Journal of Engineering Sciences
Vol 10, Issue 10, Oct /2019
ISSN NO:0377-9254

A Novel Method for Detecting Stress Levels Of Users Based On Social Interactions In OSN

AKASH KUMAR SAHUKAR
Assistant System Engineer in Tata Consultancy Services Pvt. Ltd, Mumbai

ABSTRACT

Conventional psychological wellness studies transfer information assembled through individual contact with a restorative administration. Late work has exhibited the utility of online social data for mulling over discouragement, nevertheless, there have been restricted appraisals of other mental prosperity conditions. I present an investigation of emotional wellness phenomena in openly accessible social networking sites. Initially, it characterized a lot of pressure-related literary, visual and social qualities from different perspectives; and then proposed a novel crossbreed model. By furthermore examining the social correspondence data in like manner, located a couple of captivating marvels.

KEYWORDS: Stress detection, factor graph model, social media, healthcare

1.INTRODUCTION

Psychological stress is becoming a threat to people's health nowadays. With the rapid pace of life, more and more people are feeling stressed. According to a worldwide survey reported by *Newbusiness* in 2010[1], over half of the population have experienced an appreciable rise in stress over the last two years. Though stress itself is non-clinical and common in our life, excessive and chronic stress can be rather harmful to people's physical and mental health. According to existing research works, long-term stress is related to many diseases, e.g., clinical depression, insomnia, etc. Moreover, according to a survey, suicide has become the top cause of death among Chinese youth and excessive stress is considered to be a major factor of suicide. All these reveal that the rapid increase of stress has become a great challenge to human health and quality of life. Thus, there is significant importance to detect stress before it turns into severe problems. Traditional psychological stress detection is mainly based on face-to-face interviews, self-report questionnaires or wearable sensors. However, traditional methods are reactive, which are usually labor-consuming, time-costing and hysteretic.

The rise of social media is changing people's lives, as well as research in healthcare and wellness.

With the development of social networks like Facebook or Twitter, more and more people are willing to share their daily events or moods and interact with friends through social networks. As these social media data timely reflect users' real-life states and emotions in a timely manner, it offers new opportunities for representing, measuring, modeling and mining users' behavior patterns through the large-scale social networks. Such social information can find its theoretical basis in psychology research. For example, [7] found that stressed users are more likely to be socially less active, and more recently, there have been researched efforts on harnessing social media data for developing mental and physical healthcare tools. For example, [27] proposed to leverage Twitter data for real-time disease surveillance; while [35] tried to bridge the vocabulary gaps between health seekers and providers using the community-

generated health data. There are also some research works [28] [47] using user tweeting content on social media platforms to detect users' psychological stress. Existing works [28], [47] demonstrated that leverage social media for healthcare and in particular, stress detection is feasible.

Limitations in the existing system are that stress analysis is a crucial tool for designing structurally sound shapes. However, the expensive computational cost has hampered its use in interactive shape editing tasks. I augment the existing example-based shape editing tools and propose a fast subspace stress analysis method to enable stress-aware shape editing. In particular, it is constructed by a reduced stress basis from a small set of shape exemplars and possible external forces. This stress basis is automatically adapted to the current user-edited shape on the fly and thereby offers reliable stress estimation. then introduce a new finite element discretization scheme to use the reduced basis for fast stress analysis. Some limitations exist in tweeting content based stress detection. Firstly, tweets are limited to a maximum of 140 characters on social platforms like Twitter and users do not always express their stressful states directly in tweets. Secondly, users with high psychological stress may exhibit low activeness on social networks. These phenomena incur the inherent data sparsity and ambiguity problem, which may hurt the performance of tweeting content based stress detection performance.

2.DATAMINING

Data mining is the process of discovering patterns in large data sets involving methods at the intersection of machine learning, statistics and database systems. It is an essential process where intelligent methods are applied to extract data patterns. It is an interdisciplinary subfield of computer science. The overall goal of the data mining process is to extract information from a data set and transform it into an understandable structure for further use.[1] Aside from the raw analysis step, it involves database and data management aspects, data preprocessing model and inference considerations, complexity considerations, post-processing of discovered structures, visualization and online updating.[1] Data mining is the analysis step of the "knowledge discovery in databases" process or KDD.[5]

In the 1960s, statisticians and economists used terms like *data fishing* or *data dredging* to refer to what they considered the bad practice of analyzing data without an a priori hypothesis. The term "data mining" was used in a similarly critical way by economist Michael Lovell in an article published in the Review of Economic Studies 1983. Lovell indicates that the practice "masquerades under a variety of aliases, ranging from "experimentation" (positive) to "fishing" or "snooping" (negative).[10]

The term *data mining* appeared around 1990 in the database community, generally with positive connotations. For a short time in the 1980s, the phrase "database mining"™, was used, but since it was trademarked by HNC, a San Diego-based company, to pitch their Database Mining Workstation;[11] researchers consequently turned to *data mining*. Other terms used include *data archaeology*, *information harvesting*, *information discovery*, *knowledge extraction*, etc. Gregory Piatetsky-Shapiro coined the term "knowledge discovery in databases" for the first workshop on the same topic (KDD-1989) and this term became more popular in AI and machine learning

community. However, the term data mining became more popular in the business and press communities.[12] Currently, the terms data mining and knowledge discovery are used interchangeably. The related terms data dredging, data fishing, and data snooping refer to the use of data mining methods to sample parts of a larger population data set that are (or maybe) too small for reliable statistical inferences to be made about the validity of any patterns discovered. These methods can, however, be used in creating new hypotheses to test against the larger data populations.

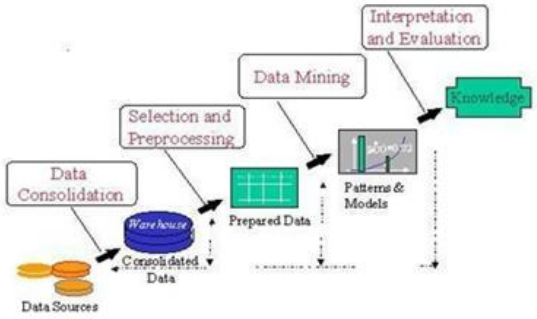


Figure 1: Structure Of Data Mining

The above figure depicts how data is mined through various processes and required useful data is obtained.

WORKING OF DATA MINING

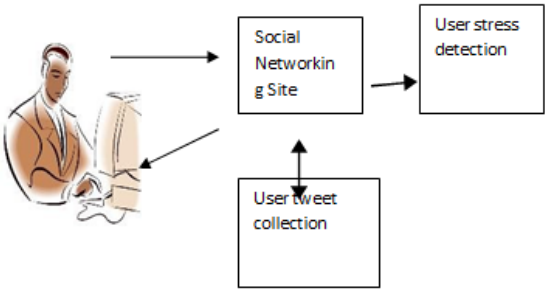
While large-scale information technology has been evolving separate transaction and analytical systems, data mining provides the link between the two. Data mining software analyzes relationships and patterns in stored transaction data based on open-ended user queries. Several types of analytical software are available: statistical, machine learning, and neural networks. Generally, any of the four types of relationships are sought:

- *Classes:* Stored data is used to locate data in predetermined groups. For example, a restaurant chain could mine customer purchase data to determine when customers visit and what they typically order. This information could be used to increase traffic by having daily specials.
- *Clusters:* Data items are grouped according to logical relationships or consumer preferences. For example, data can be mined to identify market segments or consumer affinities.
- *Associations:* Data can be mined to identify associations. The beer-diaper example is an example of associative mining.
- *Sequential patterns:* Data is mined to anticipate behavior patterns and trends. For example, an outdoor equipment retailer could predict the likelihood of a backpack being purchased based on a consumer's purchase of sleeping bags and hiking shoes.

Data mining consists of five major elements:

1. *Extract, transform and load transaction data onto the data warehouse system.*
2. *Store and manage the data in a multidimensional database system.*
3. *Provide data access to business analysts and information technology professionals.*
4. *Analyze the data by application software.*
5. *Present the data in a useful format, such as a graph or table.*

SYSTEM ARCHITECTURE:



DIFFERENT TYPES OF ANALYSIS

- *Artificial neural networks:* Non-linear predictive models that learn through training and resemble biological neural networks in structure.
- *Genetic algorithms:* Optimization techniques that use process such as genetic combination, mutation and natural selection in a design based on the concepts of natural evolution.
- *Decision trees:* Tree-shaped structures that represent sets of decisions. These decisions generate rules for the classification of a dataset. Specific decision tree methods include Classification and Regression Trees (CART) and Chi Square Automatic Interaction Detection (CHAID). CART and CHAID are decision tree techniques used for classification of a dataset. They provide a set of rules that you can apply to a new (unclassified) dataset to predict which records will have a given outcome. CART segments a dataset by creating 2-way splits while CHAID segments using chi square tests to create multi-way splits. CART typically requires less data preparation than CHAID.
- *Nearest neighbor method:* A technique that classifies each record in a dataset based on a combination of the classes of the *k* record(s) most similar to it in a historical dataset (where *k*=1). Sometimes called the *k*-nearest neighbor technique.
- *Rule induction:* The extraction of useful if-then rules from data based on statistical significance.
- *Data visualization:* The visual interpretation of complex relationships in multidimensional data. Graphics tools are used to illustrate data relationships.

3.IMPLEMENTATION

A. System Framework:

In this framework I propose a novel hybrid model - a factor graph model combined with Convolution Neural Network to leverage tweet content and social interaction information for stress detection. Experimental results show that the proposed model can improve the detection performance by 6-9% in F1-score. By further analyzing the social interaction data, I also discover several intriguing phenomena, i.e. the number of social structures of sparse connections (i.e. with no delta connections) of stressed users is around 14% higher than that of non-stressed users, indicating that the social structure of stressed users' friends tend to be less connected and less complicated than that of non-stressed users.

B. Social Interactions:

Here, analyzed the correlation of users' stress states and their social interactions on the networks, and address the problem from the standpoints of: (1) social interaction content, by investigating the content differences between stressed and non- stressed users' social interactions; and (2) social interaction structure, by investigating the structure differences in terms of structural diversity, social influence, and strong/weak tie. I find that the number of social structures of sparse

connection (i.e. with no delta connections) of stressed users is around 14% higher than that of non-stressed users, indicating that the social structure of stressed users' friends tend to be less connected and complicated, compared to that of non-stressed users.

C. Attributes categorization

First, define two sets of attributes to measure the differences between the stressed and non-stressed users on social media platforms: 1) tweet-level attributes from a user's single tweet; 2) user-level attributes summarized from a user's weekly tweets. Tweet-level attributes describe the linguistic and visual content, as well as social attention factors (being liked, commented, and retweeted) of a single tweet. Classification of words into different categories, e.g. positive/negative emotion words, degree adverbs is possible. Furthermore, extract linguistic attributes of emoticons, so as to map the keyword in square brackets to find the emoticons. Twitter adopts Unicode as the representation for all emojis, which can be extracted directly. User-Level Attributes Compared to tweet-level attributes extracted from a single tweet, user-level attributes are extracted from a list of user's tweets in a specific sampling period. One week is used as the sampling period in this paper. On one hand, psychological stress often results from cumulative events or mental states. On the other hand, users may express their chronic stress in a series of tweets rather than one. Besides, the aforementioned social interaction patterns of users in a period also contain useful information for stress detection. Moreover, as aforementioned, the information in tweets is limited and sparse. Here, there is a need to integrate more complementary information around tweets, e.g., users' social interactions with friends. There are some techniques to detect the intrusion that occurs in the cloud computing environment. They are:

1. Signature-based Detection: A set of rules that can be used to design a given pattern is that of an intruder.
2. Anomaly-based Detection: Identifying events that appear to be anomalous concerning the normal system.
3. Stateful Protocol Analysis: This intrusion detection system could know and trace the protocol states.

4.CONCLUSION

In this system, I displayed a system for distinguishing user's psychological stress states from client's week after week online networking information, utilizing tweet's substance and additionally the client's social associations. Utilizing true online networking information as the premise, I contemplated the connection between client mental anxiety states and their social communication practices. To completely use both substance and social communication data of clients' tweets, I proposed a half and half model which joins the factor graph model (FGM) with a convolution neural system (CNN).

REFERENCES

[1] Rui Fan, Jichang Zhao, Yan Chen, and Ke Xu. Anger is more influential than joy: Sentiment correlation in weibo. PLoS ONE, 2014.

[2] Zhanpeng Fang, Xinyu Zhou, Jie Tang, Wei Shao, A.C.M. Fong, Longjun Sun, Ying Ding, Ling Zhou, , and Jarder Luo. Modeling paying behavior in game social networks. In In Proceedings of the Twenty-Third Conference on Information and Knowledge Management (CIKM'14), pages 411–

420, 2014

[3] Jennifer Golbeck, Cristina Robles, Michon Edmondson, and Karen Turner. Predicting personality from twitter. In Passat/socialcom 2011, Privacy, Security, Risk and Trust, pages 149–156, 2011.

[4] Frank R Kschischang, Brendan J Frey, and H-A Loeliger. Factor graphs and the sum-
Transactions on, 47(2):498–519, 2001.

[5] . Lin, J. Jia, Q. Guo, Y. Xue, J. Huang, L. Cai, and L. Feng. Psychological stress detection from cross-media microblog data using deep sparse neural network. In proceedings of IEEE International Conference on Multimedia & Expo, 2014.

[6] Andrey Bogomolov, Bruno Lepri, Michela Ferron, Fabio Pianesi,and Alex Pentland. Daily stress recognition from mobile phone data, weather conditions and individual traits. In ACM International Conference on Multimedia, pages 477–486, 2014

.[7] Chris Buckley and Ellen M Voorhees. Retrieval evaluation with incomplete information. In Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval, pages 25–32, 2004.

[8] Xiaojun Chang, Yi Yang, Alexander G Hauptmann, Eric P Xing, and Yao-Liang Yu. Semantic concept discovery for large-scale zero-shot event detection. In Proceedings of International Joint Conference on Artificial Intelligence, pages 2234–2240, 2015.

[9] Wanxiang Che, Zhenghua Li, and Ting Liu. Ltp: A Chinese language technology platform. In Proceedings of International Conference on Computational Linguistics, pages 13–16, 2010.

[10] Chih Chung Chang And Chih-Jen Lin. Libsvm: A Library For Support Vector Machines. Acm Transactions On Intelligent Systems And Technology, 2(3):389–396, 2001.

Author's Profile

Akash Kumar Sahukar

Currently working in Tata Consultancy Services Pvt. Ltd, Mumbai as an Assistant System Engineer. The following are the technologies on which I am working: C# , .NET, SQL and Angular. Successfully graduated with a Bachelor's degree from GITAM UNIVERSITY, Visakhapatnam - in the field of Electronics and Communication Engineering in the year 2018. Contact: akashksahukar@gmail.com