A PROMINENT TEMPORAL AILMENT TOPIC ASPECT MODEL FOR HEALTH MANAGEMENT SYSTEM

Shankar Reddy P[1], Padmavathamma M[2]
[1]PG Student, Department of Computer Science, Sri Venkateshwara University Tirupati
[2]Professor, Department of Computer Science, Sri Venkateshwara University Tirupati

Abstract
Web based life has turned into a noteworthy hotspot for breaking down all parts of day by day life. Because of committed inert theme examination strategies, for example, the Ailment Topic Aspect Model (ATAM), general wellbeing would now be able to be seen on Twitter. In this work, we are keen on utilizing web-based life to screen individuals' wellbeing after some time. The utilization of tweets has a few advantages including quick information accessibility at basically no expense. Early observing of wellbeing information is correlative to post-factum contemplates and empowers a scope of utilizations, for example, estimating conduct hazard factors and activating wellbeing efforts. We plan two issues: wellbeing changes recognition and wellbeing progress forecast. We initially propose the Temporal Ailment Topic Aspect Model (TM– ATAM), another dormant model devoted to taking care of the primary issue by catching advances that include wellbeing related themes. TM– ATAM is a non-evident expansion to ATAM that was intended to extricate wellbeing related subjects. It learns wellbeing related point advances by limiting the expectation mistake on subject circulations between back to back posts at various time and geographic granularities. To tackle the second issue, we create T– ATAM, a Temporal Ailment Topic Aspect Model where time is treated as an irregular variable locally inside ATAM. Our investigations on an 8-month corpus of tweets demonstrate that TM– ATAM beats TM– LDA in assessing wellbeing related changes from tweets for various geographic populaces. We analyze the capacity of TM– ATAM to recognize advances because of atmosphere conditions in various geographic districts. We at that point show how T– ATAM can be utilized to foresee the most essential change and furthermore contrast T– ATAM and CDC (Center for Disease Control) information and Google Flu Trends.
Keywords: Ailments, Temporal Ailment Topic Aspect Model, Health Transition Prediction.

I. INTRODUCTION

Web based life has turned into a noteworthy wellspring of data for breaking down all parts of everyday life. Specifically, Twitter is utilized for general wellbeing observing to separate early pointers of the prosperity of populaces in various geographic locales. Twitter has turned into a noteworthy wellspring of information for ahead of schedule observing and forecast in zones, for example, wellbeing [1], fiasco the executives [2] and governmental issues [3]. In the wellbeing area, the capacity to display advances for infirmities and distinguish proclamations like "individuals talk about smoking and cigarettes prior to discussing respiratory issues", or "individuals talk about cerebral pains and stomach throb in any request", benefits syndromic observation and helps measure social hazard factors and trigger general wellbeing efforts. In this paper, we define two issues: the wellbeing progress location issue and the wellbeing progress forecast issue. To address the discovery issue, we create TM– ATAM that models worldly changes of wellbeing related points. To address the forecast issue, we propose T– ATAM, a novel strategy which reveals dormant sickness inside tweets by treating time as an irregular variable locally inside ATAM [4]. Treating time as an irregular variable is critical to anticipating the inconspicuous change in wellbeing related talk on Twitter. Basic afflictions are generally observed by gathering information from medicinal services offices, a procedure known as sentinel reconnaissance. Such assets limit observation, most particularly for ongoing criticism. Consequently, the Web has turned into a wellspring of syndromic observation, working on a more extensive scale, close ongoing and at for all intents and purposes no cost. Our difficulties are: (I) recognize wellbeing related tweets, (ii) decide when wellbeing related talks on Twitter advances starting with one subject then onto the next, (iii) catch diverse such advances for various geographic areas. To be sure, notwithstanding developing after some time, illness conveyances likewise advance in space. In this manner, to accomplish adequacy, we should cautiously display two key granularities, transient and geographic. A worldly granularity that is too-fine may result in meager and fake advances though a too-coarse one could miss significant illness advances. So also, a too-fine geographic granularity may deliver false positives and a too-coarse one may miss important changes, e.g., when it concerns clients living in various atmospheres. For instance, exchanges on hypersensitivity break at various periods in various states in the USA [4]. In this manner, handling all tweets starting from the USA together will miss atmosphere varieties that influence individuals' wellbeing. We contend for the need to think about various time granularities for various districts and we wish to distinguish also, show the advancement of affliction conveyances between diverse fleeting granularities. While several latent topic modelling 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 a word 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 for mining 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 distributions 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.

II RELATED WORK

Proliferation of social media platforms such as Twitter, pinterest, facebook, tumblr has led to their application to a wide array of tasks including mental health assessment [14], [15], [16], inferring political affiliation [17], [18], [19], [20], brand perception [21], [22] etc. Social media, especially Twitter, are good sources of personal health [23], [24], [25], [26]. Previous studies on public health surveillance have attempted to uncover ailment topics on online discourse [4], [27] or model the evolution of general topics [7]. In this paper, we combine the best of both worlds which leads to the discovery of disease-change-points for social-media active regions. We model the evolution of diseases within change-points and obtain significant improvement over the state-of-the-art for public health surveillance using social media.

Just like TM-LDA, TM-ATAM and T–ATAM learn topic transitions over time and not topic trends. Such transitions the purpose of answering questions such as people talk about fever before talking about stomach ache. Other complementary approaches that learn the dynamicity of word distributions or topic trends have been proposed. That is the case of [9] that models topic evolution over time as a discrete chain-style process where each piece is modeled using LDA. In [11], the authors propose a method that learns changing word distributions of topics over time and in [10], 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 [28] have been applied to mining topics for inferring citations. Other discriminative approaches [29], [30] 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.

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

There is no Mapping Tweets to Documents.

There is Uncovering Health Topics with ATAM.

II 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

- 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.

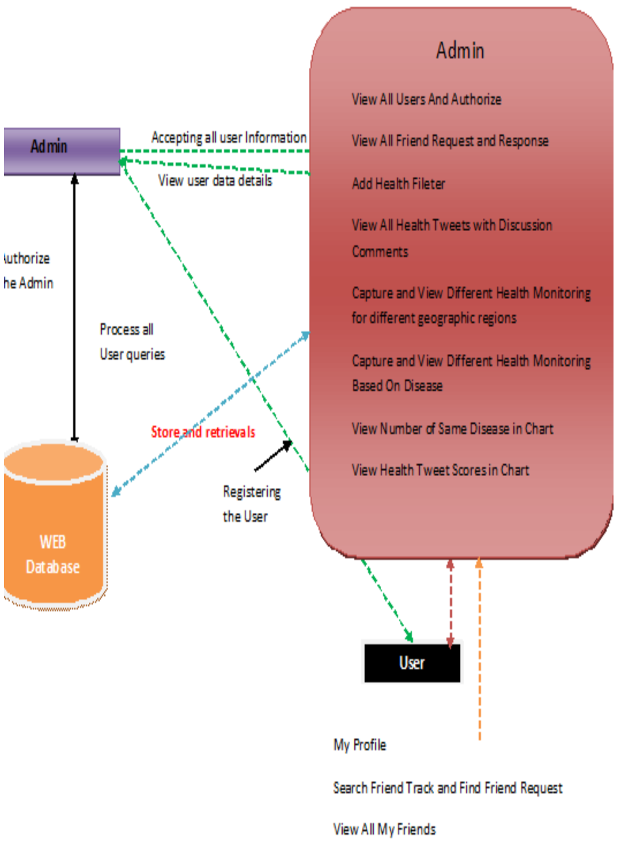
IV METHODOLOGY

TM–ATAM assumes that there is a common linear relation between all the aggregate topic distributions at a given period t and the one at the period just before. TM–ATAM fails to perform optimally when operated in regions where there are no substantial transitions in health topics. In particular, TM–ATAM does not take into account the potential seasonality effect, which maybe very different according to the disease of interest. Also, in TM–ATAM, we need to do post processing in order to come up with homogeneous time periods, with respect to health-topics discussed in tweets.

Admin

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

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.

V CONCLUSION

In this paper, 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.

VI 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 in Computing Systems, Denver, CO, USA, May 06-11, 2017., 2017, pp.170–181.

[2] S. R. Chowdhury, M. Imran, M. R. Asghar, S. Amer-Yahia, and C. Castillo, “Tweet4act: Using incident-specific profiles for classifying crisis-related messages,” in 10th Proceedings of the International Conference on Information Systems for Crisis Response and Management, Baden-Baden, Germany, May 12-15, 2013., 2013.

[3] T. Davidson, D. Warmsley, M. W. Macy, and I. Weber, “Automated hate speech detection and the problem of offensive language,” in Proceedings of the Eleventh International Conference on Web and Social Media, ICWSM 2017, Montréal, Québec, Canada, May 15-18, 2017.,2017, pp. 512–515.

[4] M. J. Paul and M. Dredze, “You Are What You Tweet: Analysing Twitter 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 Dirichlet Allocation,” Journal of Machine Learning, vol. 3, pp. 993–1022, 2003.

[7] Y. Wang, E. Agichtein, and M. Benzi, “TM-LDA: Efficient Online Modelling 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 Proceedings of the 39th International ACM SIGIR conference on Research and Development 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] C. X. Lin, Q. Mei, J. Han, Y. Jiang, and M. Danilevsky, “The Joint Inference 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-Marko Continuous-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,” Machine Learning, 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 Characterization Study,” in DH’15, 2015, pp. 43–50.

[15] M. De Choudhury, A. Monroy-Hernández, and G. Mark, “"Marco "Emotions: Affect and Desensitization in Social Media During the Mexican Drug War,” in CHI’14, 2014, pp. 3563–3572.

[16] U. Pavalanathan and M. De Choudhury, “Identity Management and 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 Most Efficient Measures for Tweets?" in PLEAD'12. ACM, 2012, pp.23–30.

[18] L. Hemphill and A. J. Roback, "Tweet Acts: How Constituents Lobby Congress via Twitter," in CSCW'14, 2014, pp. 1200–1210.

[19] A. Ceron, L. Curini, and S. M. Iacus, "Using Sentiment Analysis to Monitor Electoral Campaigns: Method Matters-Evidence from the United 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: Bayesian Ideal 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-dependent Twitter Sentiment Classification," in HLT'11, 2011, pp. 151–160.

P SHANKAR REDDY he is a master of Computer Science (M.Sc) pursuing in Sri Venkateswara University, Tirupati, A.P. He received Degree of Bachelor of Science in 2017 from Rayalaseema University, Kurnool. His research interests are Cloud Computing, Internet of Things (IoT), and Artificial Intelligence.