CubeThat: News Article Recommender

Sidharth Chhabra
University of Michigan
School of Information
sidc@umich.edu

Paul Resnick
University of Michigan
School of Information
presnick@umich.edu

ABSTRACT
The CubeThat browser extension for Chrome displays recommended additional news stories related to the same topic as the current news story. The recommended stories are organized into clusters, and clusters that the user has already sampled from are grayed out, in order to encourage users to explore multiple aspects of a story. Users can also provide feedback to improve the clustering, by dragging stories from one cluster to another.

Categories and Subject Descriptors
H.5.m Information interfaces & presentation: Miscellaneous

General Terms
Algorithms, Design, Experimentation, Human Factors.

Keywords
Browser Extension, Recommender, Diversity, News, Clustering.

1. INTRODUCTION
News, which was once heavily dependent on print media, is now consumed digitally. The Web provides much more freedom in selection of news sources and types than any other previous media. But Sunstein & others have argued that that freedom will lead to social fragmentation because self-selection of stories will lead people to only expose themselves to like-minded viewpoints [1]. Similarly, Eli Pariser has argued recommender systems, in a race to please the users, create invisible, personal universes of results around them [2]. Among other problems, this could lead people to think that their opinions are more broadly shared than they actually are [3]. However, personalization systems can be designed to deliberately inject diversity into a user's information diet too [7].

Park et. al.[4] found that when people are presented with stories organized into meaningful clusters, in their NewsCube system, it cued people to read more and diverse stories about a single topic. Eveland et. al. [5] found that structuring knowledge so that reading is not a linear activity can be effective in creating dense knowledge structures. In other words, providing affordances while a person is reading an article to traverse other parts of the same topic can lead to denser knowledge structures.

RecSys'12, September 9–13, 2012, Dublin, Ireland.
ACM 978-1-4503-1270-7/12/09.

We will demo a browser extension that encourages users to explore multiple aspects of news topics, by recommending other stories related to the current story the user is reading. The recommendations are not personalized, but as in the NewsCube system, the browser keeps track of which clusters each user has already explored and thus cues them to explore additional clusters.

2. ARCHITECTURE
Figure 1 shows the system architecture. A crawler gets stories from popular news sites, and clusters them by topic. The recommendation engine further clusters articles within each topic. Initially, it applies the algorithm used in Newscube [4] to cluster the articles; later it incorporates user feedback to modify the within-topic clustering. Clustered news articles are served through a browser extension (Chrome) which is the interface for the recommender.

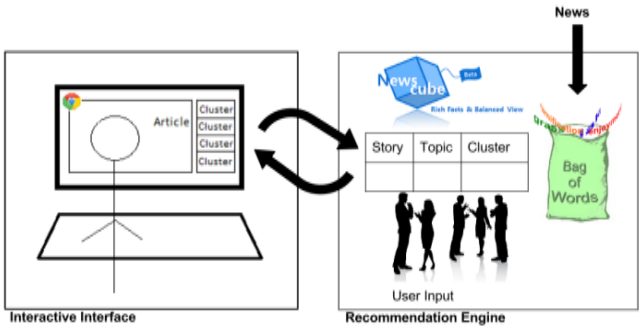

Figure 1: CubeThat Architecture

When a person is reading a news story, he/she can click on the extension icon in the browser header to get recommendations for that article. Figure 2 shows an example recommendation for a news article. For each cluster in a topic, four or less relevant stories are picked. For each story, the source and count of how many Twitter tweets link to the story are shown. A user who disagrees with the clustering of stories can drag a story from one cluster to another and this move affects the recommendations that will be shown to other users as well. If the last story is removed from a cluster, that cluster is deleted, as illustrated in Figure 3, where the sole story from cluster 5 has been moved to cluster 4.

The browser extension tracks which stories users have opened. If a user has previously opened any story in a cluster, that cluster is shaded, as in the original NewsCube interface [4]. Figure 4 shows the recommendation after another story has been read from a different cluster.

Figure 2. An example recommendation; darkened cluster shows the user has already sampled from it.

Figure 3. The recommendations after the sole story from cluster 5 is moved to cluster 4.

Figure 4. The recommendations after a story from cluster 3 has been read.

2.1 CLUSTER CHANGE

When users reassign stories to different clusters or new stories arrive that are part of an existing topic, we have to find a suitable clustering of new & existing stories that takes into account both the text features and the previous user actions. There has been a lot of research on constrained clustering [6] but here we face multiple constraints together. Again, we propose and use a simple solution to the problem. First, from the story assignment to clusters, as adjusted by the most recent user, we compute the cluster centroid for each cluster in the database. For stories that were not displayed as part of the recommendations, including new stories, we compute the distance from cluster centroids and allocate each to the closest. The simplicity of the algorithm ensures that running time is linear in the number of stories about the topic.

3. CONCLUSION & FUTURE WORK

In this paper, we proposed a novel approach to tackle the concerns regarding polarization due to selective exposure when reading news online. The proposed Chrome extension should cue people to read about more different aspects of a news story. In future work, we plan to incorporate other diversity factors besides the words used in the stories, such as the political ideology of the stories, which we have been able to classify automatically using semi-supervised learning algorithms [8]. The usability of the extension will need to be honed through iterative testing and the efficacy of the extension in encouraging people to consume more diverse news will need to be measured through field trials.

4. REFERENCES

[1] Cass R. Sunstein. 2001. *Republic.Com*. Princeton University Press, Princeton, NJ, USA.

[2] Eli Pariser. 2011. *The Filter Bubble: What the Internet is Hiding from You*. Penguin Group , The

[3] Glenn S. Sanders and Brian Mullen. 1982. Accuracy in perceptions of consensus: Differential tendencies of people with majority and minority positions. European journal of social psychology 13(1).

[4] Souneil Park, Seungwoo Kang, Sangyoung Chung, and Junehwa Song. 2009. NewsCube: delivering multiple aspects of news to mitigate media bias. In *Proceedings of the 27th international conference on Human factors in computing systems* (CHI '09). ACM, New York, NY, USA, 443-452. http://doi.acm.org/10.1145/1518701.1518772

[5] W.P. Eveland and J. Cortese. 2004. How Web Site Organization Influences Free Recall, Factual Knowledge, and Knowledge Structure Density. Human Communication Research, 30: 208–233. doi: 10.1111/j.1468-2958.2004.tb00731.x

[6] Sugato Basu, Ian Davidson, and Kiri Wagstaff. 2008. *Constrained Clustering: Advances in Algorithms, Theory, and Applications* (1 ed.). Chapman & Hall/CRC.

[7] Garrett, R. Kelly, and Paul Resnick. 2011. Resisting Political Fragmentation on the Internet. Daedalus 140(4): 108-120.

[8] Zhou, D. X. and P. Resnick (2011). Classifying the Political Leaning of News Articles and Users from User Votes Fifth International AAAI Conference on Weblogs and Social Media, Barcelona, AAAI.