iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** XII **Month of publication:** December 2024

DOI: https://doi.org/10.22214/ijraset.2024.65916

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: 7.538
Volume 12 Issue XII Dec 2024- Available at www.ijraset.com

Behavioral Factors Based Content Propagation Modelling

Rupali P. Purkar, Jagdish Y. Kapadnis

Abstract: ***In advertisement and marketing, companies may hire viral users to propagate positive content about their products, or to 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. The modelling of virality and susceptibility factors has many important applications. Also, one may detect events by tracking those mentioned by non-susceptible users and detect rumors based on susceptible user's interactions with the content. As a contribution We will like to incorporate more fine-grained factors affecting the propagation. Fine-grained sentiment is better for reflecting the opinion of the public when they are facing the social focus.***
Keywords: Microblogging, Predictive Models, Receivers, Tensile Stress, Numerical Models, Twitter, Data Models.

I. INTRODUCTION

This system addresses the first challenge by inferring user-content exposure based on the chronological order in microblogging user's timeline and their following network. To address the second challenge, we devise a multi- step heuristic method for removing noise and identifying topics of the content, coupling with the state-of-the-art topic model for microblogging content.

We construct a propagation tensor representing senders content receivers' relationship, and propose a factorization framework on this tensor to simultaneously derive the three-topic specific behavioral factors.

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 [9], in this paper we assume that most relationships among users in a microblogging site are casual and identical in strength. We therefore focus on modeling 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 the advertisement with viral content so as to maximize their reach [10]. Similarly, politicians may leverage on viral users to disseminate their messages widely or to conduct campaigning [11], [12]. Also, one may detect events by tracking those mentioned by no susceptible users [13], and detect rumours based on usucaptible users' interactions with the content [14], [15].

Inter-relationship among user virality, user susceptibility and content virality. Prior empirical researches have suggested there are inter-dependencies among the three factors. Hence, the measurement of a user's susceptibility requires the virality of topics of tweets propagated to her and the virality of users propagating the tweets. The modeling of 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 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 user's interactions with the content.

II. RESEARCH OBJECTIVE AND SCOPE

An objective of this system, to solve constrained factorization problem into an unconstrained optimization which can be solved effectively using gradient descent methods. Gradient descent is a first-order iterative optimization algorithm. To find a local minimum of a function using gradient descent, one takes steps proportional to the negative of the gradient of the function at the current point. State-of-the-art method refers to the highest level of general development, as of a device, technique, or scientific field achieved at a particular time. To predict retweets in a large Twitter dataset and show that the models out perform state-of-the-art methods.

©IJRASET: All Rights are Reserved

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue XII Dec 2024- Available at www.ijraset.com

To Experiment on synthetic datasets to verify the effectiveness of my approach in learning the three behavioral factors. The first set includes the factors external to the user network, e.g., advertising. The second set consists of internal factors due to virality of the user propagating the item, susceptibility of the infected user, and the virality of the item. I want to consider inter-relationships of the factors and their dynamics over time and topics Scope of this system, to incorporate more fine-grained factors affecting the propagation, these factors include positions in the network, linguistic features in content, and emotion factors of users.

In modeling content propagation behavior of users, I have been assuming that users' links are casual and identical in strength. Hence, a natural extension is to relax this assumption by incorporating heterogeneous pair-wise social influence among users.

III. SYSTEM ARCHITECTURE

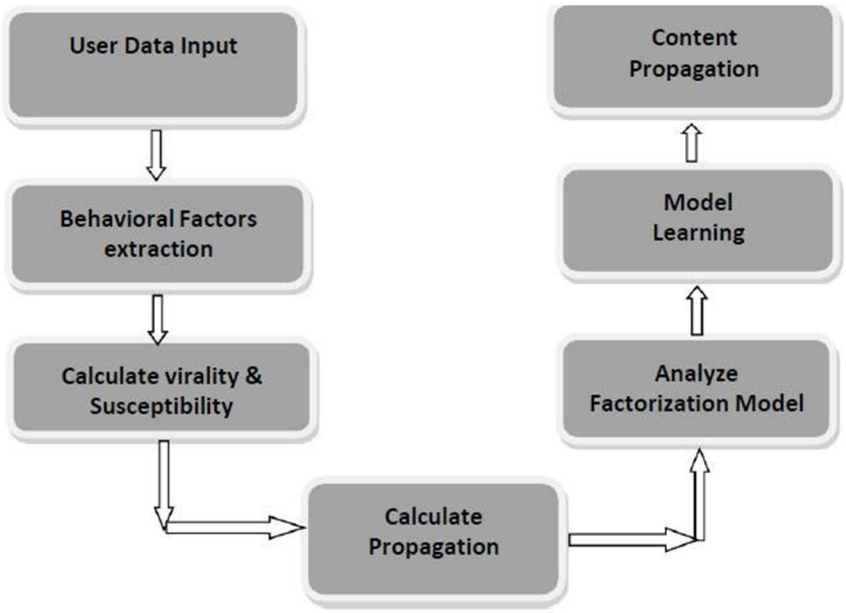


Fig.1-: System Architecture

We propose a tensor factorization framework, calledV2S framework, to model an observed content propagation dataset using three behavioral factors, i.e., topic virality, topic-specific user virality, and topic specific user susceptibility. Within this framework, we develop two factorization methods: Numerical Factorization Method and Probabilistic Factorization Method to simultaneously measure topics' virality as well as topic-specific users' virality and susceptibility.

We convert the above constrained factorization problem into a unconstrained optimization which can be solve deffectively using gradient descent methods.

We apply the V2S - based factorization models to predict retweets in a large Twitter dataset and show that the models outperform state-of-the-art methods.

We also conduct extensive experiments on synthetic datasets to verify the effectiveness of our approach in learning the three behavioral factors.

We address the first challenge by inferring user-content exposure based on the chronological order in microblogging users' timeline and their following network.

To address the second challenge, we devise a multi-step heuristic method for removing noise and identifying topics of the content, coupling with the state-of-the-art topic model for microblogging content. For the third challenge, we construct a propagation tensor representing senders – content- receiver's relationship, and propose a factorization framework on this tensor to simultaneously derive the three-topic specific behavioral factors. We develop two factorization models base on the framework so as to learn the behavioral factors effectively. Lastly, to evaluate the proposed models, we examine the performance of our models in propagation prediction tasks, comparing them with the state-of-the-art baselines. We also use synthetically generated datasets to evaluate the models and the learning algorithm.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

A. Factorization Models

We describe two factorization models built.

1) Numerical Factorization Model

In this model, I consider l(δuvm) as an approximation of δuvm, and f is the identity function. That is,

$$\delta uvm \approx \Sigma K\ k=1[Dm,k.\ Vu,k\ .Ik.\ Sv,k] \quad \ldots..(1)$$

Given the approximation in Equation 1, the loss function Rl,f (u, v,m) is then the squared loss, defined as follows.

$$Rl,f\ (u,\ v,m) = (\delta uvm - \Sigma K\ k=1[Dm,k\ .Vu,k\ .\ Ik\ .Sv,k])2 \quad \ldots(2)$$

2) Probabilistic Factorization Model

In this model, I consider l(δuvm) as the likelihood of δuvm, and f is a probability distribution. Since δuvm∈{0, 1},
I choose f to be the Bernoulli distribution with mean μ(u, v,m) = ΣK k=1[Dm,k . Vu,k .Ik .Sv,k].
That is,

$$l(\delta uvm) = \mu(u,\ v,m)_uvm\ .\ (1 - \mu(u,\ v,m))(1-_uvm) \quad \ldots..(3)$$

The loss function Rl,f (u, v,m) is now the negative log likelihood of δuvm, defined as follows: $Rl,f\ (u,\ v,m) = -\delta uvm\ .\ \ln(\mu(u,\ v,m)) -- (1 - \delta uvm).\ \ln(1 - \mu(u,\ v,m))$ ….(4)

B. Contribution

As a contribution we will like to incorporate more fine-grained factors affecting the propagation. Fine-grained sentiment is better for reflecting the opinion of the public when they are facing the social focus.
We propose a tensor factorization framework, called V2S framework, to model an observed content propagation dataset using three behavioral factors, i.e., topic virality, topic-specific user virality, and topic specific user susceptibility. Within this framework, we develop two factorization methods: Numerical Factorization Method and Probabilistic Factorization Method to simultaneously measure topics' virality as well as topic-specific users' virality and susceptibility.
We convert the above constrained factorization problem into a unconstrained optimization which can be solved effectively using gradient descent methods. We apply the V2S - based factorization models to predict retweets in a large Twitter dataset and show that the models outperform state-of-the-art methods. We also conduct extensive experiments on synthetic datasets to verify the effectiveness of our approach in learning the three behavioral factors.

IV. METHODOLOGY

A. Model Learning Algorithm

With respect to the loss defined in above Equations, the objective function L(I,V,S) defined in above is not a convex function of (I,V,S) but a convex function of I, V, and S respectively. Hence, the problem can be solved efficiently by alternating gradient descent methods. However, due to the conditions, I cannot apply the methods directly as they require variables unconstrained. To deal with these conditions, I first tried the projected gradient descent method.
However, this method only returns locally optimal solutions for the alternating optimization problems (i.e., minimizing L(I,V,S) with respect to I, V, or S), and hence results in poor solutions for Problem. Hence, I make use of the following variable transformation to transform the constrained variables into unconstrained ones.

B. Datasets

1) *Data Collection:* The dataset used in this work was collected from Twitter by a snowball sampling-based crawler. We first manually selected a set of highly followed Twitter users in Singapore. They include the accounts of local sport and entertainment celebrities, political parties, politicians, mass media and bloggers, etc. We expanded this set of users by adding more Singapore based users1 that are at most two hops away from some user in the original set. Using Twitter Stream APIs2, we then obtained all tweets and retweets by the users in the set. In this work, we use all tweets in October2014 to simulate a live tweet stream. This set includes 35,491,260 tweets and retweets posted by 525,632 users.
2) *Item Adoption and Propagation:* We again use as an item. We consider a user u adopts a hash tag when u posts an original tweet containing the hash tag. Also, if user v retweets original tweets from u that contains a hash tag h, u is said to propagate h to v. We filtered away hash tags shorter than 2 characters excluding the # symbol. These short hash tags do not have clear semantics and are often the prefix of other truncated hash tags due to 140 characters length constraint. We also excluded hash tags longer than 20 characters as such hash tags are unpopular.

V. CONCLUSION

In this paper, we 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, we propose to model the factors to topic level. We develop V2S, a tensor factorization- based framework and its associated models, to learn topic-specific user virality and susceptibility, and topic virality from content propagation data. Our experiments on a large Twitter dataset show that the proposed V2S-based models outperform baseline models significantly in propagation prediction. Our experiments on synthetic databases also show that our proposed models outperform all the other baseline methods in learning the topic-specific factors.

VI. ACKNOWLEDGMENT

I take this opportunity to express my sincere thanks and deep sense of gratitude to my team member for valuable guidance.

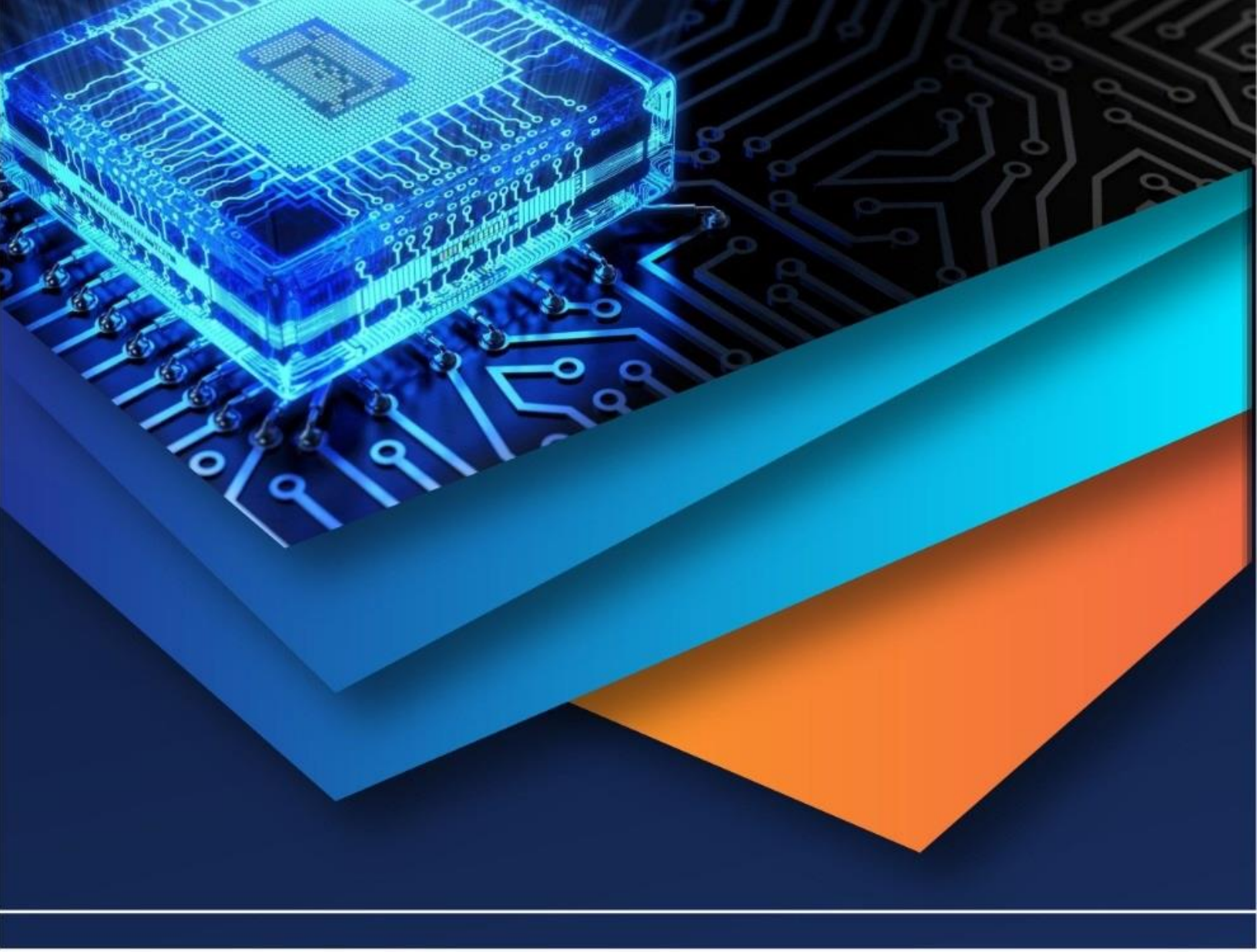
REFERENCES

[1] Tuan-Anh Hoang and Ee-Peng Lim “Microblogging Content Propagation Modeling Using Topic- specific Behavioral Factors”, in IEEE 2016.
[2] S. A.MacskassyandM. Michelson, “Why do people retweet? Antihomophily wins the day!” in ICWSM, 2011.
[3] Z. Liu, L. Liu, and H. Li, “Determinants of information retweeting in microblogging,” Internet Research, 2012.
[4] S. Stieglitz and L. Dang-Xuan, “Political communication and influence through microblogging–an empirical analysis of sentiment in twitter messages and retweet behavior,” in HICSS, 2012.
[5] T.-A. Hoang, W. W. Cohen, E.-P.Lim, D. Pierce, and D. P.Redlawsk, “Politics, sharing and emotion in microblogs,” in ASONAM, 2013.
[6] B. Suh, L. Hong, P. Pirolli, and E. H. Chi, “Want to be retweeted? large scale analytics on factors impacting retweet in twitter network,” in Social Com, 2010.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

[7] J. A. Berger and K. L. Milkman, "What makes online content viral?" Journal of Marketing Research, 2012.
[8] E. Bakshy, I. Rosenn, C.Marlow, and L. Adamic, "The role of social networks in information diffusion," in WWW, 2012.
[9] H. Kwak, C. Lee, H. Park, and S. Moon, "What is twitter, a social network or a news media?" in WWW, 2010.
[10] B. J. Jansen, M. Zhang, K. Sobel, and A. Chowdury, "Twitter power: Tweets as electronic word of mouth," JASIST, 2009.
[11] Z. Zhou, R. Bandari, J. Kong, H. Qian, and V. Roy chowdhury, "Information resonance on twitter: watching iran," in SOMA,2010.
[12] 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, 2011.
[13] 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 TMIS, 2012.
[14] C. Castillo, M. Mendoza, and B. Poblete, "Information credibility on twitter," in WWW, 2011.
[15] J. Ratkiewicz, M. Conover, M. Meiss, B. Gonalves, A. Flammini, and F. Menczer, "Detecting and tracking political abuse in social media," in ICWSM, 2011.
[16] D. Gruhl, R. Guha, D. Liben-Nowell, and A. Tomkins, "Information diffusion through blog space," in WWW, 2004.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)