2016 International Conference on Collaboration Technologies and Systems

Recommending News Based on Hybrid User Profile, Popularity, Trends, and Location

Suraj Natarajan
Department of Computer Science
San Jose State University
San Jose, CA, USA
surajnm@gmail.com

Melody Moh
Department of Computer Science
San Jose State University
San Jose, CA, USA
melody.moh@sjsu.edu

Abstract— Reading the news is a favorite hobby for many people anywhere in the world. With the popularity of the Internet and social media, users are constantly provided, or even bombarded, with the latest news around the world. With numerous sources of news, it has become a real challenge for users to follow the news that they are interested. Previous work used user profile to recommend personalized news; and used RSS feeds and latest tweets to provide popular, trendy news. In this work we combine these two methods with three enhancements. First, to personalize news recommendation we used a hybrid approach, which involved the analyses of click through, user tweets, and user Twitter friends list to build user profile, this method significantly improves the accuracy of user profile. Second, to address the importance of temporal dynamics, we add a unique new feature of location preference to the news recommendation system. Third, we allow users to choose the ratio of popular news vs. trendy news they desire. The resulting system is then evaluated based on user satisfaction and accuracy. The results show that the average user satisfaction increases from 8.6 to 9.4 when location preference is added, while the accuracy of the recommendation system is around 92-95%. We believe that the proposed system is a successful example of incorporating temporal dynamics to recommendation systems; the combination of using hybrid user profile, popularity, trends and location would have significant impact on other recommendation systems in the future.

Keywords; click-through analysis, hybrid user profile, news recommendation system, temporal dynamics, Twitter.

I. INTRODUCTION

News media has been commonly refers to as "Fourth Estate" due to its societal and political influences. Nowadays all the news organizations have their own websites and social media feeds. Due to the rapidly increasing volume of news articles available, there are vast sources for users to find out what is happening around the globe. This makes it challenging for the users to select the articles of interest. This problem has led to the evolution of the recommender systems that proactively present users with news articles related to their interests.

Many organizations, not just news media, use recommender systems to suggest specific products to targeted users. The most famous examples include Netflix recommending movies to its users, Amazon and other online retail companies recommending their products. These recommendations are usually based on users' past purchase history and preferences.

There is a long history of using recommender systems to help users navigating through the enormous amount of sources of information. Good recommender systems would provide the most relevant information to the users based on the learned or stated preferences, thus help the users to save time sifting through less relevant information.

Comparing with other recommender systems, news recommender systems possess unique challenges. Most recommender systems are limited in their ability to identify real-time, trendy stories, since typically they rely on a critical mass of user consumption before such information can be recognized [6]. Furthermore, there are many news sources, and news maybe popular or trendy. It is a tremendous task for users to select popular, trendy news tailored to their own interests.

Previous work on news recommendation focused on popularity and trends [6], personal profile [10], and both popularity and personal profile [4]. Our proposed work integrates all these methods. In addition, recognizing *temporal dynamics* as one important characteristics of the concept of interest [5], we introduced *location preference* to the proposed recommendation system, so that a user may choose the current location or another preferred location as an additional input for news recommendation. This is a continuation of our work on applying machine learning on social networks and their security and privacy issues [22-26].

The main contributions of this work include:

1. *A news recommended system that combines popularity, trends, user profile, and preferred location.*
2. *The user profile is built using hybrid analysis, a combination of implicit profiling (click-through analysis) and explicit profiling (both the user's tweets and user's Twitter friends).*

This work was supported in part by SJSU RSCA grant

978-1-5090-2300-4/16 $31.00 © 2016 IEEE
DOI 10.1109/CTS.2016.48

3. *Taking into consideration of temporal dynamics to allow users to choose preferred location (whether the current location or another preferred one).*
4. *Users are allowed to choose the ratio of popular vs. trendy news they want.*
5. *The built system is evaluated based on user satisfaction and correctness.*

The paper is organized as follows. Section II discusses related works. Section III described the main approach including system design and implementation. Section IV illustrates the experiments conducted to evaluate the system. Finally Section V concludes the paper.

II. BACKGROUND AND RELATED WORK

In the following, we describe background and related works on four relevant areas.

A. Generating User Profiles

Vanitha proposed a new idea on personalized web search based on user profile and user clicks [7]. The author used click-through analysis to analyze the user browse history, for understanding user interests [7]. Click-through analysis has also been adopted as part of our hybrid user-profile analysis.

Orlandi, Kapanipathi, Sheth, and Passant proposed a real-time model for characterizing the concept of interest composed of popularity, temporal dynamics, and specificity [5]. In particular, they measured specificity based on the DMOZ hierarchical classification of entities [17], and applied on DBpedia data [16] for classifying user interests. DBpedia is an open source community project that extracts the information from Wikipedia and categories these information into the appropriate categories. DMOZ, or the Open Directory Project (ODP), is a popular web page taxonomy. It combines the collaborative efforts of more than 96,877 volunteers helping to categorize the Web; ODP is thus one of the largest and most comprehensive human-edited Web page taxonomies [17]. This technique has also been adopted in our user-profile analysis.

Kanoje et al. described three main approaches for generating user profiles [9]. Explicit user profiling is a static process where surveys are conducted to understand the user interests. Implicit user profiling is a dynamic process where the user interests are inferred by analyzing the presence of users across various social media. Hybrid user profiling combines the advantages of implicit as well as explicit user profiling [9]. We have used hybrid user profiling to build the user profile; using click through analysis (an implicit method), tweets analysis and user following/friend list in (explicit methods) to build the user profile.

B. Finding Popular, Trending, and Personal News

One of the earliest works using Twitter for recommending real-time news is by Phelan, McCarthy, and Smyth [6]. They applied TF-IDF (Term Frequency and Inverse Document Frequency) [8] on RSS feed news articles and latest tweets to determine if a news article is trendy. TF-IDF is also used in this work to find popular, trendy news.

Jonnalagedda and Gauch used the idea of user profile to build a personalized popular news recommendation system using Twitter [4]. They proposed three approaches: popularity-based, profile-based, and hybrid that is a combination of the last two methods. Both popularity and personal profile methods used cosine similarity [14] to determine how well a news article fits popularity or personal interest. Cosine similarity has also been adopted as part of our approach.

More recently, Lee, Ho, Lim, and Choi suggested to build personal profiles using the information obtained from Twitter, which are then used to provide personalized news recommendation [10]. The idea has been adopted as part of our hybrid user profile analysis.

C. Temporal Dynamics and Locality for Recommending News

Abel, Gao, Houben, and Tao analyzed temporal dynamics of micro-blogging activities on Twitter, and evaluated the quality of the user modeling strategies in the context of personalized recommender systems [1]. They showed that those strategies that consider the temporal dynamics of the individual profiles have achieved better performance.

As temporal dynamics plays an important role in the spontaneous change of the user interests, in the proposed system we have used WOEID (Where on the Earth ID) [12] to locate the user's current location. The idea of WOEID was initially developed by GDC, a London-based geographic information company. The number is unique for each location. For example, the WOEID for United States of America is 23424977. Based on the WOEID of the location (which may be user's current or interesting location), the news are recommended accordingly by the proposed approach.

D. Evaluating of Recommendation Systems

Beel et al. specified the following three features of recommender system quality: accuracy (or the capacity to satisfy the individual user's information need), user satisfaction, and satisfaction of the recommendation provider (such as ease of implementation, cost, etc.) [3]. Avazpour, Pitakrat, Grunske, and Grundy described the dimensions and metrics for evaluating recommendation systems; most important among them are accuracy and user preference (user satisfaction) [2]. In this study we have adopted both user satisfaction and accuracy in our evaluation.

III. PROPOSED ARCHITECTURE

The proposed news recommendation system consists of the following three parts; A) user profile analysis, B) location-based analysis, and C) news recommendation. Each of them is described in the subsequent subsections.

A. Creation of User Profile

To understand user interests, the best approach is to build a user profile. As shown in Figure 1, three methods are used for building the user profile: click through analysis (browser history of the user), user tweet analysis, and user follower analysis as described below.

1) Click through Analysis: Click through analysis, shown on the left column of Figure 1, is to analyze a user's browser history [7]. This analysis will help to understand the user browsing behavior. The most frequently visited websites' URLs are then collected. The next step is to categorize these websites using DMOZ [17], as described in Section II. If the category does not exist, then a new category is created. The last step is to increment the click through count of the particular category.

2) User Tweet Analysis: With more than 320 million active Twitter users [20] and more than 500 million tweets per day [21], Twitter is clearly one of the most popular services on the Social Media, and is therefore a good place to understand users' interests. Twitter provides a feature called tweets, which allow twitter users to express their views on any issue of interest, with the maximum of 140 words for a public tweet. In this work ,only public tweets are used for the analysis, as private tweets cannot be captured using twitter API. As shown on the middle of Figure 1, we use the user's twitter name to fetch the current tweets of the user. From the tweets, nouns are extracted using Stanford Parser, developed by the Stanford Natural Language Processing Group [19]. After the nouns are extracted from tweets the next step is to categorize them to find the user interests. For categorizing the nouns DBpedia is used [16], as described in Section II. All the data in DBpedia is available in both csv and json format. There are a vast number of csv files in DBpedia, and we manually categorize them into 20 different categories including politics, sports, entertainment, etc. If the category does not exist, then a new category is added. The last step is to increment the item count of the particular category.

3) User Twitter Follower/Friend Analysis: To further understand user interest, in addition to analyzing user tweets, it is a good idea to analyze the friends, or the list of people whom the user is following in twitter [18]. The list of friends can be retrieved using Twitter API. As shown on the right column of figure-1, after obtaining the friend list, the next step is to categorize them using DBpedia data. Similar to the last two analyses, if the category does not exist, then a new category is added. The last step is to increment the item count of the particular category.

The final step is to sort the item count in each of the three lists. Then, we take one-third of the top categories from each of the three lists and combine them. This combined list is then sorted in descending order, according to the cumulative counts from the three lists. The sorted list then forms the user profile, which contains the user interests (topics) in the order of user preference.

Category Score = 1/3 (Click through count of the click-through-based category) + 1/3 (item count of the Tweets-based category) + 1/3 (item count of the Follower-based category) (1)

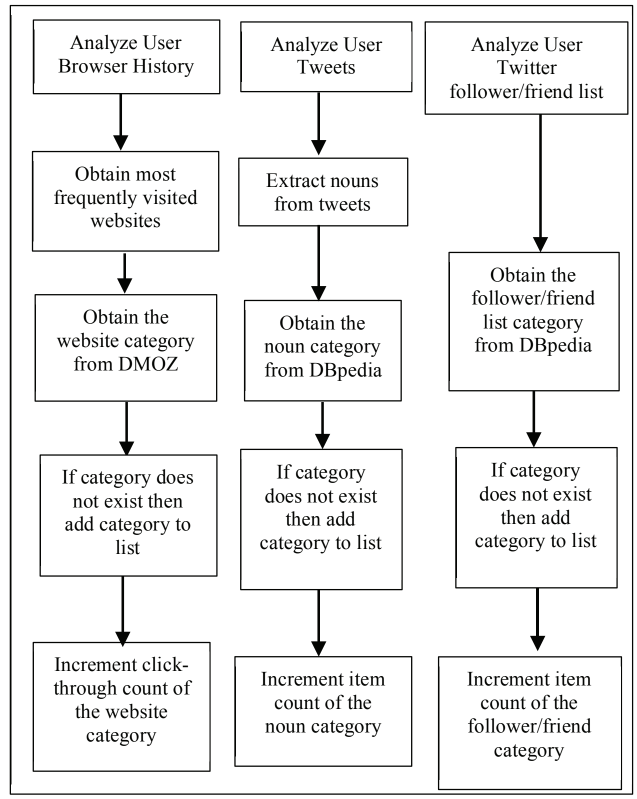

Figure 1. The User Profile Creation Architecture

B. Location Based Analysis

Observing the importance of temporal dynamics in characterizing users' interests, we add the user's preferred location (either current location or another one of the user's choice) in the recommendation system. For finding the current location we could use the IP address of the network to which the user is connected. There are many open-source RESI API available to map an IP address to its location, including the one used in this project [13]. Apart from the current location name, its state and country are also obtained.

Once we have the current (or another preferred) location, state, and country, the next step is to map the location to WOEID [12] (Where on Earth ID), as described in Section II. There are cases when the user might be interested to know more about what is happening at a particular place. WOEID may be used to find the recommendation based on any location the user might be interested in. Since this ID is unique, it is very useful, and is incorporated when retrieving news articles.

C. News Recommendation System

Using the generated user profile (described in subsection A above), a list of customized topics of interest is built. From this list users are allowed to choose any topic. Based on the user location preference (as described in Section B above) and the chosen topic, a list of popular and trendy news will need to be selected and shown to the user. Note that popular news are the news articles which are among the most read articles and are popular over a period of time, while trendy news are the news articles which are related to the topics which are popular at that instant of time.

This subsection describes the system that will recommend personalized, popular, and trendy news. After an overview, the system will be described step-by-step.

The main steps of the proposed news recommender system are outlined in Figure 2. Each step is described in detail in the next subsection.

Proposed News Recommendation System
Step 1: Collect latest news articles from RSS feed URLs.
Step 2: Collect latest tweets using Twitter API.
Step 3: Obtain popular news articles: Compute the Cosine Similarity of news articles and tweets.
Step 4: Collect twitter trends using Twitter API.
Step 5: Obtain Popular and Trendy News Articles: Compute the TF-IDF of news articles and Twitter trends.
Step 6: User to choose the ratio of popular news articles vs. trendy news articles.

Figure 2. The User Profile Creation Architecture

Step 1: Collect Latest News Articles from RSS Feed URLs

The main goal of this step is to obtain the latest popular news from various news media. These are obtained by using RSS (Really Simple Syndication) feeds. RSS is a way to syndicate or distribute news information in the form of short-updates that can be linked back to complete stories. The RSS feeds are obtained by using Google feed-find API [15]. The input to the API includes a query containing the user's selected topic and the preferred location name. (For example, if the user selects "sports" from the user profile list and selects location preference as "Bangalore" then the query will be Bangalore Sports News). After obtaining the top feed URL's related to the query, the top news articles would be obtained from these URL's using the Google feed load API.

Step 2: Collect Latest Tweets using Twitter API

In this step, we collect the latest, most popular tweets. These would be compared with the latest news articles obtained from Step 1. The news articles that are most similar to the top tweets would be considered as most trendy news. The top tweets are obtained from the twitter API, using the same query as described in Step 1. The similarity between top tweets and latest news are determined by the cosine similarity value, as described next.

Step 3: Obtain Popular News: Compute Cosine Similarity of News Articles and Tweets

After obtaining the popular articles and the tweets for a given query, the cosine similarity between each news article and each of the popular tweets is calculated. The cosine similarity between two documents on the Vector Space is a measurement that calculates the cosine of the angle between them [14]; it essentially measures the frequency a common word appears in both documents:

Cosine Similarity =

$$\left(\sum_{i=1\ldots n} A_i * B_i\right) / \left(\sqrt{\sum_{i=1\ldots n} (A_i)^2} * \sqrt{\sum_{i=1\ldots n} (B_i)^2}\right) \qquad (2)$$

Where, *A* and *B* are the documents whose cosine similarity is to be calculated. *n* represented the number of common words in both the documents. A_i represent the count of times word *i* appearing in document *A*, and B_i represent the count of times word i appearing in document *B* [14].

Finally, we sum the cosine similarity of each article with each of the popular tweet and generate the *Popular Co-Occurrence* matrix; an example is given in Table I. Note that for the example shown on Table I, since Article 4 has the top score, it will be listed as the most popular news article.

TABLE I. A SAMPLE POPULAR CO-OCCURRENCE MATRIX
T – Tweet, A – News Article

	A_0	**A_1**	**A_2**	**A_3**	**A_4**
T_0	0.12			0.81	
T_1		0.56	0.31		0.09
T_2			0.46		
T_3		0.36			0.91
T_4	0.11				
TOTAL	**0.23**	**0.92**	**0.77**	**0.81**	**1.00**

Step 4: Collect Twitter Trends using Twitter API

Our next goal is to find the trendy news. The best way to find the trendy news from the collected popular news articles is to collect the latest trends of twitter for a particular location. The top trends in twitter at a particular region can be obtained using the twitter API. The input to the twitter API is the *where on earth ID (WOEID)* [12].

Step 5: Obtain Popular and Trendy News Articles: Compute the TF-IDF of News Articles and Twitter Trends

Now we have the popular articles and the tweeter trends for a given region at a particular instant of time. With these we calculate the TF-IDF (Term Frequency - Inverse Document Frequency) between each news article with all the trends of the twitter. Finally we sum the TF-IDF of each article with each of the twitter trends and generate Trend Co-Occurrence matrix.

Recall that the TF-IDF weight is a weight often used in information retrieval and text mining. This weight is a statistical measure used to evaluate how important a word is to a document in a collection of documents [8]. In particularly, TF (Term Frequency) measures how frequently a term occurs in a document. It is calculated by using the below formula:

TF(t) = (Number of times term t appears in a document) / (Total number of terms in the document) (3)

IDF (Inverse Document Frequency) measures the importance of a term. While computing TF, all the terms are considered equally important. However it is known that certain terms, such as "is", "of", and "that", may appear a lot of times but have little importance. Thus we need to weigh down the frequent terms while scale up the rare ones, by computing the following:

IDF(t) = log_e [(Total number of documents) / (Number of documents with term t in it)] (4)

An example of TF-IDF co-occurrence matrix is shown on Table II. One can see that Article 1 has the top score, and it will be listed as the most trendy news article.

TABLE II. SAMPLE TREND CO-OCCURRENCE MATRIX

T – Tweet, A – News Article

	A_0	A_1	A_2	A_3	A_4
T_0	0.12				
T_1			0.30		
T_2		0.56	0.56		
T_3		0.36		0.89	
T_4	0.87				
TOTAL	**0.99**	**0.92**	**0.86**	**0.89**	**0.00**

Step 6: User to choose the ratio of popular news articles vs. trendy news articles

Finally, we introduced a tunable parameter *ß* that allows users to control how much the Popular and Trend Co-Occurrence matrices each contributes. The value of *ß* lies between 0.0 and 1.0. If the value of ß is closer to 1.00 then the user is interested to know more of popular news, if it is closer to 0.00 then the user is interested to know more of trendy news. So the final score is given by:

Final Recommended News =

ß (Popular News) + (1 – ß) (Trend News) (5)

The tunable parameter will help a user to choose how much popular news articles vs. trendy news articles to be recommended. Note that it would also influence if a particular news article would appear as only popular, only trendy, or both.

IV. EVALUATION

The news recommender system has been implemented on an Intel i5 processor with 64 GB RAM. For evaluating the proposed architecture, we developed a web application for carrying out the evaluation. The web application was deployed in the tomcat server on Windows 8 platform. Table III summarizes the main parameters used in the design of the news recommender systems and the evaluation experiments.

TABLE III. PARAMETER SUMMARY

Parameter	Value
Recommender System: Hardware platform	
Hardware Configuration	6GB RAM, Intel i5 processor
User Profile Analysis	
Number of categories (user interests)	20
Click-through analysis: Number of most frequent visited websites	15
User tweet analysis: Number of tweets	15
User Twitter friend analysis: Number of twitter friends	100
Popular and Trendy News	
Number of RSS Feed URLs	10
Total Number of news articles collected from RSS feeds	100
Number of tweets used to find Popular News Articles	15
Number of Twitter Trends used to find Trend News Articles	10
Evaluation System Design	
Platform	Windows 8
Server	Tomcat
Number of Users/Evaluators	25 (Randomly solicited on the SJSU campus)
ß	User input (between 0 and 1)

A. User Satisfaction

First, the user profile (interests) is built by the following:

1. Only the click-through details
2. Only the user twitter profile (tweets and friend list)
3. Both the click-through details and user twitter profile

Then we ask the users/evaluators to rate their satisfaction of the predicted user profile. The ratings can be between 1.0 – 10.0; 1.0 being the lowest score means predicted user profile is not accurate and 10.0 being the highest score means predicted user profile is accurate. The rating are obtained and the user satisfaction on the predication of user profile (interest) is calculated. Table IV shows the scores given by all the evaluators:

TABLE IV. USER SATISFACTION SCORES
A – Click through Analysis, B - Tweeter Analysis,
C – User Profile using both A and B.

User/Evaluator	A	B	C
1	8.5	8	9
2	8.75	7.25	9.25
3	8	8.25	9.5
4	8.5	7.5	9
5	8	8.5	9
6	9.6	9.1	9.9
7	8	8.7	9.4
8	8	8.7	9.3
9	9	9.1	9.3
10	8.8	8.6	9
11	9.3	8.7	9.7
12	9.2	9.5	9.8
13	9.5	8.9	9.2
14	8.75	8.4	9.7
15	8	7.5	8.4
16	9	8	9.5
17	9	9.3	9.25
18	9	9	9.5
19	8.5	8.7	9.3
20	9	8.5	9.6
21	9	9.5	9.8
22	9	9.3	9.75
23	9	8.5	9.5
24	9.4	9.5	9.7
25	9.5	9	9.7
Total Score:	220.3	216	235.05
Average Score:	**8.81**	**8.64**	**9.40**

It is clear that the user satisfaction in the prediction of the user interest by using both the click through analysis and user twitter profile is higher than using one of the concepts independently. Table V summarizes the percentage improvement. With the additional user data, there is an improvement of 6.69% against only the click-through analysis and 8.79% against only the user twitter profile.

TABLE V. PERCENTAGE IMPROVEMENT IN USER SATISFACTION OF THE PREDICTION OF USER INTEREST

Against Click Through Analysis	Against User Twitter Profile
6.69%	8.79%

B. Accuracy

To evaluate the accuracy of the evaluator satisfaction scores for the user profile, we embedded a feature in the web application, which captures the time spent in reading various news articles on each of the predicted interests. We find the top five categories based on the time spent in reading various news articles on each of the predicted interests, and check if these categories come under the top five categories of the user interest. If all the top five categories based on the time spent in reading various news articles are matching with the top five categories of the user interest then a score of 1 is given, we call this score as the *Profile Correctness Index*. Thus,

Profile Accuracy Index =
(Number of identical categories between top five recommendations and top five longest-time reading) /
Minimum [5, Total number of recommendation categories] (5)

Note that the denominator of the above equation is necessary since some users have less than 5 categories of interest (as in users 4, 9, and 15). So, if only one category based on the time spent in reading matches with the top five categories of the user interest then the Profile Accuracy Index is 0.2. We will then calculate the average of the Profile Accuracy Index, if the value is closer to 1 then the authenticity of the evaluator score is 100%.

Table VI summarizes the profile correctness of each evaluator. The average profile accuracy index is 92.68%, which shows that the evaluators spent more time on reading the news articles that are among or f the top five categories of the predicted user interests.

TABLE VI. PROFILE ACCURACY INDEX
(Note that users 4 and 15 each has only 3 categories, and user 9 has 4 categories of interest. All other users have 5 or more categories of interest.)

User/Evaluator	Profile Accuracy Index
1	1
2	1
3	1
4	0.33
5	1
6	1
7	1
8	1
9	0.5
10	1
11	1
12	1
13	1
14	1
15	0.33
16	1
17	1
18	1
19	1
20	1
21	1
22	1
23	1
24	1
25	1
Total	23.17
Average	**92.68%**

Next, we compute the percentage of time spent in the top five recommended categories. The total time spent by all the evaluators for viewing the recommended news articles considering all predicted interests was 9189 seconds (2 hours, 55 minutes). The total time spent by all the evaluators for viewing the recommended news articles considering top five interests was 8717 seconds (2 hours, 42 minutes). So the percentage of time spent on viewing the recommended news articles considering top five interests is 8717/9189 = 94.86%.

Table VII compares the profile accuracy index and the percentage of time spent on viewing the predicted news articles. The table clearly shows that the predicted user interests match the actual user interests as the users are spending more time in viewing the news articles which comes under the top five predicted user interests.

TABLE VII. PROFILE ACCURACY INDEX SUMMARY

Profile Accuracy Index	% of time spent on viewing the top 5 recommended news articles
92.68%	94.86%

C. User Location Preference Evaluation

Recall that one unique feature of the proposed system is the addition of user location choice. We therefore ask the evaluators to rate their satisfaction of the recommendation system by considering both with and without the user location choice. Recall that the location preference is used along with the user selected interest from the predicted list to search the news articles. The user satisfaction score is the same as those shown in Table IV in subsection A. The results are shown in Table VIII. It is clear that with location choice the user satisfaction scores are increase by more than 5%.

TABLE VIII. USER SATISFACTION SCORE

User/Evaluator	**Without Location preference**	**With Location preference**
1	9	10
2	8.5	9.5
3	9	9.75
4	9.25	9.75
5	9.5	10
6	10	9
7	9	9.5
8	9	9.5
9	9.2	9
10	8.5	9.2
11	8.7	9.1
12	9.7	9.8
13	9.7	9.4
14	8	9.8
15	8.1	8.5
16	8	9.7
17	8.5	8.75
18	9.1	8.9
19	8.5	9
20	9.8	9.9
21	9.6	9.8
22	8	9.3
23	8	9.5
24	9.6	9.8
25	9.1	9
Total Score:	223.35	235.45
Average Score:	**8.93**	**9.42**

To further understand how user location preference improves the recommendation system, we embedded a feature in the web application that captured the number of articles clicked by the user. We took the average of number of articles of all the evaluators. The total number of article clicks was 134 for without location preference method; with location preference method, the total number of article clicks is 284. Table IX summarizes these results.

TABLE IX. AVERAGE NUMBER OF ARTICLE CLICKS BY ALL EVALUATORS

	Without location preference	**With location preference**
Total number of article clicks	134	284
Average number of article clicks per user	**5.36**	**11.36**

The average number of news articles clicked based on the location preference is twice the average number of articles clicked without using location preference. Clearly the evaluators are more inclined towards using the user location preference, as there are more news articles that attract their attention and clicks.

V. CONCLUSION AND FUTURE WORK

We have presented the design and implementation of a news recommendation system based on popularity, trends, location and hybrid user profile. The user location preference is a new feature introduced to news recommendation systems. The system has been evaluated based on user satisfaction and accuracy. We found that using hybrid user profile has improved the accuracy of the prediction of user interests; i.e., users are more satisfied with the news recommended. To evaluate the accuracy of the recommendation system, we measured how much users spent on the news articles of their top 5 interests predicted by the system, and found the ratio, and therefore the accuracy, to be as high as 94%. The evaluation results also found that, with the addition of user location preference, users are more satisfied with the news recommended, which has also been demonstrated by many more news clicks. Future works may include adding feedback to the proposed news recommendation system based on users' satisfaction, dynamically adjust the beta parameter based on users' news clicks, improving search features of retrieving news articles, and performance comparison with existing news recommendation methods. We would also consider users' reading habits as well as develop a way for showing users how their news interests are changing over time. Finally, we would also attempt to spot the commonality between pages a user visits, and evaluate content as another measure that groups pages and as a connection between pages.

ACKNOWLEDGMENT

Suraj Natarajan is currently with IBM Silicon Valley Labs. The authors would like to thank Prof. Teng Moh of SJSU and Mr. Kevin Hite of IBM Silicon Valley Labs for many insightful comments and suggestions, and the anonymous reviewers for their helpful remarks. This work was supported in part by SJSU RSCA grant.

REFERENCES

[1] F. Abel, Q. Gao, G.-J. Houben, and K. Tao, “Analyzing Temporal Dynamics in Twitter Profiles for Personalized Recommendations in the Social Web,” Proc. of WebSci Conf., 2011.

[2] I. Avazpour, T. Pitakrat, L. Grunske and J. Grundy, “Dimensions and Metrics for Evaluating Recommendation Systems, in Recommendation Systems,” Software Engineering, Springer Berlin Heidelberg publication, pp 245-273, 2014.

[3] J. Beel, S. Langer, M. Genzmehr, B. Gipp, C. Breitinger, and A. Nürnberger, “Research Paper Recommender System Evaluation: A Quantitative Literature Survey,” Proceedings of the Workshop on Reproducibility and Replication in Recommender Systems Evaluation, at the ACM Int. Conf. on Recommender System (RecSys)., 2013.

[4] N. Jonnalagedda, and S. Gauch, “Personalized News Recommendation Using Twitter,” Proceedings of IEEE/WIC/ACM International Conferences on Web Intelligence (WI) and Intelligent Agent Technology (IAT), 2013.

[5] F. Orlandi, P. Kapanipathi, A. Sheth, and A. Passant, “Characterizing concepts of interest leveraging Linked Data and the Social Web,” Proceedings of IEEE/WIC/ACM International Conferences on Web Intelligence (WI) and Intelligent Agent Technology (IAT), 2013.

[6] O. Phelan, K. McCarthy, and B. Smyth, “Using Twitter to Recommend Real-Time Topical News,” Proceedings of Third ACM Conference on Recommender System. ACM Press, 2009.

[7] S.Vanitha, “A personalized web search based on user profile and user clicks,” Proceedings of International Journal of Latest Research in Science and Technology, October 2013.

[8] TF-IDF Internet: http://www.tfidf.com/, [October 18, 2015].

[9] S. Kanoje, S. Girase, and D. Mukhopadhyay, “User Profiling Trends, Techniques and Applications.,” Proceedings of International Journal of Advance Foundation and Research in Computer, January 2014.

[10] W.-J. Lee, K.-J. Oh, C.-G. Lim, and H.-Jin. Choi, “User Profile Extraction from Twitter for Personalized News Recommendation,” Proceedings of International Conference on Advanced Communications Technology, 2014.

[11] J. Xu and T.-Ching. Lu, “Toward Precise User-Topic Alignment in Online Social Media,” Proc. of IEEE Int. Conf. on Big Data, 2015.

[12] Yahoo Developer Network Internet: https://developer.yahoo.com/geo/geoplanet/guide/concepts.html, Retrieved October 2015.

[13] Freegeoip Internet: http://freegeoip.net/, Retrived October 2015.

[14] C.-D. Manning, P. Raghavan, and H. Schütze, “An Introduction to Information Retrieval (Draft edition),” (2009, April 1, Retrieved Oct. 2015): http://nlp.stanford.edu/IR-book/pdf/irbookonlinereading.pdf.

[15] Google Feed API Internet: https://developers.google.com/feed/, Retrieved October 2015.

[16] DBpedia Internet: http://dbpedia.org, Retrived Oct 2015.

[17] DMOZ, Open Directory Project Internet: http://www.dmoz.org, Retrieved Oct. 2015.

[18] T.-S. Moh and A. Murmann, “Can You Judge a Man by His Friends? - Enhancing Spammer Detection on the Twitter Microblogging Platform Using Friends and Followers,” Proc. of the 4th Int. Conf. on Information Systems, Technology, and Management (ICISTM), Bangkok, Mar. 2010, pp. 210-220.

[19] Stanford Natural Language Processing Group. Stanford Parser Internet: http://nlp.stanford.edu/software/, Retrieved Oct 2015.

[20] Twitter Company Internet: https://about.twitter.com/company, Retrieved March 2016.

[21] Internet Live Stats Internet: http://www.internetlivestats.com/twitter-statistics/, Retrieved March 2016.

[22] M. Moh, A. Gajjala, S. Gangireddy, and T.-S. Moh, “On multi-tier sentiment analysis using supervised machine learning,” IEEE/WIC/ACM Web Intelligence Conference, pp. 341-344, Dec. 2015.

[23] M. Moh, S. Pininti, S. Daddapaneni, and T.-S. Moh, “Detecting Web Attacks Using Multi-Stage Log Analysis,” Proc. 6th IEEE Int. Advance Computing Conference (IACC-2016), Bhimavaram, India, Feb 2016.

[24] V. Zhao, M. Moh, and T.-S. Moh, “Contextual-aware hybrid recommender system for mixed cold-start problems in privacy protection,” IEEE International Conference on Intelligent Data and Security, pp. 400-405, Apr. 2016.

[25] F. Yu, M. Moh, and T.-S. Moh, “Towards Extracting Drug-Effect Relation From Twitter: A Supervised Learning Approach,” IEEE Int. Conf. on Intelligent Data and Security, New York, Apr 2016.

[26] P. Nair, M. Moh, and T.-S. Moh, “Using Social Media Presence for Alleviating Cold Start Problems in Privacy Protection,” International Conference on Collaboration Technologies and Systems, Oct. 2016.