Mobile Personalized Recommendation of Trends for Social Networks

Gajanan Vinayak Bhole [1], Tejashwa Khare[2], Ashutosh Srivastava[3], Omkar Kadam[4]

[1]Assistant Professor, Dept. of information technology, Bharati Vidyapeeth University, Pune, Maharashtra, India
[2]UG student, Dept. of information technology, Bharati Vidyapeeth University, Pune, Maharashtra, India

---***---

Abstract - *Number of users and their information feed is increasing day by day on social networks. The users communicate through social networks, share and distribute information by sending short text messages in near real-time. As a results of the social networks, the users are often experiencing information overload as they interact with many users and stick with it reading contents on large scale. Recommendation systems are proposed to assist users accommodate information overload by prioritizing the things of user's interest. The user's preferences are s haped by personal interests. At the identical time, users are full of their surroundings, as determined by their geographically located communities. One among the approach takes into account both personal interests and native communities. These community preferences are generally reflected within the localized trending topics. The proposed dynamic recommendation system provides better customized content described with the most important tweets on social media like Twitter per his/her individual interests considering the placement diffusion. Hence the effect of change within the geographical community preference on individual user's interests is observed through this system which provides top trending tweets supported the identical*

1. INTRODUCTION

Social networks having large growth recently in number of users and their shared information. The social networks have different challenges in providing differing types of information to the user. Twitter, Face-book may be a social networking applications that allow people to brief about a broad range of topics. Personalized recommendation often a promising solution for the knowledge overload problem in social network sites. Three recommendation problems on social network sites are explored, being recommending people, recommending information, and recommending conversation Social networks became a vital source for generating recommendations. Using social networks to understand the relations between users and their friends as well because the information obtained about them can improve the knowledge about users' behaviours. Also integrating recommendation systems into social networks can provide new observations and thus decisions that cannot be achieved through using traditional recommendation systems. Research studies have also found different properties of social networks encourage the combination of recommendation systems with social network. In this paper, the study is varied and address areas like network structure, trust, information credibility, event detection, social tagging, Geo fencing etc. The recommend- systems aim is returning items that are kind of like the users' demand. To provide the user with personalized recommendations for online social network information associated with personal and community level using locations of the user. The system may be show the quantity of users they will be login within the trend set location. This technique may be useful for new friend creation. Also during this system it's easy to identify favorite category for tweets for the aim of read or write the tweets. The users' preferences are shaped by personal interests. At the identical time, users are plagued by their surroundings, as determined by their geographically located communities. Ever since the dawn of civilization, people in general have always been an element of 1 tribe or another, brought together by their shared interests and a standard thanks to communicate the identical. Capturing user's interest, which change over the time, is vital nowadays. So, specializing in suggestions provided by the social media must be improved. The social networks suggest the recent trends to the users supported their location will reflect positively on their online experience. User can show message the correspond. Hence, it's important to mine user's interest from social network. Although tweets may contain valuable information, many don't seem to be interesting to the users. an oversized number of tweets can overwhelm users since they interact with many other users and that they should read ever increasing content volume on their timeline. Thus, the problem find the "matching" users and recommending content that are of interest to users became a key challenge for social networks sites. Recommendation systems are proposed to assist users address information overload by predicting the things that a user is also fascinated by. Recommender systems became a crucial research area since the looks of the primary papers on collaborative filtering within the mid-1990s . There has been much work done both within the industry and academia on developing new approaches to recommender systems over the last decade. The interest during this area still remains high because it constitutes a problem-rich research area and since of the abundance of practical applications that help user to handle information, overload and supply personalized recommendations, content, and services to them. However, despite all of those advances, the present generation of recommender further improvements to form recommendation methods more practical and applicable to a good broader range of real-life applications. A way is proposed to spot tweets that will be of potential interest to the user. Since the user's interests in numerous topics change over time, we specialize in studying this alteration, and recommending to every user the foremost interesting tweets on the user's timeline at specific time. The users' preferences are shaped by personal interests. At the identical time, users are laid low with their surroundings, as determined by their geographically located communities. Ever since the dawn of civilization, people in general have always been a component of 1 tribe or another, brought together by their shared interests and a typical thanks to communicate the identical. Capturing user's interest, which change over the time, is vital nowadays. So, specializing in suggestions provided by the social media must be improved. The social media can suggest the trending topics to the users supported their location will reflect positively on their online experience. User can show message that correspond to dynamic interest. Hence, it's important to mine user's interest from social network.

2. REVIEW OF LITERATURE:

Study of knowledge propagation in social networks together with recent trends is vital to be considered on the premise of localities.

2.1 Recommendation Systems:

The personalized recommendation systems recommend through a ranking procedure, useful content to the users using collaborative filtering method to come up with personalized recommendations in Twitter. In paper, a model relies on different topics and therefore the history of the user activities in each topic. The limitation is rigid recommendation methods are proposed.

2.2 Topic Modelling:

Topic Modelling is vital research field within the area of text mining and statistical modelling. Topic models are wont to solve the matter of "information overload" in text and corpuses. As an example, latent Dirichlet allocation (LDA), used successfully. In paper, a numerous content-based, collaborative, knowledge and data engineering, and hybrid methods were proposed. Capabilities are improved and recommender systems are made applicable to big selection of applications. Limitation is Utilization of multi-criteria ratings isn't used.

2.3 Information and Influence Propagation in Social Networks:

The information propagation consists of study of the message's propagation and therefore the decrease time since the posting of the message. The second is that the level of interactions on geographic, demographic levels and their contextual features.

In paper, domain-specific features are studied with hemp of the user's profile and text. Predefined set of generic classes like News and document messages are made more correctly.

2.4 Trends in Social Networks:

Trends (words and phrases) appearing on the timeline of social networking site are identified per hour, day, and week. The try to analyse the relation between trends and locations isn't satisfactory.

In paper, tweets, and networks of their social graphs for Twitter is shown. Advantage is to demonstrate the potential for effective and efficient recommendation. Limitation of that noisy content are removed. In paper, chronological order is followed for tweets and users view the followers' timelines to seek out their interests. A collaborative ranking to capture personal interests is proposed and is employed to decrease the user effort.

2.5 Collaborative personalized tweet recommendation:

Twitter has rapidly grown to a well-liked social network in recent years and provides an oversized number of real-time messages for users. Tweets are presented in chronological order and users scan the followers' timelines to search out what they're inquisitive about. However, an information overload problem has troubled many users, especially those with many followers and thousands of tweets arriving daily. During this paper, we specialize in recommending useful tweets that users are curious about personally to cut back the users' effort to search out useful information. Many forms of information on Twitter are available for helping recommendation, including the user's own tweet history, retweet history and social relations between users.

2.6 Experiments on recommending content from information streams:

More and more web users sustain with newest information through information streams like the popular microblogging website Twitter. During this paper we studied content recommendation on Twitter to raised direct user attention. In an exceedingly modular approach, one explored three separate dimensions in designing such a recommender: content sources, topic interest models for users, and social voting.

3. ANALYSING USER MODELING ON TWITTER FOR PERSONALIZED NEWS RECOMMENDATIONS:

How can micro-blogging activities on Twitter be leveraged for user modeling and personalization? During this paper we investigate this question and introduce a framework for user modeling on Twitter which enriches the semantics of Twitter messages (tweets) and identifies topics and entities (e.g., persons, events, products) mentioned in tweets. Following are some advantages of various methods used:

3.1 Media Connectivity:

Easier or faster thanks to connect is media like Facebook, Twitter, LinkedIn etc. and interests are shared in several contexts.

This helps to:
1. Get new opportunities.
2. Locate assistance
3. Get product and repair referrals
4. Receive support from similar posting users
5. Share advice and data
7. Access news or posts

3.2 Similarity of Interest:

One can share, access, and acquire idea about similar interests and hence can keep updating the general information.

3.3 Information Sharing:

Social media provides way of communications and better improvements worldwide. The problem stated on develop system 'Trend Fusion', to boost user's interaction and skill in social networks through a dynamic personalized recommendation system that has with the foremost important news, events and actions supported location, using GPS like sensors. It analyses and predicts the localized diffusion of trends in social networks and recommends the foremost interesting trends to the user. Objectives are to boost user's interaction and knowledge in social networks to the user through a dynamic mobile personalized recommendation system.

Also, to investigate and predict the localized diffusion of reports, events, and actions in social networks, and recommend the foremost interesting trends to the user. And to update users by sensing the contexts (location and weather), for locating nearest events and actions taken by other users. The scope of the proposed system is restricted to the geographic region within the district (50km) to urge the notifications of nearby events and actions (activity and status) of around 1000 other users, supported their current location. Users can see the highest 10-15 recommended news on the timeline that's important supported previous inputs and interests.

4. SYSTEM ARCHITECTURE:

The system consists of 6 basic blocks as shown in fig 1. In system one can login with twitter. Once user logins the twitter account, the recent tweets may be shown. During this system user can write post or tweet. User can use Dynamic subject creation or LDA method for identify the favorite category supported user post. And fencing nearby Geo location. User can do categorization of post per different categories e.g., Social, Sport, Entertainment, NEWS etc.

These are shown in fig. below.

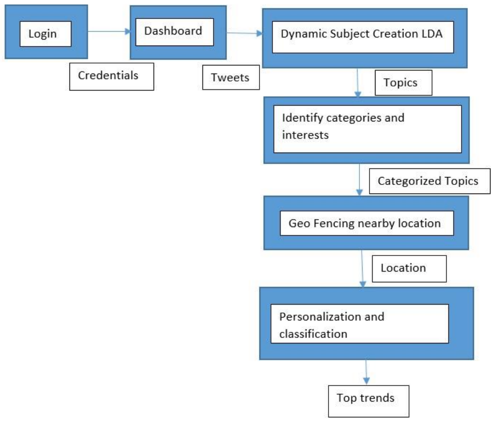

Fig. 1. Proposed System Architecture Latent Dirichlet

4.1 Allocation (LDA):

Latent Dirichlet allocation (LDA) is used widely for different micro blogs. Model involves a two-levels: LDA involves three levels.

T= {T1,T2.........Tn}, T- set of posts

L= {l1,l2,........ln}, L- set of topics

Li= {wi1,wi2,........win}, Li- set of probability values of words n for topic Li.

Activity in each day d for a topic is denoted by

Ad = {ad1,...,adk}, where ad[i] is the level of activity in topic Li for post Oe of the user for a day. Ad[i]=Sum(Oe[i])

A new tweet Oenew and the user's LoI in the tweet can be calculated as: sum (Oenew[i].sum(Ad[i]))

The cosine similarity over time for the user is given by: Ut.Pt/(MAG(Ut).MAG(Pt)) where, Ut is user vector and Pt is post vector. It is specified between vectors formed by the summation of the LoI in a topic and value of 1 means the exact distribution match, 0 means users have nothing in common.

4.2 Dynamic Subject Creation:

Important tweets are classified based on this. The model captures the user's interests over changing time, and the messages are shown based on dynamic interests. This is used to analysis the tweet from the raw tweets and give the important tweets as per the user's requirement.

System stages can be categorized as:

Group of trends from different locations: Trends should be identified from all the locations of user's interest. The trends are collected for each unit time.

Storing trending topics:

Trending topics are labelled by the location/time and they were received from/at.

4.2.1 Building cascades:

It is specified weather a trend is a beginning of a new cascade or a continuation of an old one. Hence it is important to categories those.

4.2.2 Extraction of Parameters:

The parameters are measured for each location based on the diffusion models. There are mainly 4 types:

1.Diffusion Parameters

2.Geographical Parameters

3.Historical Parameters

4.Trend Parameters

4.2.3 Model Learning:

Based on previous results and parameters set up model should be able to learn and give correct output based on different inputs and featured outputs.

Mathematical Model of a system can be stated as –

Let S = be the twitter set,

S= {T,F},

T= {T1,T2,T3,T4|Tgiventweetsfromuser},

F= {F1,F2,F3|Fgivenfeedbackonnews}

Functions: F: F = {F1(),F2(),F3(),F4(),F5(),F6(),F7()}, where F1 (S) = login with twitter (Once)

F2 (S) = Display recent tweets

F3 (S) = Dynamic subject creation

F4 (S) = Identify the favorite category based on user Post

F5 (S) = Categorization of post according to different Categories

F6 (S) = Get news feedback using geo fencing nearby location

Algorithm Building Cascades

Procedure: Build Cascades From Activations

Input: Activations List al begin

// An activation a is a record a = (trend, location, time)

Activations List algo = Order al by time for all Activation a in algo do

1. If a trend appeared in (a time – 24 hours) then

2. cas = last cascade of a trend

2.1If a location appeared in cas then

Add a time to instances of a location in cas

else if a time equals time of last step in cas then

Add a to last step of cas

3. else Add new step to cas containing a location end if

4. else Create new cascade cas Add new step to cas containing a location end if end for end.

The record with trend, location and time is need to be categorized as new or old cascade based on its time to fetch accurate recent trend as per the user's interest. Hence classification of trends is done using above algorithm.

5. Topic Model:

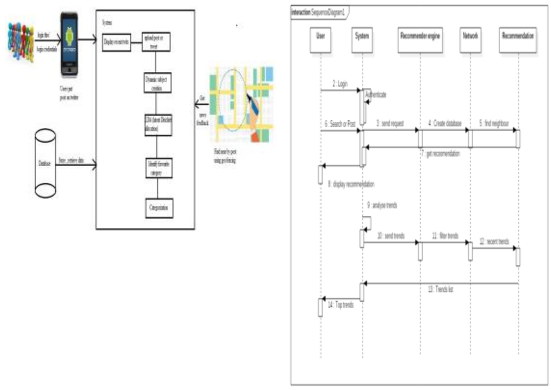

Fig 2. System flow.

Methodology LDA is preferred over Simple Dirichlet multinomial clustering model. A classical clustering model involves two-levels –Dirichlet is sampled once for a corpus. A multinomial clustering variable is selected once for each document in the corpus, and a set of words are selected for the document conditional on the cluster variable. A model restricts a document to being associated with a single topic. LDA involves three levels, and the topic node is sampled repeatedly within the document.

B. Work Flow of system Flow of the system is stated in Figure 2:

1. Number of keywords:

User will tweets on twitter. System will search key words. Based on keywords recommendation system is proposed.

The unsupervised nature of topic modeling methods makes can be evaluated based on perplexity as a quantitative method. Perplexity is a well-known standard metric used in information retrieval field. It tries to quantify the accuracy of a model by measuring how well the trained model deals with an unobserved test data. Perplexity is defined as a reciprocal geometric mean of per word likelihood of a test corpus. A lower perplexity indicates a better generalization performance.

2. Trained dataset: Trained dataset is to be prepared. Tweets related to keywords are stored in our dataset.

3. Geo Fencing: Location of users are tracked using this functionality. Nearby users who tweet post are detected. This is a

feature in a software program that uses the global positioning system (GPS) to define geographical boundaries.

C. Experimental Setup

Software requirements:

XML (Extended Markup Language) for designing layouts screen of android application.

Java language for connecting xml design part with database and to perform basic operation of database.

Google map API key to use Google map services.

Operating System: Windows 7 or above

Technology: Java, J2EE ,Android

Web Server: XAMMP server

Database: My SQL

Java Version: J2SDK 1.7 / 1.8, Android SDK 19 and above

Hardware Requirements:

Windows PC/Android Mobile (above version 2.3)

4. Functional Requirements:

Download the mobile application a user should be able to download the mobile application through any store or service and it should be free of cost.

Download and notify users of new releases User should get notification of a new version and updates should be made easily within the application.

User registration-Mobile application

After download user should be able to register to the application and should easily recover the credentials if required.

Mobile application- Search timeline

User should be able to get the recent news after refreshing his/her timeline after login. Location, interests, tags should get considered while showing the trends to the user on the timeline.

5. Recommended list to user

User should get top 10 trends as a list on his/ her timeline which are relevant to user based on the specified constraints.

6. Navigation though mobile application

A user should be able to navigate through the application. User should be able to post, search, update and delete the feeds.

6. System Analysis

After user twits on twitter, system will search keywords. Based on keywords recommendation system is proposed.

Topic models are evaluated based on perplexity. Perplexity gives the accuracy of a model on measuring unobserved test data. A lower perplexity indicates a better generalization performance. It is given by:

$$\text{Perplexity}(D_{\text{test}}|M) = e^{\frac{-\sum_{d \in D_{\text{test}}} \log P(w_d|M)}{\sum_{d \in D_{\text{test}}} N_d}}$$

Where, wd is the words in document d, M is the topic model, and Nd is the number of words in document d. Trained dataset is to be prepared. Twits related to keywords are stored in our dataset. Location of users are tracked using this functionality. Nearby users who tweet post are detected.

A new information cascade model, Snowball Cascade (SC) model is compared with General Threshold model (GT) with the help of recall and precision values based on three classifiers-logistic regression (LR), stochastic gradient descent (SGD) and random forest (RF).

The results shows that SC model performs better than GT model as shown in the below table.

	Snowball Cascade Model			General Threshold Model		
Classifiers	Logistic Regression	Stochastic Gradient Descent	Random Forest	Logistic Regression	Stochastic Gradient Descent	Random Forest
Average Recall	0.95	0.98	1	0.5	0.5	0.5
Average Precision	0.79	0.8	0.7	0.68	0.68	0.5
	Without Topics			Distance Only		
Average Recall	0.75	0.73	0.84	0.82	0.86	0.98
Average Precision	0.6	0.7	0.65	0.7	0.68	0.99

Fig. 4. Comparison of models

Implementation details:

A. Dataset

Twitter API

400k trends

Tuple form: (weird, trend0, trend1, . . . , trend9, and date/time),

where weird is Yahoo Where On Earth ID [50] trend0, . . . , trend9 are the top ten trends.

B. Results

After login user is able to see the options on dashboard. User can view the tweets uploaded by him or her on time line.

User can enter the topic of his or her choice and the location so that nearby tweets can be fetched. Based on the fetched tweets classification is done according to the relevance of users' personal interests. Positive and negative tweets can be shown using graph along with the location on the map. Based on the location and the topic nearby and relevant tweets are fetched and shown to user as per Figure 5.

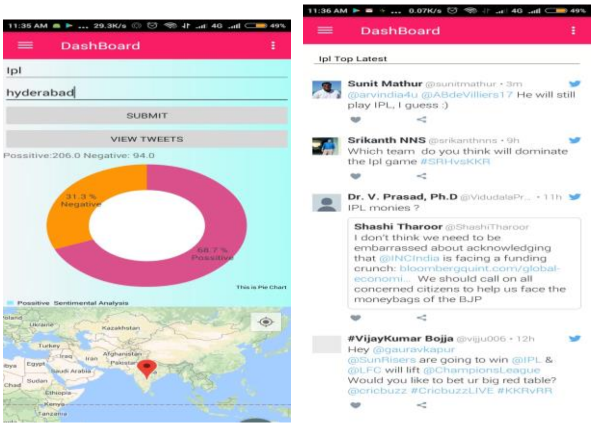

Fig 5. Results.

C. Performance evaluation parameters Top 10 trends:

Topic	IPL	Dance	Play	Food
Average precision	0.79	0.85	0.65	0.55
Average Recall	0.95	0.87	0.75	0.65
F measure	0.86	0.86	0.69	0.56
Perplexity	20	35	60	55
Location	Pune	Mum-bai	Kolkata	Bangalore

Table 1. Results.

In this way for different locations and topics based on their relevance with user, we get varied recall and precision values along with their perplexities.

Precision is the percentage true positives in the retrieved results. Recall is the percentage true negatives in the retrieved results.

Using the two values F measure is defined which is the harmonic mean of the two as:

$$F = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$

Lower perplexity denotes more accuracy and more value of recall boosts the performance of the application.

7. Conclusion:

The proposed solution categorizes trends using the author in-formation and features within the trends that provide better way of advice. Users hence are ready to filter feeds supported their interests and priorities. Dynamic LOI are beneficial for the user to acknowledge important feeds even with their short forms and in future if particular trend appearing in some city in any context together with its time. SC model used is more beneficial which is continuous and supply better results than other models.

8. Future Work

In research attempt, I have focused on the localized diffusion of trends. However, the database is growing tremendously via use of the Internet. Thus, future work would be to extend the perplexity and improved effective.

9. REFERENCES

[1] B. Sriram, D. Fuhry, E. Demir, H. Ferhatosmanoglu, and M. Demirbas, "Short text classification in Twitter to improve information filtering," in Proc. 33rd Int. ACM SIGIR Conf. Res. Develop. Inf. Retr., New York, NY, USA, 2010, pp. 841–842.
[2] Twitter Home Page, accessed on Jun. 23, 2017. [Online]. Available: http://twitter.com
[3]] D. Laniado and P. Mika, "Making sense of Twitter," in Proc. 9th Int. Semantic Web Conf. Semantic Web, vol. 1. Berlin, Germany, 2010, pp. 470–485.
[4] R. Mehrotra, S. Sanner, W. Buntine, and L. Xie, "Improving LDA topic models for micro-blogs via tweet pooling and automatic labeling," in Proc. 36th Int. ACM SIGIR Conf. Res. Develop. Inf. Retr., New York, NY, USA, 2013, pp. 889–892.
[5] Z. Ren, S. Liang, E. Meij, and M. de Rijke, "Personalized time aware tweets summarization," in Proc. 36th Int. ACM SIGIR Conf. Res. Develop. Inf. Retr., New York, NY, USA, 2013, pp. 513–522.
[6] Altingovde, I.S., Demir, E., Can, F., and Ulusoy, O. Site-based dynamic pruning for query processing in search engines. In Proc. SIGIR (Singapore, July 2008), 861-862.
[7] Banerjee, S., Ramanthan, K., and Gupta, A. Clustering short text using Wikipedia. In Proc. SIGIR (Amsterdam, the Netherlands, July 2007), 787-788.
[8] H. Baars and H.-G. Kemper, "Management support with structured and unstructured data—An integrated business intelligence framework," Inf. Syst. Manage., vol. 25, no. 2, pp. 132–148, Mar.
[9] 2008.
[10] C.L.Wang and Y.S.Chiou,"An adaptive Positioning Scheme Based on Radio Propagation Modeling for Indoor WLANs," 63rd IEEE Vehicular Technolgy Conference, Melbourme Australia, May 2006: 2676-2680.
[11] J. Martinez-Romo and L. Araujo, "Detecting malicious tweets in trending topics using a statistical analysis of language," Expert Syst. Appl., vol. 40, no. 8, pp. 2992–3000, 2013.
[12] J. Weng, E.-P. Lim, J. Jiang, and Q. He, "TwitterRank: Finding topicsensitive influential twitterers," in

Proc. 3rd ACM Int. Conf. Web Search Data Mining, New York, NY, USA, 2010, pp. 261–270.

[13] M. Steyvers, P. Smyth, M. Rosen-Zvi, and T. Griffiths, "Probabilistic author-topic models for information discovery," in Proc. 10th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, New York, NY, USA, 2004, pp. 306–315.

[14] W. X. Zhao et al., "Comparing Twitter and traditional media using topic models," in Proc. Eur. Conf. Inf. Retr. (ECIR), Berlin, Germany, 2011, pp. 338–349.