Advanced health Monitoring on Social Media

Jacob Thaliyan[1], Ancy Mariya Jacob[2], Anju Shaju[3], Rinto Antony[4], Sandra Biju[5]
[1, 2, 3, 4, 5] Dept of Computer Science
[1, 2, 3, 4, 5] De Paul Institute of Science and Technology, Angamaly

Abstract- *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 , disaster management and politics .We are applying the Temporal Ailment Topic Aspect Model (TM–ATAM), a new latent model dedicated to solving the health-related topics. TM–ATAM is a non-obvious extension to ATAM that was designed to extract health-related topics.We conduct experiments to evaluate the performance of TM–ATAM and T–ATAM on real world data. The experimental setup including the datasets and test-bench.we compare TM–ATAM and T–ATAM against state-of-the-art approaches. That is followed by a detailed study of the behavior of TM–ATAM and a qualitative analysis of TM–ATAM's results.The effect of changing parameters in T–ATAM is studied. Finally, we study the correlations between T–ATAM's results with CDC data and Google Flu Trends for the influenza rates in US. Finally, we highlight the key insights drawn from our experiments.*

Keywords- Data, Filtering health related tweets, Geolocation, Test bench

I. INTRODUCTION

Thanks to dedicated latent topic analysis methods such as the Ailment Topic Aspect Model, public health can now be observed on Twitter.We have systematically detected two problems: health transition detection and health transition prediction.

For solving the first problem we put forward a new latent model that captures transitions involving health-related topics called, TM-ATAM.TM-ATAM(Temporal Ailment Topic Aspect Model) is a non-obvious extension to ATAM(Ailment Topic Aspect Model) that was designed to extract health-related topics.It gains an understanding of health-related topic transitions by reducing the prediction error on topic distributions between consecutive posts at different time and geographic locations.

To solve the second problem, we develop T-ATAM (Temporal Ailment Topic Aspect Model) where time is considered as a random variable natively inside ATAM. Our observations on an 8-month data set of tweets show that TM-ATAM overtakes TM-LDA in calculating health-related transitions from tweets for different geographic populations.

II. EXISTING FRAMEWORK

We have systematically detected two problems: health transition detection and health transition prediction.
For solving the first problem we put forward a new latent model that captures

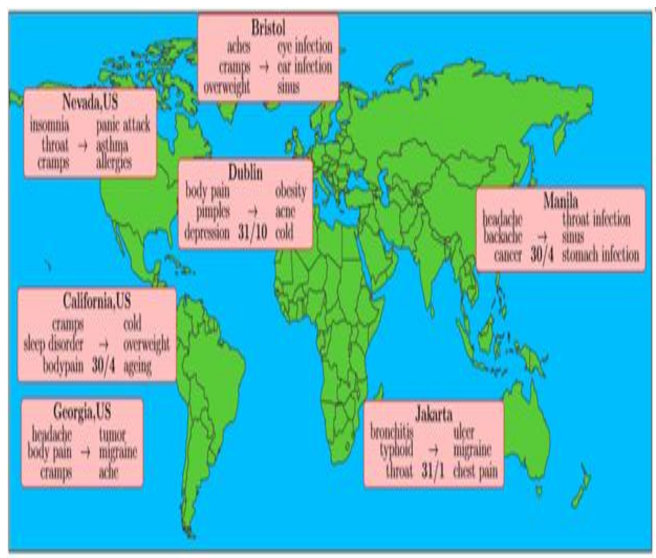

Fig. 1: One-Way ailment transitions obtained by TM–ATAM for various regions. For each location the time period is divided into two parts, preceding and following the most significant change-point discovered for that location. We show the most popular ailments on either side of this boundary.

transitions involving health-related topics called, TM-ATAM. TM-ATAM(Temporal Ailment Topic Aspect Model) is a non-obvious extension to ATAM(Ailment Topic Aspect Model) that was designed to extract health-related topics. It gains an understanding of health-related topic transitions by reducing the prediction error on topic distributions between consecutive posts at different time and geographic locations.

To solve the second problem, we develop T-ATAM (Temporal Ailment Topic Aspect Model) where time is considered as a random variable natively inside ATAM. Our observations on an 8-month data set of tweets show that TM-ATAM overtakes TM-LDA in calculating health-related transitions from tweets for different geographic populations. We observe the effect of climate conditions in different geographic regions on the ability of TM-ATAM to detect transitions. 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 syndromic surveillance and helps measure behavioral risk factors and trigger public health campaigns.

In particular, we find that prediction accuracy for health topics is higher when operating TM-ATAM on finer spatial granularity and shorter time periods. Further, we go on to discover interesting region-specific intra and inter-homogeneous time period health related transitions. While studying these transitions, we find that homogeneous time periods are continuous time periods for which people in the same region tweet about similar health issues. These results show that it is more logical to predict future ailments concerning people within the same homogeneous time period of a region than on any random health tweets.

Since in T-ATAM, time is considered a random variable following multinomial distribution, we expect it to outperform other models, TM-LDA and TM-ATAM in predicting health topics using perplexity measure.

According to our expectations, in most social-media active regions, in both US active regions and non-US active regions, T-ATAM outperforms TM-ATAM and ATAM. After analyzing T-ATAM's performance by changing various spatio-temporal parameters, we find that the prediction accuracy for health topics is higher when operating T-ATAM on finer spatial granularity and shorter time periods.

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.

III. PROPOSED SCHEME

As part of the modifications to the existing system, we have provided a platform for individuals with similar interests to group together and work towards a common goal. As such, we were able to filter out groups with the topics related to health . We provide early warnings, especially to these groups as the result of predictions of the TM-ATAM. In conjuction we provide warnings for people in the pedicted areas through Electronic Mail and by personalised notifications.

For interested groups, we provide administrative services for supervision and visualized graphical representations of the TM-ATAM. For preprocessing the discussions before passing them onto the ATAM, we identify keywords that express medical relevance and also the positivity and negativity expressed in the discussion . We also identify the affected areas through location based mapping or the nativity of the group members.

IV. 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 unravel them. These transition detections are corrected with TM–ATAM, a granularity-based model for conducting region-specific analysis that results in the identification of some 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 leads to significant improvements in modeling 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] C. Chemudugunta, P. Smyth, and M. Steyvers, “Modeling General and Specific Aspects of Documents with a Probabilistic Topic Model,” in NIPS’06, 2006, pp. 241–248.

[2] F. Bouillot, P. Poncelet, M. Roche, D. Ienco, E. Bigdeli, and S. Matwin, “French Presidential Elections: What are the Most Efficient Measures for Tweets?” in PLEAD’12. ACM, 2012, pp. 23–30.

[3] D. M. Blei and J. D. Lafferty, “Dynamic Topic Models,” in ICML’06, 2006, pp. 113–120.

[4] T. Hofmann, “Probabilistic Latent Semantic Indexing,” in SIGIR’99, 1999, pp. 50–57.

[5] D. M. Blei and J. D. Lafferty, “Dynamic Topic Models,” in ICML’06, 2006, pp. 113–120.

[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] 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.

[8] M. J. Paul and M. Dredze, "You Are What You Tweet: Analyzing Twitter for Public Health," in ICWSM'11, 2011.

[9] Sumit Sidana, Sihem Amer-Yahia, Marianne Clausel, Majdeddine Rebai, Son T. Mai, Massih-Reza Amini "Health Monitoring on Social Media over Time", Journal of l atex class files, vol. 14, no. 8, aug 2015.