International Journal of Research
Available at https://journals.pen2print.org/index.php/ijr/
e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 07 Issue 06
June 2020

Health Monitoring on Social Media over Time

Aakula Chandrakala[1], K.Charan Theja[2]
[1]P.G. Scholar, [2]Guide, Assistant Professor
[1,2] Dept. of COMPUTER SCIENCE ENGINEERING
[1,2] Geethanjali College of Engnieering and technology
Email: [1]chandrakala.aakula@gmail.com, [2]charantheja.2628@gmail.com

ABSTRACT

Social media has become a major source for analyzing all aspects of daily life. Thanks to dedicated latent topic analysis methods such as the Ailment Topic Aspect Model (ATAM), public health can now be observed on Twitter. In this work, we are interested in using social media to monitor people's health over time. The use of tweets has several benefits including instantaneous data availability at virtually no cost. Early monitoring of health data is complementary to post-factum studies and enables a range of applications such as measuring behavioral risk factors and triggering health campaigns. We formulate two problems: health transition detection and health transition prediction. We first propose the Temporal Ailment Topic Aspect Model (TM–ATAM), a new latent model dedicated to solving the first problem by capturing transitions that involve health-related topics. TM–ATAM is a non-obvious extension to ATAM that was designed to extract health-related topics. It learns health-related topic transitions by minimizing the prediction error on topic distributions between consecutive posts at different time and geographic granularities. To solve the second problem, we develop T–ATAM, a Temporal Ailment Topic Aspect Model where time is treated as a random variable natively inside ATAM. Our experiments on an 8-month corpus of tweets show that TM–ATAM outperforms TM–LDA in estimating health-related transitions from tweets for different geographic populations. We examine the ability of TM–ATAM to detect transitions due to climate conditions in different geographic regions. We then show how T–ATAM can be used to predict the most important transition and additionally compare T–ATAM with CDC (Center for Disease Control) data and Google Flu Trends.

Keywords: Social Media, TM–ATAM, CDC, TM–LDA

INTRODUCTION

Social media has become a major source of information for analyzing all aspects of daily life. In particular, Twitter is used for public health monitoring to extract early indicators of the well-being of populations in different geographic regions. Twitter has become a major source of data for early monitoring and prediction in areas such as health [1], disaster management [2] and politics [3]. In the health do main, the ability to model transitions for ailments and detect statements like "people talk about smoking and cigarettes before talking about respiratory problems", or "people talk about headaches and stomach ache in any order", benefits syndromic surveillance and helps measure behavioural risk factors and trigger public health campaigns. In this paper, we formulate two problems: the health transition detection problem and the health transition prediction problem. To address the detection problem, we develop TM–ATAM that models temporal transitions of health-related topics. To address the prediction problem, we propose T–ATAM, a novel method which uncovers latent ailment inside tweets by treating time as a random variable natively inside ATAM[4]. Treating time as a random variable is key to predicting the subtle change in health-related discourse on Twitter.

Common ailments are traditionally monitored by collecting data from health-care facilities, a process known as sentinel surveillance. Such resources limit surveillance, most especially for real-time feedback. For this reason, the Web has become a source of syndromic surveillance, operating on a wider scale, near real time and at virtually no cost. Our challenges are: (i) identify health-related tweets, (ii) determine when health-related discussions on Twitter transitions from one topic to another, (iii) capture different such transitions for different geographic regions. Indeed, in addition to evolving over time, ailment distributions also evolve in space.

Therefore, to attain effectiveness, we must carefully model two key granularities, temporal and geographic. A temporal granularity that is too-fine may result in sparse and spurious transitions whereas a too-coarse one could miss valuable ailment transitions. Similarly, a too-fine geographic granularity may produce false positives and a too-coarse one may miss meaningful transitions, e.g., when it concerns users living in different climates. For example, discussions on allergy break at different periods in different states in the USA [4].Therefore, processing all tweets originating from the USA together will miss climate variations that affect people's health. We argue for the need to consider different time granularities for different regions and we wish to identify and model the evolution of ailment distributions between different temporal granularities.

While several latent topic modeling methods such as Probabilistic Latent Semantic Indexing (PLSI) [5] and Latent Dirichlet Allocation (LDA) [6], have been proposed to effectively cluster and classify general-purpose text, it has been shown that dedicated methods such as the Ailment Topic Aspect Model (ATAM) are better suited for capturing ailments in Twitter [4]. ATAM extends LDA to model how users express ailments in tweets. It assumes that each health-related tweet reflects a latent ailment such as flu and allergies. Similar to a topic, an ailment indexes aword distribution. ATAM also maintains a distribution over symptoms and treatments. This level of detail provides a more accurate model for latent ailments.

On the other hand, while PLSI and LDA have been shown to perform well on static documents, they cannot intrinsically capture topic evolution over time. Temporal-LDA (TM–LDA) was proposed as an extension to LDA formining topics from tweets over time [7]. To address the health transition detection problem, we propose TM–ATAM that combines ATAM and TM–LDA. A preliminary version of TM–ATAM was described in a short paper [8]. We show here that it is able to capture transitions of health-related discussions in different regions (see Figure 1). As a result, the early detection of a change in discourse in Nevada, USA into allergies can trigger appropriate campaigns.

In each geographic region, TM–ATAM learns transition parameters that dictate the evolution of health-related topics by minimizing the prediction error on ailment distribution of consecutive pre-specified periods of time. Our second problem, the health transition prediction problem, is to automatically determine those periods. We hence propose T–ATAM, a different and new model that treats time as a random variable in the generative model. T–ATAM discovers latent ailments in health tweets by treating time as a variable whose values are drawn from a corpus-specific multinomial distribution. Just like TM–LDA, TM–ATAM and T–ATAM are different from dynamic topic models [9], [10], [11],as they are designed to learn topic transition patterns from temporally-ordered posts, while dynamic topic models focus on changing word distributions of topics over time.

Our experiments on a corpus of more than 500K health related tweets collected over an 8-month period, show that TM–ATAM outperforms TM–LDA in estimating temporal topic transitions of different geographic populations. Our results can be classified in two kinds of transitions. Stable topics are those where a health-related topic is mentioned continuously. One-Way transitions cover the case where some topics are discussed after others. For example, our study of tweets from California revealed many stable topics such as headaches and migraines. On the other hand, tweeting about smoking, drugs and cigarettes is followed by tweeting about respiratory ailments. Figure 1 shows example one way transitions we extracted for different states and cities in the world. Such transitions are often due to external factors such as climate, health campaigns, nutrition and lifestyle of different world populations

EXISTING SYSTEM

In the existing system, the authors propose a method that learns changing word distributions of topics over time and in the system, the authors leverage the structure of a social network to learn how topics temporally evolve in a community. TM–ATAM and T–ATAM are however different from dynamic topic models such as [9] and [10], and from the work of Wang et al. [11], as they are designed to learn topic transition patterns from temporally-ordered posts, while dynamic topic models focus on changing word distributions of topics over time.

TM–ATAM learns transition parameters that dictate the evolution of health-related topics by minimizing the prediction error on ailment distributions of consecutive periods at different temporal and geographic granularities. T–ATAM on the other hand discovers latent ailments in health tweets by treating time as a corpus-specific multinomial distribution.

Classical approaches have been applied to mining topics for inferring citations. Other discriminative approaches have been applied to do an empirical study on topic modeling and time-based topic modeling respectively. None of those are directly applicable to health data.

Disadvantages of Existing System:

There is no Mapping Tweets to Documents.

There is Uncovering Health Topics with ATAM.

PROPOSED SYSTEM

In the proposed system, the system formulates and solves two problems: the health transition detection problem and the health transition prediction problem. To address the detection problem, the system develops TM–ATAM that models temporal transitions of health-related topics. To address the prediction problem, we propose T–ATAM, a novel method which uncovers latent ailment inside tweets by treating time as a random variable natively inside ATAM. Treating time as a random variable is key to predicting the subtle change in health-related discourse on Twitter.

Advantages of Proposed System:

TM–ATAM, a model able to detect health-related tweets and their evolution over time and space. TM–ATAM learns, for a given region, transition parameters by minimizing the prediction error on ailment distributions of pre-determined time periods.

T–ATAM, a new model able to predict health-related tweets by treating time as a variable whose values are drawn from a corpus-specific multinomial distribution.

Extensive experiments that show the superiority of T–ATAM for predicting health transitions, when compared against TM–LDA and TM–ATAM, and its effectiveness against a ground truth.

Module Description:

International Journal of Research
Available at https://journals.pen2print.org/index.php/ijr/
e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 07 Issue 06
June 2020

Admin Module

In this module, the Admin has to login by using valid user name and password. After login successful he can perform some operations such as View All Users And Authorize, View All Friend Request and Response, Add Health Filter, View All Health Tweets with Discussion Comments, Capture and View Different Health Monitoring for different geographic regions, Capture and View Different Health Monitoring Based On Disease, View Number of Same Disease in Chart, View Health Tweet Scores in Chart

Friend Request & Response

In this module, the admin can view all the friend requests and responses. Here all the requests and responses will be displayed with their tags such as Id, requested user photo, requested user name, user name request to, status and time & date. If the user accepts the request then the status will be changed to accepted or else the status will remains as waiting.

User Module

In this module, there are n numbers of users are present. User should register before performing any operations. Once user registers, their details will be stored to the database. After registration successful, he has to login by using authorized user name and password. Verify finger print and Login Once Login is successful user can perform some operations like My Profile, Search Friend Track and Find Friend Request, View All My Friends, Create Your Health Tweet, View All My Health Tweets, View and Monitor All My Friends Health Tweets.

Searching Users to make friends

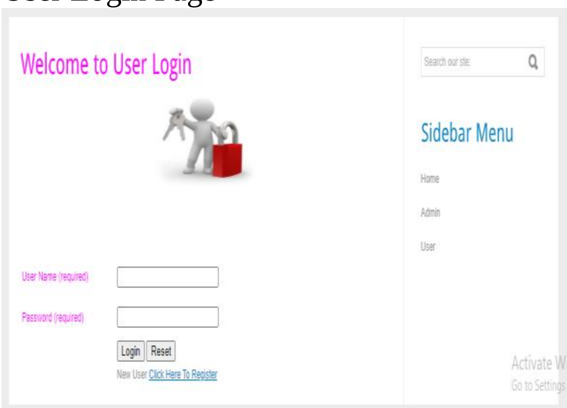
In this module, the user searches for users in Same Network and in the Networks and sends friend requests to them. The user can search for users in other Networks to make friends only if they have permission.

Home Page

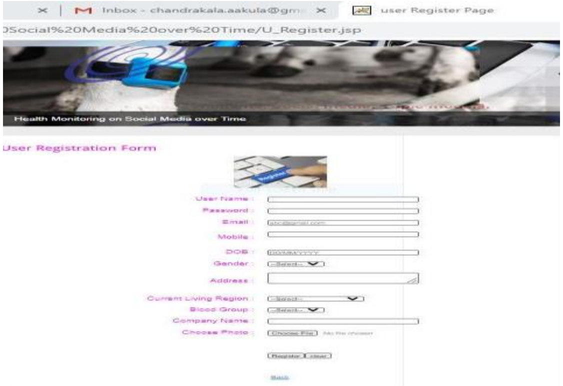
Admin login Page

Admin Page

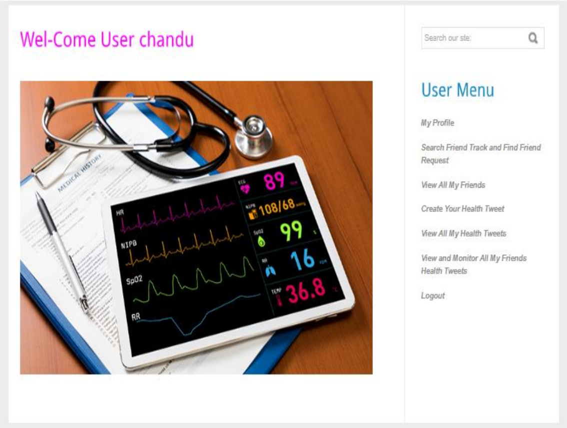
User Login Page



New User Registration

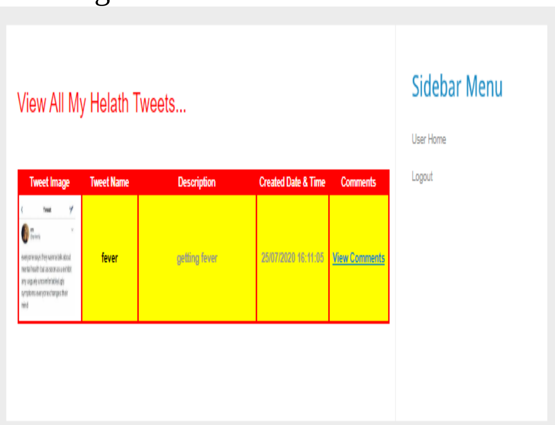


User Home Page

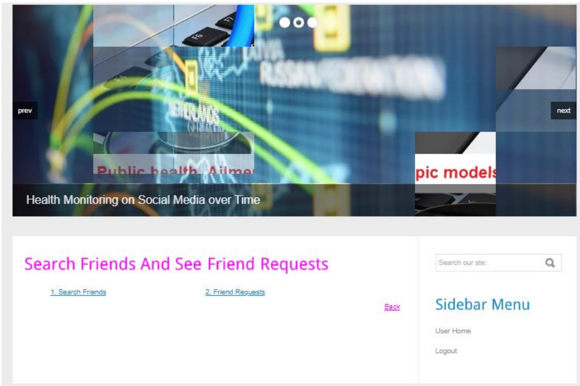


Health Tweet

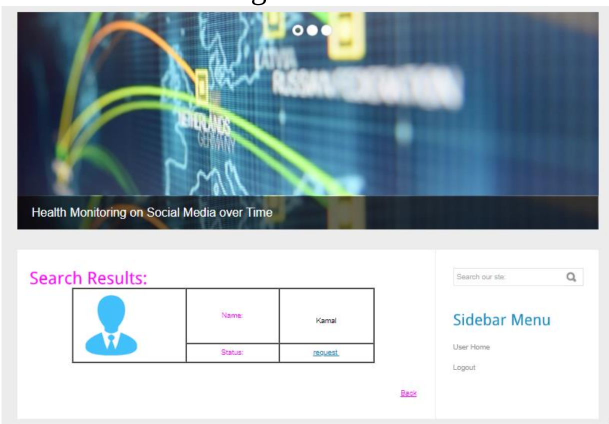
Viewing Health Tweets



Searching Friends



Results of Searching



CONCLUSION AND FUTURE WORK

We develop methods to uncover ailments over time from social media. We formulated health transition detection and prediction problems and proposed two models to solve them. Detection is addressed with TM–ATAM, a granularity-based model to conduct region-specific analysis that leads to the identification of time periods and characterizing homogeneous disease discourse, per region. Prediction is addressed with T–ATAM, that treats time natively as a random variable whose values are drawn from a multinomial distribution. The fine-grained nature of T–ATAM results insignificant improvements in modelling and predicting transitions of health-related tweets. We believe our approach is applicable to other domains with time-sensitive topics such as disaster management and national security matters.

REFERENCES

1. L. Manikonda and M. D. Choudhury, “Modeling and understanding visual attributes of mental health disclosures in social media,”in Proceedings of the 2017 CHI Conference on Human Factors inComputing Systems, Denver, CO, USA, May 06-11, 2017., 2017, pp.170–181.
2. S. R. Chowdhury, M. Imran, M. R. Asghar, S. Amer-Yahia, andC. Castillo, “Tweet4act: Using incident-specific profiles for classifyingcrisis-related messages,” in 10th Proceedings of the InternationalConference on Information Systems for Crisis Response andManagement, Baden-Baden, Germany, May 12-15, 2013., 2013.
3. T. Davidson, D. Warmsley, M. W. Macy, and I. Weber, “Automatedhate speech detection and the problem of offensive language,” inProceedings of the Eleventh International Conference on Web and SocialMedia, ICWSM 2017, Montréal, Québec, Canada, May 15-18, 2017.,2017, pp. 512–515.
4. [4] M. J. Paul and M. Dredze, “You Are What You Tweet: AnalyzingTwitter for Public Health,” in ICWSM’11, 2011.
5. T. Hofmann, “Probabilistic Latent Semantic Indexing,” in SIGIR’99, 1999, pp. 50–57.
6. D. M. Blei, A. Y. Ng, and M. I. Jordan, “Latent DirichletAllocation,”Journal of Machine Learning, vol. 3, pp. 993–1022, 2003.
7. Y. Wang, E. Agichtein, and M. Benzi, “TM-LDA: Efficient OnlineModeling of Latent Topic Transitions in Social Media,” in KDD’12,2012, pp. 123–131.
8. S. Sidana, S. Mishra, S. Amer-Yahia, M. Clausel, and M. Amini,“Health monitoring on social media over time,” in Proceedingsof the 39th International ACM SIGIR conference on Research andDevelopment in Information Retrieval, SIGIR 2016, Pisa, Italy, July17-21, 2016, 2016, pp. 849–852.
9. D. M. Blei and J. D. Lafferty, “Dynamic Topic Models,” in ICML’06,2006, pp. 113–120.
10. [10] C. X. Lin, Q. Mei, J. Han, Y. Jiang, and M. Danilevsky, “The JointInference of Topic Diffusion and Evolution in Social Communities,”in ICDM’11, 2011, pp. 378–387.
11. X. Wang and A. McCallum, “Topics Over Time: A Non-MarkovContinuous-time Model of Topical Trends,” in KDD’06, 2006, pp.424–433.
12. K. W. Prier, M. S. Smith, C. Giraud-Carrier, and C. L. Hanson,“Identifying Health-related Topics On Twitter,” in Social computing,behavioral-cultural modeling and prediction. Springer, 2011, pp.18–25.
13. C. Cortes and V. Vapnik, “Support-vector networks,” MachineLearning, vol. 20, no. 3, pp. 273–297, 1995. [Online]. Available: http://dx.doi.org/10.1007/BF00994018
14. M. De Choudhury, “Anorexia on Tumblr: A CharacterizationStudy,” in DH’15, 2015, pp. 43–50.
15. M. De Choudhury, A. Monroy-Hernández, and G. Mark, “"narco"Emotions: Affect and Desensitization in Social Media During theMexican Drug War,” in CHI’14, 2014, pp. 3563–3572.
16. U. Pavalanathan and M. De Choudhury, “Identity Managementand Mental Health Discourse in Social Media,” in WWW’15, 2015,pp. 315–321.
17. F. Bouillot, P. Poncelet, M. Roche, D. Ienco, E. Bigdeli, andS. Matwin, “French Presidential Elections: What are the MostEfficient Measures for Tweets?” in PLEAD’12. ACM, 2012, pp.23–30.
18. L. Hemphill and A. J. Roback, “Tweet Acts: How ConstituentsLobby Congress via Twitter,” in CSCW’14, 2014, pp. 1200–1210.
19. A. Ceron, L. Curini, and S. M. Iacus, “Using Sentiment Analysis toMonitor Electoral Campaigns: Method Matters-

Evidence from theUnited States and Italy," Soc. Sci. Comput. Rev., vol. 33, no. 1, pp.3–20, 2015.

20. P. Barberá, "Birds of The Same Feather Tweet Together: BayesianIdeal Point Estimation using Twitter Data," Political Analysis,vol. 23, no. 1, pp. 76–91, 2015.
21. L. Jiang, M. Yu, M. Zhou, X. Liu, and T. Zhao, "Target-dependentTwitter Sentiment Classification," in HLT'11, 2011, pp. 151–160.