



# International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 6, June 2014

## Survey on News Recommendation

Mansi Sood<sup>1</sup>, Harmeet Kaur<sup>2</sup>

Department of Computer Science, Shyama Prasad Mukherji College, University of Delhi, Delhi, India<sup>1</sup>

Department of Computer Science, Hans Raj College, University of Delhi, Delhi, India<sup>2</sup>

**ABSTRACT:** Recommender Systems have evolved as an answer to information overload problem prevalent with online users, looking for relevant information out of a huge premise of content available online. Such systems are used to provide recommendations to the users guiding them towards items that match their interest areas and choice. News Recommendation is a specific research area under recommender systems where these systems are used to suggest news articles to the users that match their reading interests and personal preferences. However, news recommendation differs from traditional recommendation needs and concerns in many aspects. This paper surveys many such systems built over past few years, and tries to present the special challenges associated with news recommendation over the traditional model of recommendation. It also explores recommendation techniques that can be used to generate news recommendations for the users.

**KEYWORDS:** News Recommendation, User Profile, Personalization, Preference, Content based Filtering, Collaborative Filtering, News Categories.

### I.INTRODUCTION

Due to tremendous and ever increasing growth of information sources available online, the World Wide Web is witnessing a rising demand of intelligent systems that can guide users to find relevant information for themselves. The challenge is to find the right content, something that will satisfy the current information needs or will match up their interest area and preferences. Search engines can help to resolve this problem to an extent, particularly if users are looking for some specific data that can be formulated as a formal query. However, in many cases, users may not even know what to look for. Often this is the case with items like news, movies etc. where users are browsing for things that might match up their interest areas. In such cases, it is better to present recommendations to users based on their interests as demonstrated by them implicitly or explicitly [1].

Recommender systems have evolved as an answer to above concerns by suggesting recommendations to the users for content suited to their needs [2]. They can be defined as systems that produce recommendations as output or has the effect of pruning large information spaces so that users are directed toward items that best meet their needs and preferences [3]. Such systems have a rising demand in an environment where the amount of online information outshines any individual's capability to survey it [4]. It is the criteria of personalized, interesting and preferred that separates the recommender systems from information retrieval systems or search engines [5]. The purpose of search engines is just to match the user query with the available items, information content etc and return the matching content ranked by the degree of match. Recommender systems are now an integral part of many e-commerce websites such as Amazon.com [1, 6]. A variety of approaches have been proposed for devising recommendations for users, like content-based, collaborative, knowledge-based etc. However, all of the known techniques have strengths and weaknesses [7, 8], and many researchers have chosen to combine these techniques in different ways leading to a hybrid method [9, 10] combining benefits of two approaches [11].

A typical area where recommender systems are gaining a lot of importance is the news industry, where news access patterns are changing and getting modernized with the advancements in technology and the ease to browse World Wide Web. Many news sources and agencies have introduced online news access portals which users can access anytime, anywhere to browse through latest news updates. This helps them to stay updated about the latest developments around them without any time delay. However, challenge remains the same or even gets bigger than the other domains i.e. to find the right and relevant content. Browsing these online news sources, users are looking for interesting, latest news amongst a huge number of articles available, something that matches their reading interests and can keep them busy. To



# International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 6, June 2014

attract a good amount of users to their websites, online news sources are increasingly employing recommender systems to improve user experience on their sites. Recommender systems are required to identify individual preferences and reading interest areas of users to present personalized tailored updates to them [12, 13].

## II. CHALLENGES ASSOCIATED WITH NEWS RECOMMENDATION

News article recommendation differs in several ways from other well-known purposes of recommender systems such as the ones for books, movies, music etc.

1. Large volume of news articles makes this domain different from other types of content available online. News articles tend to be in flood within a short period of time, requiring much more computation for recommendation [14, 15, 16, 17, 18],
2. Popularity of news articles changes dramatically along with the time, which differentiates news articles from other items like books, movies etc. Such quick changes in product's demand makes traditional recommendation methods ineffective,
3. Many news articles describe the occurrence of specific events; consist of updates about specific person, place or objects. Users may like such articles due to some special liking about the special topic but recommender systems can barely predict such preferences,
4. Recommendation techniques commonly suffer from cold start problem, which talks about generating recommendations when sufficient information is not available about products to be recommended or about users for whom recommendations are to be generated. Many times there are a lot of new articles which are being read first time, for whom ratings are not available. Also, news portals at times avoid asking users to login before reading news articles. Hence, there is a large fraction of users who appear to be cold start users [19, 15, 20, 16, 21],
5. News providers generally do not require users to create their user profiles and users consume news articles anonymously. Therefore, news recommender systems have to cope without explicit user profiles,
6. Freshness of news articles at times holds an important mark making latest news more important than relevant one. Breaking or trendy news might get a high attention from user even though it appears to be completely unrelated to the user profiles [22],
7. News articles are typically published in a rather unstructured format [18]. The unstructured format of a news story makes it more difficult to analyze than other objects with structured properties,
8. News items typically have short shelf lives. For example, few readers will be interested in weather updates of last two days; few cricket fans will be concerned with the scores of tournament that completed two days ago. In contrast, the shelf lives of products like books and movies extend several months or even years.
9. Similarity between news articles does not necessarily reflect their relatedness. For instance, two news articles might share a majority of words; still their actual topic could be very different [23].

## III. NEWS CATEGORIES

News categorization is common these days in almost all prevalent news sources like news websites, smart phone news applications etc. They classify new articles/headlines into predefined categories like Business, Sports, Technology, Health, and Politics etc. This is helpful for users looking for specific category news as they can directly access the relevant article as per their interests. This section describes Subject Reference System (SRS) to explain one of the standards used for news categorization.

International Press Telecommunications Council (IPTC), an international organization that is primarily focused on developing and publishing Industry Standards for the interchange of news data, jointly with the newspaper Association of America, has developed a coding system called Subject Reference System (SRS) [24]. The system is designed to categorize news material by identifying the general content of a news object. This is done by using a three-level hierarchy where the top level is Subject; the second level is Subject Matter and the third level is Subject Detail. There are 17 top-level Subjects; secondary Subject Matter lists have been developed for each of these subjects. To date, third-



# International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 6, June 2014

level Subject Detail lists are defined only for three Subjects (namely, Economy, Business and Finance; Politics; and Sports).

A unique eight-digit number is assigned to each entry in the three-level Subject hierarchy. This number is broken down as follows:

- The first two digits indicate the top-level Subject. The valid values are 01 through 17. (Leading zero is mandatory.)
- The next three digits indicate the Subject Matter. Default is 000, used when no Subject Matter is specified. The rest of the values (001-999) must be used in conjunction a two-digit Subject number.
- The last three digits, when preceded by valid Subject and Subject Matter numbers, indicate Subject Detail.

Thus, all references are controlled by a fixed eight digit reference number, for example Arts, Culture and Entertainment (ACE) 01000000, Crime, Law and Justice (CLJ) 02000000, Disasters and Accidents (DIS) 03000000, etc.

## IV. RELATED WORK

Many frameworks have been developed for news recommendation over the past few years. Some of them have used content based filtering; some are based on collaborative filtering or a combination of two recommendation techniques resulting in a hybrid approach to generate news recommendations for users. Researchers have also focused on creating and manipulating user profiles to capture users' reading interests so as to track their dynamically changing reading interests. Users' interests are generally classified as short term and long term interests. Short term interest usually is related to hot news events, breaking news, latest headlines etc., which changes quite frequently. However, long term interest often reflects actual user interests.

A combination of textual content analysis with machine learning techniques can be used to find similarity between information content and users' interest areas. News Dude [25], WebMate [21], and SIFT [26] are examples of systems implementing this methodology.

News Dude is a system that compiles a personalized news program. Besides representing user's short-term and long-term interests, it takes into account the news previously heard by the user to avoid presenting the same information twice. It reads news to users, supporting a series of feedback options such as "interesting", "not interesting", "I already know this", etc.

WebMate is a tool that compiles information from a list of URLs that the user wants to monitor (e.g., newspaper home pages) or from the search results using popular engines. The information is selected by its accordance with a user profile, which represents their multiple interests using vectors of terms and their weights.

SIFT is an information filtering system that also models user's interest topics using keyword vectors provided by the user which are updated automatically by relevance feedback.

LOGO [27] integrates the long term and short term reading preferences of users when recommending news items to them. Long term profile of a user is constructed based on a time sensitive weighting scheme [28] and the short-term profile by analysing the latest reading history of the user. Both of them can help determine the news recommendations for users.

Tan and Tee [29] presented a personalized news system named PIN. PIN retrieves and ranks news articles according to the user's profile, which is initially defined by the user as a list of keywords and then learned from user feedback using neural network technology. When interacting with PIN, users provide explicit feedback by rating the articles.

Hochul Jeon, Taehwan Kim and Joongmin Choi [30] proposed a model for personalized information retrieval. Users' information should be extracted to find out similarity between them and information should be recommended to users by similar user groups.



# International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

*(An ISO 3297: 2007 Certified Organization)*

**Vol. 3, Issue 6, June 2014**

A special purpose news browser for PDAs (Personal Digital Assistant), named WebClipping2, is implemented in [31]. WebClipping2 uses a Bayesian Classifier in order to calculate the probability that a specific article would be deemed interesting by the user. Rather than requiring users to provide explicit feedbacks, WebClipping2 observes the total reading time, number of lines read and some other characteristics of user behaviour to infer the user's interests.

The system described in [32] focuses on the change of user preferences. In this system, user's interests are modelled by a multi-layer tree with a dynamically changeable structure, the top layers of which are used to model user interests on fixed categories, and the bottom layers are for dynamic events. This model can track the user's reading behaviours on both fixed categories and dynamic events, and consequently capture the interest changes.

Another personal news agent, PVA [33], uses a proxy to collect user's page clicks and the browsing time, in order to construct a personal view that reflects user interests. PVA is applied and evaluated to provide personalized news access.

Fikadu Gemechu, Zhang Yu and Liu Ting [34] proposed a framework in which a user profile module is incorporated as an integral component of information retrieval process so that results returned by traditional retrieval system are filtered based on the user's profile to meet users' specific information need.

Liang and Lai [35] proposed a time based approach to build user profiles from browsing behaviour, which took into account the time spent by the user on reading the articles and the recent user activities.

Lee, Liu and Cho [36] presented a formal framework and a method to automatically learn user interest based on past click history. The learned user interest is integrated in topic sensitive Page Rank to generate personalized ranking.

## V. ANALYSIS

Many recommendation techniques have been used to build personalized news recommender systems like content based filtering, collaborative filtering, knowledge based methods or a combination of any two techniques giving a hybrid approach to recommendations. It has been observed that recommender systems built on hybrid approach are more successful than the ones based on a specific recommendation technique. This observation comes from the fact that specific requirements and challenges are associated with news recommendation. Simply identifying user preferences and choices at any given time is not sufficient for such systems. These systems should keep track of dynamic user requirements and tastes which changes quite frequently but at the same time gets reverted back to original preferences at times or may add up some special interests in user profiles. These changing interests can be classified as short term and long terms interests where short term interests generally relate to breaking news which keeps on changing, while long term interests reflects users' actual interests. Tracking user profiles keeping in mind all these aspects is absolutely necessary for news recommender systems to work efficiently and effectively.

Asking and obtaining implicit or explicit user feedback can be another important factor for such systems. Although not many users like to invest time in this process, but feedback is important not just to identify user satisfaction level for generated recommendations but also to keep track of changing user preferences and interests.

Also, such systems should take care of peculiarities related with 'news' as an item to be recommended. Article recommendations should not repeat frequently, latest updates should be flashed first; if possible news categorical preferences should be taken care of, providing articles that are relevant to users' reading interests, special events and occasions should be reported tracking user's response to them etc. All these requirements make news recommendations different from traditional recommendation models and hence require combination of multiple techniques to efficiently address all the concerns associated.

## VI. CONCLUSION

This paper surveyed and reviewed news recommender systems built over the past few years, specific requirements, concerns associated with them, and different recommendation approaches used to build such systems. We tried to observe user concerns, behaviour and expectations from such systems and concluded that recommender systems that



# International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 6, June 2014

can keep track of changing user requirements and can present recommendations after identifying user's current reading interests are considered as helpful from user's point of view. Also such systems should track which articles were recommended to users and when so that article recommendations do not repeat frequently and latest updates are flashed first to attract users.

## REFERENCES

- [1] Gediminas Adomavicius and Alexander Tuzhilin, "Towards the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions", IEEE Transactions on Knowledge and Data Engineering, , vol 17, no 6, pp: 734 – 749, June 2005
- [2] A. Ansari, S. Essegiaer, and R. Kohli, "Internet recommendations systems", Journal of Marketing Research, pp 363-375, August 2000
- [3] R. Burke, P. Brusilovsky, A. Kobsa, and W. Nejdl, "Hybrid web recommender systems", The Adaptive Web, LNCS 4321, pp. 377 – 408, 2007
- [4] A M Rashid, I. Albert, D. Cosley, S. K. Lam, S. M. McNee, J. A. Konstan, and J. Riedl, "Getting to Know You: Learning New User Preferences in Recommender Systems", Proceedings of the International Conference on Intelligent User Interfaces, 2002
- [5] N. Belkin, and B. Croft, "Information filtering and information retrieval", Communications of the ACM, Vol 35 No 12, pp 29-37, 1992.
- [6] W. W. Cohen, R. E. Schapire, and Y. Singer, "Learning to order things", Journal of Artificial Intelligence Research, Vol 10, pp 243-270, 1999.
- [7] J. L. Herlocker, J. A. Konstan, L. G. Terveen, and J. T. Riedl, "Evaluating Collaborative Filtering Recommender Systems", ACM Transactions on Information Systems, Vol. 22, No. 1, January 2004
- [8] B. Mobasher, X. Jin, and Yanzan Zhou, "Semantically enhanced collaborative filtering on the web", First European Web Mining Forum, EWMF 2003, Cavtat-Dubrovnik, Croatia, September 22, 2003.
- [9] P. Melville, R. J. Mooney, and R. Nagarajan, "Content-Boosted Collaborative Filtering for Improved Recommendations", Proceedings of the Eighteenth National Conference on Artificial Intelligence, Edmonton, Canada, 2002.
- [10] T. Tran and R. Cohen, "Hybrid Recommender Systems for Electronic Commerce", In Knowledge-Based Electronic Markets, Papers from the AAAI Workshop, Technical Report, WS-00-04, AAAI Press, 2000
- [11] R. Kumar, P. Raghavan, S. Rajagopalan, and A. Tomkins, "Recommendation Systems: A Probabilistic Analysis", Journal of Computer and System Sciences, Vol. 63, No 1, pp 42-61, 2001
- [12] G. Paliouras, A. Mouzakidis, V. Moustakas, and C. Skourlas, "PNS: A Personalized News Aggregator on the Web", studies in Computational Intelligence (SCI), pp. 175–197, 2008
- [13] A. S. Das, M. Datar, A. Garg, and S. Rajaram, "Google news personalization: scalable online collaborative filtering", Proceedings of the 16th international conference on World WideWeb, ACM, pp 271–280, 2007
- [14] F. Garcin, C. Dimitrakakis, and B. Faltings, "Personalized news recommendation with context trees", arXiv reprint arXiv, 2013
- [15] L. Li, D. Wang, T. Li, D. Knox, and B. Padmanabhan, "Scene: a scalable two-stage personalized news recommendation system" In SIGIR, pp 125–134, 2011
- [16] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, and J. Riedl, "GroupLens: an open architecture for collaborative filtering of netnews", CSCW, ACM, pp 175–186, 1994
- [17] K. G. Saranya and G. S. Sadhasivam, "A personalized online news recommendation system", International Journal of Computer Applications, pp 6–14, November 2012
- [18] Y. Lv, T. Moon, P. Kolari, Z. Zheng, X. Wang, and Y. Chang, "Learning to model relatedness for news recommendation", Proceedings of WWW, ACM, pp 57–66, 2011
- [19] H. J. Lee and S. J. Park, "Moners: A news recommender for the mobile web", Expert Systems with Applications, pp 143–150, 2007
- [20] C. Lin, R. Xie, L. Li, Z. Huang, and T. Li, "Premise: personalized news recommendation via implicit social experts", Proceedings of the 21st ACM, international conference on Information and knowledge management, pp 1607–1611, 2012
- [21] L. Chen, K.P. Sycara, "WebMate: A Personal Agent for Browsing and Searching", Proceedings of the second International Conference on Autonomous Agents, Minneapolis, 1998
- [22] B. Fortuna, C. Fortuna, and D. Mladenović, "Real-time news recommender system", Machine Learning and Knowledge Discovery in Databases, pp 583–586, 2010
- [23] J. Liu, P. Dolan, and E. R. Pedersen, "Personalized news recommendation based on click behavior", Proceedings of the 15th international conference on Intelligent user interfaces, ACM, pp 31–40, 2010
- [24] "Subject Reference System Guidelines" Online at <http://www.iptc.org> (as of 13 June 2014)
- [25] D. Billsus, M. J. Pazzani, "A Hybrid User Model for News Story Classification", Proceedings of the Seventh International Conference on User Modeling, Banff, Canada, 1999
- [26] T. W. Yan, H. Garcia Molina, "SIFT – A Tool for Wide-Area Information Dissemination", Proceedings of the USENIX Technical Conference, 1995
- [27] Lei Li, Li Zheng, Tao Li, "LOGO: A Long-Short User Interest Integration in Personalized News Recommendation", Proceedings of the fifth ACM conference on Recommender systems, pp 317–320, 2011
- [28] Y. Ding and X. Li, "Time weight collaborative filtering", Proceedings of CIKM, ACM, pp 485–492, 2005
- [29] Tan, A. and Tee, C., "Learning User Profiles for Personalized Information Dissemination", Proceedings of 1998 IEEE International Joint conference on Neural Networks, pp. 183- 188, May 1998
- [30] Hochul Jeon, Taehwan Kim, Joongmin Choi, "Adaptive User Profiling for Personalized Information Retrieval", Third International Conference on Convergence and Hybrid Information Technology, 2008
- [31] R. Carreira, J. M. Crato, D. Gonalves, J. A. Jorge, "Evaluating adaptive user profiles for news classification", Proceedings of the 9th international conference on Intelligent user interfaces, 2004
- [32] J. Wang, Z. Li, J. Yao, Z. Sun, M. Li, and W. Ma, "Adaptive User Profile Model and Collaborative Filtering for Personalized News", Proceedings of the APWeb 2006, pp. 474-485, 2006





# International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

*(An ISO 3297: 2007 Certified Organization)*

**Vol. 3, Issue 6, June 2014**

- [33] N. Good, J. B. Schafer, J. A. Konstan, A. Borchers, B. Sarwar, J. Herlocker, J. Riedl, “Combining collaborative filtering with personal agents for better recommendations”, Proceedings of the 16th national conference on Artificial intelligence and the 11<sup>th</sup> Innovative applications of artificial intelligence conference innovative applications of artificial intelligence, 1999
- [34] Fikadu Gemechu, Zhang Yu, Liu Ting, “A Framework for Personalized Information Retrieval Model”, Proceedings of IEEE transaction, 2010
- [35] T. P. Liang, and H. J. Lai, “Discovering User Interests from Web Browsing Behavior: An Application to Internet News Services”, IEEE Computer Society, Los Alamitos, CA, USA, 2002
- [36] U. Lee, Z. Liu, J. Cho, “Automatic identification of user goals in Web search”, Proceedings of the 14<sup>th</sup> international conference on World Wide Web, 2005

## BIOGRAPHY



**Mansi Sood** received her Master of Science degree in Computer Science from Department of Computer Science, University of Delhi, Delhi in 2008. She is an Assistant Professor in Department of Computer Science, Shyama Prasad Mukherji College, University of Delhi. She has about two and half years of teaching experience and has published two research papers in reputed refereed journal and conference. She is a member of ACM Computer society. Her research work has focused mainly on recommender systems and generating personalized recommendations for online web users.



**Dr. Harmeet Kaur** received her Ph.D. in Computer Science from the Department of Computer Science, University of Delhi, Delhi, India in 2007. She is an Associate Professor in the Department of Computer Science, Hans Raj College, University of Delhi. She has about 15 years of teaching and research experience and has published more than 25 research papers in National/International Journals/Conferences. Her research interests include Multi-agent Systems, Intelligent Information Retrieval Systems, Trust and Personalization.