Global Journal of Enterprise Information System

DOI: 10.18311/gjeis/2017/18138

Study of Existing Design Architectures and Frameworks in Twitter

Kamaljit Kaur[1*] and Kanwalvir Singh Dhindsa[2]

[1]*Department of Computer Science and Engineering, Sri Guru Granth Sahib World University, Punjab, India; kamaljit_cse@sggswu.org*

[2]*Department of CSE & IT, Baba Banda Singh Bahadur Engineering College, Punjab, India; kdhindsa@gmail.com*

Abstract

Micro-blogging is a way of communication where users publish messages with limited length. Twitter is one such example where the user message is known as tweet which is a post of 140 characters and published to reflect their views on an event and discussions. Large numbers of users are connected on the Twitter network and large amount of information is generated from them. Based on the different Twitter aspects various commercial and noncommercial systems have been developed. This paper aims to study these different aspects of Twitter and discusses the different design frameworks and architectures related to it.

Keywords: Follow, Followee Recommendation, HashTag Recommendation Retweet, Similar Users, Twitter, Tweet, Trend Detection

Paper Code: 17954; **Originality Test Ratio:** 21%; **Submission Online:** 27-Oct-2017; **Manuscript Accepted:** 25-Sep-2017; **Originality Check:** 30-Oct-2017; **Peer Reviewers Comment:** 01-Nov-2017; **Double Blind Reviewers Comment:** 3-Nov-2017; **Author Revert:** 22-Nov-2017; **Camera-Ready-Copy:** 10-Dec-2017)

1. Introduction to DSS

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.

2. 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.

2.1 Trend Detection

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

*Author for correspondence

Table 1. Different Recommendation Areas in Twitter and its corresponding frameworks discussed

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] Based on Twitter stream characteristics Based on Topic Models
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.	TREC[17] TweetSense[15]

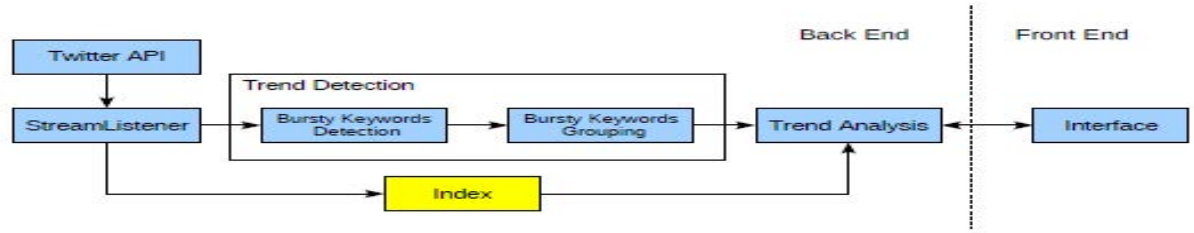

Figure 1. TwitterMonitor Architecture[12].

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. 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 is constructed which computes the total number of all tweets, the occurrence of each word and the occurrence of each word pair which is an indication of potential tweet popularity. A sketch-based topic model is developed to find the bursty topics from the data sketch.

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.

These above methods of detecting the trends are based on analyzing the bursty terms. These methods have the limitation that they fails to capture the fact that multiple terms may be involved with the same event and requires that at least one term undergoes a sufficiently high jump in relative frequency that the event can be identified.

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.

2.2 Similar Users 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. Each user is modeled as a document containing up to 200 most recent tweets of that user, then using TF-IDF metric the frequency of a given term is examined within a document multiplied by its frequency score across all the documents. Twittomender mines a large number of user profile details and content by following the links between users on the social graph. It 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's , Follower's and using all the connections.

Table 2. Comparison of Different Trend Detection Frameworks and Architecture

Framework Name	Methodology	Technique/Algorithm	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 using total number of tweets, the occurrence of each word and each word pair From this, sketch based topic model is proposed	Keyword-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

Table 3. Comparison of Similar User Recommendation Frameworks and Architecture

Framework Name	Methodology	Technique/Algorithm
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

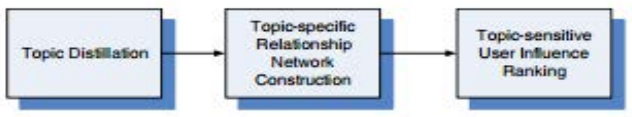

Figure 2. TwitterRank Approach[20].

TwitterRank is an extension of PageRank algorithm that is used to measure the influence of users in Twitter. TwitterRank measures the influence taking both the topical similarity between users and the link structure into account[20].

Topic Distillation automatically identify the topics that users are interested in based on the users published tweets. Latent Dirichlet Allocation (LDA) technique is used that identifies the latent topic information from large document collection. It uses a "bag of words" concept which treats each document as a vector of word counts.

Using this each document is represented as a probability distribution over some topics, while each topic is represented as a probability distribution over a number of words. The influence of a twitter user is considered as high if the sum of influence of followers is high.

2.3 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. When a Twitter user generates context less tweets it is referred as an Orphan Tweet, which have no hash tags or @mentions. It is a hashtag rectification system that does not force users to use the hash tags at the time of origin rather recommends a list of hash tags for the users who look for context.[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 fol-

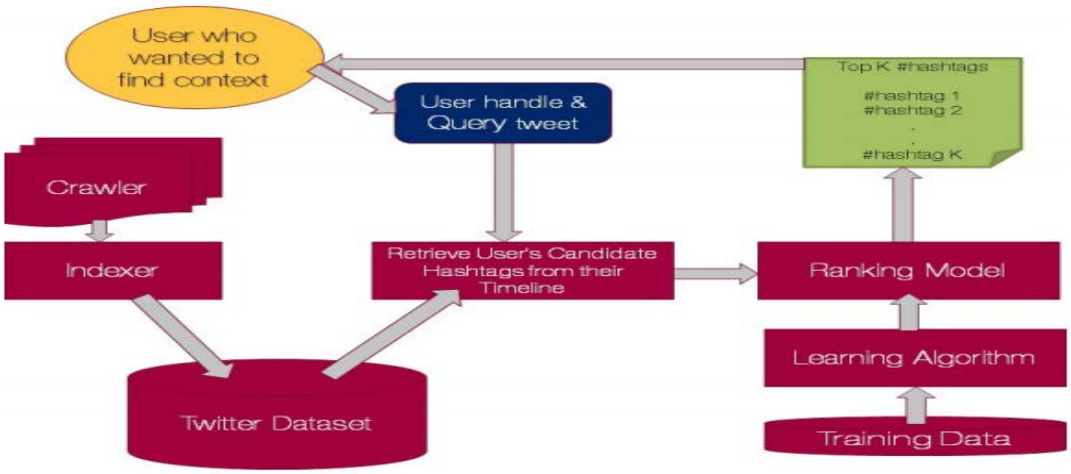

Figure 3. TweetSense Architecture[15].

Table 4. Comparison of Hashtag Recommendation Frameworks and Architecture

Framework Name	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

lower-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 top K most promising hash tags are recommended to the user.

3. 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.

4. References

1. Java XS, Finin T, Tseng B. Why We Twitter: Understanding Microblogging Usage and Communities. In Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis. ACM, 2007. https://doi.org/10.1145/1348549.1348556
2. Hannon J, Bennett M, Smyth B. Recommending Twitter Users to Follow Using Content and Collaborative Filtering Approaches. In Proceedings of the fourth ACM Conference on Recommender Systems, RecSys'10, New York, NY, USA, 2010. p. 199–206. https://doi.org/10.1145/1864708.1864746
3. Allan J, Papka R, Lavrenko V. On-line new event detection and tracking. In SIGIR. 1998; 37–45. https://doi.org/10.1145/290941.290954
4. Brants T, Chen F, Farahat A. A system for new event detection. In SIGIR. 2003; 330–7. https://doi.org/10.1145/860435.860495
5. Weng J, Lee B. Event detection in twitter. Proc of ICWSM, 2011.
6. Li C, Sun A, Datta A. Twevent: segment-based event detection from tweets. In Proceedings of the 21st ACM international conference on Information and knowledge management. 2012. p. 155–64. https://doi.org/10.1145/2396761.2396785
7. Chen C, Yin H, Yao J, Cui B. Terec: a temporal recommender system over tweet stream. Proceedings of the VLDB Endowment. 2013; 6(12):1254–7. https://doi.org/10.14778/2536274.2536289
8. Godin F, Slavkovikj V, De Neve W, Schrauwen B, Van de Walle R. Using topic models for twitter hashtag recommendation. In Proceedings of the 22nd international Conference on World Wide Web companion. 2013. p. 593–6. https://doi.org/10.1145/2487788.2488002
9. Weng J, Lim E-P, Jiang J, He Q. Twitterrank: finding topic sensitive influential twitterers. In Proceedings of the Third ACM

International Conference on Web Search and Data Mining, ACM. 2010. p. 261–70. https://doi.org/10.1145/1718487.1718520
10. Song K, Wang D, Feng S, Zhang Y, Qu W, Yu G. Ctrof: A collaborative tweet ranking framework for online personalized recommendation. In Advances in Knowledge Discovery and Data Mining, Springer. 2014; 1–12. https://doi.org/10.1007/978-3-319-06605-9_1
11. Sankaranarayanan J, Samet H, Teitler BE, Lieberman MD, Sperling J. TwitterStand: News in Tweets. Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. 2009. p. 42–51. https://doi.org/10.1145/1653771.1653781
12. Mathiousdakis M, Koudas N. TwitterMonitor: Trend Detection over the Twitter Stream. SIGMOD'10, Indianapolis, Indiana, USA. 2010 Jun 6–11.
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, 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, Gertz M. EvenTweet: Online Localized Event Detection from Twitter. The 39th International Conference on Very Large Data Bases. Italy, Proceedings of the VLDB Endowment. 2013; 6(12). https://doi.org/10.14778/2536274.2536307
17. Rahman M, Zhu Q, Yu E. TRECT: A Hashtag Recommendation System for Twitter.
18. Kywe SM, Hoang T-A, Lim E-P, Zhu F. On Recommending Hashtags in Twitter Networks. In SocInfo, Volume 7710 of Lecture Notes in Computer Science, Springer. 2012; 337–50.
19. 19. Sankaranarayanan J, Samet H, Teitler B, Lieberman M, Sperling J. TwitterStand: News in Tweets. ACM GIS '09, Seattle, WA, USA. 2009 Nov 4-6.
20. 20. Weng J, Lim EP, Jiang J, He Q. Twitterrank: Finding Topic-Sensitive Influential Twitterers. Proceedings of the Third ACM International Conference on Web Search & Data Mining. New York. Research Collection School of Information Systems. 2010 Feb 3-6; 261–70. https://doi.org/10.1145/1718487.1718520

Annexure-I

PRIMARY SOURCES

1	149.169.27.83 Internet	94 words — 5%
2	Lecture Notes in Computer Science, 2011. Crossref	69 words — 3%
3	www.wsdm-conference.org Internet	68 words — 3%
4	www.cs.sfu.ca Internet	39 words — 2%
5	Zhihui Yang, Huixin Ma, Zhenying He, X. Sean Wang. "Finding maximal ranges with unique topics in a text database", World Wide Web, 2017 Crossref	34 words — 2%
6	ijcsi.org Internet	30 words — 1%
7	ijaiem.org Internet	17 words — 1%
8	Tang, Xing, Qiguang Miao, Shangshang Yu, and Yining Quan. "A Data-Based Approach to Discovering Multi-Topic Influential Leaders", PLoS ONE, 2016. Crossref	15 words — 1%
9	www.ijsr.net Internet	13 words — 1%
10	www.3riverdev.com Internet	12 words — 1%
11	www.slideshare.net Internet	10 words — < 1%
12	Lecture Notes in Computer Science, 2013. Crossref	8 words — < 1%
13	ceur-ws.org Internet	8 words — < 1%
14	elib.uni-stuttgart.de Internet	8 words — < 1%
15	Xie, Wei, Feida Zhu, Jing Jiang, Ee-Peng Lim, and Ke Wang. "TopicSketch: Real-Time Bursty Topic Detection from Twitter", 2013 IEEE 13th International Conference on Data Mining, 2013. Crossref	6 words — < 1%

Citation:
Kamaljit Kaur and Kanwalvir Singh Dhindsa
"Study of Existing Design Architectures and Frameworks in Twitter",
Global Journal of Enterprise Information System. Volume-9, Issue-4, October-December, 2017. (http://informaticsjournals.com/index.php/gjeis)
DOI: 10.18311/gjeis/2017/18138
Conflict of Interest:
Author of a Paper had no conflict neither financially nor academically.