iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 8 **Issue:** VIII **Month of publication:** August 2020

DOI: https://doi.org/10.22214/ijraset.2020.27921

Location-Aware Personalized News Recommendation Based on Behavior and Popularity Technique

Hengane Shubham[1], Pathan N.J.[2]
[1, 2]Computer Science and Engineering, Dr. Babasaheb Ambedkar Technological University, Lonere.

Abstract: Now-a-days people can read news from several sources around the world. This paper investigates a novel user profile model to express users' preferences from different aspects. Then, considers the scope of the user's preferences for historical news, and propose a method to calculate the desire weight of historic news consistent with the user's analyzing behavior and the popularity of news. This method may want to assemble user profiles greater correctly. Additionally, represents a dynamic technique for news recommendation, wherein each short-term and long-term user preferences are taken into consideration. The contribution work is to implement location-aware personalized news recommendation with explicit semantic analysis (LP-ESA), which recommends news using both the users' personal interests and their geographical contexts. The experimental consequences show that BAP technique and LP-ESA technique can fundamentally increase the recommendation outcome.
Keywords: Location-Aware News Recommendation, User Profiling Method, News Recommendation, Personalization, User Behavior, Collaborative Filtering.

I. INTRODUCTION

As we know, an ample amount of population in the world is reading News Article over the internet by visiting News Article of their interest or by visiting popular news web portals. However immense amount of news information is available online everyday but in this busy lifestyle, it is difficult for the individual user to find the news of his interest as well as happening near to his location, so to address this issue personalized news Recommendation approach is needed in this real world.

Recommending news articles is one of the most challenging recommendation tasks. The news domain differs from other domains in many ways. For example; the popularity and recency of news articles changes so fast over time. So focusing on the recency issue becomes more challenging than it is in other domains. Also some news articles may be connected with each other that the user may want to read the previous news items related to the one he/she already reads or he/she may want to keep informed about. Only learning user preferences can be an unsatisfactory solution to news recommendation. This is because the user may want to read a news article when he/she is not really interested in the subject but he/she thinks it is important. For example; wanting to read the news about elections even if he/she is not generally interested in politics. Also considering the high number of new articles published every hour increases the complexity of other challenges.

Recommendation System can be classified broadly into several categories depending upon the information provided to it. (i) Content Based Filtering System uses Information of users as well as information / metadata of product to perform recommendation. CBF implemented in movie recommendation based on user rating and user interest. (ii) Collaborative Filtering Recommendation System uses information about a set of users and their relation with the item to provide recommendation to the active user. This approach adopt Neighbourhood method who‘s focus is on relationship between the item or alternatively between the users .CF Approach is further divided into User Based CF and Item Based CF. (iii) Hybrid Recommender System incorporates both the above mentioned techniques to overcome their drawbacks currently most of the web platforms uses this recommendation technique to provide the user the product of his sole interest.

The existing news recommendation systems can be divided into two parts: By category-wise and by recommendation type. The category-wise news recommendation contains 4 categories like, personalized [2], [6], [11], [12], in which users personal interest are considers and those interest are evaluated from their web site browsing behavior or social platforms, location-based [1], [13], [15]: users longitudinal and latitudinal components are considered for performing recommendation , content-based [9] and semantics-based [10], [14] news recommendation works on the contextual similarity of user and news item . The recommendation type contains three categories: content-based recommendation, recommendation based on collaborative filtering, and hybrid recommendation.

Content-based recommendation [8]: the recommendation system tries to find news with similar content to the news that the user has read. Content-based recommendation systems are usually easy to implement. However, in some scenarios, the profile of the user with a bag of words is not sufficient to accurately capture the user's preferences. Recommendation based on Collaborative filtering [4], [5]: the system recommends news by users' news ratings, and generally they are content free. Many users do not have sufficient historical behaviors, or the number of users in the system is not high enough, which is known as a cold-start problem. Hybrid recommendations, [3], [7]: as discussed above, content- based and collaborative filtering recommendation systems can provide meaningful results, but they have some drawbacks.

A. Motivation

1) To calculate the user's preference for the news based on user behavior and news popularity using Behavior And Popularity method.
2) To provide a high quality of dynamic recommendation results with the short-term and long-term preferences of users.
3) Users' news preferences strongly depend not only on their geographical contexts (i.e., locations), but also on their personal interests; so, both should be gives satisfactory personalized news recommendations outcomes.

This paper is organized as follows. Section 2 Literature Reviews which discusses about the state of the art methods and approaches. Section 3 Discussions: summarizes all the challenges of recommender systems in news domain which includes some common challenges with general recommender systems. Section 4 Future Directions this section describes the future development can be done in the field of news recommendation. Experimental results are discussed in Section 5 and at last Section 6 Conclusion of this implementation paper.

II. RELATED WORK

In this section, we briefly review the related work on news recommendation system and different techniques used in it.

A. Category-wise News Recommendation

1) *Personalized News Recommendation:* The paper [2] presents a rough set based collaborative filtering approach to predict a missing news category rating values of a user, and a new novelty detection approach to improve ranking of novel news items. An end-to-end system prototype that can take a collection of news articles and the user interest as input and then re-rank the news articles based on novelty and CPCC similarity between the user profile database and common news articles database to provide a personalized news recommendation to the users. Advantages are: Automatically detect the novelty of news items. Efficient approach to automatically detect the missing rating value of an active user. Disadvantages are: Does not work on dynamic community detection. In [6] paper, proposes PENETRATE, a novel Personalized News recommendation framework using ensemble hierarchical clustering to provide attractive recommendation results. Our proposed framework is beyond content-based methods and collaborative filtering, in which individual user behavior and user group behavior are simultaneously considered for recommendation. Advantages are: Improved accuracy and efficiency. Disadvantages are: Time consumption is high. The paper [11] Proposes *SCENE*, a Scalable two-stage personalized News recommendation system with a two-level representation, where the first level contains various topics relevant to users' preference, and the second level includes specific news articles. SCENE consists of three major components – Newly-Published News Articles Clustering, User Profile Construction and Personalized News Items Recommendation. Advantages are: This system supports efficient clustering on newly-published news articles. It provides high quality of recommendation results. Disadvantages are: Does not handle high volume press releases. The paper [12] presents a mobile application iNewsBox 1 enabling users to listen to news collected from the Internet. Then learn interests and needs of users from their implicit feedback in iNewsBox and play the synthetic sound of news selected by our news recommendation system to users individually. Advantages are: This system is effective and extensible. A story can be broadcasted to users by this system within 2 minutes after the media published the story through RSS.
2) *Location-based News Recommendation:* The brief representation of LP-ESA is shown in Figure 1, where a geo-tagged document $d \in Dl$, a past tweet history records of user for generating user profile U(u, l) $h \in Hu$, a news article for generating news profile N(n, l) $n \in N$, and a topic may contain several sentence s as well as many of word w and phrases p, finally all are dependent on topic $t \in T$, while the overall process and the topics depend on a location(GPS) l.

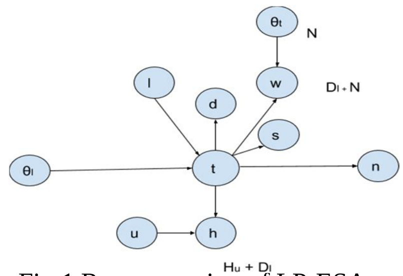

Fig.1 Representation of LP-ESA

The paper [1] proposes a hybrid method called location-aware personalized news recommendation with explicit semantic analysis (LP-ESA), which recommends news using both the users‘ personal interests and their geographical contexts(GPS). Further proposes a novel news recommendation method, called LP-DSA, to solve the huge dimensionality, sparsity, and redundancy problems in LP-ESA by deep semantic analysis. LP-DSA uses recommendation oriented deep neural networks to extract dense, abstract, low dimensional and effective feature representations for locations, users, and news. Advantages are: LP-DSA overcomes the huge dimensionality, sparsity, and redundancy problems in LP-ESA by using deep neural networks. Increase the news recommendation performance in terms of both effectiveness and efficiency. Reduces the computation cost. Disadvantages are: Need to learn a more effective abstract feature space, to further improve the performance of LP-DSA.

In [13] paper, presents Geo-Rank, efficient location-aware news feed ranking system that provides top-*k* new feeds based on (a) spatial proximity, (b) temporal proximity, and (c) user preferences. GeoRank is composed of two main modules, namely, query processor and message updater. Advantages are: GeoRank significantly reduces the user response time and system overhead. Disadvantages are: All the followers need to be selected for the most relevant message update. The paper [15] proposes a geo-topic extraction framework for geo-location inference, including location name entity recognition, location related image association and a multimodal location dependent pLSA geo-topic model. Proposes an inference framework that includes location name entity recognition, related image association and a location related fusion model (mLD-pLSA) that combines location, text and visual information. Advantages are: This fusion model improves the inference performance.

3) *Content-based News Recommendation:* The paper [9] is to develop a Web content recommender system. A user‘s long term interest and preference models are constructed based on the user‘s navigational history and integrated with the recommender system. The similarity between Web content and the user‘s models is used to determine whether the content will be provided to the user. Advantages are: It provides an autonomous navigation model that is able to relieve Web users from repetitive and tedious Web surfing, auto-classifying Web pages and improves the efficiency.
4) *Semantics based News Recommendation:* In [10] paper presented two approaches that take into account the meaning of words. The methods are based on concepts and their semantic similarities, from which derive the similarities between news items. First method, Synset Frequency - Inverse Document Frequency (SF-IDF), second method, Semantic Similarity (SS).The proposed approaches to news item recommendation have been implemented as Ceryx, an extension to the Hermes News Portal news personalization service.
5) *Advantages Are:* Performs statistically better than TF-IDF. The paper [14] proposed a localized news article recommendation with the Explicit Localized Semantic Analysis (ELSA).The proposed method recommends the news articles that are appropriate to the location by reflecting the geographical context of user. Advantages are: It stresses only the topics that are relevant to a given location. It allows a simple comparison of the user location and news articles, because it projects both locations and articles onto an identical space composed of Wikipedia topics.
6) *Recommendation Type*
a) *Content Based Recommendation:* The paper [8] proposes a scalable news recommendation method to solve the problems, including the multi-dimensional similarity calculation method, the Jaccard–Kmeans fast clustering method and the Top-N recommendation method. The multi- dimensional similarity calculation method computes the similarity between users combining abundant content feature of news with the behavior and time feature of users, solving the data sparsity problem in traditional collaborative filtering. Advantages are: The proposed scalable recommendation method can solve the scalability problem effectively. It reduces the negative effect of data sparsity, and increases the quality of news recommendation. Disadvantages are: This paper represents only content based recommendation not collaborative filtering.

b) Collaborative Filtering Method: The paper [4] analyses web server access logs of a large online news publisher to identify readership patterns on the web. In particular, the analysis is done by first developing a model, which can be used to predict most likely articles to be read by a particular user, followed by analyzing what are the most important features and interpreting the learned model. Advantages are: Time window selection can have a significant impact on the accuracy of predictions. Disadvantages are: The inclusion of Users feature, which is computationally the most expensive. The paper [5] represents the effect of user interest evolution when modeling user profiles, and represents the user‘s reading preference with seamless integration of long-term and short-term user profiles. Construct a two-stage news selection strategy, where the long-term profile is firstly utilized to differentiate news groups with specified preference, and then the short-term profile is applied to filter specific news articles to individual users. Advantages are: The short-term one is able to capture the recent/current user preferences over fine-grained news topics. To introduce more diversity in order to expand the user‘s reading interest. In[20] paper, they provide high quality news recommendation to anonymous user. When a reader starts reading news article, the user model gets updated in the background or during reading based on present browsing behavior. On this basis, it gives better recommendation.

B. Hybrid Method

In [3],[18],[19] paper, proposes personalized news recommendation, a hybrid recommendation model encompassing content- based (i.e., CB) and collaborative filtering (CF, for short) algorithms that leverages multi-dimensional domain-specific features resided in news to make recommendation. To the best of our knowledge, this is the first work to exploit simultaneously domain- specific and general news features for news recommendation.

Proposes a new CB algorithm, utilizing trends feature (domain- specific feature from news) which means that different news categories have different lifecycle and play different roles in acquiring user profiles, and popularity effect (general feature) to improve news recommendation accuracy. Advantages are: The benefits of leveraging domain-specific features from news domain to improve news recommender systems.

The hybrid strategy using deviation does better in terms of accuracy and stability than both individual methods and benchmarked hybrid strategies. The improved CF recommendations called FereBSP and FereRBML perform better than corresponding benchmarks. Disadvantages are: Limits the applicability of the proposed method to publishers who manage to track their readers over longer periods of times. A hybrid recommender based on both, user content and collaborative filtering Wesomender framework. Wesomender [7] is comprised of two main components, a collaborative-filtering component and a content-based component. Each component evaluates the news the user has not seen or rated yet, and produces independent recommendations. Advantages are: To generate context-aware recommendations in the journalism field. A context-aware adaptive recommendation engine can fulfill the needs of journalists‘ daily work when retrieving timely. Disadvantages are: The workload of the other component can need to reduce by using heuristics like eliminating very old news or those where the action occurred too far away to affect the user.

III. OPEN ISSES

As we have already discussed about the existing news recommendation systems content-based recommendation, recommendation based on collaborative filtering, and hybrid recommendation. These systems have some open issues which are as follows

1) Content-based recommendation: the recommendation system tries to find news with similar content to the news that the user has read. Content-based recommendation systems are usually easy to implement. However, in some scenarios, the profile of the user with a bag of words is not sufficient to accurately capture the user's preferences.
2) Recommendation based on Collaborative filtering: the system recommends news by users' news ratings, and generally they are content free. Many users do not have sufficient historical behaviors, or the number of users in the system is not high enough, or sometimes we don‘t have sufficient user ratings, which is known as a cold-start problem[17].
3) In many news recommendation systems, the user profiles are one-sided, and user modeling from a single perspective cannot reflect the real preferences of users.
4) There is not yet a way to assess the degree of users' preferences for historical news. In reality, users' preferences for news are quite different. Thus, treating these historical records equally to analyze a user's preferences is not reasonable.
5) While building a short-term profile, most research studies abandon the relatively early browsing records, or use only a few recent browsing records. This may cause many contingencies and an incorrect understanding of the user's preferences, or the recommendation results will be too similar to what the user just read.

6) Most common and important challenge is Recency[16], it is proposed a proactive news recommender for mobile devices. Many readers want to read newly published news articles instead of previous dated news. So the mobile user has dynamic nature and becomes a challenge to deliver most proper as well as recent information.

Finally, due to rapid growth of topics and preferences news recommendation becomes more challenging.

IV. PROPOSED METHODOLOGY

This paper proposes a new news recommendation system that extends the user profile to three stages. And we proposed a novel method called BAP to build the user profile.

A user‘s personal interest can be modeled using or tracking his/her past history records which can be, past preferences, browsing history and tweeting history etc. The method gives each historical news a corresponding weight based on user's reading *B*ehavior *A*nd the *P*opularity of news, instead of 0, 1, or some fixed value.

Furthermore, when dealing with short-term profiles, we propose a time function to adjust the user's preferences for all historical news rather than some of it. This helps us construct a more objective and comprehensive short-term profile of the user.

Preferences of user can be changed in very short time. On an average, reader requires very less time to read any news article or content.

For ex, if elections are near reader is more interested to read articles about topic election. If world- cup is coming, similarly same thing happens with topic sports, either cricket or football. Now a day, one more Coronavirus Pandemic is going on, millions of news data or tweets generated across the globe.

So, considering all these examples, time parameter plays an important role over here. We are using system‘s current time to track short- term preferences, remaining all are long-term preferences.

The system consists of four main components: news collection and processing, user profiling method, personalized news recommendation and location-aware personalized news recommendation.

A. *Recommendation Framework*

1) *News Collection and Processing:* There are three phase profile model for news. First phase, crawl the news from the Twitter API and extract keywords from news using vector space model. Second phase, extract topics by using NLP model with the help of LDA algorithm. Third phase, to identify the entities.
2) *User Profiling:* The profile of user is composed of four stages with help of user‘s browsing or reading history. The first stage represents some of the news keywords in which the user is interested. The second stage represents the topic distribution of the user's preferences. The third stage represents the named entities in which the user is interested. And fourth stage represents the users' news preferences usually includes with the change of their locations. In this paper, the user's reading behavior is divided into several types by analysis. User‘s reading behaviors, popularity and location wise weight of the each news can be calculated.
3) *Dynamic Personalized News Recommendation using BAP:* We recommend news by using the long-term and short-term preferences of users. A time-sensitive function is proposed to construct the short-term profile by adjusting the long-term profile of the user. After calculate the similarities between every news and profile of the user. And finally, the rank results of the long-term and short-term preferences, the news recommendation of the long-term and short-term profiles are retrieved.
4) *Location-Aware Personalized News Recommendation With Explicit Semantic Analysis[LP-ESA]:* A hybrid news recommendation method, called LP-ESA, which takes into account both the user's current location and also his/her personal interests. User is more interested to read the news article happing nearest to them. LP-ESA collects a set of geo-tagged documents for a given location, and estimates a local topic distribution using *given location's* geographical topics. LP-ESA models a localized user profile for each user at location according to both general user profile and local topic distribution. Finally the news recommendation provides to the user.

Advantages are

a) The profiles of users can be constructed more accurately with BAP method.
b) This system provides a high quality of dynamic recommendation results with the short-term and long-term preferences of users.

LP-ESA significantly outperforms the state-of-the-art topic-based location-aware news recommendation methods.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue VIII Aug 2020- Available at www.ijraset.com

B. Architecture

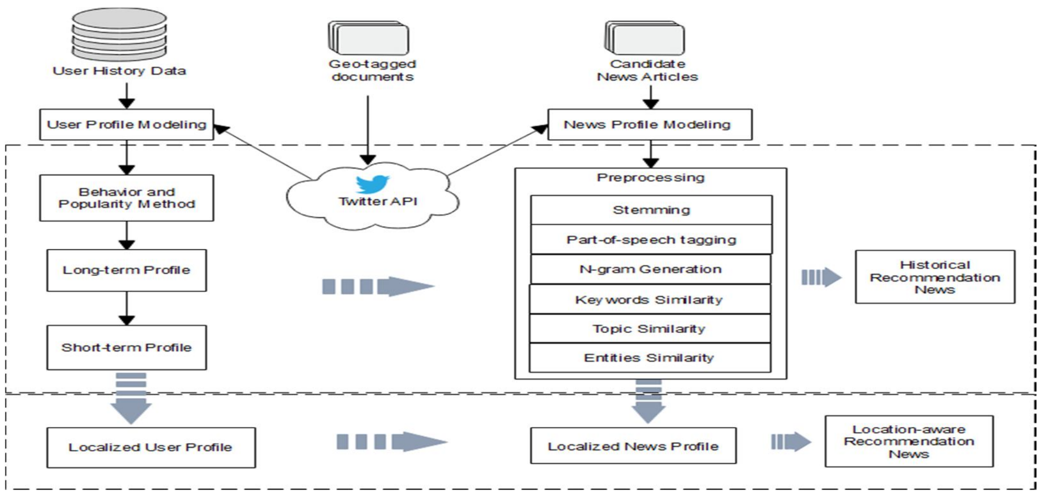

Fig.2 Proposed System Architecture

C. Mathematical Model

1) *User Profile Construction:* Construct the profile of the user by three different but related stages: News keywords, News named entities, and News topic distributions. Each profile of the user can be expressed as $V_u =< \langle F_u, G_u, E_u \rangle$ which corresponds to the three-stage model of news. However, there are some differences between this and the profile of news, where

a) $F_u = \{\langle f_1, w_1, t_1 \rangle, \langle f_2, w_2, t_2 \rangle, \ldots.\}$, F_u represents the keywords collected from the historical news that the user accessed in the past, and each entry consists of a representative word, the corresponding weight, and the last time the user accessed it;

b) $E_u = \{\langle e_1, w_1, t_1 \rangle, \langle e_2, w_2, t_2 \rangle, \ldots.\}$, E_u represents the named entities collected from the historical news that the user accessed in the past, and each entry consists of a representative named entity, the corresponding weight, and the last time the user accessed it; and

c) $G_u = \{\langle g_1, w_1, t_1 \rangle, \langle g_2, w_2, t_2 \rangle, \ldots.\}$, G_u represents the topic distribution collected from the historical news that the user accessed in the past, and each entry consists of a representative topic id, the corresponding weight, and the last time the user accessed it.

2) *News Profile Construction:* As we have seen in user profile construction, news profile model has the same stages. Additionally it has some more features like N-gram generation and POS Tagging. Then, we build a vector space model with keywords of news by the TF-IDF algorithm. A multistage profile of each piece of news can be expressed as Vn=⟨Fn,Gn,En⟩

a) Fn=⟨f1,w1⟩,⟨f2,w2,⟩,….,
where each keyword vector Fn consists of a representative news keyword and corresponding weight.

b) Many news readers prefer to read news about their favorite personality.
En=⟨e1,w1⟩,⟨e2,w2⟩,…., where each named entity vector En consists of a representative named entity word and corresponding weight.

c) Gn=⟨g1,w1⟩,⟨g2,w2⟩,….,where each entry consists of a representative news keyword and corresponding weight. Readers are more interested in reading news about similar topics.

d) N-gram is mainly used in natural language processing as well as mining of text related data for text summarization and breaking the word into different N-grams. Let‘s create an equation P(w|h), the probability of word *w*, with some given history, *h*. it adds only a small amount of improvement in recommendation engine.

e) Pos tagging- as we have limited words in any tweet, so understanding the semantic similarity between the words is very essential. Not only to extract relationship but also build knowledge graph POS tags are used. A very common example we can observe from our schooling days, ―Give me your answer‖ here answer in noun. And ―Answer the question‖ here the answer is verb. According to above sentences, a single word has two different POS tag in two different sentences. Hence, understanding the basic meaning of any sentence is very important.

3) User Profiling Method

$$V_u = \sum_{V_{n_j} \in N_r} \beta_j \cdot V_{n_j} - \sum_{V_{n_k} \in N_{nr}} \gamma_k \cdot V_{n_k}$$

In this paper, the method is adjusted to construct the profile of the user, as follows:

(1)

where V_{n_j} represents the profile of news n_j . N_r and N_{nr} represent the collection of news that the user has read and has not read, respectively. β_j represents the weight of the news n_j for user u. γ_k represents the weight of how the user *u* is uninterested in the news n_k.

4) Preference Calculation

The preference value of reading behavior of type *x* is

$$i_x = \frac{\sum_{i=1}^{a} \sum_{l=1}^{3} \left(\frac{|r_{il} \cap r_x|}{|r_{il}|} . w_l\right)}{a} \quad (2)$$

where r_{il} represents the reading record of the news with preference degree *l* marked by user u_i, r_x represents the reading record in type *x*, and *a* represents the count of users selected.

i_x needs to be fixed in order to eliminate the effect of different numbers of r_x.

$$i'_x = \left(\frac{i_x}{|r_x|}\right) . d$$

(3)

where i'_x represents the adjusted i_x, and *d* determines the range of i'_x.

Finally, the preference value based on the user's behavior is given below, where $p_x(u, n)$ represents the probability that the user's reading behavior belongs to the reading behavior type *x*, and $Behavior(u, n)$ represents the preference value of the user *u* for news *n*.

$$Behavior(u, n) = \sum_{x=1}^{m} (i'_x . p_x(u, n))$$

(4)

5) News Popularity Calculation

Proposes a new method to calculate the popularity of news.

$$Popularity(n) = \left(\frac{\left(\frac{|N_T|}{|S_T(n)|}\right) - min\left(\frac{|N_T|}{|S_T(n)|}\right)}{max\left(\frac{|N_T|}{|S_T(n)|}\right) - min\left(\frac{|N_T|}{|S_T(n)|}\right)}\right)^{\omega}$$

(5)

where $S_T(n)$ represents the collection of news that is similar to news *n* over the threshold in the time range *T*, and N_T represents the collection of all news in the time range *T*. $min\left(\frac{|N_T|}{|S_T(n)|}\right)$ represents the minimum value that $\frac{|N_T|}{|S_T(n)|}$ can be in the dataset, and $max\left(\frac{|N_T|}{|S_T(n)|}\right)$ represents the maximum value that $\frac{|N_T|}{|S_T(n)|}$ can be in the dataset. The parameter ω determines the change speed of popularity as the change of $\frac{|N_T|}{|S_T(n)|}$.

a) News Weight Calculation

The influence of user behavior and news popularity on the user's preference for the news, we calculate the weight of news *n* for user *u* as follows:

$$\beta = Behavior(u, n). Popularity(n)$$

(6)

D. *Algorithmic description*

1) *Sentiment Analysis using Sentiwordnet Dictionary Algorithm 1 Sentiwordnet Dictionary*

1: polarizedTokensList ← newList() 2: **while** tokenizedTicket.hasNext() **do**
3: token←tokenizedTicket.next()
4: lemma←token.lemma
5: polarityScore←null
6: **if** DomainDictionary.contains(lemma,pos) **then** 7: **if** SentiWordNet.contains(lemma,pos) **and**
8: SentiWorNet.getPolarity(lemma,pos) != 0) **then**
9: polarityScore ← SentiWordNet. Get
Polarity(lemma, pos)
10: **else**
11: domainDicToken←DomainDictionary.get Token(lemma, pos)
12: **if** domainDicToken.PolarityOrientation == ‖POSITIVE‖ **then**
13: polarityScore ← DefaultPolarity.positive 14: **else**
15: polarityScore ← DefaultPolarity.negative 16: **end if**
17: **end if**
18: polarizedTokensList.add(token, polarityScore)
19: **end if**
20: **end while**
21: **return** polarizedTokensList

2) *LDA Algorithm:* First and foremost, LDA provides a generative model that describes how the documents in a dataset were created. In python, we use _gensim' library for LDA model. In this context, a dataset is a collection of D documents. Document is a collection of words. So our generative model describes how each document obtains its words. Initially, let's assume we know K topic distributions for our dataset, meaning K multinomial containing V elements each, where V is the number of terms in our corpus. Let βi represent the multinomial for the ith topic, where the size of βi is V: |βi|=V. Given these distributions, the LDA generative process is as follows:

Algorithm 2 LDA generative process
1: For each document:
2: (a) Randomly choose a distribution over topics (a multinomial of length K)
3: (b) for each word in the document:
4: (i) Probabilistically draw one of the K topics from the distribution over topics obtained in (a), say topic βj
5: (ii) Probabilistically draw one of the V words from βj.

3) *Probabilistic Latent Semantic Indexing (PLSI) Algorithm 3 PLSI*

Algorithm 3 PLSI
1: **Input**: Matrix $U, X = S^{-1}$, index I_θ, query Q and parameter k;
2: **Output**: Top-r most similar sorted documents;
3: **Process:**
4: Initialize <C,S> by setting C and S as Ø;
$\bar{Q} \leftarrow XQ$;
4: **for** $t_j \in \{t_j | \bar{Q}(t_j) \neq 0\}$ **do**
5: **for** $< d_i, \mathrm{PartialSim}_\theta(d_i, t_j) > \in I_\theta(t_j)$ 6: **do**
obtain $\mathrm{PartialSim}_\theta(d_i, t_j)$ from $< d_i, \mathrm{PartialSim}_\theta(d_i, t_j) >$;
7: **if** $d_i \in C$ **then**
$$S(d_i) \leftarrow S(d_i) + \frac{\bar{Q}(j)\mathrm{PartialSim}_\theta(d_i, t_j)}{\sqrt{\Sigma_j(\bar{Q}(j))^2}};$$
8: **else**
$$S(d_i) \leftarrow \frac{\bar{Q}(j)\mathrm{PartialSim}_\theta(d_i, t_j)}{\sqrt{\Sigma_j(\bar{Q}(j))^2}};$$
$C \leftarrow C \cup \{d_i\}$;
$S \leftarrow S \cup \{S(d_i)\}$;
9: **end if**
10: **end for**
11: **end for**
return GetSortedCenter (k, < C, S >);

V. RESULTS AND DISCUSSIONS

Experiments are done by a personal computer with a configuration: Intel (R) Core (TM) i3-2120 CPU @ 3.30GHz, 4GB memory, Windows 7, MySQL 5.1 backend database and jdk 1.8. The application is web application used tool for design code in Eclipse and execute on Tomcat server. The real time news posts collection for dataset of this application using Twitter API with the help of Twitter4j-core and Twitter4j-stream jars. Some functions used in the algorithm are provided by list of jars like standfordcore-nlp jar for POS tagging etc.

Proposed work is expected to implement news posts recommendation system which collects input dataset of list of news posts from Twitter API. Keywords extracts using LDA algorithm. Topics extract using POS tagging method. And, finally entities extract using Named Entity Recognition (NER) method. Apply all the keywords extraction, topics extraction and entities extraction methods for news posts recommendation provide to user side. Expected outcome of this project is providing news recommendations to users with the help of Behavior-and-Popularity (BAP) method. Also providing news recommendations to the users with the help of collaborative filtering method using user‘s personal interests as well as their GPS tracking (geographical contexts) locations. Fig. 2 represents performance analysis of proposed method with some existing systems. we test the recommendation result with a no-weight calculated method using only interest ,only popularity, only BAP and the combined BAP+GPS+interest method. At that point, the mean values of the Precision, Recall, F1-score and Accuracy is calculated to assess the recommendation outcomes. The respective results are appeared in Fig. 2 and shows performance on combination of features for combination of BAP+GPS+Interest proposed system effectively gives recommendations to the users as compared to existing methods.

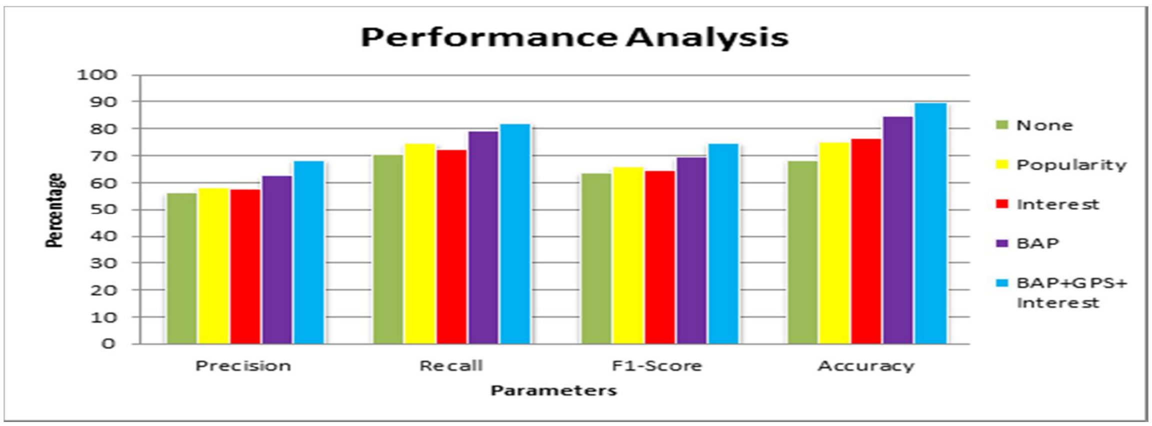

Fig. 3 Performance analysis of existing and proposed news recommendation system

The experimental results are appeared in Fig.3 and Fig.4. As per the figures, it very well may be seen that the algorithm execution is superlative when we set ω=0.33 and threshold=0.1. in this way, we can set above respective values for following studies.

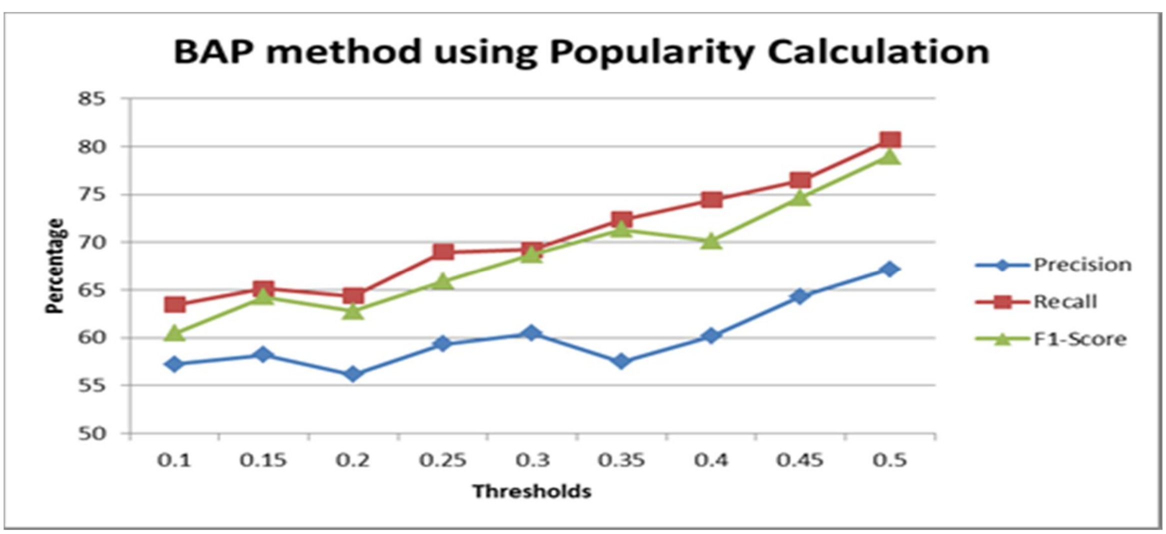

Fig. 4 Comparison of precision, recall, and F1-score of different thresholds of similar news in popularity calculation.

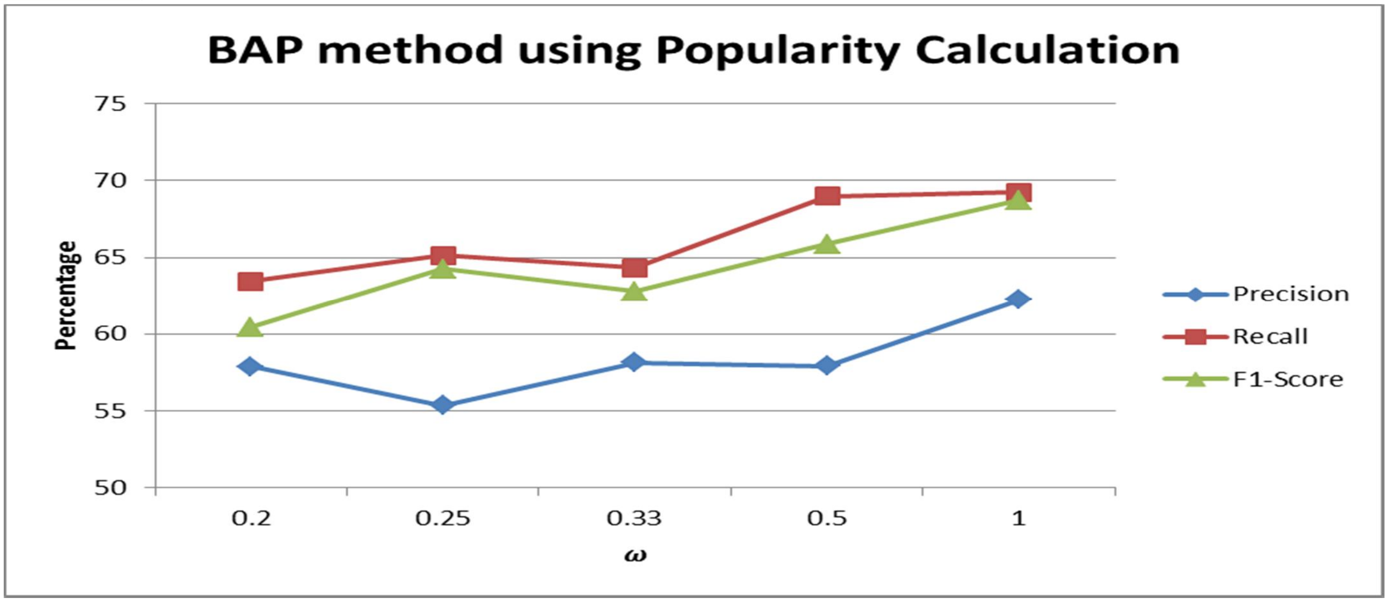

Fig. 5 Comparison of precision, recall, and F-score of different parameters ω in popularity calculation.

VI. CONCLUSION

In this paper, the BAP based user profiling method is proposed to calculate the user's preference for the news based on user behavior and news popularity. The profiles of users can be constructed more accurately with this method. This system can provide a high quality of dynamic recommendation results with the short-term and long-term preferences of users. The contribution work is a location-aware personalized news recommendation with explicit semantic analysis (LP-ESA) incorporating collaborative filtering recommendation system. Which considers both the users' location information and personal interests for news recommendation and use of collaborative filtering will get the better outcome of existing recommendation system. Experimentally, performance analysis of both BAP method and LP-ESA news recommendation system performs well. For future work, to increase heterogeneity of the recommendation outcomes and elaborate more preferences of users and also improve the quality of user experience.

VII. ACKNOWLEDGMENT

Express my true sense of gratitude, sincere and sincere gratitude to my guide to the project Prof. Pathan N. J. Sir for his precious collaboration and guidance that he gave me during my research, to inspire me and provide me with all the laboratory facilities, this it allowed me to carry out this research work in a very simple and practical way. I would also like to express my thanks and thanks to our HOD S. R. Tandale sir and all my faculty members as well as friends who, knowingly or unknowingly, helped me during my hard work.

REFERENCES

[1] C. Chen, X. Meng, Z. Xu, and T. Lukasiewicz, ``Location-aware personalized news recommendation with deep semantic analysis,'' IEEE Access, vol. 5, pp. 1624_1638, 2017.

[2] K. G. Saranya and G.S. Sadasivam, ``Personalized news article recommendation with novelty using collaborative filtering based rough set theory,'' Mobile Netw. Appl., vol. 22, no. 4, pp. 719_729, 2017.

[3] P. Lv, X. Meng, and Y. Zhang, ``FeRe: Exploiting influence of multidimensional features resided in news domain for recommendation,'' Inf. Process. Manage., vol. 53, no. 5, pp. 1215_1241, 2017.

[4] B. Fortuna, P. Moore, and M. Grobelnik, ``Interpreting news recommendation models,'' in Proc. 24th Int. Conf. WorldWideWeb (WWWCompanion), 2015, pp. 891_892.

[5] L. Li, L. Zheng, F. Yang, and T. Li, ``Modeling and broadening temporal user interest in personalized news recommendation,'' Expert Syst. Appl., vol. 41, no. 7, pp. 3168_3177, 2014.

[6] L. Zheng, L. Li, W. Hong, and T. Li, ``PENETRATE: Personalized news recommendation using ensemble hierarchical clustering,'' Expert Syst. Appl., vol. 40, no. 6, pp. 2127_2136, 2013.

[7] Montes-García, J. M. Álvarez-Rodríguez, J. E. Labra-Gayo, and M. Martínez-Merino, ``Towards a journalist-based news recommendation system: The Wesomender approach,'' Expert Syst. Appl., vol. 40, no. 17, pp. 6735_6741, 2013.

[8] M. Lu, Z. Qin, Y. Cao, Z. Liu, and M.Wang, ``Scalable news recommendation using multi-dimensional similarity and Jaccard_Kmeans clustering,'' J. Syst. Softw., vol. 95, pp. 242_251, Sep. 2014.

[9] H. Wen, L. Fang, and L. Guan, ``A hybrid approach for personalized recommendation of news on the Web,'' Expert Syst. Appl., vol. 39, no. 5, pp. 5806_5814, 2012.
[10] M. Capelle, F. Frasincar, M. Moerland, and F. Hogenboom, ``Semanticsbased news recommendation,'' in Proc. 2nd Int. Conf. Web Intell., Mining Semantics, 2012, pp. 27_36.
[11] L. Li, D.Wang, T. Li, D. Knox, and B. Padmanabhan, ``SCENE: A scalable two-stage personalized news recommendation system,'' in Proc. 34th Int. ACM SIGIR Conf. Res. Develop. Inf. (SIGIR), 2011, pp. 125_134.
[12] Y. Xie, L. Chen, K. Jia, L. Ji, and J. Wu, ``iNewsBox: Modeling and exploiting implicit feedback for building personalized news radio,'' in Proc. 22nd ACM Int. Conf. Inf. Knowl. Manage., Nov. 2013, pp. 2485_2488.
[13] J. Bao and M. F. Mokbel, ``GeoRank: An ef_cient location-aware news feed ranking system,'' in Proc. 21st ACM SIGSPATIAL Int. Conf. Adv. Geogr. Inf. Syst., Nov. 2013, pp. 184_193.
[14] J.-W. Son, A.-Y. Kim, and S.-B. Park, ``A location-based news article recommendation with explicit localized semantic analysis,'' in Proc. 36th Int. ACM SIGIR Conf. Res. Develop. Inf. Retr., Aug. 2013, pp. 293_302.
[15] Y. Zhou and J. Luo, ``Geo-location inference on news articles via multimodal pLSA,'' in Proc. 20th ACM Int. Conf. Multimedia, Nov. 2012, pp. 741_744.
[16] K. F. Yeung, Y. Yang, ―A Proactive Personalized Mobile News Recommendation System,‖ In Developments in E-system engineering ,pp.207-212 ,2010
[17] J. Liu, P. Dolan, and E. R.Pedersen ―Personalized News Recommendation Based on Click Behavior,‖ In IUI 10 proceedings of the 15th international conference on intelligent user interface,pages 31-40,(2010).
[18] B. Fortuna, C. Fortuna, and Dunja Mladenić(2010), ―Real-Time News Recommender System,‖ In 2010 European conference on Machine Learning and Knowledge Discovery in Databases, pp 583-586(2010).
[19] Chen Lin, R. Xie , X. Guan, Lei Li and, T. Li, ―Personalized news recommendation via implicit social experts,‖ In proceedings of the 21th ACM international conference on information and knowledge management, pages 1607-1611, 2014, ACM.
[20] F. Garcin, C. Dimitrakakis and B. Faltings, ―Personalized News Recommendation with Context Trees,‖ In Proceedings of the 7th ACM conference on Recommender systems (2013), pp. 105—112,2013, ACM.