that relates profile characteristics of different domains, and whose edge weights indicate the strength of relations between such characteristics.

Social tagging systems are an important source of multi-domain user preferences. In [18] Szomszor et al. propose an approach for filtering and consolidating tags from several folksonomies to build aggregated user profiles. Tags in such profiles are then linked to Wikipedia concepts creating a semantic model that reflects the user's high level interests in multiple domains. Szomszor et al.'s study on Delicious and Flickr datasets shows that a user's profiles from different folksonomies provide distinct facets of interests, and aggregated profiles enhance the captured user interests. In a related work [1], Abel et al. perform a profile semantic enhancement process by grouping tags into WordNet categories. The results of their analysis agree with Szomszor et al.'s: integrated profiles reveal further preferences by increasing their information entropy. According to their experiment results the quality of recommendations obtained with integrated profiles is higher in case of both sparse and evolved profiles.

In this context, an issue that needs to be further investigated is whether integrated user profiles really improve the generated recommendations. In [20] Winoto and Tang conduct a study in which users evaluate a number of items in twelve domains. With the retrieved user preferences, the authors build different aggregated profiles and use them as input of a collaborative filtering algorithm. Then, they compare the algorithm's rating prediction error using preferences in only the target domain, and the rating prediction error using integrated user profiles from different domains. The obtained results show that aggregating profiles from multiple domains produces less precise but more diverse recommendations, and can be of more utility for the user in cold-start situations.

4 Exploiting explicit relations between domains

An alternative technique to integrate multi-domain user preferences consists of establishing relations between domain characteristics, and exploiting them for the cross-domain recommendation task. Approaches in this line have been recently proposed in the literature. Most of them are based on **content-based domain relations**.

Azak [3] proposes a recommendation model that utilizes knowledge-based decision rules to establish relations between domains. The rules relate user preferences with certain item characteristics, and item content attributes across domains.

Graphs are a usual structure to represent relations between domains. Loizou [12] uses Wikipedia as a universal vocabulary describing items from multiple domains in a common form. Users and items are added as nodes to a graph in which edges represent user ratings and semantic item relations. With this graph a Markov probabilistic model generates recommendations based on the paths existing between users and items. In [5] Cremonesi et al. represent items belonging to different domains in a single graph whose edges link each item with its k most similar items according to their rating-based profiles. Thus, to establish inter-domain relations some users have had to rate items in several domains. In general a relative small number of items satisfy such condition, which may lead to weak relations between domains. Cremonesi et al. propose to find non-direct item relations from the square of the graph adjacency matrix, and use them for making collaborative filtering recommendations. In [6] Fernández-Tobías et al. present an approach that uses DBpedia as a multi-domain knowledge source for building a semantic network that links concepts from several domains. On such semantic network, which has the form of an acyclic directed graph, a weight spreading activation algorithm retrieves concepts (items) in a target domain that are highly related to other input concepts in a source domain.

Social tags have also been used to link domains since they can be used as an agreed vocabulary to describe items from any domain in a simple, generic way. Aiming to adapt music recommendations for places of interest, Kaminskas and Ricci [8] propose to use item profiles based on common (emotional) tags assigned to both types of items. The relevance of a music composition for a particular place of interest is determined by computing the similarity between their tag-based profiles. Shi et al. [17] utilize tags to build user-user and item-item similarity matrices. The similarity between two users/items from different domains is proportional to the number of tags shared by their annotation profiles. Computed similarities are incorporated as constraints into a probabilistic model based on matrix factorization and collaborative filtering.

5 Transfer learning across domains

Transfer learning is an active research topic in the Machine Learning area. It aims to improve a learning task in a particular domain by using knowledge transferred from other domain in which a related task is known [13]. In recommender systems transfer learning techniques have recently been applied to improve **collaborative filtering**.

In [10] Li et al. present a transfer learning approach that mitigates the sparsity problem in collaborative filtering. The approach performs a co-clustering strategy on the rating matrix of an auxiliary domain with high rating density, and finds rating patterns at the cluster level. Assuming that user rating behavior is similar in two domains, the approach establishes relations between domains based on the found rating patterns. More specifically, using a compact matrix representation called *codebook*, the patterns are transferred to a target domain with a low rating density. There they are used for estimating unknown ratings. In other work [11] Li et al. extend the previous approach by means of a probabilistic model in which a user or item belonging to a particular cluster is not binary, but is described in terms of probability density function. In this case a common model is built from the ratings of all the considered domains, without requiring a dense source domain. The main advantage of both approaches is that they do not require user/item overlap, which makes them applicable in many real situations where profiles from isolated systems have to be merged or linked. Zhang et al. [21] propose a collaborative filtering learning model that identifies rating correlations in a latent factor space. The approach transforms rating matrices from different domains into user and item latent factors. Assuming the sets of users in the considered domains are the same, the above latent factors are used to make cross-domain recommendations. An example application where this approach may be used is a movie recommender system in which each movie genre corresponds to a particular domain.

Pat et al. [15] propose another model based on latent factors taking into account the heterogeneity in rating representations of different CF systems. Differently from the above approaches, this model assumes the existence of two auxiliary domains, one of them with related users, and the other with related items. The model applies matrix factorization to find principal latent components for users and items in the auxiliary domains, and integrate them into a target domain trough a regularization technique.

6 Conclusions and future work

In this paper we have surveyed the state of the art on cross-domain recommendation, revising approaches proposed in different research areas, namely User Modeling, Information Retrieval, Knowledge Management, and Machine Learning. Aiming to characterize, classify and compare such diverse approaches, we have provided a formal, generic definition of the cross-domain recommendation task, and have proposed a new categorization of cross-domain recommendation models that takes into account and extends previous considerations of other authors. Our categorization distinguishes content-based and collaborative filtering strategies to establish relations between domains, and adaptive and collective strategies to exploit information in a source domain for making recommendations in a target domain, or to integrate information from both source and target domains into a multi-domain recommendation model.

We have shown that a large percentage of the revised approaches correspond to transfer learning techniques focused on mitigating the cold-start and sparsity problems in collaborative filtering. The main limitation of these approaches is the need of having an overlap in the sets of ratings assigned to items in different domains. We believe that hybrid approaches can enhance the multi-domain user preference space with further content-based and contextual information/relations across domains. In this context, we plan to extract multi-domain knowledge from the Social Web (e.g. from social networks such as Facebook) and the Semantic Web (e.g. from Linked Data ontologies). We are also interested in investigating novel knowledge-based recommendation approaches, exploiting diverse mechanisms such as inference rules, Bayesian networks, and ontology-based models. Preliminary results on this topic are described in [6], where we propose a semantic framework built upon DBpedia ontology for automatically creating semantic networks linking concepts in different domains. Among other issues, the semantic framework could be enhanced by using multi-domain user preferences not only expressed by means of explicit ratings, but other types of user preferences, such as social tags, contextual features, and implicit item consuming records.

Acknowledgements The work presented herein was supported by the Spanish Ministry of Science and Innovation (TIN2011-28538-C02).

References

1. Abel, F., Herder, E., Houben, G.J., Henze, N., Krause, D.: Cross-system user modeling and personalization on the social web. User Modeling and User-Adapted Interaction, 1–42 (2011)
2. Adomavicius, G., Tuzhilin, A.: Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. IEEE TKDE 17(6), 734–749 (2005)
3. Azak, M.: CrosSing: A framework to develop knowledge-based recommenders in cross domains. MSc thesis, Middle East Technical University (2010)
4. Berkovsky, S., Kuflik, T., Ricci, F.: Mediation of user models for enhanced personalization in recommender systems. User Modeling and User-Adapted Interaction 18(3), 245–286 (2008)
5. Cremonesi, P., Tripodi, A., Turrin, R.: Cross-domain recommender systems. In: Proc. of ICDMW 2011 Workshops, pp. 496–503 (2011)
6. Fernández-Tobías, I., Cantador, I., Kaminskas, M., Ricci, F.: A generic semantic-based framework for cross-domain recommendation. In: Proc. of the 2nd International Workshop on Information Heterogeneity and Fusion in Recommender Systems, pp. 25–32. (2011)
7. González, G., López, B., de la Rosa, J.L.: A multi-agent smart user model for cross-domain recommender systems. In: Proc. of IUI 2005 Beyond Personalization Workshop, 93-94 (2005)
8. Kaminskas, M., Ricci, F.: Location-adapted music recommendation using tags. In: Proc. of UMAP 2011, pp. 183–194 (2011)
9. Li, B.: Cross-domain collaborative filtering: A brief survey. In: Proc. of ICTAI 2011, pp. 1085–1086 (2011)
10. Li, B., Yang, Q., Xue, X.: Can movies and books collaborate? Cross-domain collaborative filtering for sparsity reduction. In: Proc. of IJCAI 2009, pp. 2052–2057 (2009)
11. Li, B., Yang, Q., Xue, X.: Transfer learning for collaborative filtering via a rating-matrix generative model. In: Proc. of ICML 2009, pp. 617-624 (2009)
12. Loizou, A.: How to recommend music to film buffs: enabling the provision of recommendations from multiple domains. PhD dissertation, University of Southampton (2009)
13. Pan, S.J., Yang, Q.: A survey on transfer learning. IEEE TKDE 22(10), 1345–1359 (2010)
14. Pan, W., Liu, N.N., Xiang, E.W., Yang, Q.: Transfer learning to predict missing ratings via heterogeneous user feedbacks. In: Proc. of IJCAI 2011, pp. 2318–2323 (2011)
15. Pan, W., Xiang, E.W., Liu, N.N., Yang, Q.: Transfer learning in collaborative filtering for sparsity reduction. In: Proc. of AAAI 2010 (2010)
16. Shani, G., Gunawardana, A.: Evaluating recommendation systems. In: Ricci, F., Rokach, L., Shapira, B., Kantor, P.B. (eds.) Recommender Systems Handbook, pp. 257–297 (2011)
17. Shi, Y., Larson, M., Hanjalic, A.: Tags as bridges between domains: Improving recommendation with tag-induced cross-domain collaborative filtering. In: Proc. of UMAP 2011, pp. 305–316 (2011)
18. Szomszor, M., Alani, H., Cantador, I., O'Hara, K., Shadbolt, N.: Semantic modelling of user interests based on cross-folksonomy analysis. In: Proc. of ISWC 2008, pp. 632–648 (2008)
19. Vargas, S., Castells, P.: Rank and relevance in novelty and diversity metrics for recommender systems. In: Proc. of RecSys 2011, pp. 109–116 (2011)
20. Winoto, P., Tang, T.Y.: 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), 209–225 (2008)
21. Zhang, Y., Cao, B., Yeung, D.Y.: Multi-domain collaborative filtering. In: Proc. of UAI 2010, pp. 725–732 (2010)