

Recommender Systems and the Social Web

Amit Tiroshi, Tsvi Kuflik, Judy Kay and Bob Kummerfeld

University of Haifa, Israel.

{atiroshi,tsvikak}@is.haifa.ac.il

School of Information Technologies, University of Sydney, Australia.

{judy.kay,bob.kummerfeld}@sydney.edu.au

Abstract. In the past, classic recommender systems relied solely on the user models they were able to construct by themselves and suffered from the “cold start” problem. Recent decade advances, among them internet connectivity and data sharing, now enable them to bootstrap their user models from external sources such as user modeling servers or other recommender systems. However, this approach has only been demonstrated by research prototypes. Recent developments have brought a new source for bootstrapping recommender systems: social web services. The variety of social web services, each with its unique user model characteristics, could aid bootstrapping recommender systems in different ways. In this paper we propose a mapping of how each of the classical user modeling approaches can benefit from nowadays active services’ user models, and also supply an example of a possible application.

1 Introduction

Information overload is a phenomenon that has invaded every field in our lives, from work activities (decide which books to order, which emails to read first) to leisure time ones (which movies to see, which restaurants to go). One way to ease the problem is through the use of recommender systems [1], systems that try to match users and items/entities that might interest them. There are several classic approaches for generating recommendations: collaborative filtering [2], content-based [3], case-based [4] and hybrid methods [5]. Most recommender systems require a user model to base their recommendations on and every method described earlier requires a different type of user model. Until a decade ago each system had its proprietary user model, however with the bloom of the internet and connectivity, user models sharing and bootstrapping from online sources are becoming a real possibility. One possible source for bootstrapping user models is the freely available personal information from the social web. Social web services are online services that let their users connect, communicate, share and collaborate with others. Users can link themselves to groups, individuals and causes, they can share all types of content (written, visual, audio) and they communicate both live and in a delayed manner. Each social web service has its unique characteristics which are also reflected in its user model, some let users define

their interests explicitly as a set of features (Facebook¹, LinkedIn²) others do so implicitly and in plain text (Twitter³, Blogs). Facebook only allows a bidirectional connection among users (if user A is connected to B then B is also connected to A) while Twitter users can follow without being followed (user A is linked to B, B is not linked to A). As a result, the social web contains vast amounts of personal information about users that is free and publicly available or can be made available by the users. This information may serve as a source for information used by online recommender systems to bootstrap their user models and to solve the “cold start” problem. In this paper we survey existing social web services and show how the different recommendation approaches (or user model representations) can each benefit from the social web’s available user models, and present an example in the form of a possible application.

2 Background and related work

The recommendation approaches mentioned earlier are the classical ones for handling information overload [6]. Each of them has a unique method for modeling its users’ interests and how to match items accordingly. In this section we will review the various models, starting with the collaborative filtering approach (CF). CF [2] is based on similarity of user preferences, it assumes that users that agreed in the past on items they liked will probably agree on more items in the future. For example, taking one user’s bookshelf and crosschecking it with shelves of other users, finding those with similar books will yield several possible book recommendations for that user. To carry out such an operation, ratings of items must be gathered and stored from a large number of users. This approach is called user-user CF. A variation of CF is to base filtering on similarity of items (item-item CF) rather than similarity of users. A matrix exists, which represents the relationship between each pair of items. Thus, every item listed under the active user can serve as a lead for potential related items in the matrix. Overall, the general user model of CF systems requires a matrix of users’ ratings of items. Having such a matrix is not easy and a major challenge is how to support a new user (new user problem) or how to rate a new item (first rater problem) – both are two aspects of the “cold start” problem of CF.

In the content-based approach, recommendations are made based on content analysis. The content is a set of terms representing an item (Website, Document, Email) or describing it (Movie/Music CD/Restaurant descriptions), usually extracted from the larger textual description of the item. To create a user model, the content that interests the target user is either explicitly given or implicitly learned through machine learning techniques [7]. Then, when new content becomes available, it is analyzed and compared to the user model, and recommended (or not) to the user based on that similarity. Among the most common techniques used for content similarity analysis is the vector space model that uses TF*IDF [8] weighting, Rocchio

¹ <http://www.facebook.com>

² <http://www.linkedin.com>

³ <http://www.twitter.com>

algorithm [9] and the naïve Bayesian classifier [10]. Another approach, quite similar to the content-based approach is a feature based one, where users (and items) are represented by preferences of specific features (like movie genre, book author etc). Again, these features form an n-dimensional vector where similarity of users and items may be measured by a cosine in an n-dimensional space.

The third approach, case-based [4], is another variation of the content-based approach, aimed at generating better recommendations for feature described items such as consumer products, based on past interactions with the system by similar users. In this approach, user sessions are recorded and when similar users request recommendations – similar sessions generated by similar users (users with similar preferences), are used as a basis for recommendations. An exemplifying implementation is presented in [11].

Hybrid recommender systems [5] are systems that combine two (or more) approaches together in order for them to overcome each other's shortcomings. For example, a system that combines the collaborative filtering approach with a content-based recommender can overcome the first rater problem by matching new items using content analysis, as demonstrated at [12].

The Social Web was introduced in [13] as a project in which people could create an online representation for themselves, get organized in groups and communities, share knowledge and items while interacting and collaborating with others. Since then services implementing those concepts have evolved and currently many variations can be found. Among the commonly adopted ones are Facebook, Twitter, Flickr⁴ and Blogger⁵. Facebook is a social web service (also categorized as a Social Networking Service – SNS) that focuses on personal life aspects, its users are able to create an online rich representation of themselves, containing elements varying from detailed profile attributes to personal photo albums and status sharing. Facebook users are encouraged to connect and interact with others whom they know and join groups based on shared interests. Flickr on the other hand is a social web service designed for photography hobbyists and professionals, its users can upload their works and share them publicly or to specific interest groups. Other users can then comment and tag elements in those photos and socially interact. Blogger is a service that allows its users to log their thoughts and happenings online and share them with others. Such "Posts" can then be commented on by other users, leading to social interaction. Twitter is also a blogging service, however for micro posts which do not exceed 140 characters. Twitter is characterized by an additional informal usage pattern in the form of frequent real-time updating. The user models of these services are accessible to 3rd parties through APIs with the users' consent. On the next section we will map between those services and their possible contribution to classical recommendation approaches user models. The services mentioned above and in the next section are the leading representatives of current social web services (usage wise). The given map can be used to project additional existing services' contribution to user models based

⁴ <http://www.flickr.com/>

⁵ <http://blogger.com/>

on their similarity to the chosen services. For example LinkedIn⁶ is a social web service that shares a similar concept to Facebook but aimed at connecting professionals instead of social friends. Thus in the approaches Facebook is mapped to, it would also be suitable to use LinkedIn if the context of the application is more "Professional" than social. Additional relatively less adopted social web services are: Plurk⁷ and Tumblr⁸ which belong to the micro-blogging category Twitter is part of. There are also many photo sharing services⁹ and blogging services¹⁰. One service that stands out is YouTube¹¹ being an exact match to Flickr in the map presented on the next section both in scale and properties with the only difference being that it serves videos instead of photos, a difference that does not affect its contribution relatively to Flickr.

Social web services have been used for bootstrapping user models. In an early study [14], social web service profiles were captured and mapped to a "Taste Fabric" using ontologies of books, music, movies and more. The taste fabric was constructed using machine learning techniques to infer semantic relevance among the ontologies. It was then used to recommend new items to users who share the same cluster of "Taste". In another study [15] bootstrapping a Scrutable User Modeling Infrastructure (SUMI) from fragments of the user's user model located at various social e-networking and e-commerce domains was explored. Using APIs from Facebook, Amazon, eBay and Google OpenSocial, SUMI was able to harness users' data for its own purpose of Lifelong User Modeling and personalized learning.

3 Mapping Social Web Services Contribution to Classical Recommender Systems Approaches

Social Web Services, by nature, contain large amounts of personal information about their users. Some details may be publicly available while more may be kept private and released explicitly by the users. This information can be valuable for online recommender systems seeking to bootstrap a user model for first time users, in order to overcome the "cold start" problem, where without any personal information (or interaction history) the system is unable to provide a personalized service to the new user. It can also be used to enrich existing models with complementary data from different domains.

⁶ <http://www.linkedin.com/>

⁷ <http://www.plurk.com/>

⁸ <http://www.tumblr.com/>

⁹ http://en.wikipedia.org/wiki/List_of_photo_sharing_websites

¹⁰ http://en.wikipedia.org/wiki/Category:Blog_hosting_services

¹¹ <http://www.youtube.com/>

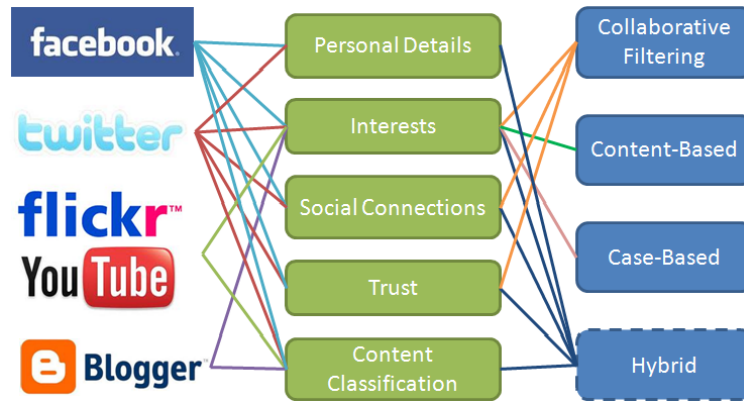


Fig. 1. Mapping of Social Web Services and their possible contribution to classical Recommender Systems User Models

Figure 1 illustrates the leading social web services and their possible contribution to the user models of classical recommender systems. We will now analyze the specific contribution to each approach starting with the CF approach. Since CF relies on user ratings of items, bootstrapping those ratings from the social web services could have a tremendous contribution. The networks offering information that resembles such ratings are Facebook, Twitter and Blogger. On Facebook, users can explicitly declare their interests through profile features, association with groups and fan pages or through status line updates. Such attributes once extracted can be mediated [16] to ratings on items, for example: a user linking her profile to 'Levis' fan page is essentially rating the brand and its products as favorable. The same process can be used for tweets (the name for a Twitter post), however methods such as sentiment analysis [17] are required in order to resolve the precise rating, since an open text sentence regarding 'Levis' for example, can be a statement of endorsement or of hate. Flickr being a visual content sharing hub is less helpful in the interests bootstrapping process and thus not linked in the mapping above. A second possible contribution social web services could have for CF is related to the social links they store as part of their user models. Social links might serve as an indicator of trust among users, and trust could be an important factor among raters in a collaborative filtering system. In a research by [18] a collaborative filtering system is demonstrated, in which users can request recommendations based on items rated by specific users whose ratings they trust. Facebook's social links along with the mutual interests of the two people connected could supply this trust factor. On Twitter the people a user follows can serve as raters in whom she trusts on the specific subjects tweeted about.

Content-based recommender systems require a set of terms representing the content the user is interested at. These terms can be extracted from the user's social web service profile, in which the text tends to be short and focused. Such interest terms can be extracted from Facebook fan pages and groups the user is associated with, the group/page names themselves are suitable (as in the 'Levis' example) and additional terms can be found in the accompanying short descriptions. Additional

short open text fields that could ease the term identification process are status lines and wall messages in Facebook, Twitter's messages which are limited to 140 characters are appropriate too. Blogger posts and blogs in general are more extensive in content than the services mentioned above, therefore their contribution would be similar to classical content based sources. Their advantage is that they already contain the user's content of interest organized in a single point of access, hence serve as a more comfortable bootstrapping source. Although Flickr is not a textual site, the tags used to annotate images (addressed in Figure 1 as "Content Classification") can possibly serve as focused terms of interests for the sharing user, this approach was explored by [19] which also surveys additional similar methods. Content classification (aka "Tagging") also exists in all the mentioned social web services and could be used in the same way mentioned, on each service the content elements which are "Taggable" varies.

Case-based recommender systems having their origins in content-based ones can benefit from social web services in the same ways mentioned above. A possible unique contribution of social web services to case-based recommendations could be in the form of bootstrapping feature weights. Instead of requiring users to rank features based on their importance (for example price vs. color), those can be retrieved using stereotypically matching user profiles to predefined weight vectors. For example if an online consumer recommending system has mapped their products to various consumer stereotypes (students vs. professionals) and set for each stereotype a preset of feature weights, now all that has to be done is find whether a user is a student or professional, a detail that is available on a social web service such as Facebook.

Hybrid recommender systems can benefit from the fact that some users have their social web service profiles linked together, hence having different representations of the user that can complete each other without needing to manually link between the two system profiles (identity linking). A user that is both a member in Facebook and Twitter for example, and has those two profiles connected using methods similar to such mentioned at [20], can permit a hybrid recommender system to use the first to bootstrap its CF user model part and the second to bootstrap its content-based user model. In case the user's social profiles were not priorly linked the recommendation system can attempt to link them automatically by using personal details features available on both social web service as shown in Figure 1, there are commercial social data aggregation services which do this, for example ZoomInfo¹². Another option for hybrid systems to enrich their models is to use classified content with identical tags across services, for example photos of a user from Flickr can be matched to textual items from other services and users, that were tagged identically, thus aiding in bootstrapping a content based user model. An issue which requires attention when using social data from multiply sources is user modeling interoperability. Each source's user model can have its unique data representation and formats, leading to a need in translation/conflict resolution/mediation methods that could integrate them all into a unified model. Such methods were surveyed in depth in a recent study [21].

¹² <http://www.zoominfo.com/>

4 Theoretical Use Case

To illustrate the potential benefits which were described in the previous section we would like to propose a theoretical example of a socially enhanced museum guidance system. The purpose of the system would be to offer personalized museum tours tailored to users' interests as reflected by their social data. Systems for personalizing the museum experience were studied in [22] [23] [24] [25] and more, various mechanism were required to initialize those systems' user models. In the proposed approach all that is required is visitors consent for the museum's personalization system to access their online social web profiles in order to bootstrap a local user model. Once the museum's system has access to the various social profiles of the visitor and its local user model is bootstrapped, exhibits of interests can be recommended using any of the classical approaches. Actual links that were manually found between real life exhibits presented in the Hecht Museum¹³ and public social profiles are also attached to demonstrate the approaches suggested.

A content-based exhibits recommendation method for example would use the user's Twitter stream as a source for terms of interests. The terms would be extracted using a method such as Bag of Words [8], and then matched against content describing the museum exhibits using content analysis methods. If the user had tweeted about cosmetics and the museum hosts exhibits related to that they would be recommended for a visit (Figure 2 and 3).



Fig. 2. A Twitter post (Right) about cosmetics and a related exhibit (Left) in Hecht Museum that could be recommended to its owner

A different approach that can be combined with the one mentioned above would make use of the users' social profiles and log of visited exhibits to personalize future visitors experience based on CF. If a user visited certain exhibits and her/his Facebook page mentions she/he is a "Fan" of certain items, those would be saved for later matching

¹³ <http://mushecht.haifa.ac.il/>

against new visitors profiles. New visitors would be recommended exhibits that were viewed by people whom they most resemble based on the items they are "Fan" of. Another interesting case for social web services user model bootstrapping would be in hybrid recommender systems. An exemplifying scenario would be museum visitors that have taken photos of exhibits they have seen and tagged them, those can then serve as a basis to identify visitors with similar interests (using the CF approach on social profiles) and recommend those tagged exhibits or similar ones, based on content-base recommendation. The recommendation process would be in the lines of: Find user profiles resembling current visitor's profile, extract tagged photos that are also related to museum's key terms, recommend exhibits relating to those. The great advantage in this case is in the fact that the two user models (CF and content based) are already linked together through the social web services, thus identity linkage is not required. Also such links between profiles allow users to maintain their partial models in the services fitting best for them, for example a user using photos could store them in a service that specializes in it such as Flickr, and link that profile to a Facebook profile which is more suitable for maintaining social relations online.

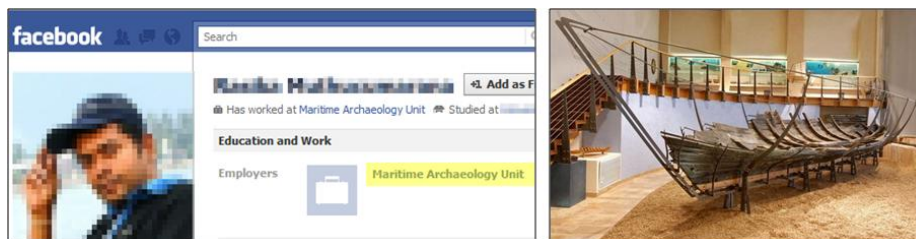


Fig. 3. A visitor whose Facebook profile (Left) states he works in a Maritime Archaeology Unit might be recommended the Ma'agan Mikhael shipwreck exhibition (Right) in Hecht Museum

Finally social web service based recommender systems can also contribute to future uses whether they will be using the same system or 3rd party systems by asking permission from the users to update their online social profile with information related to their latest usage. In our example this would be done by asking visitors for permission to update their Facebook/Twitter streams with summaries of the tour they have taken, e.g. a list of exhibits visited and personal photos taken with them, relevantly tagged. It could enrich the users' experience by giving them a memoir of their visit, and also serve other museum systems in knowing which exhibits to recommend to them.

5 Discussion

This paper surveys social web services and presents a mapping between them and possible usage of their data to enhance classical recommendation approaches' user models. We have also presented a theoretical example for a recommender system that is based on the mentioned methods, and illustrated on real publicly available social

data how it can be linked to actual exhibits. Future work will focus on concrete evaluations of the methods proposed. Also we would like to extend the mapping to modern recommendation approaches such as Social Tagging Based and Group Based.

Social data usage comes along with the responsibility to preserve its owner's privacy. Besides the elementary rules of using the user's data only for the purposes to which permission was granted and not to forward it to unauthorized parties, there are also some less obvious rules that should be taken into account (e.g., for how long can the data retrieved from a social web service be stored by a recommendation service? This is important in order to prevent the service from using outdated data that could lead to misleading/offensive recommendations). It was out of this study's scope to cover this issue, however a future study should offer a mapping of privacy risks and preservation techniques corresponding to the utilization approaches suggested.

6 Bibliography

1. Resnick, P., Varian, H.: Recommender systems. *Commun. ACM* 40(3), 56-58 (March 1997)
2. Schafer, J., Frankowski, D., Herlocker, J., Sen, S.: Collaborative filtering recommender systems. Springer-Verlag, Berlin, Heidelberg (2007) 291-324
3. Pazzani, M., Billsus, D.: Content-based recommendation systems. Springer-Verlag, Berlin, Heidelberg (2007) 325-341
4. Smyth, B.: Case-based recommendation. Springer-Verlag, Berlin, Heidelberg (2007) 342-376
5. Burke, R.: Hybrid web recommender systems. Springer-Verlag, Berlin, Heidelberg (2007) 377-408
6. Brusilovsky, P., Kobsa, A., Nejdl, W.: The adaptive web: methods and strategies of web personalization. Springer-Verlag New York Inc (2007)
7. Webb, G., Pazzani, M., Billsus, D.: Machine Learning for User Modeling. *User Modeling and User-Adapted Interaction* 11(1), 19-29 (March 2001)
8. Baeza-Yates, R., Ribeiro-Neto, B.: Modern Information Retrieval. ACM Press / Addison-Wesley (1999)
9. Rocchio, J.: Relevance feedback in information retrieval. Englewood Cliffs, NJ: Prentice-Hall (1971) 313-323
10. Duda, R., Hart, P.: Pattern Classification and Scene Analysis. John Wiley & Sons Inc (1973)
11. Nguyen, Q. N., Cavada, D., Ricci, F.: Trip@ dvice Mobile Extension of a Casebased Travel Recommender System. (2003)
12. Claypool, M., Gokhale, A., Miranda, T., Murnikov, P., Netes, D., Sartin, M.: Combining Content-Based and Collaborative Filters in an Online Newspaper. (1999)
13. Hoschka, P.: CSCW research at GMD-FIT: from basic groupware to the social

Web. SIGGROUP Bull. 19, 5-9 (1998)

14. Liu, H., Maes, P., Davenport, G.: Unraveling the taste fabric of social networks. *International Journal on Semantic Web and Information Systems* 2(1), 42-71 (2006)
15. Kyriacou, E., others: Enriching Lifelong User Modelling with the Social e-Networking and e-Commerce" Pieces of the Puzzle". (2009)
16. 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 (August 2008)
17. Pang, B., Lee, L.: Opinion Mining and Sentiment Analysis. *Found. Trends Inf. Retr.* 2(1-2), 1-135 (January 2008)
18. Goldberg, D., Nichols, D., Oki, B., Terry, D.: Using collaborative filtering to weave an information tapestry. *Commun. ACM* 35(12), 61-70 (December 1992)
19. Guy, I., Zwerdling, N., Ronen, I., Carmel, D., Uziel, E.: Social media recommendation based on people and tags. In : *Proceeding of the 33rd international ACM SIGIR conference on Research and development in information retrieval*, New York, NY, USA, pp.194-201 (2010)
20. Abel, F., Henze, N., Herder, E., Krause, D.: Linkage, aggregation, alignment and enrichment of public user profiles with Mypes., pp.11:1--11:8 (2010)
21. Carmagnola, F., Cena, F., Gena, C.: User model interoperability: a survey. *User Modeling and User-Adapted Interaction*, 1-47
22. Kuflik, T., Sheidin, J., Jbara, S., Goren-Bar, D., Soffer, P., Stock, O., Zancanaro, M.: Supporting small groups in the museum by context-aware communication services. In : *IUI*, pp.305-308 (2007)
23. Kuflik, T., Stock, O., Zancanaro, M., Gorfinkel, A., Jbara, S., Kats, S., Sheidin, J., Kashtan, N.: A visitor's guide in an active museum: Presentations, communications, and reflection. *J. Comput. Cult. Herit.* 3(3), 11-1 (February 2011)
24. Bright, A., Kay, J., Ler, D., Ngo, K., Niu, W., Nuguid, A.: Adaptively Recommending Museum Tours. In : *Proceedings of the UbiComp 2005 Workshop on Smart Environments and their Applications to Cultural Heritage* (2005)
25. Zancanaro, M., Kuflik, T., Boger, Z., Goren-Bar, D., Goldwasser, D.: Analyzing Museum Visitors' Behavior Patterns. In : *User Modeling*, pp.238-246 (2007)