iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 6 **Issue:** IV **Month of publication:** April 2018

DOI: http://doi.org/10.22214/ijraset.2018.4786

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 6 Issue IV, April 2018- Available at www.ijraset.com

Linguistic Virality Based Microblog Content Propagation Modeling Using V3sl Framework

Divya S[1], Anju J Prakash[2]
[1]*M Tech in CSE, Sreebuddha College of Engineering*
[2]*Assistant Professor in CSE, Sreebuddha College of Engineering*

Abstract: ***Microblogging is the practice of posting small pieces of digital content such as text message on the internet. It is a combination of blogging and instant messaging. Microblogs have rapidly become a significant means by which people communicate with the world and each other in near realtime. When a microblogging user adopts some content propagated to her, it affects three behavioral factors, namely, topic virality, user virality, and user susceptibility. Topic virality measures the degree to which a topic attracts propagations by users. User virality and susceptibility refer to the ability of a user to propagate content to other users, and the propensity of a user adopting content propagated to her, respectively. This paper is the study of the problem of mining these behavioral factors specific to topics and linguistic features from microblogging content propagation data. A new microblog is created for mining these factors. Then a three dimensional tensor framework is constructed to simultaneously derive the three sets of behavioral factors. Based on this framework, a five dimensional tensor framework is developed which considers fine grained factors such as linguistic features that affects virality. Parameter learning is also implemented by developing an algorithm based on SVM. The proposed system models each talk of the user based on topic specific behavior and linguistic behavior. A demonsration of this modeling is also carried out by popularity comparison using PRCC, virality analysis using behavioral factors and framework comparison based on virality prediction and time complexity.***
Keywords: micro blogging, content, propagation, behavioural modelling, linguistic features

I. INTRODUCTION

Microblogging[1] is the method of posting small pieces of digital content-which could be text, pictures, links, short videos, or other media-on the Internet. It is a combination of blogging and instant messaging. Microblogging has become popular among groups of friends and professional colleagues who frequently update content and follow each other's posts, creating a sense of online community. Twitter is currently the best-known microblogging site, its popularity supported by a growing collection of add-on applications that enable different and often more engaging microblog updates, such as TwitPic for uploading pictures or Polly Trade for buying and selling stocks. Meanwhile, a number of competing microblog applications-some open source, many aimed at specific interest groups-continue to challenge Twitter's popularity. This resulting profusion of tools is helping to define new possibilities for this type of communication. Some other popular microblogging platforms are tumblr, instagram, vine etc.The benefits of microblogging platforms are fast content development, fast content consumption, opportunity for more frequent posts, easy to share and easy to communicate with followers.

In mainstream culture, microblogging has become an extremely popular channel for both professional and personal pursuits. Friends use it to keep in touch, business associates use it to coordinate meetings or share useful resources, and celebrities and politicians (or their publicists) microblog about concert dates, lectures, book releases, or tour schedules. For higher education, microblogging is an increasingly important tool for communities of practice, enabling scholars to communicate informally on subjects of shared interest and to open windows into their own projects, sparking interest and discovery among peers. Some universities are considering using microblogging in the curriculum to emphasize timeliness, student engagement, and aggregation of artifacts relevant to course content and experience. At some institutions, faculty offer course-centric microblogging streams to create a backchannel among students in the classroom. Stephen Prothero, professor of religion at Boston University, has set himself the challenge of using Twitter to sum up eight major religions, in a maximum of 140 characters per post. The microblogs he offers will feed into a book that he is writing on the same topic. Microblogging is one of the application of Datamining.

Microblogging is from the area of Text mining which is a subarea of Data mining. Data mining or Knowledge Discovery is the process of analysing data from different perspectives and summarizing it into useful information - information that can be used to increase revenue, cuts costs, or both. Data mining software is one of a number of analytical tools for analysing data. It allows users to analyse data from many different dimensions or angles, categorize it, and summarize the relationships identified. Text mining,

also referred to as text data mining, roughly equivalent to text analytics, refers to the process of deriving high-quality information from text. High-quality information is typically derived through the devising of patterns and trends through means such as statistical pattern learning. Text mining usually involves the process of structuring the input text (usually parsing, along with the addition of some derived linguistic features and the removal of others, and subsequent insertion into a database), deriving patterns within the structured data, and finally evaluation and interpretation of the output. 'High quality' in text mining usually refers to some combination of relevance, novelty, and interestingness. Typical text mining tasks include text categorization, text clustering, concept/entity extraction, production of granular taxonomies, sentiment analysis, document summarization, and entity relation modeling (i.e., learning relations between named entities). Text analysis involves information retrieval, lexical analysis to study word frequency distributions, pattern recognition, tagging/annotation, information extraction, data mining techniques including link and association analysis, visualization, and predictive analytics. The overarching goal is, essentially, to turn text into data for analysis, via application of natural language processing (NLP) and analytical methods. A typical application is to scan a set of documents written in a natural language and either model the document set for predictive classification purposes or populate a database or search index with the information extracted. In NLP, Latent Dirichlet allocation (LDA) is a generative model that allows sets of observations to be explained by unobserved groups that explain why some parts of the data are similar. For example, if observations are words collected into documents, it posits that each document is a mixture of a small number of topics and that each word's creation is attributable to one of the document's topics.The components of text analysis includes Information retrieval or identification of a corpus is a primary step: collecting or identifying a set of textual materials, on the Web or held in a file system, database, or content corpus manager, for analysis.

Content propagates among microblogging users through their follow links, from followees to followers. The former are the senders, and the latter are known as the receivers. A receiver may adopt the content exposed to her based on a number of factors, namely the: (a) virality of the sender [2],[3], [4], (b) susceptibility of the receiver [5], [6], (c) virality of the content topic [7], [8], and (d) strength of relationships between sender and receiver [9]. User virality refers to the ability of a user in getting others to propagate her content, while user susceptibility refers to the tendency of a user to adopt her followees' content. Topic virality refers to the tendency of a topic in getting propagated. Since Microblogging has been shown rather an information source than a social networking service [10], in this paper it is assumed that most relationships among users in a microblogging site are casual and identical in strength. The proposed system therefore focus on modelling[1] the user and content factors that drive content propagation without considering the pairwise relationships among users. The modeling of the virality and susceptibility factors has many important applications. In advertisement and marketing, companies may hire viral users to propagate positive content about their products, or to attach the advertisement with viral content so as to maximize their reach [11]. Similarly, politicians may leverage on viral users to disseminate their messages widely or to conduct campaigning [12], [13].Also, one may detect events by tracking those mentioned by non-susceptible users [14], and detect rumors based on susceptible users' interactions with the content [15], [16].

Microblogging has been shown rather an information source than a social networking service, most relationships among users in a microblogging site are casual and identical in strength. So users and content factors to be modelled that drive content propagation without considering the pairwise relationships among users.In advertisement and marketing, companies may hire viral users to propagate positive content about their products, or to attach the advertisement with viral content so as to maximize their reach . Similarly, politicians may leverage on viral users to disseminate their messages widely or to conduct campaigning,Also, one may detect events by tracking those mentioned by non-susceptible users, and detect rumors based on susceptible users' interactions with the content.

When a microblogging user adopts some content propagated to her, it can attribute that to three behavioral factors, namely, topic virality, user virality, and user susceptibility. Topic virality measures the degree to which a topic attracts propagations by users. User virality and susceptibility refer to the ability of a user to propagate content to other users, and the propensity of a user adopting content propagated to her, respectively. In this work, the problem of mining these behavioral factors specific to topics is studied from micro blogging content propagation data. Then it is possible to construct a three dimensional tensor for representing the propagation instances and these factors and distinguishing them from other related factors. These tensor can be extended to five dimensional by incorporating linguistic features and network virality. The modelling of the virality and susceptibility factors has many important applications. In advertisement and marketing, companies may hire viral users to propagate positive content about their products, or to attach the advertisement with viral content so as to maximize their reach . Similarly, politicians may leverage on viral users to disseminate their messages widely or to conduct campaigning.

In existing system, virality of content is not utilized as a factor for business or marketing data propagation.Existing models measure the three behavioral factors separately. That is, they
measure a user's virality by aggregating propagations on her content without considering the virality of content and susceptibility of the receivers. Again, similar remarks are applicable to existing works that measure users' susceptibility and topics' virality. Such simplistic approaches may lead to less accurate modeling results. The main demerits of the existing system are:

1) Content Propagation is modeled from simple virality and susceptibility.
2) Does not use topic specific behavioral factors.
3) Does not consider fine grained factors.
4) Does not cover all the three factors in a common framework.
5) Does not consider inter relationship of factors.
6) Needs a lot of side information to measure the importance of topic.

The proposed system jointly models the content propagation data set using three behavioral factors,i.e., topic virality, topic -specific user virality, and topic-specific user susceptibility. In micrblogging retalk is the common form of content propogation Virality can be evaluated from the number of talks and retalks based on the topic.. Content propogation modeling is done in four steps: preprocessing, topic modeling, topic-specific behavioral modeling, Virality Analysis and Popularity Comparison using Pearson Rank Correlation Coefficient. Preprocessing removes stop words from the corpus data. Topic modeling is done using LDA Algorithm. With the help of Collapsed Gibbs Sampling LDA generates topics with their strengths. From these topics Topic-specific behavioral modeling is done using three Topic-specific behavioral factors. Then the Virality analysis is done using charts. Finally comparison of generating popularity ie popularity of talks and propogating popularity ie popularity of retalks are done using Pearson Rank Corelation Coefficient. The main advantages of the proposed system are

a) Three Topic-specific behavioral factors can be derived simultaneously.
b) Helps companies or authors to find the popularity of their working.
c) More effective recommendation method can be derived from topic based algorithm.
d) The exact followers are traced out easily.
e) Fine grained factors can be used to model the propagation.

II. METHODOLOGY

Fig.3.1 shows the system architecture of proposed system. Input is obtained from the corpus which is the collection talks. Preprocessing is done on this dataset which is the stop word removal process. On this preprocessed data topic modeling is done which results in the discovery of topics. Topic modeling involves topic generation and virality calculation which is done using LDA Algorithm. LDA Algorithm generates probability of the topic with the help of Collapsed Gibbs Sampling. After topic modeling three topic specific behavioral factors Topic virality, Topic-specific User virality and Topic-specific user susceptibility are calculated. Now a three dimensional V2S tensor framework is generated using these factors. The corpus generated after talk preprocessing is used to perform linguistic feature identification. NLP and Senti strength algorithms are used for this process. Using these fine grained factors a five dimensional V3SL framework is generated. In order to get the accurate prediction SVM based prediction is done on the framework. Thus each talk is categorized to any of the three categories which are high virality, medium virality or low virality. The proposed system is designed with the creation of a microblog platform and the virality is analysed from the users of this platform.

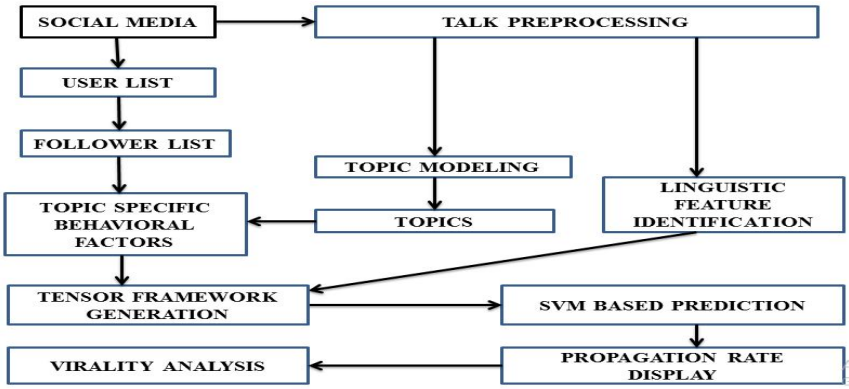


Fig 1 System Architecture

The proposed system mainly consist of two modules: User module and Admin module. Table 3.1 shows the main activities of the User module which are registration, upload profile, update profile, talk, retalk, view followers, unfollow, view virality analysis and view popularity comparison. User is also able to view the status of their follower's talks.

TABLE 1
User Privileged Tasks

User Privileged Task	Description
Registration	A new user can register to the microblog site with his username and password. Registered user can log in with his username and password.
Upload profile	New user can upload his personal details
Update profile	Registered user can update his profile
Talk	User can talk his thoughts. Talk are stored with talktid, author's emailid, date and time of talk.
Retalk	User can retalk author's talks.Retalk details are stored with authorid, followerid, retalkid, date and time of retalk.
View Followers	User can view his followers at any time. Follow details are stored with authorid, followerid, date and time of follow.
Unfollow	User can unfollow an untrusted follower.
View Virality Analysis	User can view the Virality Analysis
View Popularity Comparison	Populariy is compared using Pearson Rank Correlation Coefficient
View Popularity Analysis	Popularity graph is generated with topic on x axis and popularity on y axis.

The main privileged tasks of the admin module are user management, talk preprocessing, topic modeling, topic specific behavioral factor derivation, framework generation, linguistic behavioral modeling, virality analysis, compare popularity, analyse popularity and compare frameworks.

TABLE 2
Admin Privileged Tasks

Admin Privileged Task	Description
User Management	Admin can delete an untrusted user.
Talk Preprocessing	Includes Stopword management and Slang management.
Topic Modeling	Performs Topic Discovery and Topic strength calculation.
Topic Specific Behavioral Factor Derivation	Performs Virality calculation which involves topic virality,user virality and user susceptibility.
Framework Generation	Generate the Framework based on the Topic Specific Behavioral Factors
Linguistic Behavioral Modeling	dentifies the linguistic features in the talks and models content propagation.
Virality Analysis	Analyses the different virality measures
Compare Popularity	Popularity is compared using Pearson Rank Correlation Coefficient
Analyse Popularity	Popularity graph is generated with topic on x axis and popularity on y axis.
Compare Frameworks	Frameworks are compared based on virality prediction and time complexity.

III.IMPLEMENTATION

Implementation is one of the most important tasks in a project. Implementation is the phase, in which one has to be cautious, because all the efforts undertaken during this project will be fruitful only if the software is properly implemented according to the plans made. Implementation is the stage in the project where the theoretical design is turned into a working system. The crucial stage is achieving successful new system and giving the users confidence in that the system will work effectively and efficiently.
It involves careful planning, investigation of the current system and its constraints on implementation and design of methods to achieve changeover. Apart from these, the major task of preparing for implementation is education and training of users and system testing.
The proposed system has the following phases:

1) Talk Preprocessing
2) Topic Modeling
3) Topic Specific Virality Factor Derivation
4) Popularity Comparison
5) Framework Generation
6) Linguistic Modeling
7) Model learning
8) Virality Analysis
9) Framework Comparison

A. Talk Preprocessing

Talk preprocessing involves the tasks like stopword removal and slang management. The data may be processed further to remove stop words like auxillery verbs, prepositions etc. The corpus data will be having only relevant terms mainly containing nouns and verbs. A dictionary based approach is utilized to remove stopwords from document. A generic stopword list containing 75 stop words created using hybrid approach is used . The algorithm is implemented as below given steps.

1) *Step 1:* The target document text is tokenized and individual words are stored in array.
2) *Step 2:* A single stop word is read from stopword list.
3) *Step 3:* The stop word is compared to target text in form of array using sequential search technique.
4) *Step 4:* If it matches , the word in array is removed , and the comparison is continued till length of array.
5) *Step 5:* After removal of stopword completely, another stopword is read from stopword list and again algorithm follows step 2. The algorithm runs continuously until all the stopwords are compared.
6) *Step 6:* Resultant text devoid of stopwords is displayed, also required statistics like removed, no. of stopwords removed from target text, total count of words in target text, count of words in resultant text, individual stop word count found in target text is displayed.

B. Topic Modeling

The experiment automatically identify the topics of every original talk. This step is conducted for every time window, independently from each others. The proposed system first remove all retalks and non-informative talks, e.g., talks generated by third party applications like Foursqure or Instagram. Then remove from remaining talks all stop words, slang words and non-English phrases. Next, we iteratively filter away words, talks, and users such that: each word must appear in at least 3 remaining talks, each talk contains at least 3 remaining words and each user has at least 20 remaining talks. These minimum thresholds are designed to ensure that for each user, talk, and word, the system have enough observations to learn the latent topics accurately. A set of topics will be generated from each talks, which will help to assign the area of topic to each talks in the domain. These data will be used to find the content virality .The algorithm adopted for topic modeling is LDA(Latent Dirichlet Allocation). LDA represents documents as mixtures of topics that spit out words with certain probabilities. It assumes that documents are produced in the following fashion:

When writing each document

1) Decide on the number of words N the document will have (say, according to a Poisson distribution).
2) Choose a topic mixture for the document (according to a Dirichlet distribution over a fixed set of K topics). Generate each word in the document by:

3) First picking a topic .

4) Then using the topic to generate the word itself.

5) Assuming this generative model for a collection of documents, LDA then tries to backtrack from the documents to find a set of topics that are likely to have generated the collection.

C. Topic Specific Behacioral Factor Mining

Topic-specific behavioral modeling is done using three Topic-specific behavioral factors Topic virality, Topic-specific user virality and Topic-specific user susceptibility. In microblogging, retalk is the most common form of content propagation. Therefore use retalk to define propagation in the remaining part of this section. That is each original talk m is considered as a content item, and user v is exposed to m if (a) v follows m's author, and (b) v receives and reads m. m is said to be propagated from its author u to v if (i) v follows u and (ii) v retalks m. Topic-specific behavioral factors can be calculated using the database tables talk which includes talkid, talk, authorid, date and time of talk, follow which includes authorid, followerid, date and time of follow, retalk which includes authorid, followerid, retalkid, date and time of retalk.

Topic virality: This refers to the ability of a topic to attract propagation. Every topic k is associated to a virality score $I_k \in [0,1]$ indicating how viral the topic is, i.e. how likely a content about the topic will get propagated. Topic virality is modelled using topicid, topic and topic virality value. Topic virality can be taken as the ratio of total number of followers of each topic to the total number of followers.

Topic Specific User Virality: This refers to the ability of a user to get her content propagated for a specific topic. Every user u is assigned a topic-specific user virality vector $V_u = (V_{u,1}, V_{u,2}, \ldots\ldots V_{u,k})$ where $V_{u,k} \in [0,1]$ for $\forall k=1,2,\ldots\ldots,K$. For topic k, Vu,k denotes how viral user u is for the topic, i.e.,how likely u gets propagations for her content with topic k.

Topic-specific User Virality Calculation is modelled using userrid, topic and userviralityvalue. Topic-specific User Virality can be taken as the ratio of total number of followers of each topic written by the specific user to the total number of followers.

Topic-specific user susceptibility: This refers to the tendency of a user to adopt content propagated to her for a specific topic. Each user v is associated with a topic-specific user susceptibility vector $S_v = (S_{v,1}, S_v, 2, \ldots\ldots\ldots S_{v,k})$ where $S_{v,k} \in [0,1]$ for $\forall k=1,2,\ldots\ldots,K$, and $S_{v,k}$ indicates how susceptible user v is to topic k, i.e., how likely v adopts a content about the topic k after being exposed to the content. Topic-specific User Susceptibility Calculation is modelled using userid,topic and user susceptibility value.Topic-specific User Virality can be taken as the ratio of total number of authors of each topic followed by the specific user to the total number of followers.

D. Topic Popularity Comparison

To compare the likelihood of getting retalked across topics, in each time window and for each topic k, the system derive the relative popularities of topic k among the set of all original talks and the bag of retalks in the time window. The former is called generating popularity of the topic, G_k. The later is called propagating popularity, denoted by P_k. The two popularities are defined based as follows:

$$G_k = \frac{1}{|\mathcal{M}|}\sum_{m \in \mathcal{M}} D_{m,k}$$

$$P_k = \frac{1}{\sum_{m \in \mathcal{M}} p_m}\sum_{m \in \mathcal{M}} [p_m . D_{m,k}] \qquad (1)$$

where, in each time window, $\mathcal{M}$ is the set of all content items, and p_m is number of time m is propagated successfully.Since the system use talks and retalks to define content and propagation respectively, $\mathcal{M}$ is the set of original talks while pm is number of m's retalks.

To examine the difference between the two popularities of topics, use their Pearson rank correlation coefficientPRCC, defined as below:

$$PRCC = \frac{\sum_{k=1}^{K}(r_G(k)-\bar{r})(r_P(k)-\bar{r})}{\sqrt{\sum_{k=1}^{K}(r_G(k)-\bar{r})^2}\sqrt{\sum_{k=1}^{K}(r_P(k)-\bar{r})^2}} \qquad (2)$$

where $r_G(k)$ is the rank of generating popularity of topic k (i.e., the rank of G_k in $G_1, \ldots, G_K$), $r_P(k)$ is the rank of propagating popularity of topic k (i.e., the rank of P_k in $P_1, \ldots, P_K$), and r is the mean rank: $r = (1+K)/2$. Certainly,PRCC ε [-1, 1]. PRCC is close to 1 (respectively -1) if the two popularities are strongly correlated (respectively invert correlated), and PRCC is close to 0 if the two popularities not correlated.

E. Topic Specific Virality Analysis

Topic Specific Virality Analysis is done by comparing the virality values of three topic specific behavioral factors. It is done with the help of graph. Topic virality is analysed with topic on x axis and topic virality value on y axis. User virality is analysed with userid on x axis and user virality value on y axis. User susceptibility is analysed with user id on x axis and user susceptibility on y axis.

F. Framework Generation

V2S framework[1] represents each content propagation observation by a tuple *(u,v,m)* where *m* is a content item generated by user *u*, and exposed to user *v*. We use a binary variable δ_{uvm} to denote whether *v* adopts *m* (δ_{uvm}=1) or otherwise (δ_{uvm}=0). We call a propagation observation positive or negative when δ_{uvm} =1and 0 respectively. In V 2S framework, δ_{uvm} depends on topic-specific virality of *u*, topic specific susceptibility of *v*, and the topics' virality as follows. Consider a propagation observation *(u,v,m)*, assume that the likelihood that *v* adopts *m* is determined by: (a) m's topic distribution $D_m = (D_{m,1}, D_{m,2},\ldots\ldots\ldots D_{m,k})$ (b) u's topic specific user virality V_u; (c) topic virality I; and (d) v's topic specific user susceptibility S_v.

G. Linguistic Behavioral Modeling

In microbogging, content propagation is also affected by more fine grained factors. These factors include user's positions in the network, linguistic features in content, and emotion factors of users. User's position on the network represents number of followers of each user. Content propagation is also affected by the linguistic feature of the microblog content or talk. There are mainly two phases for the linguistic behavioral modeling. Former is the linguistic feature identification on talks. Latter is SVM based prediction. There are mainly four categories of linguistic features. These are understand ability, level of details, punctuations and emotions and cognition indicators. Understand ability represents readability and word familiarity. Level of details represents parts of speech like verb, adverb and use of functional words like conjunction, preposition, quantifiers etc. Punctuations and emotions are represented by words like wow, oh etc. Cognition indicators are discrepancy words like should, may etc, tentative words like perhaps, guess etc, casual words like because, hence etc, insight words like think, consider etc, motion words like arrive, go etc. These linguistic features are collected from each talks and SVM based prediction is done. This results in the generation of V3SL framework.

NLP is the algorithm used for linguistic feature identification. NLP provides interactions between human language and computers . NLP is a way for computers to analyze, understand, and derive meaning from human language in a smart and useful way. NLP considers the hierarchical structure of language: several words make a phrase, several phrases make a sentence and, ultimately, sentences convey ideas. By analyzing language for its meaning, NLP systems have long filled useful roles, such as correcting grammar, converting speech to text and automatically translating between languages. NLP is used to analyze text, allowing machines to understand how human's speak. This human-computer interaction enables real-world applications like automatic text summarization, sentiment analysis, topic extraction, named entity recognition, stemming, and more. NLP is commonly used for text mining, machine translation, and automated question answering.

Support Vector Machines(SVM) are supervised learning models with associated learning algorithms that analyse data used for classification. Given a set of training data, each marked as belonging to one or the other of two categories, an SVM training algorithm builds a model that assigns new inputs to one category or the other. An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New inputs are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall.

IV. RESULT AND DISCUSSIONS

The proposed system generates V2S and V3SL frameworks using topic specific and fine grained factors. Using these frameworks microblog content propagation is modeled and each talk is categorised as highly viral, medium viral and less viral based on the topic specific behavioral factors. The proposed system also models each talk as good or bad based on the fine grained factors. In the talkon platform this modeling is represented using different colors and icons. Analysis of the proposed system is carried out by considering topic specific behavioral factors, popularity and framework comparison.

A. Topic Specific Virality Analysis

Here the proposed system conduct an evaluation on topic-specific behavioral factors using virality analysis.In Fig4.1 the three factors are analysed and compared using a graph with five topics on x-axis and virality values on y-axis. User virality and user susceptibility values for this five topics are determined by considering a user u.

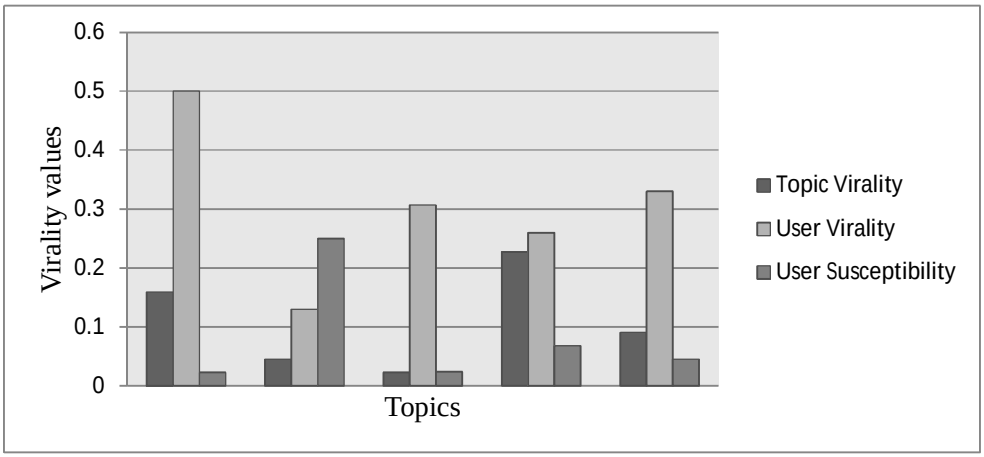


Fig.2 Topic Specific Virality Analysis

B. Popularity Comparison

The proposed system performs the comparison of generating topic popularity and propogating topic popularity by using Pearson Rank Correlation Coefficient.This comparison is evaluatd by considering the topic popularity pair on x-axis and prcc on y-axis.

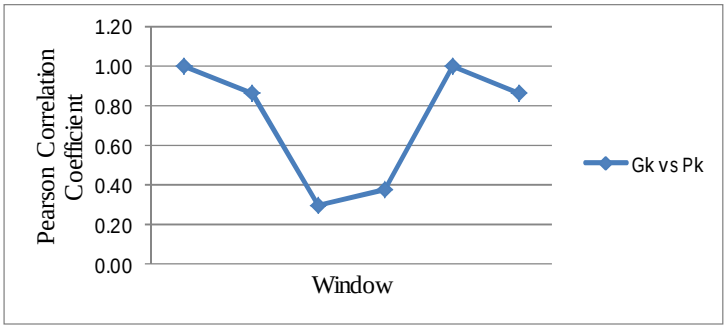


Fig 3 Topic Popularity Comparison using PRCC

C. Performance Comparison

The performance of the two frameworks can be compared by considering the virality prediction and timecomplexity. Virality prediction is carried out with the number of viral talks. Time Complexity is determined by analysing the time needed for the execution of the two frameworks in milliseconds.

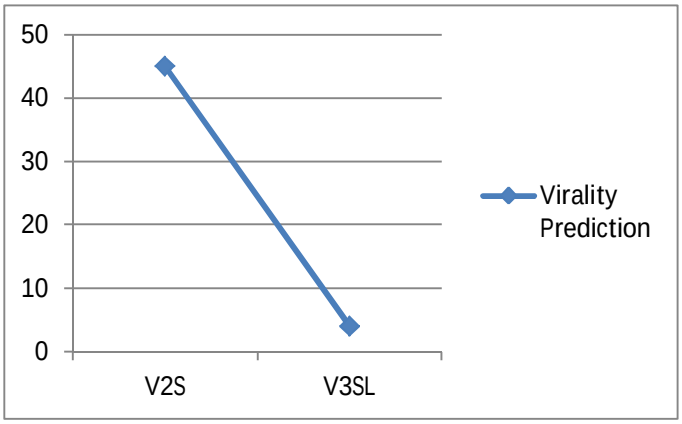


Fig.4 Virality Prediction Comparison on V2S and V3SL

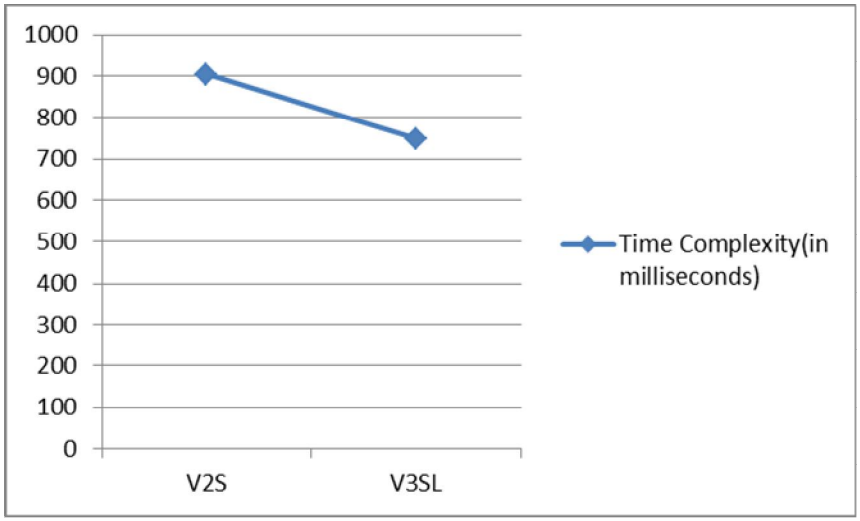


Fig.5 Time Complexity Comparison on V2S and V3SL

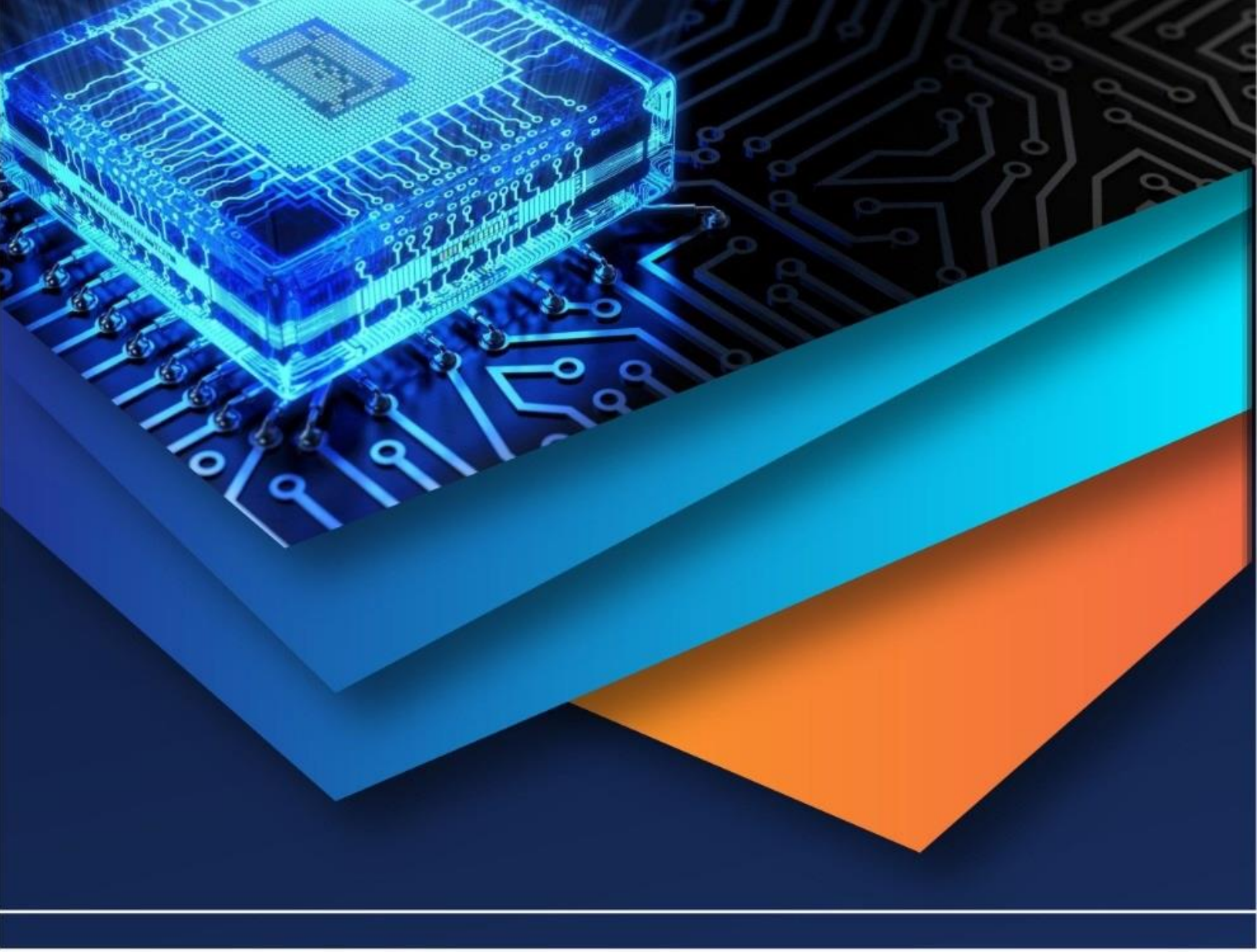
V. CONCLUSIONS

The proposed system study user and content factors underlying content propagation in microblogging. Motivated by an empirical studying showing that different topics have different likelihood of getting propagated at both network and individual levels, the system propose to model the factors to topic level. From these factors a three dimensional V2S tensor framework is developed to learn topic-specific user virality and susceptibility, and topic virality from content propagation data. This V2S framework is extended to a five dimensional V3SL framework which incorporates the fine grained factors such as user's position on the network and linguistic feature in content. This helps to analyse the talks of the user's based on virality and linguistic behavior. The major objectives of the work were achieved and could trace out the advantages found with virality detection like virality and susceptibility helps companies or authors to find the popularity of the writing. The exact followers are traced out easily. Also the different popularities of the topic can be compared easily. In future, the virality model can be used to perform event detection and sentiment analysis.

REFERENCES

[1] Tuan-Anh Hoang and Ee-Peng Lim,"Microblogging Content Propogation Modeling using Topic-specific Behavioral Factors",IEEE Transactions on Knowledge and Data Engineering,Vol.28,N0.9,September 2016.
[2] D. Bakshy, J. M. Hofman, W. A. Mason, and D. J. Watts,"Everyone's an influencer: Quantifying influence on twitter," in Proc. 4th ACM Int. Conf. Web Search Data Mining, 2011, pp. 65–74.
[3] S. A. Macskassy and M. Michelson, "Why do people retalk? Antihomophily wins the day!" in Proc. 5th Int. AAAI Conf. Weblogs Social Media, 2011, pp. 209–216.
[4] Z. Liu, L. Liu, and H. Li, "Determinants of information retalking in microblogging," Internet Res., vol. 22, pp. 443–466, 2012.
[5] S. Stieglitz and L. Dang-Xuan, "Political communication and influence through microblogging–an empirical analysis of sentiment in twitter messages and retalk behavior," in Proc. 45th Hawaii Int.Conf. Syst. Sci., 2012, pp. 3500–3509.
[6] T.-A. Hoang, W. W. Cohen, E.-P. Lim, D. Pierce, and D. P. Redlawsk, "Politics, sharing and emotion in microblogs," in Int. Conf. Adv. Soc. Netw. Anal. Mining, 2013, pp. 282–289.
[7] B. Suh, L. Hong, P. Pirolli, and E. H. Chi, "Want to be retalked?large scale analytics on factors impacting retalk in twitter network," in Proc. IEEE 2nd Int. Conf. Social Comput., 2010,pp. 177–184.
[8] J. A. Berger and K. L. Milkman, "What makes online content viral?" J. Marketing Res., vol. 49, pp. 192–205, 2012.
[9] E. Bakshy, I. Rosenn, C. Marlow, and L. Adamic, "The role of social networks in information diffusion," in Proc. 21st Int. Conf.World Wide Web, 2012, pp. 519–528.
[10] H. Kwak, C. Lee, H. Park, and S. Moon, "What is Twitter, a social network or a news media?" in Proc. 21st Int. Conf. World Wide Web,2010, pp. 591–600.
[11] B. J. Jansen, M. Zhang, K. Sobel, and A. Chowdury, "Twitter power: Talks as electronic word of mouth," J. Amer. Soc. Inform.Sci. Technol., vol. 60, pp. 2169–2188, 2009.
[12] Z. Zhou, R. Bandari, J. Kong, H. Qian, and V. Roychowdhury,"Information resonance on twitter: watching iran," in Proc. 1st Workshop Social Media Anal., 2010, pp. 123–131.
[13] J. H. Parmelee and S. L. Bichard, Politics and the Twitter Revolution:How Tweets Influence the Relationship Between Political Leaders and the Public. Lexington Books, Lanham, MD, 2011.
[14] P. Achananuparp, E.-P. Lim, J. Jiang, and T.-A. Hoang, "Who is retweeting the tweeters? modeling, originating, and promoting behaviors in the twitter network," ACM Trans. Manage. Inform.Syst., vol. 3, 2012, Art. no. 13.

[15] C. Castillo, M. Mendoza, and B. Poblete, "Information credibility on Twitter," in Proc. 20th Int. Conf. World Wide Web, 2011,pp. 675–684.
[16] J. Ratkiewicz, M. Conover, M. Meiss, B. Gonalves, A. Flammini, and F. Menczer, "Detecting and tracking political abuse in social media," in Proc. 5th Int. AAAI Conf. Weblogs Social Media, 2011,pp. 297–304.
[17] D. Gruhl, R. Guha, D. Liben-Nowell, and A. Tomkins,"Information diffusion through blogspace," in Proc. 13th Int. Conf. World Wide Web, 2004, pp. 491–501.
[18] J. Weng, E.-P. Lim, J. Jiang, and Q. He, "Twitterrank: finding topic-sensitive influential twitterers," in Proc. 3rd ACM Int. Conf. Web Search Data Mining, 2010, pp. 261–270.
[19] L. Liu, J. Tang, J. Han, M. Jiang, and S. Yang, "Mining topic-level influence in heterogeneous networks," in Proc. 19th ACM Conf. Inf. Knowl. Manage., 2010, pp. 199–208.
[20] M. Cha, H. Haddadi, F. Benevenuto, and P. K. Gummadi,"Measuring user influence in Twitter: The million follower fallacy," in Proc. Int. AAAI Conf. Weblogs Social Media, 2010,pp. 10–17.
[21] G. Szabo and B. A. Huberman, "Predicting the popularity of online content," Commun. ACM, vol. 53, pp. 80–88, Aug. 2010.
[22] D. Romero, W. Galuba, S. Asur, and B. Huberman, "Influence and passivity in social media," in Proc. 20th Int. Conf. Companion World Wide Web, 2011, pp. 113–114.
[23] P. Cui, F. Wang, S. Liu, M. Ou, S. Yang, and L. Sun, "Who should share what?: item-level social influence prediction for users and posts ranking," in Proc. 34th Int. ACM SIGIR Conf. Res. Develop. Inform. Retrieval, 2011, pp. 185–194.
[24] J. L. Iribarren and E. Moro, "Affinity paths and information diffusion in social networks," Social netw., vol. 33, pp. 134–142, 2011.
[25] C. Wang and D. M. Blei, "Collaborative topic modeling for recommending scientific articles," in Proc. 17th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2011, pp. 448–456.
[26] F. Bonchi, C. Castillo, and D. Ienco, "Meme ranking to maximize posts virality in microblogging platforms," J. Intell. Inform. Syst., vol. 40, pp. 211–239, 2013.
[27] M. Guerini, A. Pepe, and B. Lepri, "Do linguistic style and readability of scientific abstracts affect their virality?" in Proc. 6th Int.AAAI Conf. Weblogs Social Media, 2012, pp. 475–478.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY