

The KeepUP Recommender System

Andrew Webster

Department of Computer Science
University of Saskatchewan
Saskatoon, SK, Canada, S7N 5C9

asw292@mail.usask.ca

Julita Vassileva

Department of Computer Science
University of Saskatchewan
Saskatoon, SK, Canada, S7N 5C9

jiv@cs.usask.ca

ABSTRACT

In this short paper, we describe our RSS recommender system, KeepUP. Too often recommender systems are seen as black box systems, resulting in general perplexity and dissatisfaction from users who are treated as passive, isolated consumers. Recent literature observes that recommendations rarely occur within such isolation and that there may be potential within more socially-orientated approaches. With KeepUP, we outline the design of a recommendation process that is based on an implicit social network where the relevancy and meaning of information can be negotiated not only with the recommender system but also with other users. Our overall goal is to support the traditional notion of “word of mouth” rather than attempting to completely automate it.

Categories and Subject Descriptors

H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval – *information filtering*.

General Terms

Design, Human Factors, Theory.

Keywords

Online communities, social networks, social visualization, RSS, collaborative tagging, word of mouth.

1. INTRODUCTION

Our main claim throughout this paper is that users should play a more active role in the recommendation process rather than being limited to a passive, isolated role. While conversational recommender systems (e.g. [3]) do engage users in an active role by means of a *dialog*; however, users are still left separately negotiating the meaning and relevancy of their recommendations with the system and not with their peers. It has been asserted that “recommendations are not delivered within a vacuum, but rather cast within an informal community of users and social context” [6, p.131]. This assertion has emerged from the observation that many recommender systems create *implicit social networks* as a side-effect of the recommendation process and that these networks have important, exploitable properties including self-organization. An implicit social network is inferred from system data and does not necessarily reflect any “real” relationship, such as friendship,

between users. While there have been a number of existing graph-based techniques used in recommender systems, such as *spreading activation* [5] and *horting* [2], these techniques target specific issues (e.g. sparsity). The theoretical work presented here is a broad approach that attempts to capture the flow of information in social networks and support the informal user communities that may form around specific topics. We are not necessarily interested in improving the system’s recall and precision measures, etc.

We developed a graph-based approach that was evaluated against the well-known *100K MovieLens* data set [10]. Briefly, an implicit social network is created between two or more users whenever a shared interest is determined. A directed edge between users, u and v , represents the strength at which u influences v within the respective network regarding the recommendation of information items. Influence simply represents the probability that whatever u likes, v will like as well. Recommendation of items is accomplished through two interrelated processes: push and poll. *Push* seeds an item into the network(s) of users that the item is initially deemed most relevant to. The item then spreads according to *diffusion of innovation* [7] models. *Poll* queries adjacent users whether the item should be recommended for the active user, given a certain *activation threshold*. Feedback from users (explicit and implicit) reshapes the network, affecting the spread and activation of subsequent items. Since items are “pushed” through these networks, the presence and strength of edges between users are crucial factors in determining recommendations. Thus, for each network, the active user can view her *neighbours* (i.e. adjacent users) through an interactive visualization. In addition to displaying the degree of influence each neighbour exerts on the active user’s recommendations (and vice-versa), the active user can manually adjust neighbours’ influence, triggering KeepUP to instantly “re-recommend” a small set of items which appear along with the visualization. This is similar to work done in [1] where e-commerce users are allowed to restrict the collaborative filtering process to a set of manually selected *contacts* each of whom have a level of *credit* or trust that is factored into the final recommendation of items. It was shown that these “local” recommendations were better than those made by unrestricted k-nearest neighbour collaborative filtering.

In our earlier work [10], we demonstrate that the push-poll approach outperforms a straightforward collaborative filtering algorithm on predicting user ratings on movies. However, predicting item ratings on a numerical scale is not the objective of push-poll which is more advantageous in directing new information to interested users who can then collaborate on further classifying the information. Also, a valid critique of such an evaluation is whether users actually notice a difference within an (albeit small) improvement. We propose a multi-stage

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

RecSys’07, October 19–20, 2007, Minneapolis, Minnesota, USA.
Copyright 2007 ACM 978-1-59593-730-8/07/0010...\$5.00.

evaluation that begins with investigating the effectiveness of push-pull at building and maintaining implicit social networks, distributing RSS items through these networks, and making recommendations. Next, we intend to evaluate what impact the visualization has on user behaviour, whether users find it beneficial, and if prediction accuracy tends to increase due to manual influence adjustments.

The rest of the paper is as follows: KeepUP's key features and UI are highlighted in Section 2.1 before outlining our design for supporting informal communities in Section 2.2. The neighbour visualization is shown and briefly discussed in Section 2.3. Section 3 concludes with an outline of our proposed evaluations.

2. KEEPUP

KeepUP is a *Rich Site Summary* (RSS) recommender system, developed at the University of Saskatchewan and available online at <http://keepup.usask.ca>. It is considered a hybrid recommender system as it mixes collaborative filtering with content-based analysis. RSS is a popular standard to publish content to the Web and is often used by web logs (*blogs*) and news services to alert subscribers to new content. RSS entries or *items* follow well-known XML formats and usually include a headline, a short description, and a URL to the full item of interest. A RSS feed is simply a Web accessible XML document that contains 1 or more items and is updated regularly.

The name "KeepUP" implies that users are able to "keep up-to-date" with personally relevant news and events. The breadth of topics and overwhelming number of RSS feeds presents an exciting challenge for a recommender system that must manage many new and diverse items per day. After a 90-day period, over 220,000 items have been indexed by KeepUP, yet there are only approximately 110 registered RSS feeds.

2.1 Channels

Recommendation lists, or what we call *channels*, are the building blocks of KeepUP. It is helpful to imagine a RSS feed as an unending stream of items where channels are filters that select items from feeds based on certain criteria. There are three types of channels:

1. **Feed Channel:** collects items from select RSS feeds.
2. **Tag Channel:** collects items that match a set of specified keywords, or *tags*. For example, the channel in Figure 1 is based on the tags *climate change*, *Kyoto*.
3. **Person Channel:** collects items that were rated positively by specified users.

Each type of channel suits a different information-gathering purpose. For example, imagine a user who has a number of favourite Web sites and would like to stay current with their updates. This user may want to have the respective RSS feed of each Web site grouped into a single *feed* channel that will show the new items appearing at any of the Web sites. Or, a user may be more interested in a specific topic and less concerned where the information comes from. In this case, a *tag* channel would be more appropriate. The tag channel will select all items that contain the matching tag or tags regardless of the RSS feed they come from. Finally, a *person* channel allows users to see what items are liked by their friends and colleagues.

Channels display recommended items as a list of headlines (Figure 1), and multiple channels can be displayed on a single

page. Note that all items appearing in a channel have been recommended (Section 2.3) to the user. Users can then quickly scan each channel for items of the most interest.

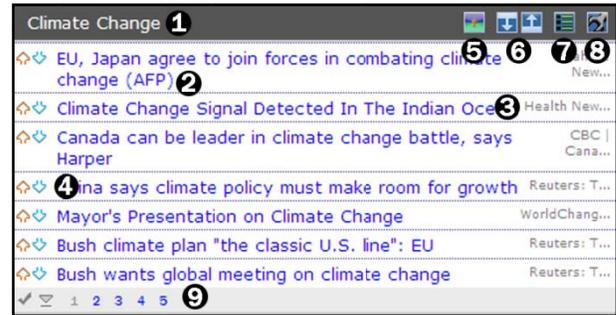


Figure 1. An example channel regarding "climate change."

The highlighted areas in Figure 1 are described as follows:

1. **Channel Title:** user-defined title.
2. **Item Headline:** in its collapsed state, an item shows only its headline. Clicking a headline expands the respective item (Figure 2) and marks the item as read.
3. **Item Source:** truncated name of the source RSS feed.
4. **Rate Item:** users can immediately rate an item positively (up arrow) or negatively (down arrow).
5. **Channel Neighbours:** indicates that there are other users sharing the current channel. Clicking the icon takes the user to the interactive neighbour visualization (Section 2.3).
6. **Expand or Collapse All Items:** a shortcut to expand or collapse all the items (does not mark the items as read).
7. **Sort Items:** items can be sorted by recommendation, date, title, or Web site (i.e. RSS feed).
8. **Channel Options:** expands to reveal additional options including deleting the channel, setting the maximum number of items to display at a time, etc.
9. **Misc. Options:** the user can mark all items as read (checkmark icon), refresh the channel (triangle icon), or move to a different page to see more items.



Figure 2. An expanded item in the channel shown in Figure 1.

The highlighted areas in Figure 2 are described as follows:

1. **Link to Full Story:** opens a new Web browser window which displays the complete story.
2. **Link to Item Source:** opens a new Web browser window which displays the feed's associated Web page.

3. **Time of Index:** the amount of time that has elapsed since KeepUP first indexed the item.
4. **Item Description:** depending on the feed, varies from a single sentence summary to the complete story including graphics, videos, etc.
5. **Add to Favourites:** bookmark items for later reading.
6. **Item Tags:** the 5 most popular tags currently applied to the item (users are encouraged to add their own).

Whenever a user creates a new channel, it is available for all other users to subscribe to (Figure 3). For example, multiple users may “join” the Climate Change channel of Figure 1, adding it to their list of channels. When this occurs, the newly subscribed user is “weaved” into the existing network of currently subscribed users. This process involves computing an initial correlation value between the new user and each existing user to use as the edge weight (i.e. influence) between the pair. If there are not enough previously rated items in common to perform a Pearson correlation, then we compare user profiles for similarity. Finally, if no edges can be established, then the new user is randomly connected to a subset of the existing users. It is important that the new user have a least one incoming edge from the network to begin receiving items within the respective channel. In the deployed system, there will be privacy controls which will allow users to anonymously create or join channels. For now, we ignore privacy issues.

Channel	Type	Started By	Description	# Users
join ESB	(Feed) 'ebioQ - Enterpr...'	██████		2
join Madmuc's Channel	(Tag) madmuc	██████	Tag stories with 'madmuc' if relevant to others in the Madmuc lab	1
join Social Computing and Web 2.0	(Feed) 'Many-to-Many' 'Meetings about T...' 'Sobleizer' 'Mc...' 'Webloggrad'	██████		4
join Tech News	(Feed) 'An Technica' 'CML.com - Technu...' 'Google News Can...' 'TechCrunch'	██████	New gadgets, software, cool gizmos.	1
join The Interesting Channel	(Tag) interesting	██████	Tag a story with 'interesting' if you wish to share that story with others in this channel	6

Figure 3. A list of channels that users can potentially join—each channel is displayed with its type, who started it, an optional description and the current number of subscribed users.

2.2 Towards Supporting Communities

We believe recommender systems have a large, *supportive* role to play in the exchange of information between users. For instance, a challenge was put forth in [9] to apply a recommender system to the task of forming and supporting communities of interest. One potential opportunity is to consider a large online community such as MySpace where a single user will only ever see a tiny fraction of all available content. Collaborative filtering is difficult to perform at the level of millions of users and items. However, push-poll offers a strategy to use implicit (or explicit) social networks to target subgroups (i.e. informal communities) of users and then track how particular items are received and categorized by each subgroup. The goal is to make users confident that while they are not actively searching for information items, personally relevant items are continually searching for *them*, especially items from “unknown” parts of the network. KeepUP is partly a response to this challenge and takes initial steps in this direction. We believe that the design and structure of channels lets users self-organize and steer the spread of information from the ground up. Tagging allows for a shared vocabulary to emerge in the self-organized communities that spring out of users creating and joining channels of their choice [8].

When deciding which channels to add, users are presented with a choice. One consideration of this choice is that a single tag, feed or person can exist across multiple channels. For example, in a system with hundreds of users, the possibility of a popular feed, e.g. Slashdot, appearing in more than one channel is high. It is likely that Slashdot is mixed in with a number of other science and technology-related feeds. The user can choose the mix that contains her most preferred feeds (thereby joining a channel where the preferences of the existing users are more aligned with hers). Alternatively, the user can create an entirely new feed channel that uses Slashdot mixed with other feeds, for example, a feed of a little-known blogger who often discusses Slashdot articles (thereby creating a new association between Slashdot and the unknown blog), thus expanding the diversity of channels and the choice-options for other users.

Another consideration that a user makes in choosing a channel is the number of users who are “subscribed” to the channel. A user may choose to join a channel that has some undesirable tags, feeds, and people, but has more subscribers than other similar channels, assuming she believes a channel with more users results in better recommendations (in most cases this should hold true). Therefore, the user is exposed to some content that she normally would not consider interesting, but which may turn out to be interesting and useful.

Tagging and tag channels are another form of self-organization as users evolve a shared vocabulary [8]. For instance, the *Interesting Channel* in Figure 3 is based on the tag *interesting*. KeepUP is tasked with initially tagging new RSS items; however, it only considers significant terms present within the item’s text as potential tags. It is unlikely the term “interesting” would be considered significant, and items would not appear in the Interesting Channel automatically. Therefore, users must “power” the channel themselves by tagging items that appear in *other* channels as *interesting*. KeepUP then automatically pushes the newly tagged item into the Interesting Channel. And, as users provide feedback on what they personally find interesting and not interesting by rating the items, we hypothesize the channel network will adapt accordingly, forming clusters which represent persistent, “niche” communities of interest. In turn, these clusters of users could be given extra support (e.g. their own discussion space, incentives to contribute additional information, etc) or targeted with specific items for their feedback.

Unfortunately, KeepUP presently does not offer means for users to communicate with each other (e.g. forum or commenting system). A discussion system based on the principles of push-poll would be an interesting avenue for future work. For example, discussion regarding a certain item could be confined to one subgroup of the network and as the discussion builds and evolves, more and more potentially interested users could be made aware of the discussion by the system.

2.3 Neighbour Visualization

The presence and strength of edges between users (neighbours) in channel networks are crucial factors in determining recommendations. Therefore, we believe it is important for users to be aware of their network “position” and allow them to make manual adjustments to the strength of incoming edges (i.e. their neighbour’s influence).

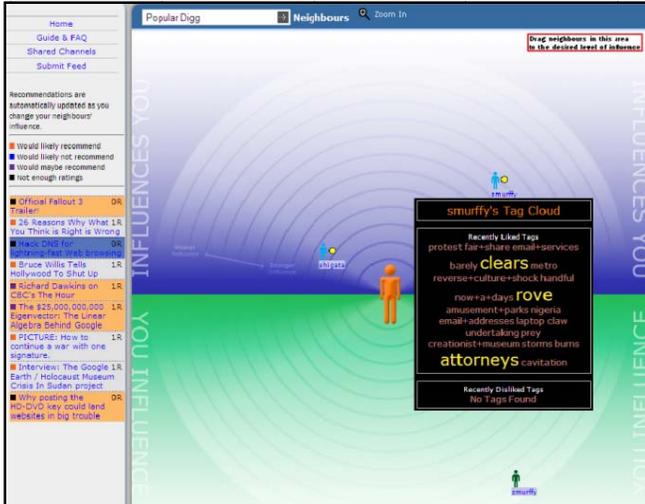


Figure 4. Visualization example with 2 neighbours.

Figure 4 shows an instance of our interactive neighbour visualization for the channel *Popular Digg* (all items appearing in this channel come from Digg, a popular URL-sharing Web site). The visualization is separated into two areas: the top area depicts the amount of influence the active user's neighbours are exerting on her (the closer to the center figure, the more influence); the bottom area depicts the amount of influence the active user is exerting on her neighbours. Neighbours appearing in the top area (person icon with a little dot) can be *dragged* within the top area to the desired level of influence (neighbours in the lower section cannot be moved, i.e. a user cannot adjust the amount of influence she exerts on others). As the active user drags a neighbor her set of recommendations along the left-hand side are automatically "re-recommended" based on the new influence value. This allows the user to see what impact individual neighbours are having on her recommendations and whether this impact is desirable or not.

Finally, if the active user holds the cursor over a neighbour's pseudonym, a *tag cloud* is shown for that neighbour. The tag cloud shows the set of recent tags that are liked and disliked by the neighbour and the relative degree of preference to each tag (i.e. tags in larger font are liked/disliked relative to tags in a smaller font). The tag cloud gives the active user additional information about the neighbour and whether the influence value of that neighbour should be adjusted. For instance, in Figure 4, the example tag cloud shows the neighbour, *smurffy*, has recently liked a number of articles concerning an American political scandal. If the active user is also interested in this particular story, then *smurffy* could be granted greater influence within this channel, increasing the probability that future items concerning the scandal will be recommended to the active user.

3. PROPOSED EVALUATION

Our first evaluation study will involve 20-30 self-selected participants who will each be asked to rate 200 RSS items that fall under 4 broad topics: arts and entertainment, world news, science and technology, and sports. Our main goal with this initial user study is to collect feedback data that will be used to test and optimize KeepUP's ability at forming and maintaining implicit social networks. That is, given a number of users concentrated in a single channel, does clustering tend to occur? And what properties of these clusters and the network in general can be

exploited to provide better recommendations? An interesting aspect of KeepUP is that we can also perform network analysis on the resulting networks and investigate the impact of metrics like *closeness* and *betweenness*? on recommendation accuracy. Finally, participants will use the neighbour visualization in order to qualitatively assess whether participants believe that their recommendations improve as a result of taking direct control.

Later evaluations would focus on how real users (i.e. people who sign-up with KeepUP) use channels and tagging to organize content, whether informal communities develop and what strategies can be adopted to support their development.

Acknowledgement: This work has been supported by the Natural Sciences and Engineering Research Council of Canada through the the LORNET Network and the discovery grant of the second co-author.

4. REFERENCES

- [1] Aïmeur, E. and Mani-Onana, F.S. Better control on recommender systems. *IEEE Joint Conference on E-Commerce Technology (CEC'06)*, 2006, 297-306.
- [2] Aggarwal, C.C., Wolf, J.L., Wu, K-L., and Yu, P.S. Horting Hatches an Egg: A New Graph-Theoretic Approach to Collaborative Filtering. *Proc. Fifth ACM SIGKDD Int'l Conf. on KDD*, 1999, 201-212.
- [3] Burke, R., Hammond, K., and Young, B.C. The FindMe Approach to Assisted Browsing. *Journal of IEEE Expert 12(4)*, 1997, 32-40.
- [4] Golder, S.A. and Huberman, B.A. Usage Patterns of Collaborative Tagging Systems. *Journal of Information Sciences 32(2)*, 2006, 198-205.
- [5] Huang, Z., Chen, H., and Zeng, D. Applying associative retrieval techniques to alleviate the sparsity problem in collaborative filtering. *ACM Trans. Info. Syst. 22(1)*, 2004, 116-142.
- [6] Perugini, S., Gonçalves, M., and Fox, E. Recommender Systems Research: A Connection-Centric Survey. *Journal of Intelligent Information Systems 20(2)*, 2003, 131-160.
- [7] Rogers, E. *Diffusion of Innovations*, 5th Edition. Free Press, New York, 2003.
- [8] Sen, S., Lam, S.K., Rashid, A.M., Cosley, D., Frankowski, D., Osterhouse, J., Harper, F.M., and Riedl, J. tagging, communities, vocabulary, evolution. In *Proceedings of Computer Supported Collaborative Work 2006 (CSCW'06)*, ACM Press, 2006, 181-190.
- [9] Terveen, L. and Hill, W. Beyond Recommender Systems: Helping People Help Each Other. *HCI In The New Millenium*, Addison-Wesley, 2001.
- [10] Webster, A. and Vassileva, J. Push-Poll Recommender System: Supporting Word of Mouth. *Proc. User Modeling 2007 (UM 2007)*, Springer-Verlag, Berlin, 2007, 288-297.