Study of Existing Design Architectures and Frameworks in Twitter

Kamaljit Kaur
Research Scholar
IKG Punjab Technical University
Kapurthala, Punjab, India
kamaljit_cse@sggswu.org

Dr. Kanwalvir Singh Dhindsa
Professor, Department of CSE & IT
Baba Banda Singh Bahadar Engineering College
Fatehgarh Sahib, Punjab, India
kdhindsa@gmail.com

***Abstract*—Millions of users are active on Twitter and they consume and disseminate messages resulting in collection of massive information. In order to provide user with personalized information service, the recommendation system is considered as an effective method that can alleviate the information overload problem. The different aspects that can be recommended in Twitter are explored in this paper along with different available commercial and noncommercial architectures and frameworks. The paper aims to provide a brief overview behind the design and compares them based on their followed recommendation approach.**

***Keywords*—Twitter, Tweet, Retweet, Follow, Trend Detection, Similar Users, Followee Recommendation, Hashtag Recommendation.**

I. INTRODUCTION

In Twitter users can post messages called tweets which are limited to 140 characters and provides a network which connects the users with similar interest. The different functions that user can perform in Twitter includes writing a new message called as Tweet, or publishing the existing posted message called as Retweet function , finding new users to Follow and including users in the tweet as Mention. These functions can be used to define the different recommendation tasks. The different components in Twitter where a recommender system can be used to recommend are described as follows:

1. Tweet Recommendation which recommends what should the user tweet about.
2. URL Recommendation that recommends URL that user can include in a new tweet.
3. Hash Tag Recommendation recommends hash tags that user can include in a new tweet.
4. Retweet Recommendation that recommends which tweet should the user retweet
5. Mention Recommendation which recommends who the user should mention in his tweet.
6. Followee Recommendation that recommends who the user should follow.

Based on these functions various frameworks have been designed and different architectures are proposed in Twitter.

II. EXISTING TWITTER FRAMEWORKS AND ARCHITECTURES

Twitter allows us to convey our message with the similar minded people. The established connections also open up many areas for recommendation. In this paper the different frameworks available in Twitter in recommending similar users and hashtags along with detection of trends are discussed.

TABLE I. DIFFERENT RECOMMENDATION AREAS IN TWITTER ALONG WITH DISCUSSED FRAMEWORKS AND ARCHITECTURES

Recommendation Area	*Description*	*Discussed Techniques*
Trend Detection	Keywords that suddenly appear at a very rate due to topics or popular events forms Trends. Detecting trends in real time is one of the challenging areas.	TwitterMonitor [12] TopicSketch [14] Twevent [21]
Similar User Recommendation	Users are connected on the social network. Such links which are in the network topology or beyond can be suggested to find users with similar interests.	Twittomender [2] TwitterRank [20]
Hashtag Recommendation	Due to limited character length in Twitter, additional features like hashtags are included which can be recommended to the user based on the content of the tweet.	TRECT [17] TweetSense [15]

III. TREND DETECTON

A trend is considered as frequently occurring keywords that suddenly appear in tweets at high rate. Large number of users is connected on the Twitter network and each of these active users publishes their information. Trend Detection is important in order to find which information is currently popular on Twitter. Trends are generated by the popular events and topics that is the point of communication in the Twitter network. Trend detection has its applications in reporting the news, marketing experts, opinion mining related companies.

TwitterMonitor [12] is system that is used to perform trend detection in real time on Twitter stream data and analyses them to identify description of each topic. The figure 1 shows the architecture of TwitterMonitor that works in three steps where trend is first detected and then grouped and analyzed in the last step [12]. QueueBurst algorithm is used to detect bursty keywords. GroupBurst algorithm is used to groups these bursty keywords into trends by checking their occurrences in the current tweets.

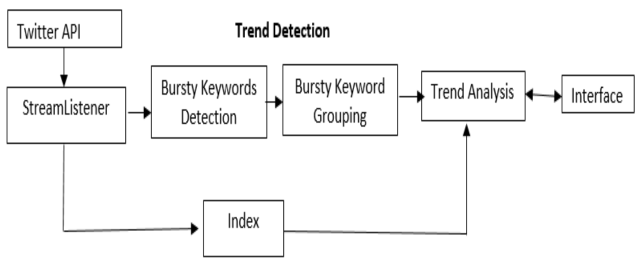

Fig. 1. TwitterMonitor Architecture [12]

After a trend is identified as set of bursty keywords, TwitterMonitor attempts to identify an accurate description associated with that trend.

TopicSketch [14] is another real time bursty topic detection method which works in two stages. Initially a data sketch indicating the tweet popularity is constructed based on tweets and word occurrence in it. In the second stage a topic model is developed to find the bursty topics from the data sketch.

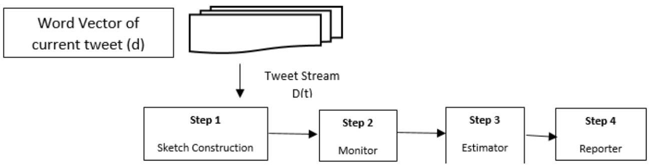

Fig. 2. TopicSketch Framework

Twevent [21] event detection architecture is composed of three segments where first the tweets are split into different segments by the process of tweet segmentation which are then stored in segment index. The bursty segments are detected by the event segment detection process and the similar event segments are grouped by the event segment clustering component.

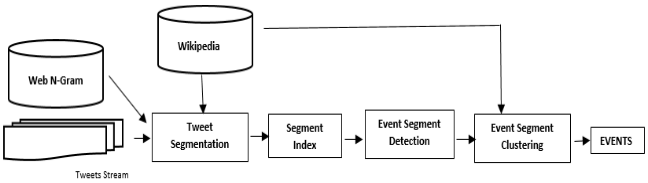

Fig. 3. Twevent Architecture [21]

Another method to detect trend and bursty keywords is based on the properties of Twitter stream data which works in three stages where initially the keywords are collected. Then the Keyword Merger module merges the acronyms, reduced keywords and typo and spacing keywords from which at last the different trends are detected.

Topic models have been used to capture events using clustering of terms along with updating the changes in terms occurring together than just focusing on the number of times the term is occurring (term frequency).Latent Dirichlet Allocation (LDA), Online LDA (OLDA) techniques have been used to learn the topics from the documents.

TABLE II. COMPARISON OF DIFFERENT TREND DETECTION FRAMEWORKS AND ARCHITECTURE

Framework	*Methodology*	*Technique*	*Approach*
TwitterMonitor	Identification of Trends Grouping of Trends Trend Analysis	Bursty keywords are detected using QueueBurst algorithm. Bursty keywords are grouped together by GroupBurst algorithm	Keyword-based approach
TopicSketch	Sketch-based topic model is used to find bursty topics.	Data sketch is proposed from tweet and from this sketch topic model is proposed	Topic Model Based approach
Based on Twitter stream characteristics	Keyword Collector Keyword Merger Trend Detection	Bursty keywords are based on the term frequency	Keyword-based approach
Based on Topic Models	Captures events by clustering of terms	Latent Dirichlet Allocation (LDA), Online LDA (OLDA)	Topic Model Based approach

IV. USER RECOMMENDATION

Finding similar users is another recommendation task in Twitter for which different frameworks have been proposed. Twittomender [2] and TwitterRank [20] have been discussed in the paper.

Twittomender framework is designed as a web service where the profile information of the user is retrieved in order to form a user document. The recent user tweets are used to model the user which is then used to examine the frequency of terms appearing in them. Twittomender uses seven recommendation strategies; these include the content and connections of a user. The content based method is based on users tweets, tweets of followees, tweets of followers, and combination of users and his followees and followers tweets, while the Collaborative based method uses the connections of Followee, Follower and using all the connections[2].

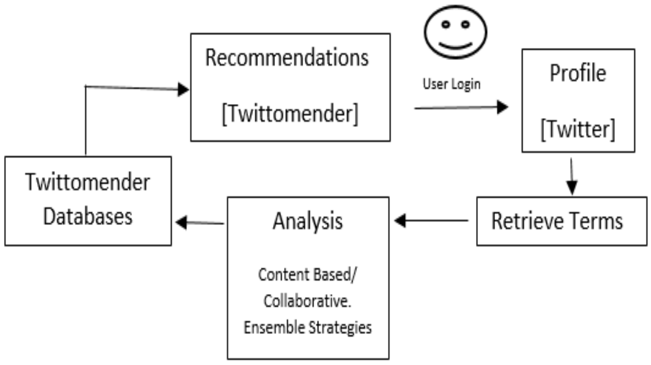

Fig. 4. Twittomender Control Flow Diagram [2]

TwitterRank is used to measure the influence of users in Twitter by considering both the user similarity based on the similar topics and the links they follow [20]. Earlier the influence of user was measured by his follower count or by incorporating the network link structure as in PageRank. The limitations of in-degree and PageRank are overcome in this paper by considering both these factors. The approach works by first identifying the topics user is interested in based on the tweets posted by him. Based on Latent Dirichlet Allocation technique that is used to identify the topic information then topic-specific relationship network is formed. From the formulated network the influence of a twitter user is considered. The influence is high if the sum of influence of followers is high.

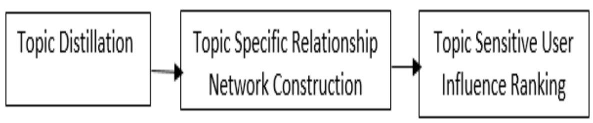

Fig. 5. TwitterRank Approach [20]

TABLE III. COMPARISON OF SIMILAR USER RECOMMENDATION FRAMEWORKS AND ARCHITECTURE

Framework	*Methodology*	*Technique*
Twittomender	User Profile is collected as a document and tf-idf mechanism frequency of term is calculated.	Based on Content and Collaborative Techniques
TwitterRank	User Influence is calculated by measuring topical similarity between users and the link structure	Latent Dirichlet Allocation

V. HASHTAG RECOMMENDATION

Hashtags can be another component that can be recommended based on the content of the tweet. TRECT and TweetSense are the two frameworks discussed in this approach.

TRECT provides an effective suggestion mechanism for hash tags from the pool of existing ones to the users of the same interest. It considers the impact of the frequency of hash tags appearing together in a Twitter message to generate the score to rank a hashtag to recommend it to be used with relevant ones[17]. Initially Hashtag Graph is constructed by using all the messages containing hash tags for a small time slice. Same set of hash tags can repeatedly appear in different messages including re-tweets and replies. This develops a multi-graph where each node stands for a number of occurrences of that specific pair of hash tags and corresponding edges represent each such occurrence. Using this system discovers the possibility of new "ties" based on already existing strong ones. This link prediction mechanism uses a fitness metric to measure the strength of the new tie. Weighted-Tag-Similarity metric is used to rank the recommendations. Network of hash tags is created, and the principle of triadic closure is applied on it. The output indicates that the different hash tags that are recommended according to their relevance which is calculated using weights.

TweetSense tries to sense out the context from the tweets. The tweets which do not have hashtags in them are called as orphan tweets. It is a hashtag rectification system that recommends the users with a list of hash tags. [15]

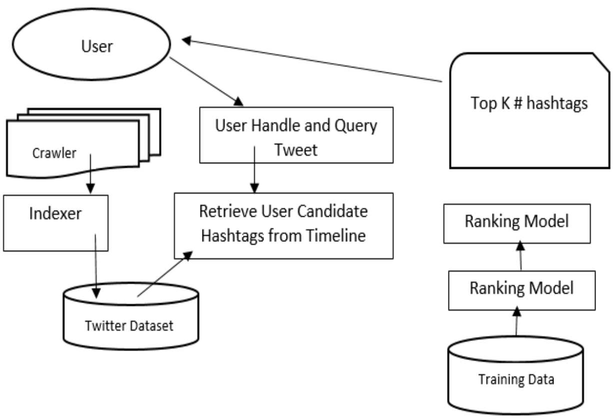

Fig. 6. TweetSense Architecture [15]

The system starts with choosing the right candidate set of hash tags from the user timeline followed by extraction of the tweet content features such as test similarity, recency of the tweet and popularity and the social signal of the users such as mutual friends, mutual followers, recent favorites, recent direct replies, and follower-following relationship. Then logistic regression model is used to union all the features that were extracted to recommend the most suitable hash tags. Based on their probabilities the most promising hash tags are recommended to the user.

TABLE IV. COMPARISON OF HASHTAG RECOMMENDATION FRAMEWORKS AND ARCHITECTURE

Framework	*Methodology*	*Technique/Algorithm*
TRECT	Hashtag Graph is constructed Link prediction mechanism is used	Hash tag network is created and using weighted tag similarity hash tags are recommended according to their relevance
TweetSense	Finds the context of the tweets and suggests lists of hashtags to orphaned tweets.	Logistic Regression Model Top K Hashtags are recommended

VI. CONCLUSION

Twitter presents a unique environment for data mining and natural language processing due to the limited message length, as well as the users and their intentions. This paper discusses the different research design frameworks in Twitter related to these different perspectives. The different areas covered in this paper are systems related to find trending topics and events are studied which identifies the emerging news in real time. Twittomender which recommend similar users and Twitterrank which identifies the influential users are discussed. Interesting news articles are recommended to users based on their profile and tweet messages.

Similar design frameworks can be extended, or new systems can be proposed to recommend the different twitter functions.

REFERENCES

[1] Java, X. Song, T. Finin, and B. Tseng, "Why We Twitter: Understanding Microblogging Usage and Communities," Proceedings of the 9th Web KDD and 1st SNA-KDD 2007 Workshop On Web Mining And Social Network Analysis. ACM, 2007.

[2] J. Hannon, M. Bennett, and B. Smyth, "Recommending Twitter Users to Follow Using Content and Collaborative Filtering Approaches," Proceedings of the Fourth ACM Conference on Recommender Systems, RecSys'10, New York, NY, USA, 2010, , pp. 199–206.

[3] J. Allan, R. Papka, and V. Lavrenko, "On-Line New Event Detection And Tracking," SIGIR, 1998, pp. 37–45.

[4] T. Brants, F. Chen, and A. Farahat, "A System For New Event Detection" SIGIR, 2003 pp. 330–337.

[5] J. Weng and B.Lee, " Event Detection In Twitter", Proceedings of ICWSM, 2011.

[6] C. Li, A. Sun, and A. Datta, " Twevent: segment-based event detection from tweets", Proceedings of the 21st ACM International Conference On Information And Knowledge Management,2012, pp. 155–164, 2012.

[7] Chen Chen, Hongzhi Yin, Junjie Yao, and Bin Cui,"Terec: A Temporal Recommender System Over Tweet Stream," Proceedings of the VLDB Endowment, 6(12), 2013, pp.1254–1257.

[8] Fréderic Godin, Viktor Slavkovikj, Wesley De Neve, Benjamin Schrauwen, and Rik Van de Walle, "Using Topic Models For Twitter Hashtag Recommendation," Proceedings of the 22nd International Conference On World Wide Web Companion, 2013, pp.593–596.

[9] Jianshu Weng, Ee-Peng Lim, Jing Jiang, and Qi He,"Twitterrank: Finding Topic Sensitive Influential Twitterers," Proceedings of the Third ACM International Conference On Web Search And Data Mining, ACM, 2010, pp 261–270.

[10] Kaisong Song, DalingWang, Shi Feng, Yifei Zhang,Wen Qu, and Ge Yu. Ctrof, "A Collaborative Tweet Ranking Framework For Online Personalized Recommendation,"Advances in Knowledge Discovery and Data Mining, Springer, 2014, pp.1–12.

[11] Sankaranarayanan, J., Samet, H., Teitler, B.E., Lieberman, M.D., and Sperling, J., "TwitterStand: News in Tweets," Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, pp. 42-51, 2009.

[12] Mathiousdakis M and Koudas N: "TwitterMonitor: Trend Detection over the Twitter Stream" SIGMOD'10, June 6–11, 2010, Indianapolis, Indiana, USA.

[13] Suleimenov A,"Twitter News: Harnessing Twitter To Build an Article Recommendation System", MS Thesis Princeton University.

[14] Xie W, Zhu F, Jiang J, Lim E and Wang K, "TopicSketch: Real-time Bursty Topic Detection from Twitter"

[15] Vijaykumar M,"TweetSense: Recommending Hashtags for Orphaned Tweets by Exploiting Social Signals in Twitter", MS Thesis, Arizona State University, 2014.

[16] Abdelhaq H, Sengstock C and Gertz M, "EvenTweet: Online Localized Event Detection from Twitter" The 39th International Conference on Very Large Data Bases, 2013, Italy, Proceedings of the VLDB Endowment, Vol. 6, No. 12

[17] Rahman M, Zhu Q and Yu E, "TRECT: A Hashtag Recommendation System for Twitter"

[18] S. M. Kywe, T.A. Hoang, E.P. Lim and F. Zhu: "On Recommending Hashtags in Twitter Networks" In SocInfo, Volume 7710 of Lecture Notes in Computer Science, Springer, 2012 pp 337–350.

[19] Sankaranarayanan J, Samet H, Teitler B, Lieberman M and Sperling J "TwitterStand: News in Tweets" ACM GIS '09, November 4-6, 2009. Seattle, WA, USA

[20] W.Jianshu,L.E Peng,J. Jing; and H.E. Qi, "Twitterrank:Finding Topic-Sensitive Influential Twitterers," Proceedings of the Third ACM International Conference on Web Search & Data Mining: February 3-6, 2010, New York, pp. 261-270 .

[21] L.Chenliang, A.Sun, and A.Datta,"Twevent: segment-based event detection from tweets," Proceedings of the 21st ACM international conference on Information and knowledge management. ACM, 2012.