Social-based Collaborative Filtering

Kostas Stefanidis, Eirini Ntoutsi, Haridimos Kondylakis and Yannis Velegrakis

Synonyms

Social-based Recommendations
Collaborative Filtering Using Social Data

Glossary

Recommendation: A suggestion or proposal to a user for an item, e.g., book, movie, video, news article, that is potentially interesting for the user.
Recommender system: A system or engine that produces recommendations by predicting the preferences of users for certain items.
Ratings matrix: Assume a set of users U and a set of items I in the recommender system. A user $u \in U$ might provide her preference for an item $i \in I$ in form of a rating denoted by $rating(u,i)$, which typically takes values in [1, 5]. The preferences of users for individual items are represented by a ratings matrix R, where the $R_{u,i}$ entry corresponds to $rating(u,i)$.

Kostas Stefanidis
School of Information Sciences, University of Tampere, Finland, e-mail: kostas.stefanidis@uta.fi

Eirini Ntoutsi
Department of Electrical Engineering and Computer Science, Leibniz University Hannover, Germany, e-mail: ntoutsi@kbs.uni-hannover.de

Haridimos Kondylakis
Institute of Computer Sciende, Foundation for Reaserch and Technology - Hellas, Greece, e-mail: kondylak@ics.forth.gr

Yannis Velegrakis
Department of Information Engineering and Computer Science, University of Trento, Italy, e-mail: velgias@disi.unitn.eu

Collaborative filtering: Given a ratings matrix R, representing the preferences of users U for items I, recommend to each user a list of items in descending order of their relevance for the user. The relevance scores are estimated based on ratings of similar users.
Social network: A structure, nowadays typically implemented as an online platform, that enables content sharing by allowing users to post information and interactions between users by allowing them comment on each other's posts, exchange messages etc.
Social content: The content that is generated in a social network as a result of activities and interactions between users.

Definition

Collaborative filtering is a special category of recommender systems that generate recommendations to the users about certain items, e.g., products or services, by relying on preferences of similar users. Social-based collaborative filtering exploits the vast amount of data generated in the social networks to improve the quality of recommendations. Such data are utilized for different purposes, e.g., to deal with traditional recommendation problems, such as the cold start problem, which is caused due to lack of user-related data in the recommender, to derive a better quality user neighborhood by integrating user-related information from the social networks and to enrich the recommendations by capturing different aspects of the users as they are manifested through the content they share and their interactions in the social networks.

Introduction

With the growing complexity of the Web, users find themselves overwhelmed by the mass of choices available. To facilitate the user selection process, recommender systems provide suggestions on data items of potential interest to the users. The interest of a user for an item is inferred from the user history, e.g., user purchases or browsing history. A key challenge in recommenders is the so-called sparsity data problem; typically, users rate only a few items, due to the huge amount of offered items and the low engagement of the users with the applications.

Nowadays, stunning opportunities are offered for dealing with the lack of data in recommender systems, as users publicly share their preferences and provide reviews and other information on certain items in several social networks. This data though is not explicit for a recommender, rather it is implicitly given in the context of the social networks. As such, except for its volume and velocity, the information is blur, unstructured, diverse, uncertain and incomplete and therefore, exploiting this data for recommendations is a big challenge.

This entry focuses on how the collaborative filtering recommender systems are reshaped by looking out-of-the-recommender-box and hunting for relevant information in the social networks. This way, from the traditional collaborative filtering approaches that utilize within-the-recommender data, i.e., the ratings matrix mainly, we are moving towards an open system that would potentially result in high-quality recommendations by enriching information for users, items and ratings through the available social network data. We illustrate such a scenario in the following example.

Example 1. Assume a user, say Sophia, and a recommender system, say MinoS, that produces recommendations for travel destinations by using a collaborative filtering approach that exploits Sophia's social data. For doing so, summarily, MinoS locates Sophia and the set of destinations she has visited in the social networks she used. Then, it focuses on finding users, other than Sophia, and the places they have visited. Having collected the visiting history of the users, MinoS computes the similarity between Sophia and every other user; the users visiting history contains the necessary information for computing similarities between them. MinoS keeps only the users with similarity greater than a threshold value to Sophia, namely the highly similar users to Sophia. Intuitively, at the next step, it performs a composition between Sophia and her highly similar network users, and users and the places they have visited. This way, for each of Sophia's similar user, who has visited a destination, we make a connection between Sophia and that destination. In overall, the relevance score of each such destination is calculated by taking into account the ratings and reviews of the users that have visited the destination, as well as their similarity with Sophia. On the basis of these scores, MinoS recommends the best destinations to Sophia. □

Key Points

Recommender systems have become indispensable for several Web sites, such as Amazon, Netflix and Google News, helping users to navigate through the infinite number of available choices, like products, movies and news articles, respectively. Recently, social networks offer opportunities for generating better recommendations, as more than ever before, users publicly share their preferences and provide information for certain items. In this entry, focusing on social-based collaborative filtering, we target the following issues:

- how can we enrich the three main entities in a recommender system, namely, users, items and ratings, by integrating content- and structure-related data from social networks
- how can we construct social-enhanced user profiles, by exploiting the integrated data, and

- how the collaborative filtering engine is reshaped due to the volume and the variety of the available data.

Historical Background

Collaborative filtering predicts user preferences for data items by keeping track of their likes and dislikes. Based on the assumption that similar items will be of interest for similar users, i.e., users with tastes in common, via collaborative filtering, users can help each other to choose products.

More specifically, assume a recommender system where I is the set of items to be rated and U is the set of users in the system. A user $u \in U$ might rate an item $i \in I$ with a score $rating(u,i)$ in $[1,5]$. Typically, the cardinality of the item set I is high and users rate only a few items. For the items unrated by the users, recommender systems estimate a relevance score, denoted as $relevance(u,i)$, $u \in U$, $i \in I$. In general, there are different ways to estimate the relevance score of an item for a user. In the content-based approach (e.g., [25]), the estimation of the rating of an item is based on the ratings that the user has assigned to similar items, whereas in collaborative filtering systems (e.g., [19]), this rating is predicted using previous ratings of the item by similar users. Typically, similar users are located via a *similarity function* $simU(u,u')$ that evaluates the proximity between $u,u' \in U$ by considering their shared dimensions. We use N_u to denote the set of the most similar users to u, hereafter, referred to as the *neighbors* of u.

Given a user u and his neighbors N_u, if u has expressed no preference for an item i, the relevance of i for u is estimated as:

$$relevance_{N_u}(u,i) = \frac{\sum_{u' \in N_u} simU(u,u') rating(u',i)}{\sum_{u' \in N_u} simU(u,u')}$$

Typically, after estimating the relevance scores of all unrated items by a user, the top-k rated items are recommended to the user.

Using Social Content for Collaborative Filtering

Most recommender systems nowadays are based, in practise, on large datasets. As a result, the ratings matrix used for collaborative filtering could be extremely large and sparse, which affects the quality of the produced recommendations. Social-based collaborative filtering exploits data that users publicly share in social networks in order to deal with the sparsity problem. This way, from the traditional collaborative filtering approaches that utilize mainly the ratings matrix, we are moving to an open system that would potentially result in high-quality recommendations by enriching information for users, items and ratings through the available social data, appearing out-of-the-recommender-box.

Data Enrichment and Integration

The main entities involved in a recommendation application, i.e., users, items and ratings, can be semantically enhanced with information from social networks.

With respect to the users, instead of using the plain information that a user gives for himself to a recommender system, we can exploit information available in numerous external sources, such as Facebook, LinkedIn, Google+, Fourthsquare and Amazon. The motivation behind this, is that a user describes himself differently in different networks, depending on the domain, so we can identify different interests, user activities, information about places he visited, and so forth. A challenge towards this direction is to integrate the user's social profile, as well as to integrate, or expand, the social graph to bring together different social networks. Approaches like [31] try to construct a complete user profile by finding, extracting, and fusing the semantic-based user profile from the Web, whereas approaches like [3] collect, integrate and discover profile information based on social content sites. In [32], the schema is first extracted out of each social site, transformed to a common representation and instantiated using possibly overlapping data. Another popular technique for collecting and understanding user data is the analysis and creation of topic models within the user-generated content [26], especially Twitter data [33]. Such solutions can be used for encountering the cold start problem as well.

At item level, one can consider that information about items can be enhanced with semantic information. In addition to the items descriptions, in which temporal characteristics of the items, such as popularity and freshness [29], can be maintained in real-time , we can exploit information retrieved from the Web, such as published results and reports, Web pages, thesauri or ontologies. The plethora of well-organized information over the Web in collectively maintained knowledge repositories, such as Wikipedia and LibraryThing, can be used for correlating and computing similarities between data items [8]. In addition, items can be annotated using terms from ontologies and other semantic resources to enhance the description quality of specific items [18].

Regarding user preferences for certain items, we can rely on users' online activities that involve such items. For instance, a user may tweet about a movie, might rate its trailer in YouTube and share the link and comment in Facebook. Except for filling missing ratings, existing ratings could be also enhanced in terms of their context (e.g., time or place) and ratings criteria (e.g., director vs actors). *Contextual recommendations* have been already studied [2, 30], however they rely on user explicit feedback. An indirect inference of the context though, is an interest area for further exploration. *Multi-criteria ratings* [1] rely also on user explicit subratings for different aspects of the items and are suffering from lack of data as the item space is further expanded due to the subcriteria. With the abundance of free text reviews nowadays, we can implicitly extract both the aspects of the items that are of interest to a user and their associated ratings/sentiment, using NLP and sentiment analysis techniques [34]. In an *active learning* manner [12], we can explicitly ask users to provide more ratings, by giving them incentives. The challenge here is the selection of a few informative items for which the user should be asked.

Even if we are able to identify additional information for users, items and preferences, appearing outside the recommender system, it is important to understand which pieces of information refer to the same entities, so as to integrate them, in order to manage and further process them. The problem of entity resolution aims to identify different descriptions that refer to the same entity, and emerges as a central data-processing task for an entity-centric organization of Web data [8]. It is needed to enrich interlinking of entities descriptions, so that Web data can be accessed by machines as a global data space allowing the use of standard languages. Although entity resolution has attracted significant attention in information systems, database and machine-learning communities, there are new challenges stemming from the Web openness in describing a multitude of entity types across domains. The scale and diversity of descriptions challenge the core entity resolution tasks, namely, (i) how descriptions can be effectively compared for similarity and (ii) how resolution algorithms can efficiently filter the candidate pairs of descriptions that need to be compared.

User Profiling

A fundamental ingredient of every successful recommendation is the ability to accurately model the user preferences and habits. This model is typically known as *profile*. A profile is actually structure that represents the principal characteristics of a user which are turned into preferences.

One way of building profiles is to have them explicitly provided by the end user, but this brings along a number of limitations. First of all, users may not be aware of all the characteristics they have, or may not be willing to provide all of them. Thus, it is better if the profile is built automatically by observing the user actions, which is exactly what traditional recommendation techniques are doing, i.e., they are monitoring the previous user interactions with the system, and based on them, they are building a user model. Unfortunately, profiles generated in this way may end up highly dependent, restricted, and very sensitive to the user interactions with the system. This means that unless a user exposes one of her characteristic with some specific action, that characteristic can never become known to the system. Furthermore, an action performed by a user, e.g., a purchase, may be recorded and leave its trace in the user preferences for a very long time even if it has been only an occasional action or an action performed for someone else, for instance, the purchase of a gift for a friend. All these mean that the user interaction history with the system, although a valuable resource, may not provide the complete information that a system would like to know about its user.

Social media constitute a great additional source of information for building user profiles since they may expose user characteristics that are not recorded in the user interaction history with the system. The scale and spectrum of activities for which the social media are used nowadays is unprecedented. Through these activities the users leave their footprint that can be used to build a rich user description, i.e., an accurate and complete profile. There are two types of information that can be exploited in social media. One is the content and the other is the social network. Exploitation of the content means that the profiling of the users is based on the content that the users publish in the social media, e.g., the tweets they send or the posts they make. Exploitation of the social network means that the profiling is constructed by using information on the way users are connected to each other to establish friendships and communications, or to follow the activity of others. Based on the above, one can distinguish two main profiling methods, the *content*-based and the *network*-based.

Generating profiles from social media is a challenging. The content is user generated and is done through some new non-traditional forms of communication. As a consequence, it has no controlled vocabulary, no restrictive syntax, no specific rules, and is full of shorthands and jargon. For what concerns the social network information, among the challenging issues is the fact that not all the connections are of equal value. Users may follow actively only a selective set of their connections or connections may exist with central non-real users like news channels or group accounts. Finally, users may use different accounts for different purposes which makes hard the recognition of the actions with user provenience.

The idea behind the content-based profiling is to see the content as a vector of terms. These terms can play the role of features for which a classifier can be trained or some discriminating score can be computed [14]. Of course, not all the words are equally important. Techniques like LDA [5] can be used to summarise the content or to identify those words that actually matter and then use only these words [7] for the profiling task. Alternatively, an information theoretic approach can be employed to identify the amount of information that each word communicates in relationship with the rest of the content and use that information of deciding the importance of each word [13]. An important characteristics of social data affecting significantly the results is sparsity. To overcome sparsity, smoothing has to often be applied, with Laplace smoothing [15] being the most prevalent technique.

The intuition behind the network-based profiling, on the other hand, is similar to that of the collaborative filtering. A user is likely to like things that are similar to what her friends like. However, instead of the collaborative filtering idea in which similarity is used to propagate features of one user directly to another, in the network based profiling, the social network is the main carrier. For instance, considering a user as a node in the network graph, a simple counting and aggregation of the features of the reachable nodes is required. The features that end up to be more popular are used to form the user profile [17]. Instead of a simple counting and aggregation, more complex probabilistic models can be used that take into consideration the distance on the social graph to adjust the weight of each feature appearing on the connected nodes [4][28].

Unfortunately, neither the content-based, nor the network-based profiling techniques work perfectly all the time. Naturally, there have been techniques that combine them to improve the quality of the results. One approach is use a single model for both the network and the content [22]. However, these two parameters can be also considered independently and their results be combined [21] using some mathematical model like those used to combine different lists [11], or considered orthogonal factors studied together using some bi-clustering technique [27].

Social-based Collaborative Filtering

Traditionally recommenders rely upon the rating matrix, i.e., the explicitly given (numerical) ratings of users to certain items. In collaborating filtering, for a query user $u \in U$ her similar users in U are located using the ratings matrix and some appropriate similarity function, like Cosine Similarity, Pearson Correlation or Spearman Rank Correlation [9]. Recently though expect for the ratings matrix, other sources of information, such us the network information and reviews accompanying the user ratings are employed in order to improve the quality of recommendations.

In particular, the network information, i.e., explicit connections between the users like friends in Facebook, following/followers in Twitter etc, can be employed in order to select a better user neighborhood for a given user and also for dealing with the cold start problem. For example, [16] replaces the traditional notion of user neighborhood, consisting of users with similar ratings to the query user, by that of the trust-neighborhood derived from the trust network where the nodes correspond to the users and the edges to trust statements.

In a different direction, textual reviews which typically accompany user ratings nowadays have been recently explored for recommendations. User reviews comprise a rich source of information as they justify users decisions on certain items and moreover they reveal which aspects of the items the users liked/ disliked. Combining numerical ratings with textual reviews for recommendations was first introduced in [23]. The authors propose the Hidden Factor Model which aligns hidden factors in product ratings with hidden factors in product reviews (discovered through LDA). The key idea is to link latent factors in ratings to hidden factors in review texts, where topics discussed in review texts for a certain product correspond to products having a certain property represented in the latent factor model. In a follow up work, the same authors also used reviews to model personal evolution or experience for recommendations [24].

Textual reviews have been also employed for extracting context, such as location and accompanying people, which allows for contextual recommendations. Explicit user-defined context is hard to acquire but usually such information is contained in the reviews, which are typically freely available. For example, [6] implicitly extracts such sort of information, by employing online reviews.

Employing heterogeneous data for recommendations, such as the network and the reviews, definitely alters the recommendation process. In a different direction, the process is also altered due to the long term tracking of users, items and their preferences, which call for online methods that are able to identify drifts in users preferences and periodicity in their habits. Data ageing is a typical way to deal with drifts in user profiles by downgrading historical obsolete data and paying more attention to recent ones that reflect the current user profile best (e.g., [10], [29]). However, approaches that discard past instances have been criticized as loosing too much signal and, although more elaborate methods exist e.g., [20] that separates transient factors from lasting ones, what to forget and what to remember is still a challenge. Moreover, a long term user monitoring implies an extensive knowledge about user tastes and preferences, which might result in privacy risks for the user. This is especially critical for mobile app recommenders.

Key Applications

Traditionally, recommendations are produced within a domain, i.e., when asking for movies, the suggestions consist only of movies. Examples of domain-specific social networks recommenders include Flixster (https://video.flixster.com/) for movie recommendations, and Epinions (http://www.epinions.com/) for a wide range of product recommendations. This paradigm can be extended so as to support cross-domain recommendations. For example, packet recommendations produce composite items consisting of a central item, possibly in the main domain of interest for a user, and a set of satellite items from different domains compatible though with the central item. Compatibility can be assumed either as soft (e.g., other books that are often purchased together with the movie being browsed) or hard (e.g., a travel destination that must be within a certain distance from the main destination). The notion of cross-domain recommendations can be extended, so as to support data items outside of the data repository of the recommender. There are already such sort of aggregators in the Web, which act as wrappers over items from different stores. For example, users in Polyvore (http://www.polyvore.com/) mix and match fashion items from different brands.

Well-established applications of social-based collaborative filtering appears in the domain of social media. Unlike traditional media in which few editors set the guidelines, in the era of social media, we may have a very big number of editors, and content data improves its quality as the number of contributors increases. Typical examples include YouTube (https://www.youtube.com/), Last.fm (http://www.last.fm/), and Reddit (https://www.reddit.com/).

In a different scenario, given a social community, a collaborative filtering application is to suggest compelling data items as judged by the community. For example, consider the news aggregator Digg (http://digg.com/) that in its front page shows stories as they are rated positively by the community. Larger and more diverse communities offer stories that better reflect the average interest of the community participants.

The well-used social networks offer recommendations as well. Either general-purpose social networks, e.g., Facebook (https://www.facebook.com/) and Twitter (https://twitter.com/), or domain-specific ones, e.g., Linkedin (https://www.linkedin.com/), offer suggestions about friends, people to follow and jobs you may be interested in.

Finally, from a different point of view, there are social networks that provide recommendations by exploiting review text to uncover user's implicit tastes and item's properties. The reviews comprise a rich source of information as they justify a user's decision on a certain choice. Such piece of information is used, for example, by Booking.com (http://www.booking.com/) for making hotel suggestions.

Future Directions

We consider that the next day of recommenders is to put the users in the foreground and try to exploit their social interactions to fulfil their needs, as opposed to approaches focusing more on the companies viewpoints. Next, we highlight services towards this direction.

Interactive exploration: New forms of data exploration and interaction become increasingly more attractive to aid users navigate through the information space and overcome the challenges of information overload. The interaction between users and recommenders can be driven directly by the interpretation of users needs. Users have to peruse the suggested results, and systems have to be able to react to the on-the-fly changes in the users demands. Although long challenged by works, such as the berrypicking model, common systems still assume that the user has static needs, which remains unchanged during the seeking process.

Visualization: Techniques for visualization contribute towards helping users perceive an overview of the data items included in the suggestions produced for them. Explanations can be used as a means for visualization to assist users identify the what, where, when, how and who of a data item. That is, explanations target at telling the story that the data has to say, aiming at minimizing the browsing effort of the users.

Seeking your past: As data and knowledge bases get larger and accessible to a more diverse and less technically-oriented audience, new forms of data seeking become increasingly more attractive. Re-finding is a different form of exploring data, aiming to locate suggestions seen in the past; here we face the task of recovery. Explicit (given by a user) or implicit (extracted, for instance, by his online traces, e.g., via Foursquare) feedback on suggestions, content and users can significantly increase the quality of recommendations and searching features of a system.

Guessing the future: Modern systems use the past as a mean to guess where the user aims at, so that the system can make the suggestions that will drive towards the fulfilment of the goal(s) as fast as possible. Existing techniques are based mainly on agents and libraries. There is a great deal of opportunities in adapting these techniques into a recommendation model that adjust dynamically the suggestions as the user provides more feedback and the goals become more clear.

Cross References

[1] Recommender Systems: Models and Techniques, id: 88, Ricci, F.
[2] Recommender Systems based on Linked Open Data, id: 110165, di Noia, T., Tomeo, P.
[3] Recommender Systems based on Social Networks, id: 110163, Burke, R.
[4] Recommender Systems in Crowd Sourcing, id: 110206, Wang, H.
[5] Social Search and Querying, id: 110208, Koloniari, G., Lionakis, P., Stefanidis, K.

References

1. G. Adomavicius, N. Manouselis, and Y. Kwon. Multi-criteria recommender systems. In *Recommender Systems Handbook*, pages 769–803. Springer, 2011.
2. G. Adomavicius, R. Sankaranarayanan, S. Sen, and A. Tuzhilin. Incorporating contextual information in recommender systems using a multidimensional approach. *TIS*, 23(1):103–145, 2005.
3. S. Amer-Yahia, L. V. S. Lakshmanan, and C. Yu. Socialscope: Enabling information discovery on social content sites. In *CIDR*, 2009.
4. L. Backstrom, E. Sun, and C. Marlow. Find me if you can: improving geographical prediction with social and spatial proximity. In *Proceedings of the 19th International Conference on World Wide Web, WWW 2010, Raleigh, North Carolina, USA, April 26-30, 2010*, pages 61–70, 2010.
5. D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. *Journal of Machine Learning Research*, 3(4-5):9931022, 2003.
6. G. Chen and L. Chen. Recommendation based on contextual opinions. In *UMAP*, 2014.
7. Z. Cheng, J. Caverlee, and K. Lee. You are where you tweet: a content-based approach to geo-locating twitter users. In *Proceedings of the 19th ACM Conference on Information and Knowledge Management, CIKM 2010, Toronto, Ontario, Canada, October 26-30, 2010*, pages 759–768, 2010.
8. V. Christophides, V. Efthymiou, and K. Stefanidis. *Entity Resolution in the Web of Data*. Synthesis Lectures on the Semantic Web: Theory and Technology. Morgan & Claypool Publishers, 2015.
9. C. Desrosiers and G. Karypis. A comprehensive survey of neighborhood-based recommendation methods. In *Recommender Systems Handbook*, pages 107–144. Springer, 2011.
10. Y. Ding and X. Li. Time weight collaborative filtering. In *CIKM*, 2005.
11. R. Fagin. Combining fuzzy information: an overview. *SIGMOD Record*, 31(2):109–118, 2002.
12. Y. Fu, X. Zhu, and B. Li. A survey on instance selection for active learning. *Knowl. Inf. Syst.*, 35(2):249–283, 2013.
13. B. Han, P. Cook, and T. Baldwin. Geolocation prediction in social media data by finding location indicative words. In *COLING 2012, 24th International Conference on Computational Linguistics, Proceedings of the Conference: Technical Papers, 8-15 December 2012, Mumbai, India*, pages 1045–1062, 2012.

14. B. Hecht, L. Hong, B. Suh, and E. H. Chi. Tweets from justin bieber's heart: the dynamics of the location field in user profiles. In *Proceedings of the International Conference on Human Factors in Computing Systems, CHI 2011, Vancouver, BC, Canada, May 7-12, 2011*, pages 237–246, 2011.
15. L. R. Herrmann. Laplacian-isoparametric grid generation scheme. *Journal of the Engineering Mechanics Division*, 5(102):749756, 1976.
16. M. Jamali and M. Ester. Trustwalker: A random walk model for combining trust-based and item-based recommendation. In *SIGKDD*, KDD, pages 397–406. ACM, 2009.
17. C. A. D. Jr., G. L. Pappa, D. R. R. de Oliveira, and F. de Lima Arcanjo. Inferring the location of twitter messages based on user relationships. *Trans. GIS*, 15(6):735–751, 2011.
18. H. Kondylakis, L. Koumakis, E. Kazantzaki, M. Chatzimina, M. Psaraki, K. Marias, and M. Tsiknakis. Patient empowerment through personal medical recommendations. In *MEDINFO*, 2015.
19. J. A. Konstan, B. N. Miller, D. Maltz, J. L. Herlocker, L. R. Gordon, and J. Riedl. Grouplens: Applying collaborative filtering to usenet news. *Commun. ACM*, 40(3):77–87, 1997.
20. Y. Koren. Collaborative filtering with temporal dynamics. In *SIGKDD*, 2009.
21. R. Li, C. Wang, and K. C. Chang. User profiling in an ego network: co-profiling attributes and relationships. In *23rd International World Wide Web Conference, WWW '14, Seoul, Republic of Korea, April 7-11, 2014*, pages 819–830, 2014.
22. R. Li, S. Wang, H. Deng, R. Wang, and K. C.-C. Chang. Towards social user profiling: unified and discriminative influence model for inferring home locations. In *The 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '12, Beijing, China, August 12-16, 2012*, pages 1023–1031, 2012.
23. J. McAuley and J. Leskovec. Hidden factors and hidden topics: Understanding rating dimensions with review text. In *RecSys*, 2013.
24. J. J. McAuley and J. Leskovec. From amateurs to connoisseurs: modeling the evolution of user expertise through online reviews. In *Proceedings of the 22nd international conference on World Wide Web*, pages 897–908. ACM, 2013.
25. R. J. Mooney and L. Roy. Content-based book recommending using learning for text categorization. In *ACM DL*, 2000.
26. M. Pennacchiotti and S. Gurumurthy. Investigating topic models for social media user recommendation. In *WWW*, pages 101–102, 2011.
27. A. Prelic, S. Bleuler, P. Zimmermann, A. Wille, P. Bühlmann, W. Gruissem, L. Hennig, L. Thiele, and E. Zitzler. A systematic comparison and evaluation of biclustering methods for gene expression data. *Bioinformatics*, 22(9):1122–1129, 2006.
28. A. Sadilek, H. A. Kautz, and J. P. Bigham. Finding your friends and following them to where you are. In *Proceedings of the Fifth International Conference on Web Search and Web Data Mining, WSDM 2012, Seattle, WA, USA, February 8-12, 2012*, pages 723–732, 2012.
29. K. Stefanidis, E. Ntoutsi, M. Petropoulos, K. Nørvåg, and H. Kriegel. A framework for modeling, computing and presenting time-aware recommendations. *Large Scale Data and Know. Centered Sys.*, 10:146–172, 2013.
30. K. Stefanidis, N. Shabib, K. Nørvåg, and J. Krogstie. Contextual recommendations for groups. In *Advances in Conceptual Modeling - ER 2012 Workshops*, pages 89–97, 2012.
31. J. Tang, L. Yao, D. Zhang, and J. Zhang. A combination approach to web user profiling. *TKDD*, 5(1):2, 2010.
32. M. Wischenbart, S. Mitsch, E. Kapsammer, A. Kusel, B. Pröll, W. Retschitzegger, W. Schwinger, J. Schönböck, M. Wimmer, and S. Lechner. User profile integration made easy: model-driven extraction and transformation of social network schemas. In *WWW*, 2012.
33. W. X. Zhao, J. Jiang, J. Weng, J. He, E. Lim, H. Yan, and X. Li. Comparing twitter and traditional media using topic models. In *ECIR*, pages 338–349, 2011.
34. M. Zimmermann, E. Ntoutsi, and M. Spiliopoulou. Discovering and monitoring product features and the opinions on them with OPINSTREAM. *Neurocomputing*, 150:318–330, 2015.

Recommended Reading

[1] Yue Shi, Martha Larson and Alan Hanjalic. Collaborative Filtering beyond the User-Item Matrix: A Survey of the State of the Art and Future Challenges. ACM Comput. Surv., 47(1), 3:1–3:45, 2014.
[2] Kostas Stefanidis, Eirini Ntoutsi and Haridimos Kondylakis. Information Hunting: The Many Faces of Recommendations for Data Exploration. ACM SIGMOD Blog http://wp.sigmod.org/?p=1580.