Tweets: Review of Micro-Blog Based Recommendation Systems (RS) for News Recommendation (NR)

Nithya. K, Krishnamoorthi. M, Kalamani. M

Abstract: Micro-blogs are an approach for social media users to articulate their views and interests by posting the short messages (Tweets) as news. In general, an enormous quantity of news tweets generated everyday for the person who does not have enough time to read all the online news based on their interest. The objective of the News Recommendation Systems (NRS) is to classify and filter out unrelated online news and to suggest only the relevant news to the users based on their preferences. The NRS is working based on her/his social media activities, news reading history, location and other online activity histories. This research article will present a wide-range of analysis about the social media NRS from various research articles in the literature by considering the approaches used for recommendation, various performance measures and the data sets used.

Index Terms: News Recommendation System, Social Media activities

I. INTRODUCTION

Recommendation systems (RS) collects information from different social networking sites based on users preferences. It aims to provide users with suggestions of news, travel destination, movie, photos, videos, or other web content they might like. A recommender system performs data analysis on users' information from different sources and provides the recommendation of items to users. Social media sites, such as Twitter and Facebook, have become predominant in recent years. In the analysis of social media content,increasing this content on the web has become a most important factor. Nowadays with the continuous use of the internet, people conveying their views and interests in the form of comments on social networking sites whenever they think. Such information typically contains multimedia contents, such as text, image, and video. The data collected from different social media websites and analyzed using analytics tools. The most common use of social media analytics is to understand customer preferences and to support marketing and various customer recommendation services. Sentiment Analysis is a technique used to extract some useful information from User Comments. There are two different approaches to the keyword or informative term extraction: Supervised framework, unsupervised framework.

In Supervised framework effective extraction of keywords is done using various machine learning classification algorithms. The dataset is usually called as training dataset. Training datasets are already classified and labeled. Supervised learning test data is given as input to the system then it identifies the label. Unsupervised framework, identify the hidden patterns in unlabeled input data. Text information is converted into a graph with nodes and edges. Each node represents individual text units like words and phrases. Each edge represents the relationships between these text units. Recommendation system mainly based on users personalized information. Some of the traditional recommendations systems are based on sources like User profiles, User online history and User location history. It tracks the mode of the public about a product or happening event. In the Figure 1 RS recommends user interesting topics or items to the users.
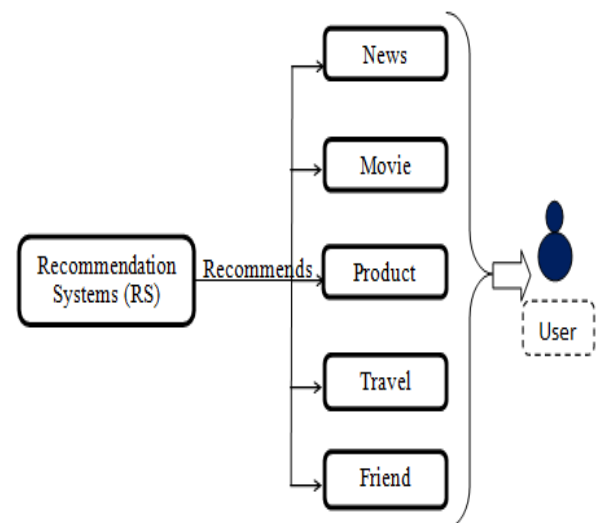

Figure 1- User Preference Based Recommendation System

The increasing use of social media some unverified contents may be posted by the user it often leads to spread out some rumors the time of posting. Rumors are categorized into long-standing rumors and newly emerging rumors. In long-standing rumors, the information or message circulate for a long time in social media and the emerging rumors spawn only for a minimum amount of time as breaking news [5]. Rumor detection and rumortracking are the components of rumor classification system and it is clearly categories the original content and rumors. Our recommender system based analysis is carried out as follows: Section-2 and Section-3 describes different classifications andreviews of news recommendation systems, respectively.

Section-4 provides the details of the datasets and disadvantages and Section-5 discuss the result and Section-6 provides the conclusion.

II. RECOMMENDATION SYSTEM-CLASSIFICATION

Tremendous growth of social media usage initiated the recommendation systems as a self-regulating research area and many researchers started their research work on recommendation problems based on user preferences. It is noted from the literature that recommendation approaches are generally categorized into three main approaches and one hybrid approach as shown in the Figure 2.
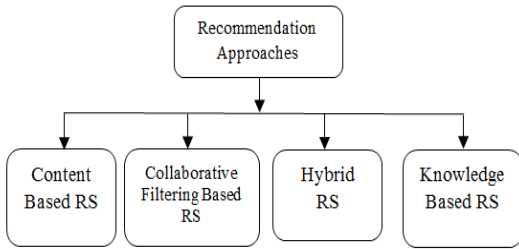

Figure 2 - Recommendation System Classification

Content-Based (CB) approaches are developed to create different representation of contents based on the comments and feedbacks acquired from the users' preferences. Collaborative Filtering (CF) approach automatically recommends user's preference by compiling similar interest from various users. Hybrid approach is most extensively used in social media recommendation system, as it blends both CB and CF approaches and acquires the benefits of both the approaches. In Knowledge-Based (KB) filtering approach offers clear facts about the user preferences and criteria used for recommendation.

III. NEWS RECOMMENDATION SYSTEM

News and social media together play a combined role. Social media sites such as Twitter have taken an important role of consumption and spreading of news. News editors have faith on various social media contents for predicting their audience's concentration on news. Twitter is a fast and ever growing social media platform that has taken a role of consumption of information, production and spreading of that information as news. Twitterposts—called tweets—are restricted to 140 characters and can be sent from any portable device. NEWS media collects news Articles from various content providers like bbc.com and cbsnews.com

Davis *et al.* suggested an unsupervised SociRank system which categorizes news facts which is more familiar in together social media streams and the popular news media. Further, topics are ranked by relevance using their degrees of Media Focus (MF), User Attention (UA), News Media & Social Media User Interaction (UI) as shown in Figure 3. Based on the events popularity, professional news media collects efficiently verified happenings; further social media suggest the audience interest in these areas. Additional information is given to a specific topic for news media with the help of social media sites. This method increases the quality and trust of news recommendation system by identifying user interested news topics and discovering the hidden popular topics. But, it only considers user attention from social media without bothering the other

practices of user attentions such as user search history and search engine click-through rates.
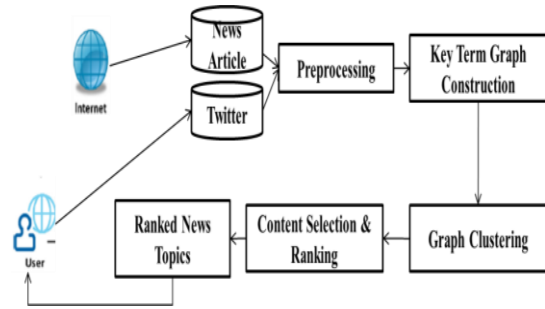
Figure 3- Social Ranking

Jiang *et. Al* developed a WEAN method to identify the association among different words. In social media big data analytics this Word Emotion Association Network is used to calculate the emotions of news incidents. In addition, this also supports for many real-world applications, such as customer attitude monitoring and news recommendation for government and various news Websites respectively. The word in dissimilar situations may have different sentiments. So, there is a need for computation of word emotion at a particular point of time. This method consists of two techniques: word emotion association network for term sentiment computation and standard emotion thesaurus for term sentiment enhancement. In word emotion association network nouns, verbs and adjectives in sentences are treated as nodes and the association among words are represented as links of nodes. The sentiment of a word is computed using WEAN and it is refined using modern refinement algorithm. But, this work not focus on the emotion distance and pattern into text sentiment computation.

ZhiweiJin *et. al* recommended fake news detection method using visual and statistical image features of various micro-blogs. Generally, without validating the reliability of news, some false news would blow out quickly through social media network and results cruel effect. News verification usually depends on text features extracted from public tweets. However, like texts content, images have a very good influence on micro-blogs news broadcast. Also, real and fake news events have dissimilar image distribution patterns. Automatic news verification on micro-blog was done using image content with 5-visual features and 7-statistical features as shown in Table 1. It sense fake news visually and statistically. If this method applied to more number of features then accuracy may be improved.

Table 1- Visual and Statistical Image Features (12 Features)

Visual Features	Statistical Features
Clarity Score(CS)	Count
Coherence Score(CoS)	Image Ratio
Similarity Distribution Histogram	Image Ratio II
Diversity Score(DS)	Multi-image Ratio
Clustering Score(ClS)	Multi-image Ratio II
-	Hot Image Ratio
-	Long Image Ratio

Bichen Shi *et.al* proposed recommendation model which includes learning-to rank approach for news recommendation and data collection approach for feature computation. The Hashtagger in Twitter hashtag function create a link between news articles posted by leading news web sites and the tweets posted by various users, by obviously identifying the stories, combination the conversation of news and marking trending topic. Hashtagger+ is an efficient and effective learning-to-rank framework model for integration of real time news and social media streams, by suggesting various hashtags to news articles as shown in Figure 4. For considering the news headlines as a rich tweet, Hashtag recommendation is used for tweets.

Recommendation Algorithm

Input:

   Set of news articles
   Public Twitter stream
   Recommendation Model

Output: Recommended hashtags for each article.

Method:

1) New article is chosen for processing
2) Set of keywords are extracted from the article and it degrade the tweet quality and volume of computing features
3) Set of streamed tweets for particular article
4) Update the tweet bag of article with new terms
5) Compute features and normalize feature vector with respect to the article
6) Calculate relevance score
7) If relevance score above threshold then recommend hashtag
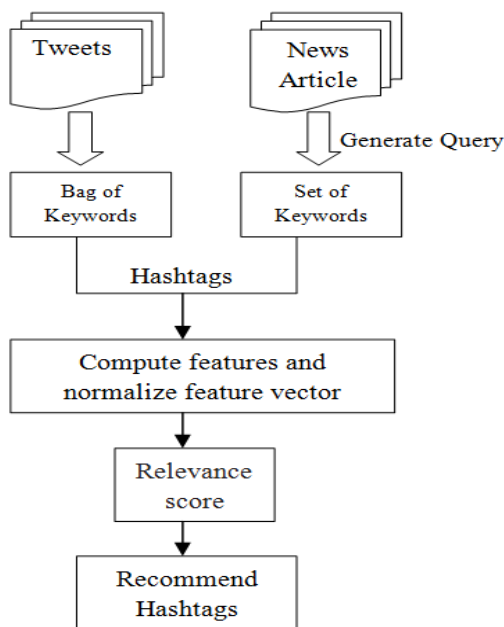

Figure 4 - Hashtagger+ Recommendation System

Stuart E. Middleton *et. Al* presented about the social media in which it is the room for both reporters and the general public when new news incident arrives. Offers a eyewitness real-time source images and videos through platforms like YouTube and Instagram. Now, content value of recording and circulated flash news is negotiation by the false claims and increasing of content misuse in social media. In order to maintain the trust value in social media of flash news, the low cost social-computing solutions are crucial (Anitha Anandhan et. al 2018). Cheng Chen *et. al* proposed location dependent news recommendation system, which considers current geographical location and users personal preferences based on that it recommends news to the users. The location based ESA and DSA both are personalized one. In the location-dependent personalized ESA (Explicit Semantic Analysis) both the user specific interest and location information are considered for news recommendation. In Explicit Semantic Analysis method each Wikipedia concept is observed as a topic, and since there are so many concepts on Wikipedia, the resulting topic suffers from the problems of dimensionality, sparsity, and redundancy. Deep Semantic Analysis method overcomes all these problems with the help of deep neural networks. It map the topic space into various feature space (ie. abstract, dense, and low dimensional) due to that irrelevant news are minimized and the similarities between the users and their target news are increased.

Shaymaa Khater *et.al* introduced vibrant modified user interest based recommendation system that offers the user with the most important tweets. It also integrated along with novel model called 'TrendFusion'. This method enhances user suggestion provided by social media. It analyze the topic which fade quickly and also detect he topic which is more interesting. Estimate trends of the particular location before even users in that location start realizing that trend. TrendFusion model can be applied to various levels like individual level and the community level (location based).

XIAO Yingyuan *et al.* proposed Time-Ordered Collaborative Filtering (TOCF) for News Recommendation. The time sequence characteristics of users browsing behavior are not considered most of the time in the traditional news recommendation systems. In order to alter this traditional system, time sequence characteristic based similarity measure is used. This measure is used for computing the similarity among different users. In addition, this also combined into the recommendation algorithm for news recommendation system.

Rita Georgina Guimarães*et.al* introduced age group classification method using deep learning. It identifies some parameters related to age through the search history and user profile information. Using machine learning algorithms these parameters was analyzed and then it predicts the age group category. DCNN reaches the best result for age group classification. But, it only considers age group as two classes like teenagers and adults. It does not identify the exactage.

IV.   BACKGROUND DATASET DESCRIPTION&PERFORMANCE MEASURE

Table 2- Background Dataset Description and Performance Measure

News Recommendation Systems	Data sets	Details of Dataset		Main short comings	Performance Metrics
SociRank[1]	Tweets: Public timeline of Twitter News articles: News websites	Tweets: 1st November 2013-28th February 2014 (71731730 English Tweets)	News Articles: bbc.com, cbsnews.com (105856 News Articles)	Other forms of User Attentions like search engine click-through rates and search history are not considered in this method	Highest Precision 0.9
Word Emotion Association Network [WEAN] [2]	SinaMicroblog:Malaysian Airlines MH-370 on March 8, 2014	SinaMicroblog: From March 8, 2014 to April 6, 2014 (48,396 microblogs)	-	In text sentiment computation sentiment distance metrics and word emotion patterns are not considered	Accuracy of 80%. Highest Precision 0.85
Micro-BlogsVisual Features and Statistical Image Features [3]	Dataset: SinaWeibo (Multimedia)	Tweets: 50287 Images:25953 Users: 42441	News Articles: 146 news Events	Only 12 features were considered	News Verification Accuracy of 83.6%. Highest Precision 0.855
Hashtagger+[4]	Tweets: Twitter public timeline News articles: leading news websites	Tweets: #Hashtag1- 790,872 #Hashtag2- 176,523 #Hashtag3- 93,414 #Hashtag4- 92,687 #Hashtag5- 83,667	News Articles: RSS feeds: 16th June 2016 - 1st July 2016 (6 different news organizations 10300 articles in total)	Traditional text Mining models can be replaced effective one News story detection and tracking can be included	Highest-precision 0.89 High-coverage recommendations 77 %
LP-ESA[8]	Twitter: Tweets posted by 1,619 users	Tweets: 2,31,6204	One Million Tweets contains URLs to the news. 63,485 – News Articles From 2,366 locations	LP-ESA is time-consuming, high dimensionality, sparsity, and redundancy	High-precision 0.56
LP-DSA [8]	Twitter: Tweets posted by 1,619 users	Tweets: 2,31,6204	One Million Tweets contains URLs to the news. 63,485 – News Articles From 2,366 locations	Not considered more contextual information More sophisticated neural network may be applied	High-precision 1.162
TrendFusion[9]	Twitter: Trending topics	Tweets: August 2014 to September 2014	Users: 20000 Tweets :5 Million	More analysis needed on Topic spreading location identification	High-precision 0.84
TOCF[10]	Real dataset: Popular news website in China	Dataset: Users:10,000 Browsing history: 1st March,2014 to 31stMarch, 2014	After Preprocessing: Users:7192 User rating behaviors: 63,714	Extensive tests and Performance analysis need to be conducted	High-precision 0.87

Age Groups Classification [12]	Tweets crawled from Twitter public timeline	Twitter: 7000 sentences Valid sentences: 6280 were considered for age group classification.	Training Model: 80% of sentences (5024) Test Model: 20% of sentences (1256)	Only considering teenagers and adults age groups	High-precision 0.95

V.   RESULT AND DISCUSSION

We compare the performance of various News Recommendation Systems (NRS) using the performance metrics precision. The value of precision is calculated using the following formula:

$$Precision = \frac{Relevant\ Doc \cap Retrieved\ Doc}{Retrieved\ Doc}$$

Out of retrieved documents, which are relevant to the topic are identified based on that precision is calculated. Many News Recommendation Systems identifies various innovative models for effective News Recommendation. The recommendation mainly focused on location and trending topic. Figure 5 shows the precision value of different methods, in this Personalized Location dependent Deep Semantic Analysis gives the highest precision value of 1.162 and the DCNN method have the next highest precision value 0.95.The various recommendation methods tries to conclude that, the greater precision value means that less number users get the incorrect recommendation or suggestions.
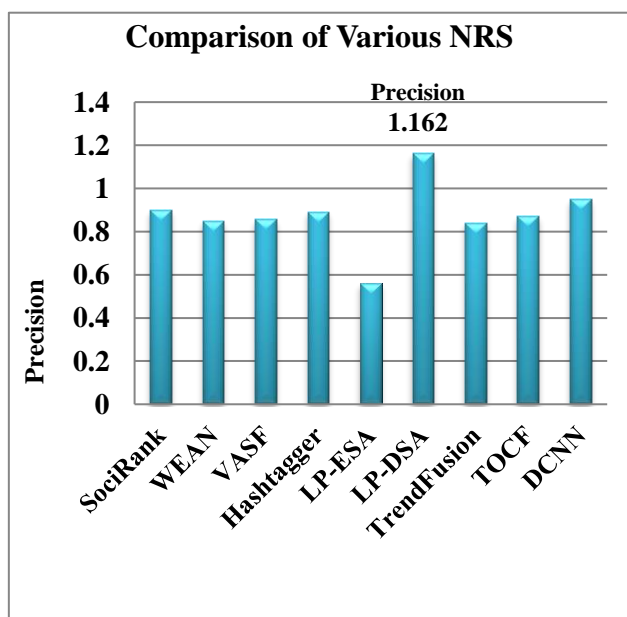

Figure 5 - Comparison of different NRS using the measure precision

VI.   CONCLUSION

Our result shows that Twitter is the main source of user opinion or preference identification. Results presented in Table 2 and Figure 5 clearly describes the location-aware method, personalized method and age group based classification method these three gives the better recommendation result compare to other recommendation systems. In future, these selected methods can combine and new hybrid method may be developed.

REFERENCES

1. Derek Davis, Gerardo Figueroa, and Yi-Shin Chen, "SociRank: Identifying and Ranking Prevalent News Topics Using Social Media Factors" IEEE Transactions on Systems, Man, And Cybernetics: Systems, Vol. 47, No. 6, June 2017.
2. Dandan Jiang, XiangfengLuo, JunyuXuan, and Zheng Xu, "Sentiment Computing for the News Event Based on the Social Media Big Data" IEEE Access on Intelligent Sensing On Mobile and Social Media Analytics, Vol. 5, March 2017 pp. 2373-2382.
3. ZhiweiJin, Juan Cao, Yongdong Zhang, Jianshe Zhou, and Qi Tian, "Novel Visual and Statistical Image Features for Microblogs News Verification", IEEE Transactions on Multimedia, Vol. 19, No. 3, March 2017 pp.598-608.
4. Bichen Shi, GevorgPoghosyan, Georgiana Ifrim, and Neil Hurley, "Hashtagger+: Efficient High-Coverage Social Tagging of Streaming News", IEEE Transactions on Knowledge and Data Engineering, Vol.30, No.1, January 2018 pp.43-58.
5. ArkaitzZubiaga, Ahmet Aker, KalinaBontcheva, Maria Liakata and Rob Procter, "Detection and Resolution of Rumours in Social Media: A Survey", ACM Computing Surveys, Vol. 51, No.2, Article 32. February 2018, pp.32:1– 32:36
6. Stuart E. Middleton, Symeon Papadopoulos and YiannisKompatsiaris, "Social Computing for Verifying Social Media Content in Breaking News", IEEE Internet Computing, March/April 2018 pp. 83-89
7. AnithaAnandhan, LiyanaShuib, MaizatulAkmar Ismail and GhulamMujtaba, "Social Media Recommender Systems: Review and Open Research Issues", IEEE. Translations and content mining, Volume 6, 2018, pp.15608 -15628
8. Cheng Chen, XiangwuMeng, ZhenghuaXu, and Thomas Lukasiewicz, "Location-Aware Personalized News Recommendation With Deep Semantic Analysis", IEEE. Translations and content mining, Vol. 5, July 2017, pp.:1624-1638.
9. ShaymaaKhater, Denis Graˇcanin, Senior Member, IEEE, and Hicham G. Elmongui, "Personalized Recommendation for Online Social Networks Information: Personal Preferences and Location-Based Community Trends", IEEE Transactions on Computational Social Systems, Vol. 4, No. 3, September 2017, pp.:104-119.
10. XIAO Yingyuan, AI Pengqiang, Ching-Hsien Hsu1, WANG Hongya and JIAO Xu1, "Time-Ordered Collaborative Filtering for News Recommendation", China Communications, December 2015, pp.: 53-62.
11. Twitter Home Page [Online] Available: http://twitter.com
12. Rita Georgina Guimarães, Renata L. Rosa, Denise De Gaetano, Demóstenes Z. Rodríguez, (Senior Member, Ieee), and GraçaBressan, "Age Groups Classification in Social Network Using Deep Learning", IEEE.Translations and content mining, Vol. 5, 2017, pp. 10805-10816.
13. Nithya K, Kalaivaani PCD and Thangarajan R, "An enhanced data mining model for text classification", International Conference on Computing, Communication and Applications (ICCCA - 2012), IEEE Xplore, Print ISSN: 2325-6001.
14. Nithya K, Saranya M and Dhivyaa C R, "Concept Based Labeling of Text Documents Using Support Vector Machine", International Journal on Recent and Innovation Trends in Computing and Communication, Vol. 2 Issue. 3, pp. 541-544.