News Recommendation Systems Using Web Mining: A Study

Dr. M. Durairaj[#1], K. Muthu Kumar[*2]

[#1]Assistant professor, [*2]Research Scholar
[#1,*2] School of Computer ScienceEngineering and Applications, Bharathidasan University, Trichy, India

Abstract **- News reading has changed with the advance of the World Wide Web (www), from the traditional model of news consumption via physical newspaper subscription to access to thousands of sources via the internet. Online news reading has become very popular as the web provides access to news articles from millions of sources around the world web users are undergoing a transformation and they are now expressing themselves in the form of sharing their opinions on an item through ratings and reviews or comments, through sharing and tagging content, or by contributing new content. Recommendation, filtering, and summary of Web news has received much attention in Web intelligence, aiming to find interesting news and summarize concise content for users. In this paper, we surveyed different proposals and approaches that take users' collective intelligence through their interactions with the contents, their contribution and navigation patterns, and finally suggests best recommendations. This paper also compares the various models used to create a solution for the problems of news recommendation.**

Keywords **— Web mining, Lexical Chains, Bayesian Framework for User Interest Prediction, Topic Analysis Model, Keyword Extraction Algorithm, Newsletters System and collective intelligence.**

I. INTRODUCTION

A critical problem with news service websites is that the volumes of articles which can be overwhelming to the users. The challenge is to help users find news articles that are interesting to read. Information filtering is a technology in response to this challenge of information overload in general. Based on a profile of user interests and preferences, systems recommend items that may be of interest or value to the user. Information filtering plays a central role in recommender systems, as it is able to recommend information that has not been rated before and accommodates the individual differences between users. Information filtering has been applied in various domains such as email news and web search. In the domain of news, this technology aims at aggregating news articles based on the user interests and creating a sort of "personalized news service" for each user.

In current scenario, recommendation system should not only present contextually relevant items or personalized items, but also show items which are hot among other users over the Web [1]. The recommendation has been identified among the top trends to watch in recent years. The two Internet companies that have been the most outstanding in using recommendations are Amazon.com and Netflix. Others, such as Google, have used it, but more as a background enabling technology.

In this paper, the first section discusses different News Recommendation System and its functionalities as well as the technologies involved in these systems. The second section extensively different topic analysis models and technologies involved to develop these models. The third section discusses the advantages and disadvantages of these systems as described by the respective authors. Finally, the paper is concluded with the suggestions and recommendations for building an effective news recommendation system based on the observations made from the extensive reviews and study.

II. NEWS RECOMMENDATION SYSTEM

Many major players, companies, in this field are trying to overcome some of the major problems to build an effective News Recommendation System to satisfy users' needs. The following five problems are the major problems to be addressed to build an effective recommender system.

Lack of Data: Obviously the biggest issue faced by recommender systems is that they need a lot of data to effectively make recommendations. A good recommender system mainly needs item data, and then it must capture and examine user data, so that the algorithm does its job.

Changing Data: Systems are usually biased towards the old and have difficulty showing new. Past behavior of users is not a convincing tool because the trends are always changing.

Changing User Preferences: The issue here is that while today user have a particular intention when browsing one site - tomorrow user might have completely different intention. The system which takes user preferences as a base for recommendation, may incorrectly label users.

Unpredictable Items: This kind of problem generally happens in the case of some eccentric video or music. These types of items are difficult to make recommendations on, because the user reaction to them tends to be diverse and unpredictable.

Too Complex: It takes lots of variables to do even simplest recommendation. In this paper, an extensive discussion is made on the algorithm which is proposed as a novel algorithm for recommendation system. This algorithm uses concepts of collective intelligence, handles most of the problems mentioned above and generic enough to be applied to any type of item. The system takes three main inputs resulted from the collective intelligence of users for recommending new items:

1. Tags contributed by users (or extracted from text)

2. Community opinion about the item

3. Co-occurrence pattern of users' activities around the item

Types of News Recommendation Systems

The news recommendation system is classified into different types, they are

i) Content-based recommendation,
ii) Collaborative recommendation,
iii) Demographic recommender,
iv) Utility-based recommendation,
v) Knowledge-based recommendation.

*Content-Based Rec*ommendation

A content-based recommender learns a profile of the user's interests based on the features present in objects the user has rated. Content-based recommendation systems may be used in a variety of domains ranging from recommending web pages, news articles, restaurants, television programs, and items for sale. Although the details of various systems differ, content-based recommendation systems share in common a means for describing the items that may be recommended, a means for creating a profile of the user that describes the types of items the user likes, and a means of comparing items to the user profile to determine what to recommend.

Collaborative Recommendation

Collaborative recommendation is probably the most familiar, most widely implemented and most mature of the technologies. Collaborative recommender systems aggregate ratings or recommendations of objects, recognize commonalities between users on the basis of their ratings, and generate new recommendations based on inter-user comparisons. A typical user profile in a collaborative system consists of a vector of items and their ratings, continuously augmented as the user interacts with the system over time [10].

Demographic recommender

Demographic recommender systems aim to categorize the user based on personal attributes and make recommendations based on demographic classes. The user's responses were matched against a library of manually assembled user stereotypes. Some more recent recommender systems have also taken this approach. A short survey is used to gather the data for user categorization. In other systems, machine learning is used to arrive at a classifier based on demographic data. The representation of demographic information in a user model can vary greatly. Demographic techniques form people-to-people correlations like collaborative ones, but use different data. The benefit of a demographic approach is that it may not require a history of user ratings of the type needed by collaborative and content-based techniques [10].

Utility-based recommendation

Utility-based recommenders do not attempt to build long-term generalizations about their users, but rather base their advice on an evaluation of the match between a user's need and the set of options available. Utility-based recommenders make suggestions based on a computation of the utility of each object for the user. Of course, the central problem is how to create a utility function for each user [10].

Knowledge-based recommendation

Knowledge-based recommendation attempts to suggest objects based on inferences about a user's needs and preferences. In some sense, all recommendation techniques could be described as doing some kind of inference. Knowledge-based approaches are distinguished in that they have functional knowledge: they have knowledge about how a particular item meets a particular user need, and can therefore reason about the relationship between a need and a possible recommendation [10].

In this section, we discussed different News Recommendation Systems proposed by different authors for effective recommendation of news as per users' need.

2.1 Personalized News Filtering and Summarization (PNFS) System:

Xindong Wu, Fei Xie, Gongqing Wu [1] designed a dynamic and robust ranking function that helps in choosing the best recommendations. As author emphasized, this recommendation system often compares the users' profile to some reference characteristics, and tries to predict more information which they may be interested in. These characteristics may be obtained from the information directory known as the content-based approach or user's social environment known as the collaborative filtering approach. The author pointed out that the News Recommendation is a potential field since the numbers of consulting topics and users as well as the number of used session are much higher than other objects consultancy.

To identify the useful information that satisfies a user's interest, the filtering and summarization of personalized Web news has drawn much attention in Web intelligence. The filtering and summarization of personalized Web news refers to the recommendation, extraction, and summarization of interesting and useful information from Web pages, which can be widely used to promote the automation degree in public opinion investigation, intelligence gathering and monitoring, topic tracking, and employment services. The author proposed a system called personalized news filtering and summarization (PNFS) system that works on news Web pages.

Process in the PNFS:

The first task of this system is to recommend interesting news to users. The process involved in this is to dynamically obtain Web news from the Google news website (http://news.google.com), and then recommend

personalized news to the users according to their interest preferences. A news filter is applied in the system to provide high quality news content for analyzing. The nature of news reading makes news information filtering distinctive from information filtering in other domains. When visiting a news website, the user is looking for new information, information that user did not know before, that may even surprise the user. Since user profiles are inferred from past user activity, it is important to know how users' news interests change over time and how effective it would be to use the past user activities to predict their future behavior.

2.2 Lexical Chains

Morris and Hirst [3] introduced a concept of lexical chains to segment text. Later, lexical chains are used in many tasks, such as text retrieval and information extraction. The construction of lexical chains needs a thesaurus for determining relations between words. They constructed the lexical chains using the thesaurus-based word similarity and the word co-occurrence model. Two thesauruses, including WordNet and HowNet, were respectively used to compute word similarity in English and in Chinese. The word co-occurrence model was adopted to solve the problem that it was observed as difficult to compute the semantic relations between words not in the thesaurus.

The notion of cohesion is a device for "sticking together" different parts (i.e., words, sentences, and paragraphs) of the text to function as a whole. Lexical cohesion occurs not only between two terms, but also among sequences of related words, called lexical chains.

The author opined that the word co-occurrence is an important model based on statistics which is widely used in natural language processing that reflects the relatedness of the words in a document. The frequency of two words co-occurring in the same window unit (i.e., a sentence or a paragraph) can be computed without a thesaurus. HowNet is a commonsense knowledge base that unveils inter-conceptual and inter-attribute relations of concepts as connoting in lexicons of the Chinese language and their English equivalents. There are two important terms in HowNet: concept. A concept is the semantic description of phrases. Each phrase has several concepts. A concept is defined by a kind of knowledge representation language named sememe that is the smallest basic semantic unit.

2.3 Bayesian Framework for User Interest Prediction

Jensen [1], proposed Bayesian framework for user interest prediction. In this work, the log analysis reveals that the click distributions of individual users are influenced by the local news trend. For example, Spanish users read more sports news during Euro Cup. Similar phenomena were also reported in a user study of the lifecycle of news interests. Based on these findings, author classifies user's news interests into two parts: users' genuine interests and the influence of local news trend. The user's genuine interests originate from the personal characteristics of the user, such as gender, age, profession, etc. and are thus relatively stable over time. On the other hand, when deciding what to read, users are also influenced by the news trend in the location that they belong to. This kind of influence produces short-term effects and changes over time. The genuine interests and news trend influence correspond to the "long-term" and "Short-term" interests discussed. However, distinct methods were used to predict the user's news interests. More importantly, the "short-term" interests is modeled from the perspective of news trend using the click patterns of the general public, instead of only using the user's own feedbacks. They developed an approach using Bayesian frameworks to predict users' current news interest based on the click patterns of the individual users and the group of users in the country. The predicted interests are used in news information filtering. The approach works as follows: first, the system predicts user's genuine news interests regardless of the news trend, using the user's clicks in each past time period; second, the predictions made with data in a series of past time periods are combined to gain an accurate prediction of the user's genuine news interests; finally, the system predicts the user's current interests by combining their genuine news interests and the current news trend in her location.

2.4 Using Collective Intelligence

Satnam Alag, et. al. [6] suggested a technique of using collective intelligence for building an effective news recommendation system. When a group of individuals collaborate or compete with each other, intelligence or behaviour that otherwise didn't exist suddenly emerges this is commonly known as collective intelligence. It is an active field of research where researchers from the fields of sociology, mass behaviour, and computer science have contributed a lot. The actions or influence of a few individuals slowly broadens across the community until the actions become the custom of the community. A crowd's collective intelligence will produce better results than those of a small group of experts.

Extracting intelligence from tags:

Tagging is the process of adding words or small phrases, to items. These keywords or tags can be attached to videos, photos, audios, articles, books, products or even to users. These tags can then be used to find similar items. Tags in their raw form are usually large in number and hence there is a need for applying tag selection process to refine the raw tags and select the most significant among them based on weight, name entity relationship and other known techniques.

Extracting intelligence from community opinion:

Users generally express their opinions about the item by performing some actions. For example, if user likes particular video, music or photo then he may rate it high in the scale from 1 to 5. In case of news articles, users may forward it to other friends, write comments, or bookmark it. Similar case happens for book or product, where user may add it to shopping cart, buy it or write a review about it. In this case, gathered information of aggregated users' activities for an item helps to take decision on recommending it.

Extracting intelligence from co-occurrence patterns of users' activities:

Users' way of navigation conveys lots of important things. For example, if most of the users who are viewing item 'A' also view item 'B', then there may be some co-relevance between these two items. In similar fashion, a user uploading photos of his tours indirectly indicate that all those photos are correlated. Another example of such co-occurrence pattern can be seen during e-shopping, where users add the number of items into their shopping cart before purchasing them. If collectively, many users are adding a particular set of items in their shopping cart, there can be a strong association among these items.

2.5 Learning a User Model

D. Billsus and M. Pazzani [8] proposed a technique in which creating a model of the user's preference from the user history is a form of classification learning. The training data of a classification learner is divided into categories, e.g., the binary categories "items the user likes" and "items the user doesn't like." This is accomplished either through explicit feedback in which the user rates items via some interface for collecting feedback or implicitly by observing the user's interactions with items. For example, if a user purchases an item, that is a sign that the user likes the item, while if the user purchases and returns the item that is a sign that the user doesn't like the item. In general, there is a tradeoff since implicit methods can collect a large amount of data with some uncertainty as to whether the user actually likes the item. In contrast, when the user explicitly rates items, there is little or no noise in the training data, but users tend to provide explicit feedback on only a small percentage of the items they interact with. Users can click on the thumbs up or thumbs down buttons to indicate whether they like the program that is recommended.

2.6 Newsletters System

D. Chakrabarti, et. al. [9] used the newsletters for the net-works are available as tabs on the top of the page to develop news recommendation system. The newsletter is presented as a list of articles, each with its title, news synopsis, links to original article and the locally cached page. Users can rate an article on a scale of 1 to5. In the newsletter, the clusters are sorted on the score of the most relevant document in each cluster. Only one result per cluster is presented to the user, but the user can look at the other results by navigating through the "Similar pages" link. Only the top 5 results are presented to the user. The users can also recommend particular news articles to their friends through a recommendation button provided beside each news result.

2.7 Keyword Extraction Algorithm

X. Wu, et. al. [11] proposed a technique known as Keyword extraction, which is applied for document retrieval, web page retrieval, document clustering, summarization, text mining, and so on. By extracting appropriate keywords, we can easily choose which document to read to learn the relationship among documents. Keyword extraction from text data is a common tool used by search engines and indexes alike to quickly categorize and locate specific data based on explicitly or implicitly supplied keywords.

Keywords are commonly used for search engines and document databases to locate information and determine if two pieces of test are related to each other. Reading and summarizing the contents of large entries of text into a small set of topics is difficult and time consuming for a human, so much so that it becomes nearly impossible to accomplish with limited manpower as the size of the information grows. As a result, automated systems are being more commonly used to do this task [12].

1: Non-news content on the news Web page is filtered. Words are segmented and stemmed (for English words), and stop words are removed.

2: Compute the TFIDF of each word using formula (4).

3: Select the top n words by TFIDF as candidate words.

4: Build the disambiguation graph in which each node is a candidate word that is divided into several senses (concepts), and each weighted edge connects two word senses.

5: Perform the word sense disambiguation for each candidate word, and the one sense with the highest sum of similarities with the other word senses is assigned to the word.

6: Build the actual lexical chains. An edge connects two words if the word similarity (using the assigned word sense) exceeds the threshold *t*4 or the word co-occurrence frequency exceeds the threshold *t*5.

7: Compute the weight of each candidate word *wi*as follows:

$$Weight(wi) = a \times TFIDFi + b \times |chaini| + c \times |relatedi| \quad (7)$$

where a, b, and c are parameters that can been adjusted. When a certain feature is used, the corresponding parameter is set to 1; otherwise, it is set to 0. |chaini| is the length of the chain in which *wi* is, and |relatedi| is the number of related words linked with *wi*.

8: Select the top m words as the keywords extracted from the candidate words by their weights.

III. TOPIC ANALYSIS MODEL

Topic models [12] are probabilistic models to reveal the underlying semantic structure of a document collection based on a hierarchical Bayesian analysis of the original texts. By discovering patterns of word use and connecting documents that demonstrate similar patterns, topic models have emerged as a powerful new technique for finding useful structure in an otherwise unstructured collection as well as modeling text corpora. They are based upon the idea that each document is a probability distribution over topics and each topic, in turns, is a mixture distribution over words.

Representing words and documents as probability distribution has some important advantages in compared with a simple space model. It can provide a probability distribution over words that can pick out correlated terms.

4. System Architecture of Effective News Recommendation Model

The architecture of the proposed Effective News Recommendation Model is as shown in Figure 1. Assuming that tags are available to the system, either from user contribution or extracted through automatic tool, tag selector module selects few important tags among the raw tags. A new user is required to register with an initial interesting topic category or keywords. Once a registered user logs in, the system returns personalized Web news to the user. When the user clicks on his/her interesting news items, the recently browsing history is updated. The user can either browse the original news Web page or read the filtered news content with summarized keywords. A keyword model is maintained to store the topic-distinguished keywords and the keywords selected from the browsed news stories. The user can also modify the keyword model to improve the recommendation performance.

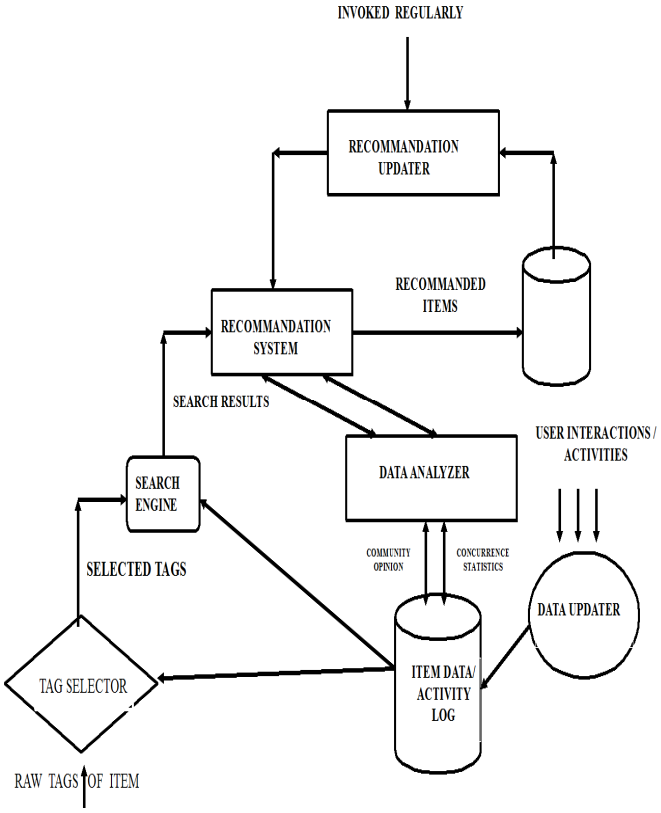

Fig.1. Models or System Architecture

Functionalities

- Tag selector – This module selects few important tags among these raw tags.
- Search Engine– In this module selected tags are then queried on search engine to get relevant set of items.
- Data Analyzer– It also analyses the information regarding community opinions for the items that are outcome of the search and co-occurrence analysis.
- Recommendation Updater– module updates pre-stored recommendations regularly.
- Data Updater– It is essentially external to recommendation system and its task is to log users' activities and update item information, which in turn helps data analyzer and subsequently recommendation system to update recommendations.

IV. RESULTS AND DISCUSSION

The observation made out from this literature study is that the news recommendation is challenging due to the rapid evolution of topics and preferences. The comparison and analysis are carried out on different techniques and methods used for mining the news recommendation models. This section discusses various methods which are used to enrich the news recommendation system.

i. PNFS Model

It often compares the users' profile to some reference characteristics, and tries to predict more information which they may be interested in.

Advantage:

- This model is used to represent the robust and effect new recommendation.
- The major tasks in the PNFS are to filter out the news stories that are uninteresting to the user.
- It also filters out non-news parts on news Web pages.

Drawbacks:

- Proper understanding of the user preference is lacking since it takes amount of data and log entries which occupy large space and cost is high.

ii. Lexical Chains Model

Lexical cohesion occurs not only between two terms, but among sequences of related words, called lexical chains. The construction of lexical chains needs a thesaurus for determining relations between words.

Advantage:

- This model provides an effective way for text retrieval and information extraction.
- This model filtering and summarization of web news refers to the recommendation, extraction, and summarization of interesting and useful information from Web pages.

Drawbacks:

- Lexical chains need a thesaurus for determining relations between words. Using thesaurus this model justifies the words in information retrieval.
- The word co-occurrence model was adopted to solve the problem that it was observed as difficult

to compute the semantic relations between words not in the thesaurus.

iii. Collective Intelligence

When a group of individuals collaborate or compete with each other, intelligence or behaviour that otherwise didn't exist suddenly emerges this is commonly known as collective intelligence.

Advantage:

- It recommends the particular news or item mainly based on tags, community opinion, and co-occurrence activity.
- The main collective intelligence is to tags help to search for relevant contents, community feedback aids to know the popularity of the items and co-occurrence patterns assist to find items having indirect relevance.
- All these components collectively decide the scores for recommended item

Drawbacks:

- The time consumption is main disadvantage for collective intelligence because it mainly based on tags, community opinion, and co-occurrence activity.
- Sometimes it provides the non-reliable data and improper information because many people involved in rating the news content.

iv. Bayesian Framework

In this work, the log analysis reveals that the click distributions of individual users are influenced by the local news trend. Author classifies user's news interests into users' genuine interests and the influence of local news trend based on the proposed framework.

Advantage:

- This framework provides "long-term" and "short-term" interests. That is genuine interests correspond to long-term and news trend influence related to short-term.
- The objective of this framework to analysis the individual news is long-term or short term.

Drawbacks:

- To understand the interest of the user is difficult since the genuine interests are based on personal characteristics, such as gender, age, profession, etc.
- The news trend influence is based on the location that they belong to.

v. Learning a User Model

Creating a model of the user's preference from the user history is a form of classification learning. The training data of a classification learner is divided into categories, like the binary categories "items the user likes" and "items the user doesn't like."

Advantage:

- This model is mainly based on user interest and user preference by likes and dislikes.
- It also provides efficient technique by rating on watching the particular news and articles influenced by user preference.

Drawbacks:

- According to this model to understand the user model based on this way we understand how to just the news content to other or based upon the rating by user .It required more time for understand the user model.

vi. Based on the literature reviews carried out, we come to the conclusion that Lexical Chains Model technique outperforms than other similar techniques in the context of forming better news recommendation system. Also, it is observed from the literature reviews that improvement is needed in existing news recommender systems which represent user preferences for the purpose of suggesting items to purchase or examine. The news recommendation system becomes fundamental applications in electronic commerce and information access, providing suggestions that effectively prune large information spaces.

V. CONNCLUSION AND FUTURE WORKS

All existing recommender systems employ one or more of a handful of basic techniques: content-based, collaborative, and knowledge-based. A survey of these techniques shows that each technique has complementary advantages and disadvantages. The content-based approach recommends items based on the profile which is built by analyzing the content of articles that a user has read in the past. In collaborative filtering approach, the system uses known preferences of a group of users to make recommendation for other users.

As a future work, the extension of this work is proposed to develop a hybrid news recommender system by combining content-based approach recommendation and collaborative filtering approach recommendation in order to increase the efficiency of the recommendation system. The proposed news recommendation system gets data from various news sites and filters unwanted and improper news to increase user's interest.

REFERENCES

[1] Xindong Wu, Fei Xie, Gongqing Wu, Wei Ding, "Personalized News Filtering and Summarization on the Web," In Proceedings of the Twenty Third IEEE International Conference on Tools. 2011.

[2] Florent Garcin, Kai Zhou, Boi Faltings, Vincent Schickel, "Personalized News Recommendation Based on Collaborative Filtering," In Proceedings of IEEE/WIC/ACM International Conferences on Web Intelligence. 2012.

[3] J. Morris and G. Hirst, "Lexical cohesion computed by thesaural relations as an indicator of the structure of text," *Computational Linguistics*, 17(1): 21-48, 1991.

[4] H. Peat and P. Willet, "The limitations of term co-occurrence data for query expansion in document retrieval systems," *Journal of American Society for Information Science*, 42(5): 378–383, 1991.

[5] Jensen, V. Bayesian Networks and Decision Graphs.Springer, 2001

[6] Satnam Alag, "Collective Intelligence in Action", October, 2008, ISBN: 1933988312

[7] Jiahui Liu, Peter Dolan and Elin Rønby Pedersen, "Personalized News Recommendation Based on Click Behavior,".

[8] D. Billsus and M. Pazzani, "Adaptive news access," In: P.Brusilovsky, A. Kobsa, and W. Nejdl (eds.): *The Adaptive Web: Methods and Strategies of Web Personalization*, Springer, 2007.

[9] D. Chakrabarti, R. Kumar, and K. Punera, "Generating succinct titlesfor web URLs," *KDD-2008*, pages 79-87, Las Vegas, Nevada, USA,August 24-27, 2008.

[10]Z. Dong and Q. Dong, "HowNet and the computation of meaning,"*Singapore: World Scientific Publishing Company*, 2006.

[11] X. Wu, G. Wu, F. Xie, Z. Zhu, X. Hu, H. Lu, and H. Li, "News filtering and summarization on the web," *IEEE Intelligent Systems*,25(5): 68-76, 2010.

[12] Mai-Vu Tran, Xuan-Tu Tran, Huy-Long Uong," User Interest Analysis With Hidden Topic In News Recommendation System".

[13] D.J. Hand and K. Yu, "Idiot's Bayes: not so stupid after all?" *Internat.Statist. Rev.* 2001, 69, 385-398.

[14] J. Konstan, B. Miller, D. Maltz, J. Herlocker, L. Gordon, and J. Riedl,"GroupLens: Applying collaborative filtering to Usenet news,"*Communications of the ACM 40*, 3: 77-87, 1997.

[15] S. Li, H. Wang, S. Yu, C. Xin, "Research on maximum entropymodel for keyword indexing," *Chinese Journal of Computers*, 27(9):1192-1197, 2004.

[16] H. Suo, Y. Liu, and S. Cao, "A keyword selection method based onlexical chains," *Journal of Chinese Information Processing*, 20(6):25–30, 2006.

[17] A. Tan and C. Tee, "Learning user profiles for personalizedinformation dissemination," in *Proceedings of the IEEE InternationalJoint conference on Neural Networks*, pages 183-188, May 1998.

[18] P. D. Turney, "Learning to extract keyphrases from text," NationalResearch Council, Canada, *NRC Technical Report ERB-1057*, 1999.[21] I. H. Witten, G.W. Paynter, E. Frank, C. Gutwin, and C.G. Nevill-Manning, "KEA: Practical automatic keyphrase extraction," in*Proceedings of the 4th ACM Conference on Digital Libraries*, pages254-256, Berkeley, California, US, 1999.

[19] H. Zhang, Q. Liu, X. Cheng, H. Zhang, and H. Yu, "Chinese lexicalanalysis using hierarchical hidden markov model," in *Proceedings ofthe Second SigHan Workshop*, pages 63-70, August 2003.

[20] Toby Segaran, "Programming CollectiveIntelligence, Building Smart Web 2.0Applications", 2007

[21] S. Debnath, N. Ganguly, P. Mitra, "Featureweighting in content based recommendationsystem using social network analysis" , InWWW '08: Proceedings of the 17th Int'l Conf.on World Wide Web (2008), pp. 1041-1042.

[22] R. Chung, D. Sundaram, A. Srinivasan,"Integrated personal recommender systems", In ICEC '07: Proceedings of the ninth Int'l Conf. onElectronic commerce (2007), pp. 65-74.

[23] J. Wang, A. P. de Vries, M. JT Reinders,"Unifying user-based and item-based collaborative filtering approaches by similarity fusion", In SIGIR '06: Proc. of the 29th annual int'l ACM SIGIR Conf. on R&D in information retrieval (2006), pp. 501-508.

[24] G. Adomavicius, ErTuzhilin, "Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions", vol. 17, no. 6, pp. 734-749, 2005.

[25] Gentili, G., Micarelli, A., Sciarrone, F.: Infoweb: An AdaptiveInformation Filtering System for the Cultural Heritage Domain.Applied Artificial Intelligence 17(8-9) (2003) 715-744.

[26] Pretschner, A., Gauch, S.: Ontology Based Personalized Search.In: Proceedings of the 11th IEEE International Conference onTools with Artificial Intelligence (ICTAI) November 8-10 (1999)391-398.

[27] Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichletallocation.Journal of Machine Learning Research, 3, 993–1022[10] H Misra, O Cappe, and F Yvon. 2008. Using LDA to detect semantically incoherent documents. In Proc. of CoNLL 2008,pages 41–48, Manchester, England.

[28] Cam-Tu Nguyen, Xuan-HieuPhan, Thu-Trang Nguyen, Quang-Thuy Ha, Susumu Horiguchi, Web Search Clustering and Labelingwith Hidden Topics, The ACM Transactions on Asian LanguageInformation Processing, Vol. 8, pages 1--40, August, 2009.

[29] PhanXuanHieu, Susumu Horiguchi, Nguyen Le Minh. Learningto Classify Short and Sparse Text & Web with Hidden Topics fromLarge-scale Data Collections.In 17th International World WideWeb Conference, 2008.

[30] Chen, L., Sycara, K.: A Personal Agent for Browsing andSearching. In: Proceedings of the 2nd International Conference onAutonomous Agents, Minneapolis/St. Paul, May 9-13, (1998) 132-139.