Preference Based Personalized News Recommender System

Mansi Sood[1], Harmeet Kaur[2]

Abstract

News reading has changed from the traditional model of hardcopy newspapers to online news access. Thousands of news sources are available on internet each having millions of articles to choose from, leaving users tangled to find out a relevant article that matches their interests and liking. Recommender Systems can be used as a solution to this information overload problem by identifying the interest areas of a user by creating user profiles, maintaining those profiles to keep accommodating changing user interests and presenting a set of recent news articles formed as recommendations based on those user profiles. This paper presents an algorithm, which requests one time input from users (during the signup) about their preference of news categories (like Sports, Entertainment etc.), which they would like to subscribe and creates a personalized profile for each user. Subsequently, it requests an optional feedback on the recommended articles, to intelligently update user profiles, and recommend relevant articles to them, based on their changing interests. The paper also presents a simulation of the proposed algorithm on various use cases to depict the correctness and robustness of the algorithm. Also, it gives a brief idea about implementation details and challenges associated with the algorithm.

Keywords

News Recommender System, User Profile, Preference, News Categories

1. Introduction

News access has evolved with the advancements in the technology and the way technology is used to read news online. Now to get latest updates on almost anything around us, we need not gaze at a big hardcopy or even switch on any television channel

or tune into some radio station, we just need to have internet connectivity on a smart device like PDA's (Personal Digital Assistant), mobile phones etc. It has evolved from the traditional model of news consumption via newspaper subscription to access to thousands of sources, websites, via the internet [1]. But, this advancement has brought forward a serious issue with online news sources presenting a huge number of news articles to the users. The challenge is to help users find news articles that are interesting to read.

Information filtering and Recommender Systems based on it have emerged in response to the above challenge, providing users with recommendations of content suited to their needs [2]. Information filtering has been applied in various domains, such as email [3], news [4], and web search [5]. Since different information needs and queries arise in varying contexts with different intentions, research has started to focus on delivering tailored, adapted and personalized information to users [6]. Based on a profile of user interests and preferences, systems recommend items that may be of interest or value to the user. An accurate profile of users' current interests is critical for the success of such systems [7]. Some systems require users to manually create and update profiles, which places an extra burden on users, something very few are willing to take on [8,9]. Instead, systems can construct profiles automatically from users' interaction with the system. Approach discussed in this paper will adopt both the above mentioned ways to build up user profiles, so as to reduce the burden on user in the first method and take advantage of accurate information received from user about their preferences as done in the second method [10].

The nature of news reading and its access patterns 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 he did not know before, that may even surprise him. 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 [11].

Dynamic user profiling that can accommodate the changes in the user's preferences and requirements into user profile shall be used in this case so that user profiles can adapt with the changing preferences [2].

As already stated, Recommender Systems have emerged as a solution to the above challenges. Recommender Systems form a specific type of information filtering technique that attempts to present information items (movies, music, books, news, images, web pages) that are likely of interest to the user. Recommender Systems use a number of different techniques [12, 13, and 14]. We can classify these systems into three broad groups: content based systems, collaborative filtering systems and hybrid systems.

Content-based systems examine properties of the items already consumed to recommend items that are similar to the ones that user liked in the past. For instance, if a user has watched many suspense thriller movies, then recommend a movie classified in the database as having the thriller or suspense genre. Collaborative filtering systems recommend items based on similarity measures between users and/or items. The items recommended to a user are those preferred by similar users. This sort of Recommender System needs guidelines to establish similarity between users and/or items [15].

However, these techniques suffer with various issues like cold start problem in collaborative filtering. Cold start problem occurs when it is not possible to make reliable recommendations due to an initial lack of ratings of a particular community or group of similar users. Alternative can be to adopt a hybrid approach between content-based matching and collaborative filtering. A hybrid system combining these two techniques tries to use the advantages of one to fix the disadvantages of other. For instance, cold start problem in collaborative filtering method can be overcome in combination with content based systems as content-based approach predicts new items based on their description (features) that are typically easily available. Given these two basic techniques, several ways have been developed in past for combining them to create a new hybrid system [16-19].

2. News Categorization

Almost all top news sources, websites, smart phone applications classify new articles/headlines into predefined categories like Entertainment, Sports, Health, Politics etc. This is helpful for users looking for specific category news as they can directly access the relevant article as per their interests.

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) [20]. The system is designed to be used in any situation where news material needs to be categorized. It provides for coding of subject of the content of a news item. SRS consists of more than 1000 categories divided hierarchically into 3 levels: Subject, Subject matter and Subject detail. There are 17 top-level Subjects having secondary Subject Matter lists for each of these. 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.

3. The Proposed Algorithm

Algorithm proposed in this paper exploits the predefined categorization done by news sources and websites to first identify user interests and preferences and later to generate recommendations for them. Steps given below explain the algorithm in detail.

Step I - Signup
During the one time signup process, user would be asked to subscribe to one or more of the **K** news categories (for example Sports, Entertainment, Politics etc.) defined by the news source/website. This would require user to provide Preference Score p_i (1<=i<=K) for each category on a scale of 0 to 10, where value 0 indicates lowest preference and 10 indicates highest preference for a given category.
User would also indicate the approximate number of articles to be recommended (**N**) each time he signs in to the site. N should be greater than or equal to the number of categories subscribed by the user. The actual number of articles recommended by the proposed algorithm will range between N and N+K.
Using these inputs, a personalized profile, capturing user preferences is created, which will be used to recommend news articles to the user.
Sample Illustration
Available Categories (K = 3): Business, Technology, Sports

User Inputs:

- Preference Scores (p_i) for each category, $p_1 = 3$, $p_2 = 8$, $p_3 = 6$
- Approximate number of articles to be recommended, N = 10

Step II   - Generate Recommendations

Based on initial preference scores provided by user at Step I, algorithm will calculate the number of articles (n_i) to be recommended from each category, where $1 <= i <= K$. n_i will be calculated as:

$$\text{Ceil} ((N*p_i)/(p_1+p_2+...+p_K)) \qquad (1)$$

Total recommendations given by this algorithm would be summation of all n_i's i.e. $n_1+n_2+n_3+....+n_k$.

Sample Illustration (continued)

$n_1 = \text{Ceil}((N*p_1)/(p_1+p_2+p_3)) = (10*3)/(3+8+6) = 2$
$n_2 = \text{Ceil}((N*p_2)/(p_1+p_2+p_3)) = (10*8)/(3+8+6) = 5$
$n_3 = \text{Ceil}((N*p_3)/(p_1+p_2+p_3)) = (10*6)/(3+8+6) = 4$

Hence, total recommendations given by this algorithm will be 2+5+4 = 11 (ranging between N(10) to N+K (13)). Now, the most recent n_i articles from each category would be picked up and a collated list would be presented/recommended to the user.

Step III – Request and Accommodate Feedback

In this step, request the users to give a positive or negative feedback of recommended articles to find out if they were relevant to their taste. Positive feedback carries a value of +1, negative feedback -1 and no feedback carries a value of 0. Effective feedback for each category is calculated by adding all the feedbacks received on that particular category.

To accommodate this effective feedback into user profile, recalculate p_i i.e. preference score of each news category by adding the corresponding Preference Score p_i and calculated effective feedback (as shown in Figure 1). Store these new preference scores into user profile so that they can be used next time to generate recommendations for the user.

Sample Illustration (continued)

Categories	n_i	Feedback					Effective Feedback	Original p_i	Updated p_i
		Article 1	Article 2	Article 3	Article 4	Article 5			
Business	2	1	0	N/A	N/A	N/A	1	3	4
Tech	5	1	0	1	0	-1	1	8	9
Sports	4	-1	-1	0	1	N/A	-1	6	5

Figure 1: Updated p_i based on user feedback

Proposed algorithm performs following two additional steps while updating Preference Scores (p_i) for each category:

1. If the updated p_i value for a category increases beyond 10 (say by x), due to positive feedback, subtract x from p_i values of all categories.
2. If the updated p_i value for a category decreases below 1 (say by y), due to negative feedback, add y to p_i values of all categories.

It is important to prevent p_i value from becoming 0, to make sure that algorithm generates at least one recommendation for each category where user had indicated an initial interest during signup. Similarly, performance score beyond 10 for a category is taken care of by reducing the scores of other categories to accommodate highly positive feedback for this category. The above two steps, provide robustness to the algorithm, by ensuring that Preference Scores (p_i) for each category remains in range of 1 to 10, even after multiple feedback iterations.

Step IV – Future Recommendations

Whenever user logs in for the next time, Step II is followed to generate and present news recommendations to the user and Step III is executed to collect feedback on the recommended articles to continuously improvise user profile information and quality of recommendations.

4. Simulation of Proposed Algorithm

Proposed algorithm was simulated to estimate its behaviour, correctness and robustness on various use cases. Simulation was done as follows:

Step I

Initial user preference scores p_i (where $1 <= i <= 3$), for K = 3 categories, namely, Business, Technology and Sports were manually collected from user as part of signup data. Also, user input was required to find out approximate number of articles to be recommended i.e. N. N was given as 10 by the user.

Step II

Based on user input received at Step I, initial number of articles (n_i, $1 <= i <= 3$) to be recommended from each category was calculated (as specified by (1)) and n_i ($1 <= i <= 3$) most recent articles were then selected from website http://news.google.co.in under the above mentioned categories (namely, Business, Technology and Sports). These selected articles, ranging from 10 to 13 i.e. N to N+K in number, were presented as initial recommendations to the user.

Step III
User was then asked to provide an optional feedback for recommended articles so as to improvise the user profile and quality of recommendations generated. Positive feedback received from user was stored as +1, negative feedback as -1 and no feedback as zero. Effective feedback was then calculated for each category by adding individual article feedbacks received within that category.
Based on the effective feedbacks calculated above, new preference score p_i ($1<=i<=3$), was calculated for each category by adding effective feedback and existing preference score for each category respectively. These p_i ($1<=i<=3$), were then stored as new preference scores of user into user profile.

Step IV
Step II and Step III were executed 4 times (shown as four iterations, including sign up) to generate and present recommendations to the user and then request user feedback for recommended articles of each category, to update user profile. This was done to show that algorithm works correctly during multiple sign in news reading sessions done by user.

Use Cases
Feedback received from users can vary according to their satisfaction level from generated recommendations. They may or may not like the articles presented to them as recommendations.
The effective feedback for a news category could be highly positive indicating high satisfaction level for recommended articles under that category; could be highly negative indicating a low satisfaction level or could be a proportionate sum of positive and negative responses indicating an average satisfaction level. The three types of possible feedback and satisfaction level of users for the recommendations generated by the proposed algorithm can be depicted by three use cases as described below:
Use Case I – Highly positive effective feedback received from user for recommended articles under all categories.

Use Case II – Highly negative effective feedback received from user for recommended articles under all categories.

Use Case III – Mix of positive and negative feedback received from user for recommended articles under all categories.
Figure 2-10 capture following information obtained from four steps of simulation mentioned above for all

the three use cases. As mentioned before, inside Step 4 in each use case, Step II and III were executed four times (including sign up session), shown as four iterations (sign up and three other iterations) in these figures.
1. Initial preference scores (p_i, $1<=i<=3$) received from the user at the time of sign up,
2. Initial number of articles (n_i, $1<=i<=3$) to be recommended from each category,
3. Effective feedback received for each category in every iteration,
4. New preference scores (p_i, $1<=i<=3$) after accommodating effective feedback received for each category in every iteration.

Use Case I – When effective feedback is positive
In this use case, positive feedbacks are received from user in almost every category (Figure 4), hence the total number of recommendations generated by algorithm in each category (Figure 3) almost remains same.

Use Case 1 Preference Scores (Pi)			
Iterations	**Business (P₁)**	**Tech (P₂)**	**Sports (P₃)**
Sign Up Data	3	8	6
Iteration 1	3	10	7
Iteration 2	4	10	6
Iteration 3	3	10	7

Figure 2: Initial & updated preference scores

Use Case 1 Recommendations to be genereated (Ni)				
Iterations	**Business (N₁)**	**Tech (N₂)**	**Sports (N₃)**	**Total Recommendations**
Sign Up	2	5	4	11
Iteration 1	2	5	4	11
Iteration 2	2	5	3	10
Iteration 3	2	5	4	11

Figure 3: Number of articles to be recommended

Use Case 1 Effective Feedback from User (EF)			
Iterations	**Business (N₁)**	**Tech (N₂)**	**Sports (N₃)**
EF for sign up	0	2	1
EF for iteration 1	1	0	-1
EF for iteration 2	-1	0	1
EF for iteration 3	-1	-1	2

Figure 4: Effective user feedback received

Use Case II – When effective feedback is negative
In this use case, based on majority of negative feedbacks received from user in various categories (Figure 7), the total number of recommendations generated by algorithm gets adjusted accordingly for each category (Figure 6).

Use Case 2 Preference Scores (Pi)			
Iterations	Business (P₁)	Tech (P₂)	Sports (P₃)
Sign Up Data	3	8	6
Iteration 1	4	6	5
Iteration 2	4	5	3
Iteration 3	3	2	2

Figure 5: Initial & updated preference scores

Use Case 2 Recommendations to be genereated (N)			
Iterations	Business (N₁)	Tech (N₂)	Sports (N₃)	Total Recommendations
Sign Up	2	5	4	11
Iteration 1	3	4	4	11
Iteration 2	4	5	3	12
Iteration 3	5	3	3	11

Figure 6: Number of articles to be recommended

Use Case 2 Effective Feedback from User (EF)			
Iterations	Business (N₁)	Tech (N₂)	Sports (N₃)
EF for sign up	1	-2	-1
EF for iteration 1	0	-1	-2
EF for iteration 2	-1	-3	-1
EF for iteration 3	-2	-1	-1

Figure 7: Effective user feedback received

Use Case III - Effective feedback is a mix of positive & negative feedback
In this use case, based on mixed feedback received from user (negative for Sports and positive for Technology, Figure 10), total number of recommendations generated decreases for Sports from 4 to 1 and increases for Technology from 5 to 9 (Figure 9).

Use Case 3 Preference Scores (PI)			
Iterations	Business (P₁)	Tech (P₂)	Sports (P₃)
Sign Up Data	3	8	6
Iteration 1	3	10	7
Iteration 2	1	10	4
Iteration 3	1	10	1

Figure 8: Initial & updated preference scores

Use Case 3 Recommendations to be genereated (Ni)			
Iterations	Business (N₁)	Tech (N₂)	Sports (N₃)	Total Recommendations
Sign Up	2	5	4	11
Iteration 1	2	5	4	11
Iteration 2	1	7	3	11
Iteration 3	1	9	1	11

Figure 9: Number of articles to be recommended

Use Case 3 Effective Feedback from User (EF)			
Iterations	Business (N₁)	Tech (N₂)	Sports (N₃)
EF for sign up	0	2	1
EF for iteration 1	-1	3	0
EF for iteration 2	2	4	1
EF for iteration 3	1	6	0

Figure 10: Effective user feedback received

5. Simulation Results

Graphs shown in Figure 11 to 13 demonstrate how number of recommendations (n_i) to be generated from each category changes after accommodating effective user feedback received during multiple sign in news reading sessions done by user. Three different graphs show the difference in algorithm behavior on obtaining different types of user feedback as described by three use cases. This also proves that the proposed algorithm takes care of changing user preferences by changing the number of recommendations generated from each category accordingly.
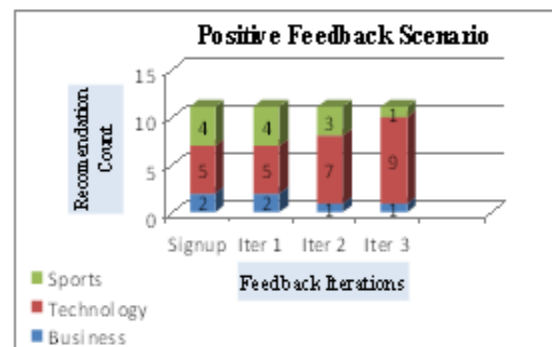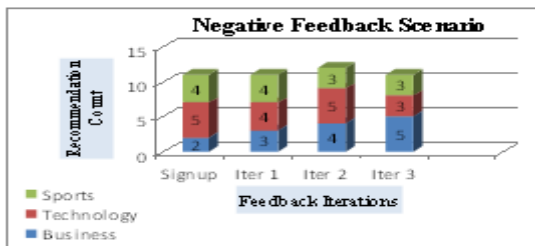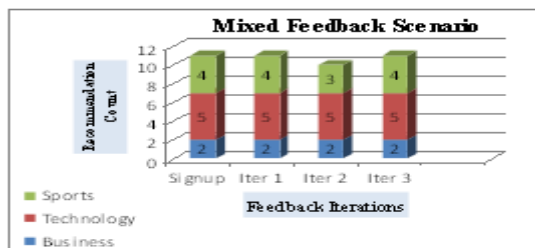

Figure 11: Result graph for Use Case I

Figure 12: Result graph for Use Case II

Figure 13: Result graph for Use Case III

6. Deployment aspects of proposed algorithm

Deployment of the News Recommender System using the proposed algorithm requires use of a reverse proxy server system. A reverse proxy is a service placed between a client and a server in a network infrastructure. Incoming requests are handled by the proxy, which interacts on behalf of the client with the desired server or service residing on the server. Using a reverse proxy server system ensures that we can provide access to news articles available on a news website to the users without giving them direct access to that site. This is done to ensure that user does an initial sign up on the developed web server so that a user profile can be created for them. Thereafter, depending on his profile information, N articles can be selected from the predefined categories of the news source and forwarded to the user as N recommendations. An optional feedback form would be associated with every recommended article to understand users' liking about the article. This feedback will be accommodated into existing user profiles to keep track of changing user interests.

7. Future Work

The proposed algorithm can maintain and update user profiles only if users give some feedback about articles recommended to them. However, in practical situations, obtaining feedback from users is a challenging task as none wants to invest time in this. This problem can be resolved in future to some extent by using additional implicit feedback in algorithm like embedded links clicked by users from a recommended article, indicating their positive response.

Another challenge associated with this algorithm is that on receiving a negative effective feedback for articles from a particular category, we decremented its preference score; however that negative feedback could be valid for some particular articles of that category only. For example, a user having preference score of 5 for Sports category might be more interested in cricket news than hockey or football articles, but presenting articles about a football match update and receiving negative feedback on it would decrement their preference score for cricket articles as well. However, this problem can be resolved by extending the algorithm in future to associate the negative feedback with specific subject inside a category and reducing preference scores for subject specific articles only in that particular category.

8. Conclusion

This paper discusses the changes in the way news is accessed these days and the concerns associated with these changed patterns. A critical problem in accessing news websites online is the volume of news articles can be overwhelming to the users. Recommender Systems have evolved as an answer to this information overload. This paper proposes an algorithm to build a Recommender System, which first develops a user profile and then recommends recent news articles to the users according to their categorical news subscription. It also requests and accommodates an optional user feedback to continuously improvise user profile and provide better recommendations. Simulation results show that the proposed algorithm takes care of changing user preferences by changing the number of recommendations generated from each subscribed news category based on feedback received.

References

[1] Jiahui Liu, Peter Dolan, Elin Ronby Pedersen, "Personalized News Recommendation Based on Click Behavior", Proceedings of the 15th international conference on Intelligent user

interfaces IUI'10, pp. 31-40, 2010.

[2] Toine Bogers, Antal van den Bosch, "Comparing and Evaluating Information Retrieval Algorithms for News Recommendation", RecSys'07, Minnesota, USA, October, 2007.

[3] Pattie Maes, "Agents that reduce work and information overload", Communications of the ACM, Volume 37-No 7, pp 31-40, July 1994.

[4] Chien Chin Chen, Meng Chang Chen, "PVA: a self-adaptive personal view agent system", Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining, 2001.

[5] Fang Liu, Clement Yu, Weiyi Meng, "Personalized Web Search For Improving Retrieval Effectiveness", IEEE Transactions on Knowledge and Data Engineering, 2004.

[6] Saranya.K.G, G.Sudha Sadhasivam, "A Personalized Online News Recommendation System", International Journal of Computer Applications, Vol. 7-No.18, November 2012.

[7] Yi-Shin Chen, Cyrus Shahabi, "Automatically improving the accuracy of user profiles with genetic algorithm", Proceedings of IASTED International Conference on Artificial Intelligence and Soft Computing, Cancun, Mexico, May 2001.

[8] Ah-Hwee Tan, Tee, C, "Learning User Profiles for Personalized Information Dissemination", Proceedings of IEEE International Joint conference on Neural Networks, pp. 183- 188, May 1998.

[9] Daniel Billsus, Michael J. Pazzani, "A hybrid user model for news story classification", Proceedings of the Seventh International Conference on User Modeling, 1999.

[10] Kazunari Sugiyama Kenji Hatano Masatoshi Yoshikawa, "Adaptive web search based on user profile constructed without any effort from users", Proceedings of 13th International Conference on World Wide Web, 2004.

[11] Shankar Prawesh, Balaji Padmanabhan, "Probabilistic News Recommender Systems with Feedback", RecSys'12, Dublin, Ireland, September, 2012.

[12] Will Hill, Larry Stead, Mark Rosenstein and George Furnas, "Recommending and evaluating choices in a virtual community of use", Proceedings of CHI'95.

[13] Paul Resnick, Neophytos Iacovou, Mitesh Suchak, Peter Bergstrom and John Riedl, "GroupLens: An open architecture for collaborative filtering of Netnews", Proceedings of ACM Conference on Computer Supported Cooperative Work, Chapel Hill, pp 175-186, 1994.

[14] Upendra Shardanand and Pattie Maes, "Social information filtering: Algorithms for automating word of mouth", Proceedings of the Conference on Human Factors in Computing Systems, 1995.

[15] Abhinandan Das, Mayur Datar, Ashutosh Garg, "Google News Personalization: Scalable Online Collaborative Filtering", World Wide Web Conference, Banff, Alberta, Canada, May 2007.

[16] Ivan Cantador, Alejandro Bellogín, Pablo Castells, "Ontology-based Personalised and Context-aware Recommendations of News Items", ACM International Conference on Web Intelligence and Intelligent Agent Technology, 2008.

[17] Nasraoui O., Soliman M., Saka E., Badia A., "A web usage mining framework for mining evolving user profiles in dynamic web sites", IEEE Transactions on Knowledge and Data Engineering, Vol 20(2), 2008.

[18] Vozalis E., Margaritis G.K., "Analysis of recommender systems' algorithms", Proceedings of the Sixth Hellenic-European Conference on Computer Mathematics and its Applications, 2003.

[19] Swearingen K., Sinha R., "Interaction design for recommender system", Proceedings of the Designing Interactive Systems, London, 2002.

[20] "Subject Reference System Guidelines" Online at http://www.iptc.org (as of 20 May 2014).

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.