International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

SOCIAL MEDIA MINING FOR PUBLIC HEALTH MONITORING AND SURVEILLANCE

T.RAJANI[1], M.PRADEEP[2]
[1] PG SCHOLAR, DEPT OF CSE, ST.MARY'S GROUP OF INSTITUTION, GUNTUR, AP, INDIA.
[2]ASST. PROFESSOR[M.TECH], DEPARTMENT OF CSE, ST.MARY'S GROUP OF INSTITUTION, GUNTUR, AP, INDIA.

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.

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 domain, 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 syndromes' surveillance and helps measure behavioral 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 syndrome 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 US 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 identifying 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 asflu and allergies. Similar to a topic, an ailment indexes award distribution. ATAM also maintains a distribution over symptoms and treatments. This level of detail provides amore accurate model for latent ailments.

LITERATUREREVIEW

Modeling and understanding visual attributes of mental health disclosures in social media by L. Manikonda and M. D. Choudhury

Content shared on social media platforms has been identified to be valuable in gaining insights into people's mental health experiences. Although there has been widespread adoption of photo-sharing platforms such as Instagram in recent years, the role of visual imagery as a mechanism of self-disclosure is less understood. We study the nature of visual attributes manifested in images relating to mental health disclosures on Instagram. Employing computer vision techniques on a corpus of thousands of posts, we extract and examine three visual attributes: visual features (e.g., color), themes, and emotions in images. Our findings indicate the use of imagery for unique self-disclosure needs, quantitatively and qualitatively distinct from those shared via the textual modality: expressions of emotional distress, calls for help, and explicit display of vulnerability. We discuss the relationship of our findings to literature in visual sociology, in mental health self disclosure, and implications for the design of health interventions.

Tweet4act: Using incident-specific profiles for classifyingcrisis-related messages by S. R. Chowdhury, M. Imran, M. R. Asghar, S. Amer-Yahia, andC. Castillo

We present Tweet4act, a system to detect and classify crisis-related messages communicated over a microblogging platform. Our system relies on extracting content features from each message. These features and the use of an incident-specific dictionary allow us to determine the period type of an incident that each message belongs to. The period types are: pre-incident (messages talking about prevention, mitigation, and preparedness), during-incident (messages sent while the incident is taking place), and post-incident (messages related to the response, recovery, and reconstruction). We show that our detection method can effectively identify incident-related messages with high precision and recall, and that our incident-period classification method outperforms standard machine learning classification methods.

Automatedhate speech detection and the problem of offensive language by T. Davidson, D. Warmsley, M. W. Macy, and I. Weber

A key challenge for automatic hate-speech detection on social media is the separation of hate speech from other instances of offensive language. Lexical detection methods tend to have low precision because they classify all messages containing particular terms as hate speech and previous work using supervised learning has failed to distinguish between the two categories. We used a crowd-sourced hate speech lexicon to collect tweets containing hate speech keywords. We use crowd-sourcing to label a sample of these tweets into three categories: those containing hate speech, only offensive language, and those with neither. We train a multi-class classifier to distinguish between these different categories. Close analysis of the predictions and the errors shows when we can reliably separate hate speech from other offensive language and when this differentiation is more difficult. We find that racist and homophobic tweets are more likely to be classified as hate speech but that sexist tweets are generally classified as offensive. Tweets without explicit hate keywords are also more difficult to classify.

You Are What You Tweet:

3 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

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
IJARST
ISSN: 2457-0362

modeling respectively. None of those are directly applicable to health data.
There is no Mapping Tweets to Documents.
There is Uncovering Health Topics with ATAM.

4.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.
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.

5.SYSTEM ARCHITECTURE:

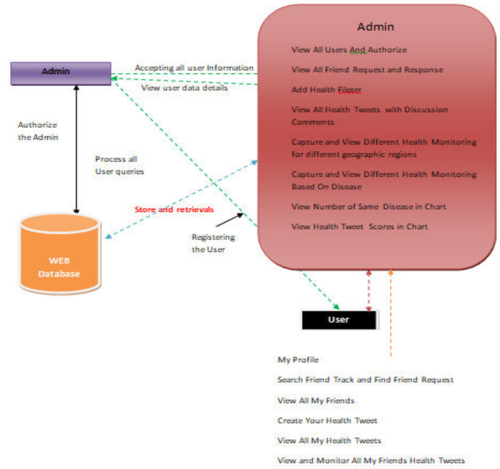

Fig 1 Architecture Diagram

6.IMPLEMENTATION

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

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

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.

7.SCREEN SHOTS

8.CONCLUSION

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, 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 modeling and predicting transitions of health-related tweets. We believe our approach inapplicable to other domains with time-sensitive topics such as disaster management and national security matters.

BIBILOGRAPHY

[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.

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

[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] 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.