A Survey on Various Techniques of Personalized News Recommendation System

Bhumika D Karwa[1], Madhushree B[2]
[1, 2] Department of Computer Engineering
[1, 2] L. J. Institute of Engineering and Technology, Ahmedabad, Gujarat, India

Abstract- *The large volume of information available on the web, the broad coverage of the web content, the phenomenal number of web users and their continued rapid growth has presented a major challenge to the web community. Identifying and accessing relevant information suited to individual needs is called personalized information retrieval and access. Recommendation systems are software agents that elicit the interests and preferences of individual users and make recommendations accordingly. 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. 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 personalized recommendation techniques that can be used to generate news recommendations for the users.*

Keywords- News Recommendation, Ontology, Personalization, Recommender Systems, User Profile

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. Personalization can be defined as any set of actions that can tailor the Web experience to a particular user or set of users. 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. 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]. 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].

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 help 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. Recommender systems are required to identify individual preferences and reading interest areas of users to present personalized tailored updates to them.

In this paper, we summarize the advances in this very special application domain of recommender systems which is news recommender systems. The paper is organized as follows: Section II is dedicated for the personalization techniques and its comparison. Section III describes challenges in news recommendation. Section IV gives a literature survey. Section V describes conclusion and ends with references.

II. PERSONALIZATION TECHNIQUES

Personalization can be done using data mining approaches, filtering techniques, soft computing models or ontology models (Figure. 1) as described in this section.

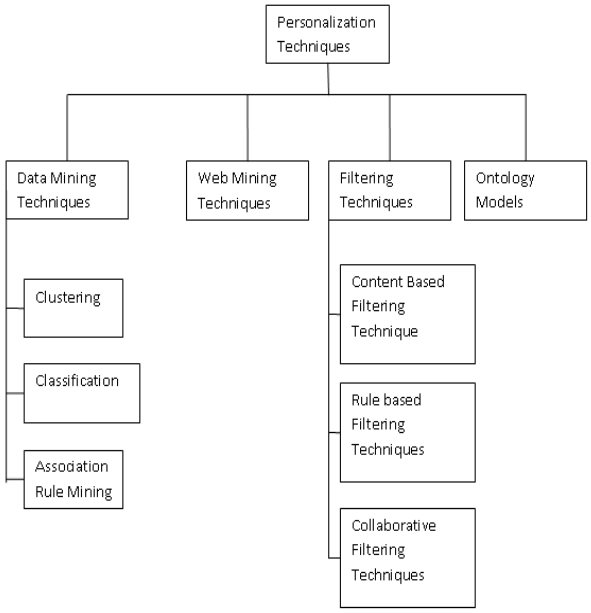

Figure 1. Taxonomy of Personalization techniques

1. Data Mining Techniques:

Data mining is a technique used to draw useful information from a large database to aid discovery of knowledge. Some of the approaches that can be used for personalization include clustering, association rule mining and classification.

Clustering aims to divide a data set into groups or clusters where inter-cluster similarities are minimized while the similarities within each cluster are maximized. It can be used to identify group of users with common interest based on user's preferences. Association rule-based systems have difficulty producing recommendations when the database is sparse as larger item sets are unable to meet minimum support constraints. To alleviate this issue, dimensionality reduction techniques can be applied to the dataset [2]. Classification relies on a set of "training data". Each category based on browsing history, demographics of consumer is given a label and this is used to train the model.

2. Web Mining Techniques

Web mining is the process of extracting interesting patterns from web information repositories [3]. The various techniques used for web mining is shown in figure 6.

Web content mining is classified into two categories namely web page mining and web search results mining. Web page mining is used to discover patterns directly from the contents of web pages. Web structure mining reveals the structure of web sites and how they are connected. Links in web structure mining are classified as internal link and external link .

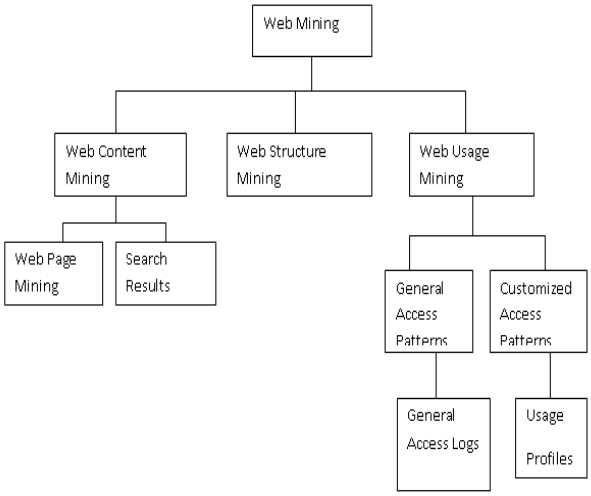

Figure 2. Taxonomy of Web Mining Techniques.

Web usage mining techniques read server log files to extract web site patterns. Hence this technique is used to discover interesting user navigation patterns and can be applied to many real world problems.

3. Ontology Based Personalization Technique

Ontology [4] is a formal description and specification of knowledge as a set of concepts and their relationships within the domain. It provides a common understanding of topics to be communicated between users and systems. Personalized ontology's [4] are a conceptualization model that formally describes and specifies user background knowledge. Web users might have different expectations for the same search query.

Ontology-based user modeling system [4] integrates user ontology that specifies users and their relationships, domain ontology that captures application specific concepts and their relationships and log ontology that specifies the semantics of the user interaction with the system.

4. Filtering Techniques

Filtering techniques are a technique used to remove unwanted data from large volume of data using automated methods prior to recommend the data to the end user.

A. Content Based Filtering Technique

Content-based filtering systems are solely based on individual user's preferences [5]. The system tracks each

user's behaviour and recommends them that are similar to the user likes in the past.

Some of the challenges in Content Based Filtering includes:

- Difficulty in Attribute identification for items difficult.
- Obtaining the user feedback.

B. Rule Based Filtering Technique

A decision tree is constructed to represent various rules. In rule-based filtering the users are asked to answer a set of questions derived from a decision tree. This approach is unable to make recommendations for patterns that do not appear in association rules.

C. Collaborative Filtering Techniques

This technique predicts the opinion of the user and recommends items based on the users opinions and the opinions of the other likeminded users. Collaborative filtering techniques are classified as follows.(Figure.3).

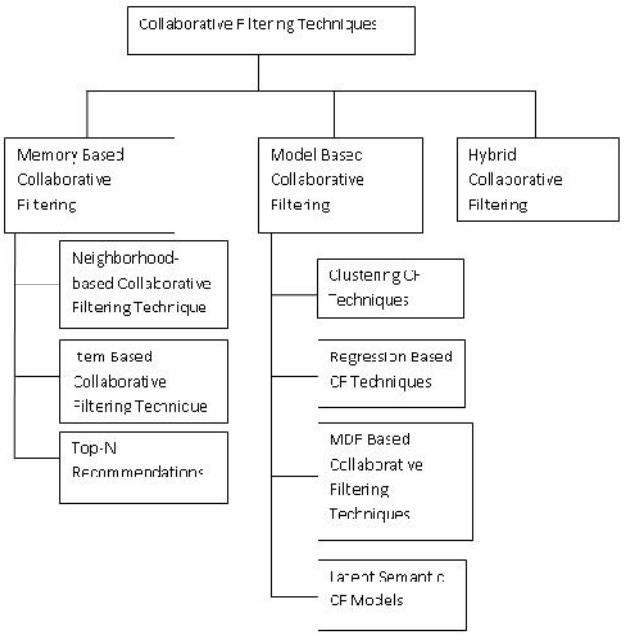

Figure 3 Taxonomy of collaborative filtering techniques

i) Memory Based Collaborative Filtering (Cf) Techniques

It use the user-item database to generate a prediction. By identifying the neighbours of a user, user preferences can be predicted. Memory based collaborative filtering techniques can be classified as follows.

a) The neighbourhood-based CF where, a subset of nearest neighbours of the active user is chosen based on their similarity.

b) Item-based CF approach where, user rating of items is used to find their similarity to the target item.

c) Top-N Recommendations where, a set of N top-ranked items are recommended based on user interest.

ii. Model Based Collaborative Filtering Techniques

Models can be trained to recognize complex patterns and then make intelligent predictions for the collaborative filtering tasks based on their learning [10]. The following approaches can be used.

A. Clustering based CF Algorithms

A cluster is a collection of data objects that are similar to one another within the same cluster and are dissimilar to the objects in other clusters [102].

Clustering methods can be classified into three categories: partitioning methods, density-based methods [10], and hierarchical methods [10]. A commonly-used partitioning method, k-means [10], has two main advantages namely, efficiency and ease of implementation. Density-based clustering methods typically search for dense clusters of objects separated by sparse regions that represent noise [10]. Hierarchical clustering methods, such as BIRCH [8], create a hierarchical decomposition of the set of data objects.

B. Regression-Based CF Algorithms

A regression method uses an approximation of the ratings to make predictions based on a regression model.

C. Markov Decision Process (Mdp) -Based Cf Algorithms.

An MDP is a model for sequential stochastic decision problems, which uses an agent to influence its surrounding environment through actions [6].

D. Latent Semantic CF Models

A Latent semantic CF technique relies on a statistical modeling technique that introduces latent class variables in a mixture model setting to discover user communities and prototypical interest profiles.

iii. Hybrid Collaborative Filtering Technique

Hybrid CF systems combine CF with other recommendation techniques like content-based systems to make predictions or recommendations [7].

TABLE.1. Comparison of different recommendation techniques

S.No	Example Recommender System	Technique used	Advantages	Disadvantages
1.	DBSCAN, OPTICS,BIRCH .	Clustering Algorithms	• Faster Recommendation. • Better performance	• Not accurate
2.	Movie Lens	Association rules	• Provide accurate Prediction about the user interest. • Fast to implement. • Not much storage space required. • Fast to execute.	• Difficulty in producing the recommendations when the data base is sparse. • Not suitable if preferences change rapidly. • Rules can be used only when enough data validates them.
3.	Levis., Netflix .	Memory Based Recommender Systems.	• Easy implementation • New data can be added easily and incrementally • Need not consider the content of the items being recommended. • Scale well with correlated items	• Dependent on human ratings • Performance decrease when data are sparse. • Cannot recommend for new users and items • Limited scalability
4.	GroupLens .	Model based Collaborative Filtering.	• Better at addressing the sparsity and scalability • Improve prediction performance	• Expensive model building. • Lose useful information for dimensionality reduction.
5.	TAN-ELR, NB-ELR .	Classification Technique	• Accurate. • Deals Sparsity problem effectively	• Scalability. • Time Consuming in model training.
6.	Personality Diagnosis.	Hybrid Recommenders	• Improved prediction performance. • Overcomes sparsity problem.	• Increased complexity. • Expensive. • Need external information that usually not available.
7.	Newsweeder, LIBRA	Content-Based Filtering	• Efficiently address Cold start problem and Sparsity Problem	• Requires content that can be encoded as meaningful features. • User interest must be represented as a learnable function. • Overspecialization problem.
8.	ATHENA [8].	Ontology	• No cold start issue. • More accurate	• Knowledge engineering is required. • Expert opinion may not match with user preferences.
9.	PEN recsys.	Web usage mining	• More accurate	• Much storage space is required.

III. CHALLENGES IN NEWS RECOMMENDATIONS

In the news domain, because of the dynamic properties of news items there are some challenges which are to be noticed.

A. Cold-start

The cold start problem occurs when too little rating data is available in the initial state. The recommendation

system then lacks data to produce appropriate recommendations. Two cold start problems are new user problem and new item problem.

B. *Data Sparsity*

In most cases for recommendation systems, the count of ratings already obtained is very small related to the count of ratings that need to be predicted. But collaborative filtering techniques have difficulties when the space of ratings is sparse (few users have rated the similar items). Sparsity in the user-item rating matrix degrades the quality of the recommendations.

C. Recency

Recency is one the most important challenges in news recommendation domain. Most of the users want to read fresh news instead of old dated articles. So the importance of news items decreases in time. On the other hand, some news articles may be connected with each other that the user may want to read the previous news items related to the one she already reads or he/she may want to keep informed about that subject (Liet al.,2011).

D. Implicit User Feedback

User feedbacks are quite important to make more precise recommendations. Without explicit feedbacks it may not be possible to understand if the user liked the article she read or not (Fortunaetal.,2010).But it is not practical for the system to interact with the user continuously. So the system should be able to collect implicit feedbacks effectively while protecting the user privacy.

E. Changing Interests of Users

Another key challenge is predicting the future interests of users for better recommendations because people may have changing interests (Liuetal., 2010). Also some people may read the news not because he/she interested in the topic in general but because she found it important.

F. Scalability

Recommender systems are aimed to serve many users, sometimes millions of users(Das etal.,2007) at a time. So in this dynamic environment to news, the news recommender system should have a fast and real time processing capabilities (Lietal.,2011).

G. Unstructured Content

For the systems which require content information, it is hard to analyze the content, especially for the news domain. For better news recommendations, news items should be structured and machine readable (Saranya.K.G 2012).

IV. LITERATURE SURVEY

Following papers were based on news recommendation system based on ontology.

Cantador et al(2008)[15] proposed news recommendation system that makes use of ontology's to provide online news recommendation services. Domain ontology's are used to provide the concept framework for news contents and user preferences.

Kim Schouten et al. (2010)[13] propose a Hermes framework which is an ontology-based framework for building news personalization services able to read news items and extract the valuable economic information from them and is able to update its knowledge base to always represent the current state of the world.

Wouter IJntema et al. (2010)[8]developed a recommendation system, Athena, to provide ontology-based recommendation for the news feed system. It extends the Hermes framework , a framework used to build a news personalization service, with the help of ontology to determine the semantic relations between terms and concepts. It uses ontology to store concepts and their relationships to the news items.

Frank Goossen et al. (2011)[11] proposes CF-IDF method for news recommendation which is based on TF-IDF, but instead of using all the terms of a text, this method only looks at the key concepts found in this text.

Kang and Choi et al(2011)[14] proposed an ontology based Recommendation system in which the ontology is used to encode the long term and short term preference information. The user preference ontology is constructed from the concepts of the general domain ontology together with the documents that the user visited. Recommendation is made based on the similarity between ontological concepts and terms.

Jiahui Liu et al(2010)[9] describes a research on developing a personalized news recommendation system based on profiles learned from user activity in Google News. Based on the analysis of Google News users click logs, Bayesian framework is being developed for predicting users' current news interests from the activities of that particular user and the news trends demonstrated in the activity of all users. By

combining the content-based recommendation mechanism which uses learned user profiles with an existing collaborative filtering mechanism personalized news recommendations is being generated.

Michel Capelle et al(2012)[12]proposed Ceryx framework, an extension to the Hermes News Portal news personalization service. Ceryx makes use of the SF-IDF and SS methods, a user profile, and a set of news items in order to suggest interesting unread news articles to the user.

Lingling Zhang et al(2015)[10] presented OF-IDF method to represent the unstructured text data in the form of key concepts, synonyms and synsets which are all stored in the domain ontology. For users, the recommendation algorithm builds the profiles based on their behaviours to detect the genuine interests and predict current interests automatically and in real time by applying the thinking of relevance feedback.

V. CONCLUSION

Many recommendation techniques have been used to build personalized news recommender systems. It has been observed recommender uses domain ontology's to enhance the personalization: on the one hand, user's interests are modeled in a more effective and accurate way by applying a domain-based inference method; on the other hand, the matching algorithm used by our content-based filtering approach, which provides a measure of the affinity between an item and a user, is enhanced by applying a semantic similarity method. Recommendation systems can take advantage of semantic reasoning-capabilities to overcome common limitations of current systems and improve the recommendations quality. Moreover, in a good recommendation system, heterogeneous information from multiple sources is usually required. Ontology can integrate the use of heterogeneous information and guide the recommendation preference.

REFERENCES

[1] Adomavicius, G., Tuzhilin, A. Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions. IEEE Transactions on Knowledge and Data Engineering 2005,17(6):734–749.

[2] Chen, M., Han, J., and Yu, P. (1996) Data Mining: An Overview from a Database Perspective. IEEE Transactions on Knowledge and Data Engineering. 8, 866-883.

[3] Eirinaki, M. and Vazirgiannis, M. (2003), "Web mining for web personalization", ACM Transactions on Internet Technology 3(1), 1–27

[4] S. Staab and S. R., editors. "Handbook on Ontologies"

[5] Van Meteren, R., & Van Someren, M. (2000). Using content-based filtering for recommendation. Proceedings of the Machine Learning in the New Information Age: MLnet/ECML2000.

[6] G. Shani, D. Heckerman, and R. I. Brafman, "AnMDP-based recommender system," Journal ofMachine Learning Research, vol. 6, pp. 1265–1295, 2005.

[7] P. Melville, R. J. Mooney, and R. Nagarajan, "Content boosted collaborative filtering for improved recommendations," in Proceedings of the 18th National Conference on Artificial Intelligence (AAAI 02), pp. 187–192, Edmonton, Canada, 2002

[8] IJntema W, Goossen F, Frasincar F, et al. Ontology-based news recommendation. Proceedings of the 2010 EDBT/ICDT Workshops. ACM, 2010: 16.

[9] Liu J, Dolan P, Pedersen E R. Personalized news recommendation based on click behaviour. Proceedings of the 15th international conference on Intelligent user interfaces. ACM, 2010: 31-40.

[10] Rui Ren, Lingling Zhang, Limeng Cui, Bo Deng, Yong Shi, "Personalized Financial News Recommendation Algorithm Based on Ontology" in Procedia Computer Science 55 (2015) 843 – 85

[11] Goossen F, IJntema W, Frasincar F, et al. News personalization using the CF-IDF semantic recommender. Proceedings of the International Conference on Web Intelligence, Mining and Semantics. ACM, 2011: 10.

[12] Capelle, M., Moerland, M., Frasincar, F., Hogenboom, F.Semantics-Based News Recommendation. In: Akerkar, R., B_adic_a, C., Dan Burdescu, D. (eds.) 2nd International Conference on We Intelligence, Mining and Semantics (WIMS 2012). ACM (2012)

[13] F. Frasincar, J. Borsje, and L. Levering. A Semantic Web-Based Approach for Building Personalized News Services. International Journal of E-Business Research, 5(3):35{53, 2009.

[14] Fikadu Gemechu, Zhang Yu, Liu Ting, "A Framework for Personalized Information Retrieval Model", Proceedings of IEEE transaction, 2010

[15] K. G. Saranya and G. S. Sadhasivam, "A personalized online news recommendation system", International Journal of Computer Applications,pp 6–14, November 2012