HEALTH MONITORING ON SOCIAL MEDIA OVER TIME

SUBIYA TAZEEN, H.no: 21S41D5815,Mtech (CSE), Department of Computer Science, VAAGESWARI COLLEGE OF ENGINEERING Thimmapur,Karimnagar, Telangana, INDIA, Email-id: subiatazeen25@gmail.com.

DR. V. BAPUJI, Associate Professor, Department of Computer Science, VAAGESWARI COLLEGE OF ENGINEERING Thimmapur,Karimnagar, Telangana, INDIA, Email-id: bapuji.vala@gmail.com.

ABSTRACT

Social media is become a valuable resource for examining various facets of everyday life. The observation of public health on Twitter is now possible due to specialised latent topic analysis techniques like the Ailment Topic Aspect Model (ATAM). Our research focuses on the long-term monitoring of people's health using social media. One of the many advantages of using tweets is that data is available instantly and at almost no cost. In addition to post-factum research, early health data monitoring allows for a variety of applications, including the measurement of behavioural risk factors and the launch of health initiatives. We construct two problems: the prediction of health transitions and the detection of health transitions. First, we provide a novel latent model called the Temporal Ailment Topic Aspect Model (TM–ATAM), which is intended to solve the first difficulty by identifying transitions involving health-related topics. A non-obvious ATAM extension called TM-ATAM was created with the purpose of extracting subjects pertaining to health. By minimising the prediction error on topic distributions across subsequent articles at various temporal and geographic granularities, it learns transitions relevant to health topics. In order to address the second issue, we create T-ATAM, a Temporal Ailment Topic Aspect Model, which natively treats time as a random variable inside ATAM. Our tests on an 8-month twitter corpus demonstrate that, for various geographic populations, TM–ATAM performs better than TM–LDA in predicting health-related transitions from tweets. We investigate TM–ATAM's capacity to identify climate-related changes across several geographical areas. Next, we demonstrate how T-ATAM may be used to forecast the most significant transition and compare it to data from the Centres for Disease Control and Google Flu Trends.

Index Terms:—People’s health, Ailments, Social media, Temporal Ailment Topic Aspect Model (TM–ATAM).

I.INTRODUCTION

Social media is become a valuable information resource for studying many facets of everyday life. Twitter is specifically utilised for monitoring public health, specifically to identify early indications of population well-being across various geographic locations. Twitter has emerged as a significant data source for early monitoring and prediction in a number of fields, including politics, health, and disaster management [1, 2, 3]. In the health domain, syndromic surveillance, behavioural risk factor measurement, and public health campaign initiation are aided by the capacity to model transitions for ailments and identify statements such as "people talk about headaches and stomach ache in any order" or "people talk about smoking and cigarettes before talking about respiratory problems." The health transition detection challenge and the health transition prediction problem are the two issues we pose in this study. We create TM–ATAM, which simulates the temporal transitions of health-related subjects, to solve the detection issue. In order to tackle the prediction issue, we provide T-ATAM, an innovative technique that uses ATAM's intrinsic handling of time as a random variable to reveal hidden illness inside tweets[4]. Anticipating the gradual shift in health-related conversations on Twitter requires treating time as a random variable.

Traditionally, sentinel surveillance—the collection of data from healthcare facilities—has been used to track common disorders. These resources restrict monitoring, particularly in the case of real-time feedback. Because of this, syndromic monitoring has expanded to a larger scale, almost always at no cost, and using the Web. Our difficulties are: (i) Recognise tweets on health; (ii) Ascertain when talks about health on Twitter shift from one issue to another; and (iii) Record distinct transitions for various geographical areas. Indeed, disease distributions not only change over time, but also change spatially.

Consequently, we need to precisely model the temporal and spatial granularities in order to achieve efficacy. A temporal granularity that is too coarse might miss important illness transitions, while one that is too fine can produce sparse and misleading transitions. Comparably, an excessively fine geographic granularity might result in false positives, while an excessively coarse one can overlook significant transitions, such as those involving individuals residing in disparate climates. For instance, talks about allergy breaks throughout various times in various US states [4].As a result, analysing all tweets coming from the USA at once would overlook climatic fluctuations that have an impact on people's health. Our aim is to discover and simulate the development of disease distributions across distinct temporal granularities. We believe that different time granularities should be taken into account for different locations.

While a number of latent topic modelling techniques, including Latent Dicrichlet Allocation (LDA) [6] and Probabilistic Latent Semantic Indexing (pLSI) [5], have been proposed to successfully cluster and classify general-purpose text, research has shown that specialised techniques, like the AilmentTopic Aspect Model (ATAM), are more effective at capturing ailments in Twitter [4]. ATAM models how people communicate their illnesses in tweets by extending LDA. It makes the assumption that every tweet about health is a latent illness, like allergies or the flu. Like a theme, an illness correlates with a word distribution. Additionally, ATAM keeps a distribution of symptoms and remedies. For latent diseases, this amount of information yields a more accurate model.

However, while pLSI and LDA have shown good performance on static texts, they are not able to capture topic change over time in their core. An extension to LDA formining themes from tweets across time was proposed: temporal-LDA (TM–LDA) [7]. We suggest TM-ATAM, which integrates ATAM and TM-LDA, as a solution to the health transition detection issue. A brief study [8] detailed an early form of TM–ATAM. Here, we demonstrate its capacity to record shifts in conversations about health across several geographies (refer to Figure 1). Therefore, timely campaigns might be started in Nevada, USA, upon the early identification of a shift in the discourse around allergens.

In each geographical area, TM–ATAM minimises the prediction error on the disease distributions of successive pre-specified time intervals to learn transition parameters that govern the development of health-related subjects. The task of automatically identifying such times is our second challenge, the health transition prediction problem. As a result, we provide T–ATAM, a novel and distinct model that handles time in the generative model as a random variable. By considering time as a variable whose values are taken from a corpus-specific multinomial distribution, T-ATAM finds hidden illnesses in health tweets. In the same way as TM–LDA, TM–ATAM, and

T–ATAMare are distinct from dynamic topic models [9], [10], and [11], they are made to identify topic transition patterns from posts that are arranged chronologically, while dynamic topic models concentrate on how subject word distributions change over time.

Our research demonstrates that TM–ATAM performs better than TM–LDA in predicting temporal topic transitions of various geographic populations, using a corpus of more than 500K health-related tweets gathered over an 8-month period. Our findings may be divided into two categories: transitions. When a health-related subject is brought up often, it's considered stable. When certain subjects are covered after others, One-Way transitions take care of such situation. For instance, our analysis of Californian tweets identified a number of consistent themes, such migraines and headaches. Conversely, tweeting about respiratory conditions comes after tweeting about drugs, cigarettes, and smoking. Example one-way transitions that we derived for several global states and cities are shown in Figure 1. These shifts are often caused by outside variables including the weather, public health initiatives, global population patterns, and dietary and lifestyle choices.

II.LITERATURE SURVEY

- L. Manikonda and M. D. Choudhury describe an approach 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.

- S. R. Chowdhury, M. Imran, M. R. Asghar, S. Amer-Yahia, and C. Castillo presented 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.

- Thomas Davidson, Dana Warmsley, Michael Macy and Ingmar Weber proposed 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.

- Y. Wang, E. Agichtein, and M. Benzi describe an approach on latent topic analysis has emerged as one of the most effective methods for classifying, clustering and retrieving textual data. However, existing models such as Latent Dirichlet Allocation (LDA) were developed for static corpora of relatively large documents. In contrast, much of the textual content on the web, and especially social media, is temporally sequenced, and comes in short fragments, including microblog posts on sites such as Twitter and Weibo, status updates on social networking sites such as Facebook and LinkedIn, or comments on content sharing sites such as YouTube. In this paper we propose a novel topic model, Temporal-LDA or TM-LDA, for efficiently mining text streams such as a sequence of posts from the same author, by modeling the topic transitions that naturally arise in these data. TM-LDA learns the transition parameters among topics by minimizing the prediction error on topic distribution in subsequent postings. After training, TM-LDA is thus able to accurately predict the expected topic distribution in future posts. To make these predictions more efficient for a realistic online setting, we develop an efficient updating algorithm to adjust the topic transition parameters, as new documents stream in. Our empirical results, over a corpus of over 30 million microblog posts, show that TM-LDA significantly outperforms state-of-the-art static LDA models for estimating the topic distribution of new documents over time. We also demonstrate that TM-LDA is able to highlight interesting variations of common topic transitions, such as the differences in the work-life rhythm of cities, and factors associated with area-specific problems and complaints.

- C. X. Lin, Q. Mei, J. Han, Y. Jiang, and M. Danilevsky describe an approach the prevalence of Web 2.0

techniques has led to the boom of various online communities, where topics spread ubiquitously among user-generated documents. Working together with this diffusion process is the evolution of topic content, where novel contents are introduced by documents which adopt the topic. Unlike explicit user behavior (e.g., buying a DVD), both the diffusion paths and the evolutionary process of a topic are implicit, making their discovery challenging. In this paper, we track the evolution of an arbitrary topic and reveal the latent diffusion paths of that topic in a social community. A novel and principled probabilistic model is proposed which casts our task as an joint inference problem, which considers textual documents, social influences, and topic evolution in a unified way. Specifically, a mixture model is introduced to model the generation of text according to the diffusion and the evolution of the topic, while the whole diffusion process is regularized with user-level social influences through a Gaussian Markov Random Field. Experiments on both synthetic data and real world data show that the discovery of topic diffusion and evolution benefits from this joint inference, and the probabilistic model we propose performs significantly better than existing methods.

III. EXISTING SYSTEM

The authors provide an approach that uses the structure of a social network to learn how topics temporally change in a community. This method uses the current system to learn changing word distributions of subjects over time. However, TM–ATAM and T–ATAM differ from dynamic topic models like those found in [9] and [10], as well as from Wang et al.'s work [11], in that the former are made to learn topic transition patterns from temporally-ordered posts, whereas the latter concentrate on how topics' word distributions change over time.

TM–ATAM minimises the prediction error on the illness distributions of successive periods at various temporal and geographic granularities, therefore learning transition parameters that govern the development of health-related subjects. T–ATAM, on the other hand, uses time as a corpus-specific multinomial distribution to identify latent illnesses in health tweets.

Traditional methods have been used to mine themes for citation inference. To conduct an empirical investigation on topic modelling and time-based topic modelling, respectively, several discriminative methodologies have been used. None of them directly relate to information about health.

DISADVANTAGES OF EXISTING SYSTEM:

- Mapping Tweets to Documents is not available.
- Uncovering Health Topics with ATAM is one of them.

IV PROPOSED SYSTEM:

The suggested system formulates and solves two problems: the issue of health transition prediction and the problem of health transition detection. The system creates TM-ATAM, which simulates the temporal transitions of health-related subjects, to solve the detection issue. In order to tackle the prediction issue, we provide T-ATAM, an innovative technique that uses ATAM's intrinsic handling of time as a random variable to reveal hidden illness inside tweets.

Anticipating the gradual shift in health-related conversations on Twitter requires treating time as a random variable.

ADVANTAGES OF PROPOSED SYSTEM:

- TM-ATAM is a model that can identify tweets on health and track how they change over time and place. By minimising the prediction error on the illness distributions of pre-specified time periods, TM–ATAM determines the transition parameters for a specific area.
- T-ATAM is a novel model that use time as a variable with values derived from a corpus-specific multinomial distribution, enabling it to forecast tweets about health.
- Numerous tests that demonstrate T-ATAM's superiority over TM-LDA and TM-ATAM in terms of health transition prediction, as well as its efficacy when compared to a ground truth.

V. SYSTEM DESIGN

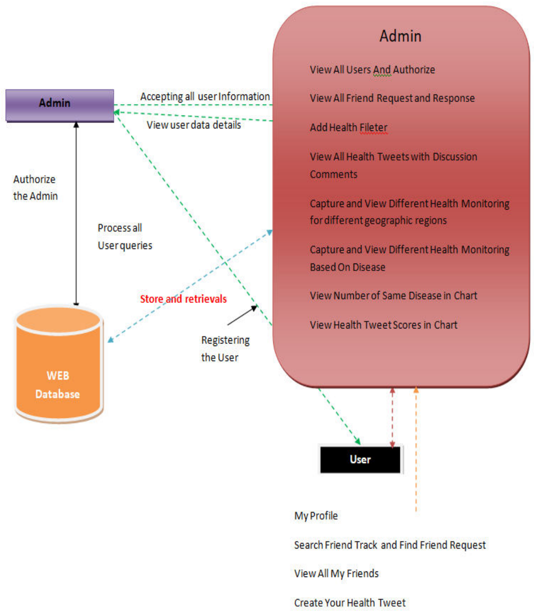


Fig1: Architecture of system.

VI.MODULE DESCRIPTION:

- **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

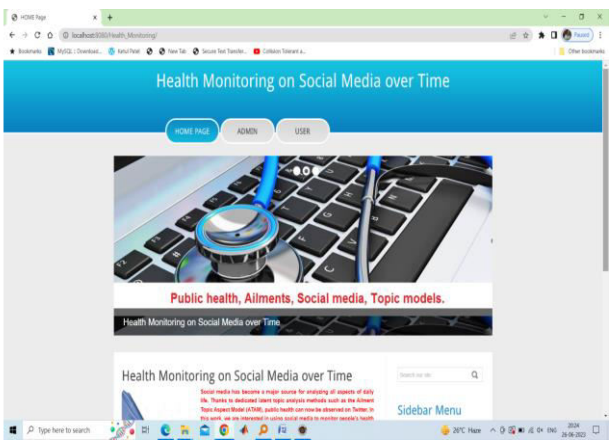
My Health Tweets, View and Monitor All My Friends Health Tweets.

- **Searching Users to make friends**

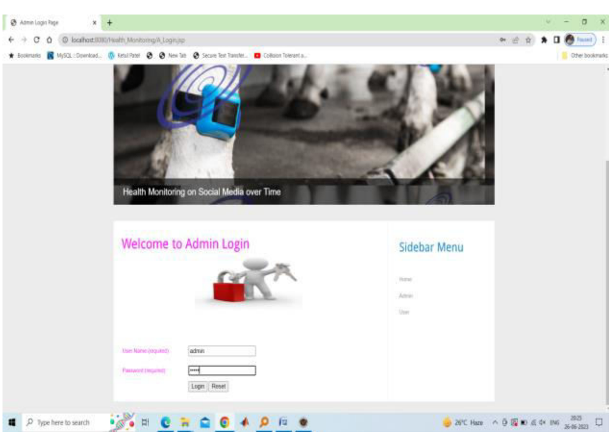
In this module, the user searches for users in Same Network and in the Networks and sends friend requests to them. The user can search for users in other Networks to make friends only if they have permission.

VII. RESULT:

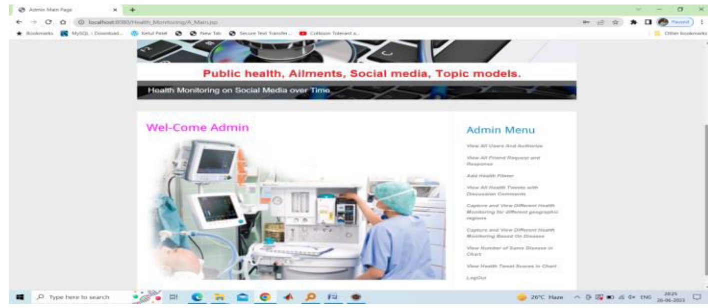
Index.html



AdminLogin.jsp



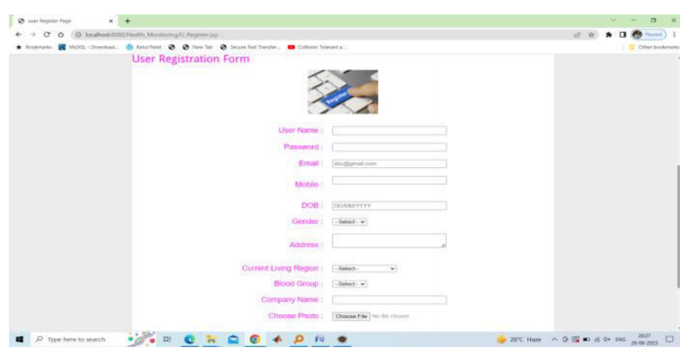
Adminmain.jsp



View_AllTweets_RegionBased.jsp

UserLogin.jsp

UserRegistration.jsp



VIII. CONCLUSION

Over time, we create techniques to identify illnesses using social media. We developed two models to address the challenges of health transition detection and prediction. TM-ATAM, a granularity-based model, is used to address detection. It allows for area-specific analysis, which identifies temporal periods and characterises the discourse around homogenous diseases, per region. T-ATAM, which considers time as a random variable with values taken from a multinomial distribution, is used to address prediction. The fine-grained structure of T-ATAM leads to negligible gains in tweet transition modelling and prediction. We think our method may be used in other fields where there are time-sensitive subjects, such national security and catastrophe management.

IX. FUTURE ENHANCEMENT

We proved the construction is "feasible" via performance assessment. We leave it to future effort to increase efficiency even more while preserving the system's lovely characteristics.

X. 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: Analyzing 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 Modeling 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.