A Personalized Hybrid Web Recommender System

A. Janet Rajeswari[1] and Shanmugasundaram Hariharan[2]
[1]Research and Development Centre, Bharathiar University, Coimbatore- 641046, India
[2]Department of CSE, TRP Engineering College, Trichy-621105, India
janetcharles1@hotmail.com, mailtos.hariharan@gmail.com

Abstract: Personalized recommender system has attracted wide range of attention among researchers in recent years. There has been a huge demand for development of web search apps for gaining knowledge pertaining to user's choice. A strong knowledge base, type of approach for search and several other factors make it accountable for a good personalized web search engine. This paper presents the state of art, challenges and other issues in this context, thereby providing the need for an improved personalized system. The paper describes an approach integrating the news feeds and users opinion on web news using content, collaborative and hybrid approaches. Experiments carried out shows the effectiveness of the proposed system using popular dataset such as MovieLens.

Keywords: Personalization, web search, recommender system, user

1. Introduction

Information extraction, knowledge sharing via social media has gained commercial interest among wide range of people, especially the community of young generation. Searching information repository and making decisions out of it has become increasingly difficult as the amount of information and the number of choices increases. In today's modern era, these social platforms are of greater importance depending on user's interest. Hence personalizing the web search process has become a stronger area of research.

With the explosion of Web 2.0, social applications such as blogs, professional networks and various other types of social media have bloomed to a greater extent. These online information sources have of knowledge flood users have posed a great challenge in terms of information overload. Millions of users spend huge amount of time on these sites and hence forth a recommender system can help users deal with information overload problem. Personalized news recommendation has become a promising area of research with a variety of news recommender systems providing personalization services.

There has been a plenty of online news reading services to help readers in finding out interesting stories that maximally match their reading appetites, which is called as personalized news recommendation. In general, existing news recommender systems can be categorized into three groups [17] as content-based methods (i.e., model users' reading preferences by the content of users' historical consumed news stories) [2, 31], collaborative filtering (i.e., exploiting user feedbacks without content analysis) [9, 28] or hybrid version of several techniques (i.e., combining both content and user feedback to generate recommendations [18, 1].

Recommendation systems analyze purchasing behavior (e.g., item ratings) of users to learn about their preferences and recommend products or services that may be of interest to them. However, as new users require time to become familiar with recommendation systems, the systems usually have limited information about newcomers and have difficulty providing appropriate recommendations. This so-called new user cold start phenomenon has a serious impact on the performance of recommendation systems [7].

Web-based personalized recommender system [8], agent-based approach [24] , rating-based collaborative filtering [15], sequential pattern analysis [6], social network combined with semantic concept analysis [33], multimodal interactions [5] , collaborative filtering [12], ontology-based [23], product attribute-specific weights [25], association rule mining [22], Partial Credit Model and Bayesian Rough Set [1] were predominant in recommender systems.

At the end, profile merging and personalization algorithms are covered to handle the challenges of aggregate multiple profiles and personalized recommendations based on distributed (portable) profiles [30]. Ratings play a vital role [27], Web Mining [36], ant colony optimization [32], Support Vector Machine [3], construction of personalized Web

page recommendation system [21], mobile tourism [34]. Recommender systems have gained commercial importance in recent years and have significant importance among research community. Popular news portals, such as Google News[1] and Yahoo News[2] have gained increasing attentions for online news reading.

The rest of the paper is structured as follows. Section 1 presents the basics on recommender systems, while related works carried out on personalized recommender systems were discussed in Section 2. Section 3 highlights the proposed framework. Experimental results and investigations were discussed in section 4. Section 5 presents conclusion and future work.

2. Literature review on recommender systems:

Personalized news recommendation has become a promising research direction [19] as the Internet provides fast access to real-time information. A variety of news recommender systems based on different strategies have been proposed in the past, utilizing the implicit ''social'' factors is a new adaptation. Feasibility of integrating content-based methods, collaborative filtering and information diffusion models extracted from implicit feedbacks has been investigated and it is successful. Handling the so-called cold-start problem is also a tricky and challenging task.

Crowd sourcing provides informational products or services with diversified group of people. Individuals contribute to such systems by selecting among a wide range of open tasks [13]. In recent years, explosive growth of information makes the users confused in making appropriate decisions. As a result, it is a challenging issue to help the user identify what she/he prefers. Recommender systems are a typical solution to discover such implicit interests in user's mind based on the usage logs. To alleviate problems namely cold-start, first-rater, sparsity and scalability, novel recommender system namely FRSA (Fusion of Rough-Set and Average-category-rating) is proposed. This integrates multiple contents and collaborative information to predict user's preferences [29].

User interaction is an important feature in many Web-based news services. This identifies the most relevant news and enables collaborated information sharing among users with relevant feedback [20]. Recommender systems have been developed in variety of domains, including asynchronous discussion group. It is quiet hard to retrieve exact information from multiple sources because of information overload. Therefore, recommender systems play an important role in filtering and customizing the desired information. Collaborative and content-based filtering is of greater importance in today's growing world which adopts techniques being utilized in recommender systems. To overcome the drawbacks of the aforementioned techniques, a hybrid recommender system combines two or more recommendation techniques to obtain more accuracy [16].

1 http://news.google.com.

2 news.yahoo.com.

Recommender systems incorporating association rule mining technique has tremendous influence in discovering similar users, related posts etc. Semantic related concepts using Word Sense Disambiguation based on WordNet lexical database were exploited which shown noticeable improvement on the accuracy of useful posts has recommended to the users as compared to content based and the collaborative filtering techniques [7].

WWW creates opportunities for businesses to provide personalized online services to their customers all around the globe. Recommender systems are designed to automatically generate personalized suggestions of products/services to customers. Because various uncertainties exist within product and customer, it is a challenge to achieve high recommendation accuracy [35]. A hybrid recommendation approach combining user-based and item-based collaborative filtering techniques with fuzzy set techniques were studied by the authors. The work has involved intelligent recommender system software called Fuzzy-based Telecom Product Recommender System (FTCP-RS) to select the most suitable products or services.

Real-time web (RTW) services such as Twitter allow users to express their opinions and interests, often expressed in the form of short text messages [10]. The recommender systems research contains useful consumer reviews on products, services and brands. Evaluation is performed on different product datasets from the Blippr service which shows the potential of type of recommendation knowledge, accuracy etc. Capturing and understanding user interests have been an important part of social media [11]. Users in such context belong to multiple interest communities whose interest constantly changes over time. Therefore, modeling and predicting dynamic user interests poses great challenges to personalized recommendation in social media. The authors have investigated this research problem by developing a temporal overlapping community detection method based on time-weighted association rule mining.

3. Recommender System Framework:

Recommender systems are automatically making helpful recommendations about various products and services to customers. Such systems can make recommendations according to user profiles or preferences, or they can rely on the choices of other

people who could be useful referees. Table 1 presents the types of users in recommender systems.

Recommender systems are broad classified based on the approach provided for recommendations based on the information filtering mechanism as content-based methods, collaborative methods and hybrid methods. Content-based (CB) recommenders provide recommendations to user, which are very similar to the ones preferred in the past. Collaborative Filtering (CF) recommenders recommend are those which the mentors had liked in the past. The basic assumption of recommender systems is that the mentors are people with tastes and preferences are similar to each other. Hybrid recommender systems fuses both the content and collaborative schemes while simultaneously avoiding the limitations of each of them. This section highlights the recommender system framework followed by experimental investigations and analysis.

Table 1: Types of users and their characteristics

Type	Name of users	Characteristics
1	Regular users	starts new thread or add comments regularly
2	Casual users	contribute to discussions infrequently
3	Regular favourite marker users	express their opinions about other users' posts regularly
4	Casual favorite marker users	express their opinions about other users' posts infrequently
5	Passive users	neither contribute to discussions nor express their opinions

3.1 Dataset:

The experiments illustrated were presented based on MovieLens dataset[#]. MovieLens data sets were collected by the GroupLens Research Project [14]. Table 2 presents the different datasets, domain of study, rating scale, number of ratings available and total no. of items. This data set consists of 100,000 ratings provided with a score of 1-5. All together there were 943 users on 1682 movies. Each user has rated at least 20 movies. The rater's been professionally sound and has good domain knowledge. It is assumed that the users who rate are trusted ones.

Table 2: Dataset and characteristics of GroupLens

Dataset	Domain	Rating Scale (Min/Max)	Total ratings	Total Items
MovieLens	Movie	1/5	1,000,000	3500
EachMovie	Movie	1/5	2,811,983	1628
BookCrossing	Books	0/10	1,149,780	271,379
Jester	Jokes	-10/+10	73,421	100

From the dataset 80% of the data were chosen for training set and rest of the 20% for test data. The time stamps were recorded in unix seconds since 1/1/1970 UTC. Table 2 presents the sample user ID, item ID, rating scale and time stamp. Several other datasets related to recommender systems are available and available in public forums[$].

Table 2: Sample User Info of an Item

User Id	Item ID	Rating	Time stamp
5	1	4	875635748
5	2	3	875636053
5	17	4	875636198
5	98	3	875720691
5	110	1	875636493
5	225	2	875635723
5	363	3	875635225
5	424	1	875635807
5	439	1	878844423
5	454	1	875721432

[#] http://www.grouplens.org/

[$] http://recsyswiki.com/wiki/Datasets

3.2 Evaluation metrics:

Each article is represented as a term vector by using the tf-idf approach [26] to calculate the weight of term 'i' in an article 'j', as defined in Eq. (1 & 2):

$$W_{i,j} = f_{i,j} * \log \frac{N}{n_i} \quad (1)$$

$$f_{i,j} = \frac{freq_{i,j}}{\max_l (freq_{l,j})} \quad (2)$$

where N is the number of articles; n_i is the number of articles that contain term i; $f_{i,j}$ is the normalized frequency of term i in article j; $freq_{i,j}$ is the frequency of term i in article j; and maxl(fl_j) is the frequency of term l which has the maximum frequency in article j. There are various metrics which are applied in the collaborative filtering techniques to find the users' similarities.

Precision and recall have been widely used in the field of information retrieval, to evaluate recommendation accuracy. These metrics have been adapted to evaluate the accuracy of a set of recommended products and are defined as follows:

$$\text{Precision} = \frac{|T \cap R|}{|R|} \quad (3)$$

$$\text{Recall} = \frac{|T \cap R|}{|T|} \quad (4)$$

where 'T' is the test set and 'R' is the recommended set of items for each user, respectively. An F1-metric can be used to balance the trade-off between precision and recall.

4. Experiments and illustrations:

The architecture of the proposed recommender system is presented in Fig 1.

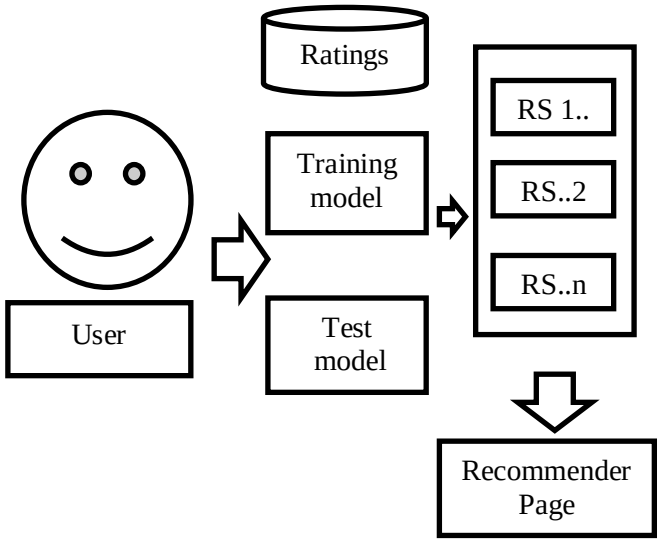

Fig 1: Architecture of Recommender System

$$F_1 = \frac{2xPxR}{P+R} \qquad (5)$$

We use the Pearson correlation coefficient as a measure of correlation between rating and predictions. The recommendation page is obtained by sum of all individual scores provided by the users. Table 3 presents the classification accuracy of different recommender systems using different techniques.

Table 3: Classification accuracy of recommender systems

Metric	TF	TF-IDF	TF+IDF+Stemming
Precision	0.298	0.511	0.713
Recall	0.191	0.245	0.396
F1	0.332	0.423	0.511

5. Conclusion and Future Work:

Personalization has emerged as a major shift for e-commerce based scenario. Recommendation systems provide customized information services to online users there by improving online commodity. We showed that the combination of term frequency combined with inverse document frequency outperforms well (after pre-processing). The state-of-the-art tag recommendation algorithms with respect to a broad variety of performance metrics. Our future work includes analysis of new metrics and techniques considering aspects of novelty and diversity. Also we focus on investigate the biasness in manual evaluation of the scores recommended by volunteers.

Acknowledgement:

The authors would like to express our sincere thanks to anonymous reviewers for their insightful suggestions. The authors indeed would extend their acknowledgement for the support by all well wishers, faculty members and others who helped out in achieving this research task to get completed.

References

1. Abbas .A and Liu. J, "Designing an Intelligent Recommender System using Partial Credit Model and Bayesian Rough Set", International Arab Journal of Information Technology, Vol.9, No.2, pp.179-187, 2012.
2. Ahn. J, Brusilovsky. P, Grady. J, He. D and Syn. S.Y, "Open user profiles for adaptive news systems: help or harm?", In: Proceedings of 16th International World Wide Web Conference, pp. 11–20, 2007.
3. Bo.Y, "Personalized Web Information Recommendation Algorithm Based on Support Vector Machine", Proceedings of International Conference on Intelligent Pervasive Computing, pp. 487 – 490, 2007.
4. Burke. R., "Hybrid systems for personalized recommendations", In: Bamshad Mobasher, Sarabjot Singh Anand (Eds.), Intelligent Techniques for Web Personalization Lecture Notes in Computer Science, vol. 31, Springer, Berlin/Heidelberg, pp. 133–152, 2005.
5. Costa. A. F. D, Domingues. M.A., Rezende. S.O., and Manzato. M.G., "Improving Personalized Ranking in Recommender Systems with Multimodal Interactions", In: Proceedings of the 2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT), Vol.1, pages 198-204,2014.
6. Choi.K, Yoo.D, Kim.G and Suh.Y, "A hybrid online-product recommendation system: Combining implicit rating-based collaborative filtering and sequential pattern analysis", Electronic Commerce Research and Applications, Vol. 11, pp. 309–317, 2012.
7. Chen.C.C, Wan. Y.H, Chung. M.C and Sun.Y.C, "An effective recommendation method for cold start new users using trust and distrust networks", Information Sciences, Vol. 224, pp.19–36, 2013.
8. Chen. D.N, Hu. P.J.H , Kuo. Y.R and Liang. T.P, "A Web-based personalized recommendation system for mobile phone selection: Design, implementation, and evaluation", Expert Systems with Applications, Vol. 37, pp. 8201–8210, 2010.
9. Debnath. S, Ganguly. N and Mitra. P, "Feature weighting in content based recommendation system

using social network analysis", In: Proceedings of the 17th International Conference on World Wide Web, pp. 1041–1042, 2008.

10. Esparza. S.G , Michael P., Mahony.O and Smyth.B, "Mining the real-time web: A novel approach to product recommendation", Knowledge-Based Systems, Vol. 29, pp. 3–11, 2012.

11. Feng. H, Tian.J, Wang. H.J., and Li.M, "Personalized recommendation based on time-weighted overlapping community detection", Information and Management, Vol.52, Issue. 7, pp.789-800, 2015.

12. Garcin. F, Zhou. K, Faltings.B., and Schickel.V., "Personalized News Recommendation Based on Collaborative Filtering", In Proceedings of IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology, pp.437-441, 2012.

13. Geiger .D and Schader. M, "Personalized task recommendation in crowdsourcing information systems Current state of the art", Decision Support Systems,Vol. 65, pp. 3–16, 2014.

14. Herlocker, J., Konstan, J., Borchers, A., and Riedl, J., "An Algorithmic Framework for Performing Collaborative Filtering", In: Proceedings of the 1999 Conference on Research and Development in Information Retrieval, 1999.

15. Jiang.Y., "An Effective Web Service Recommendation Method Based on Personalized Collaborative Filtering", Proceedings of IEEE International Conference on Web Services (ICWS), pp. 211 – 218, 2011.

16. Kardan. A.A and Ebrahimi.M, "A novel approach to hybrid recommendation systems based on association rules mining for content recommendation in asynchronous discussion groups", Information Sciences, Vol. 219,pp. 93–110, 2013.

17. Li. L , Wang. D, Li. T, Knox . D and Padmanabhan B, "Scene: a scalable two-stage personalized news recommendation system", In: Proceedings of the 34th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 125–134, 2011.

18. Li. L, Chu. W, Langford. J and Schapire R.E, "A contextual-bandit approach to personalized news article recommendation", In: Proceedings of the 19th International Conference on World Wide Web, pp. 661–670, 2010.

19. Lin. C, Xie. R, Guan. X, Li.L and Li.T, "Personalized news recommendation via implicit social experts", Information Sciences, Vol. 254, pp. 1–18, 2014.

20. Li.Q, Wang.J, Chen. Y.P., and Lin.Z, "User comments for news recommendation in forum-based social media", Information Sciences, Vol. 180, pp. 4929–4939, 2010.

21. Lian.R, "The construction of personalized Web page recommendation system in e-commerce", Proceedings of International Conference on Computer Science and Service System (CSSS), pp. 2687 – 2690, 2011.

22. Mican.D and Tomai.N., "Association Rules Based Recommender System for Personalization in Adaptive Web-Based Applications", Current Trends in Web Engineering, Lecture Notes in Computer Science Vol. 6385, pp. 85-90, 2010.

23. Moreno.A, Valls.A., Isern. D, Marin. L and Borr.J, "SigTur/E-Destination: Ontology-based personalized recommendation of Tourism and Leisure Activities", Engineering Applications of Artificial Intelligence, Vol. 26, pp. 633–651, 2013.

24. Neves. A.R.M, Carvalho. A. M. G., and Ralha. C. G., "Agent-based architecture for context-aware and personalized event recommendation", Expert Systems with Applications, Vol. 41, pp. 563–573, 2014.

25. Oh. J, Jeong. O.R and Lee. E, "A Personalized Recommendation System Based on Product Attribute-specific Weights and Improved User Behavior analysis", In: Proceedings of the 4th International Conference on Uniquitous Information Management and Communication, Article No. 57, 2010.

26. Paik. J and Mizugai .H, "Recommendation System using Weighted TF-IDF and Naïve Bayes Classifiers on RSS Contents", Journal of Advanced Computational Intelligence and Intelligent Informatics, Vol.14, No.6, pp.631-637, 2010.

27. Qumsiyeh.R., and Ng. Y.K, "Predicting the ratings of multimedia items for making personalized Recommendations", Proceedings of the 35th International ACM SIGIR conference on Research and development in information retrieval, pp. 475-484, 2012.

28. Seroussi. Y, Bohnert F., and Zukerman I., "Personalised rating prediction for new users using latent factor models", In: Proceedings of the 22nd ACM Conference on Hypertext and hypermedia, pp. 47–56, 2011.

29. Su. J.H, Wang. B.W, Hsiao. C.Y and Tseng. V. S, "Personalized rough-set-based recommendation by integrating multiple contents and collaborative information", Information Sciences, Vol. 180,pp.113–131, 2010.

30. Uhlmann. S and Lugmayr. A, "Personalization algorithms for portable personality", In: Proceedings of the 12th international conference on Entertainment and media in the ubiquitous era, pp.117-121, 2008.

31. Wang. J, Li. Q, Chen. Y.P, Liu. J, Zhang. C and Lin. Z, "News recommendation in forum-based social media", In: Proceedings of the 24th Conference on Artificial Intelligence, pp. 1449–1454, 2010.

32. Wang. F.H, "Personalized recommendation for web-based learning based on ant colony optimization with segmented-goal and meta-control strategies", Proceedings of IEEE International Conference on Fuzzy Systems (FUZZ), pp. 2054 - 2059, 2011.
33. Xu.Y, Guo.X, Hao.J ,Ma.J, Lau. R.Y.K and Xu.W, "Combining social network and semantic concept analysis for personalized academic researcher recommendation", Decision Support Systems, Vol. 54, pp. 564–573, 2012.
34. Yu. C.C., and Chang.H.P, "Personalized Location-Based Recommendation Services for Tour Planning in Mobile Tourism Applications", E-Commerce and Web Technologies, Lecture Notes in Computer Science Volume 5692, pp 38-49 , 2009.
35. Zhang.Z, Lin.H, Liu.K, Wu.D, Zhang.G and Lu.J, "A hybrid fuzzy-based personalized recommender system for telecom products/services", Information Sciences , Vol. 235, pp.117–129, 2013.
36. Zhang.L., "Personalized Instructing Recommendation System Based on Web Mining", Proceedings of 9th International Conference for Young Computer Scientists, pp. 2517 – 2521, 2008.

Author biography:

Janet Rajeswari born on 26th June, 1976 at Tirchirappalli, Tamil Nadu, India. She completed her Undergraduate degree in Computer Science, from Bharathidasan University in 1997 and Postgraduate degree from Alagappa University, Karaikudi in 2003. She has also done M.Phil., in Computer Science in 2006 from Bharathidasan University. She is currently pursuing Ph. D., in Computer Science in Bharatiar University, Coimbatore. She is currently working as an Assistant Professor in the Department of Computer Science in Christhuraj College, Trichirappalli from 15th June, 2012. Her areas of research interest are data mining, cloud computing and cryptography. She has published good number of research papers in reputed national/ international journals and conferences.

Dr.S.Hariharan received his B.E degree specialized in Computer Science and Engineering from Madurai Kammaraj University, Madurai, India in 2002, M.E degree specialized in the field of Computer Science and Engineering from Anna University, Chennai, India in 2004. He holds his Ph.D degree in the area of Information Retrieval from Anna University, Chennai, India. He is a member of IAENG, IACSIT, ISTE, CSTA and has 12 years of experience in teaching. Currently he is working as Associate Professor in Department of Computer Science and Engineering, TRP Engineering College, India. His research interests include Information Retrieval, Data mining, Opinion Mining, Web mining. He has to his credit several papers in referred journals and conferences. He also serves as editorial board member and as program committee member for several international journals and conferences.