Discovering Stress Based on Social Interaction on Social Networking Sites

Sudhir Anakal[1] Chandrasekhar Uppin[2] Ambresh Bhadrashetty[3]

[1, 3] Department of Studies in Computer Applications, Visvesvaraya Technological University, Postgraduate Centre, Kalaburagi, India.

[2] Department of Computer Science, Faculty of Computing and Applied Sciences, Baze University, Jabi Abuja, Nigeria.

Abstract: Stress is essentially humans' response to various types of desires or threats. This response, when working properly, can help us to stay focused, energized and intellectually active, but if it is out of proportion, it can certainly be harmful leading to depression, anxiety, hypertension and a host of threatening disorders. Cyberspace is a huge soap box for people to post anything and everything that they experience in their day-to-day lives. Subsequently, it can be used as a very effective tool in determining the stress levels of an individual based on the posts and status updates shared by him/her. This is a proposal for a website which takes the Twitter username of the subject as an input, scans and analyses the subject's profile by performing Sentiment Analysis and gives out results. These results suggest the overall stress levels of the subject and give an overview of his/her mental and emotional state. The tool used for analysis of the social media account is Rapidminer. Rapid miner is an environment for various data mining and machine learning procedures with a very effective and simple GUI.

Index Terms – CNN, Stress, Social Networking Sites.

I. INTRODUCTION

Mental health illness is one of the major concern in today's era. Most of the people are stressed there may be several reasons for it. Stress is something which cannot be cured or treated. The person suffering from stress may get into lot of troubles like heart attack or may suffer from depression. A person should not be effected by lot of stress level as it may lead to death. People suffering from stress often end of committing suicide, or facing restless nights or which may also lead to heart attack. In past if a person is suffering from stress it was detected by the physician by asking certain questions or conducting interviews to the patients to know the level of stress they are suffering with. As it was very hard to detect the level of stress in the past as it was purely based on the experience of the physician. The answers given by the patient are taken into the consideration for detecting the level of the stress in the past. In the past the stress was understood by the behavior or the person or looking at the way they communicated. It was very hard to get to know the level of stress they underwent or going through. Stress does not have any scientific calculator to calculate or to know whether the person is suffering from stress or not as it is very hard to detect the stress level. Today we have social networking platforms which helps us to know the level of stress the person is undergoing looking at their tweets they post or the tweet they repost or reply or like the post. It helps us to identify whether the person is suffering from stress. There are many cases where people have lost their lives due to stress. People have committed suicide for this reason. The research says that in every ten person count there are seven peoples who are suffering from stress. The stress may be due to work schedules or due to any external problem or unsatisfied with the work environment. Therefore it is very important to detect the level of stress and make sure the person is treated well so that he or she may not loose there life as life is very precious.

II. PROBLEM STATEMENT

The problem can be defined as finding the stress level of the user through the social platforms seeing the kind of tweets they post or the kind of tweets they like or the kind of people they are friend with. We are using tweet-level emotion detection by collecting the data posted by the users. But there series of tweets were not taken into the account. Only the tweets or posts were collected for the day which did not have enough evidence to come to conclusion that the person is really under stress because a single tweet or post cannot be used for considering the emotional level of a person. Though several techniques were used previously but they did not lead into the sufficient consideration. Hence we can say that there were loopholes in the existing system therefore it is required to go for proposed study.

III. LITERATURE SURVEY

Paper 1: Detecting Stress Based on Social Interactions in Social Networks -Huijie Lin, Jia Jia*, Jiezhong Qiu, Yongfeng Zhang, Guangyao Shen, Lexing Xie, Jie Tang, Ling Feng, and Tat-Seng Chua

Abstract: With the popularity people are used to sharing their daily activities and interacting with friends on social media platforms, making it feasible to leverage online social network data for stress detection. In this paper, we find that users stress state is closely related to that of his/her friends in social media, and we employ a large-scale dataset from real-world social platforms to systematically study the correlation of users' stress states and social interactions. We first define a set of stress-related textual, visual, and social attributes from various aspects, and then propose a novel hybrid model - a factor graph model combined with Convolutional Neural Network to leverage tweet content and social interaction information for stress detection.

Methodology Used: The system includes the following phases: • Support Vector Machine (SVM): It is a popular and binary classifier that is proved to be effective on a huge category of classification problems. In our problem we use SVM with RBF kernel. • Random Forest (RF): it is an ensemble learning method for decision trees by building a set of decision trees with random subsets of attributes and bagging them for classification results. • Gradient Boosted Decision Tree (GBDT): it trains a gradient boosted decision tree model with features associated with each user. • Deep Neural Network (DNN): for user-level stress detection: it is proposed to deal with the problem of user-level stress detection problem with a convolutional neural network (CNN) with cross autoencoders. This is the real baseline method that we can compare our proposed model with.

Paper 2: Sentiment analysis in twitter using machine learning techniques-M. S. Neethu,R. Rajshri

Abstract: Sentiment analysis deals with identifying and classifying opinions or sentiments expressed in source text. Social media is generating a vast amount of sentiment rich data in the form of tweets, status updates, blog posts etc. Sentiment analysis of this user generated data is very useful in knowing the opinion of the crowd. Twitter sentiment analysis is difficult compared to general sentiment analysis due to the presence of slang words and misspellings. The maximum limit of characters that are allowed in Twitter is 140. Knowledge base approach and Machine learning approach are the two strategies used for analyzing sentiments from the text. In this paper, we try to analyze the twitter posts about electronic products like mobiles, laptops etc using Machine Learning approach. By doing sentiment analysis in a specific domain, it is possible to identify the effect of domain information in sentiment classification. We present a new feature vector for classifying the tweets as positive, negative and extract people's' opinion about product.

Methodology Used: Steps followed are as follows:

1.Symbolic Techniques: In July 2013, Neethu M S and Rajasree R proposed that Symbolic techniques also known as knowledge based approach. In this technique, available lexical resources are used. In this sentiment analysis approach, bag-of-words approach is used. The BOW model focuses on the words list, or says string of words, it cannot check the context of the sentence. This model contains a list of words that have own value when found in the given text. This model totally focuses on the words and take care nothing about the language fundamentals.

2. Machine Learning Techniques: In contrast to Knowledge based approaches, Machine Learning techniques are not using any lexicon resources list , instead a training set and a test set is used in order to classify them. Training set contains input vectors and corresponding class labels for training the network. After that, test set is used to validate the given model by checking the class labels to unknown feature vectors. There are different machine learning techniques like SVM, maximum entropy and Naïve Bayes etc. This allows the algorithm to remain dynamic in the face of ever changing social network language lexicons. In this methodology, a classification model is developed using a training set, which tries to classify the input feature vectors into corresponding class labels. Use the results from the knowledge based techniques and those of the machine learning techniques to ensure a thorough analysis of the dataset.

Paper 3: Detecting Emotions in Social Media: A Constrained Optimization Approach-Yichen Wang,Aditya Pal

Abstract: Emotion detection can considerably enhance our understanding of users' emotional states. Understanding users' emotions especially in a real-time setting can be pivotal in improving user interactions and understanding their preferences. In this paper, we propose a constraint optimization framework to discover emotions from social media content of the users. Our framework employs several novel constraints such as emotion bindings, topic correlations, along with specialized features proposed by prior work and well-established emotion lexicons. We propose an efficient inference algorithm and report promising empirical results on three diverse datasets.

Methodology Used: 1.Sentiment Analysis: Sentiment analysis aims at discovering the contextual polarity of the documents [Pang and Lee, 2008]. [Li et al., 2009] proposed a Non-negative Matrix Factorization (NMF) approach which leverages lexical knowledge for sentiment classification. Recent work [Bollen et al., 2011; Golder and Macy, 2011] has focused on mining temporal and seasonal trends of sentiment. Sentiment analysis is a closely related problem, however emotions are much more expressive than sentiments. Moreover, emotions need not contain a sentiment and vice-versa. 2. Emotion Detection: Emotion models are primarily of two types [Ekkekakis, 2013]: (i) dimensional, and (ii) categorical. Dimensional models represent emotions on three dimensions: valence, arousal and dominance 3.Convex Sub-Problem

IV. METHODOLOGY USED

Twitter has over 328 million monthly active users. This makes up for a very large database to perform Natural Language Processing on. To study these tweets, we use an algorithm which is made up of 3 algorithms. These algorithms are: 1. The first part of the algorithm aims at retrieving the tweets which have certain keywords in them. 2. The second part of the algorithm performs Sentiment Analysis on the tweets found in the first step. 3. The third and final step calculates the resultant score based on the intensity values assigned to the keywords.

The algorithm works as follows:

1. Tweet retrieval: There will be a dataset of keywords and emoticons which will be pre-saved into our system. This step will retrieve all those tweets which contain any one or more of these keywords and/or emoticons. The reason for including emoticons is that they have become a very famous means of dialogue and quite a lot of times, people just reply or communicate only using them. To carry out the process of retrieving the tweets from Twitter, a password, known as Twitter API key will be required which can be found from Twitter at request. Given below is a sample of these keywords and emoticons in Fig. 1.

2. Sentiment Analysis: The same words can be used in a variety of places to mean a lot of different things. We will need to determine what a particular keyword means in the context of the particular tweet or conversation. We will also be required to stem the words in order to determine the root words which will then be used for the purpose of analyzing. The keywords will be segregated into three categories: positive, negative and neutral. Then, based on these scores, a final score will be computed which will be known as the complex result.

3. Result declaration: The complex result calculated in the above step will be the one which will be taken into account while declaring the final result about the stress of the particular user. There will be three categories in which these scores will be divided: unstressed, tensed and stressed. The highest scores will correspond to the unstressed category where the user, according to the tweets in analysis, is happy and satisfied. The tensed category will correspond to the moderate scores and it will mean the user is not stressed at the moment but proactive care needs to be taken. The stressed category will correspond to lowest scores and the users who will be in this category will be considered to tweet such messages or simply, tweets, which indicate that he is highly worried and stressed in his/her life.

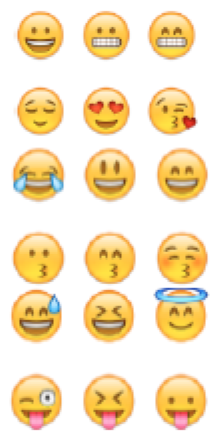
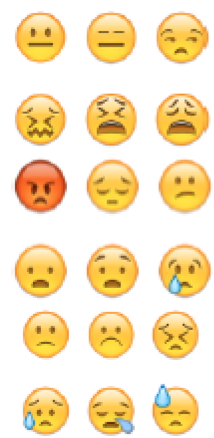
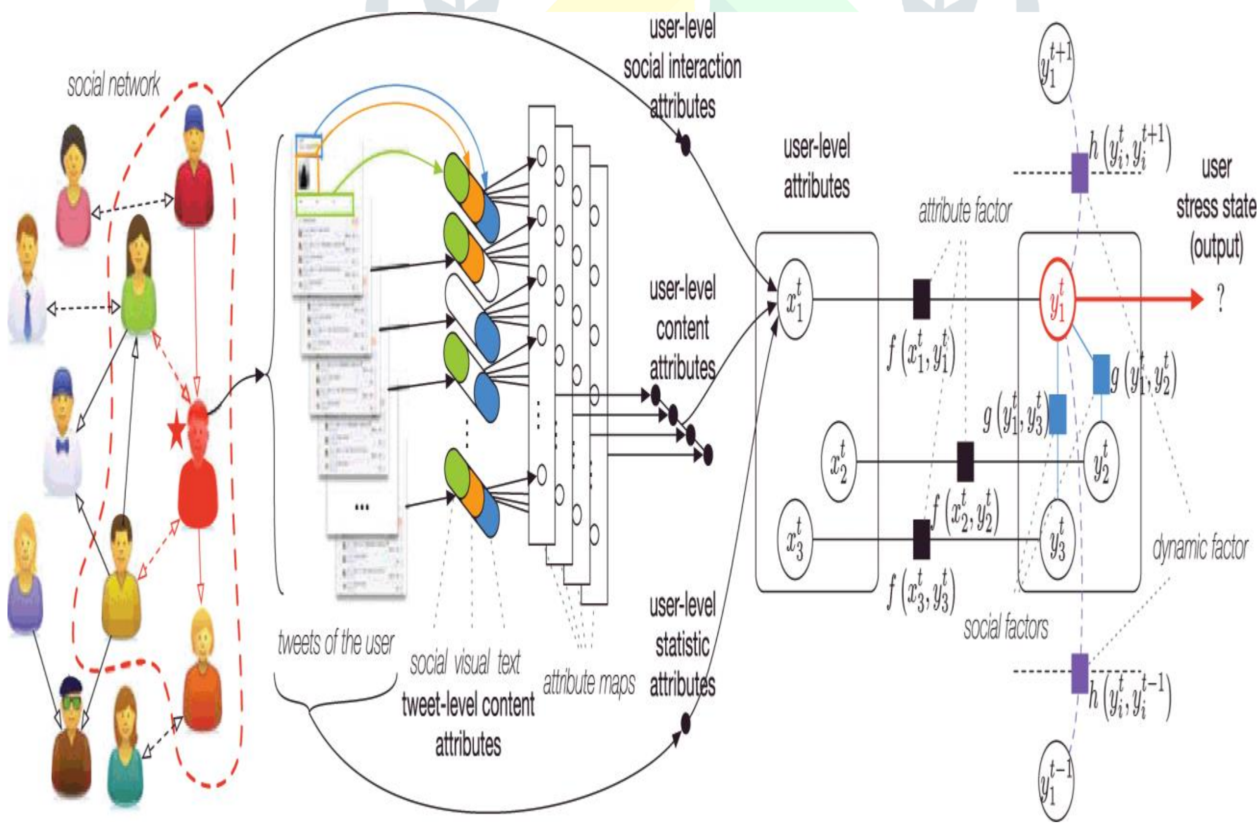
POSITIVE	NEGATIVE
Happy Excited Joyous Gleeful Satisfactory Funny Beaming	Unhappy Sorrowful Dejected Glum Gloomy Dismal Blue
POSITIVE	**NEGATIVE**
	

Fig 1. Keywords and Emoticons

V. ARCHITECTURE OF THE PROPOSED SYSTEM



Many studies on social media based emotion analysis are at the tweet level, using text-based linguistic features and classic classification approaches. A system called *MoodLens* to perform emotion analysis on the Chinese micro-blog platform Weibo, classifying the emotion categories into four types, i.e., angry, disgusting, joyful, and sad. An existing system studied the emotion propagation problem in social networks, and found that anger has a stronger correlation among different users than joy, indicating that negative emotions could spread more quickly and broadly in the network. As stress is mostly considered as a negative emotion, this conclusion can help us in combining the social influence of users for stress detection. Inspired by psychological theories, we first define a set of attributes for stress detection from tweet-level and user-level aspects respectively: 1) tweet-level attributes from content of user's single tweet, and 2) user-level attributes from user's weekly tweets.

The tweet-level attributes are mainly composed of linguistic, visual, and social attention (i.e., being liked, retweeted, or commented) attributes extracted from a single-tweet's text, image, and attention list. The user-level attributes however are composed of: (a) posting behavior attributes as summarized from a user's weekly tweet postings; and (b) social interaction attributes extracted from a user's social interactions with friends.

In particular, the social interaction attributes can further be broken into: (i) social interaction content attributes extracted from the content of users' social interactions with friends; and (ii) social interaction structure attributes extracted from the structures of users' social interactions with friends.

VI. RESULTS

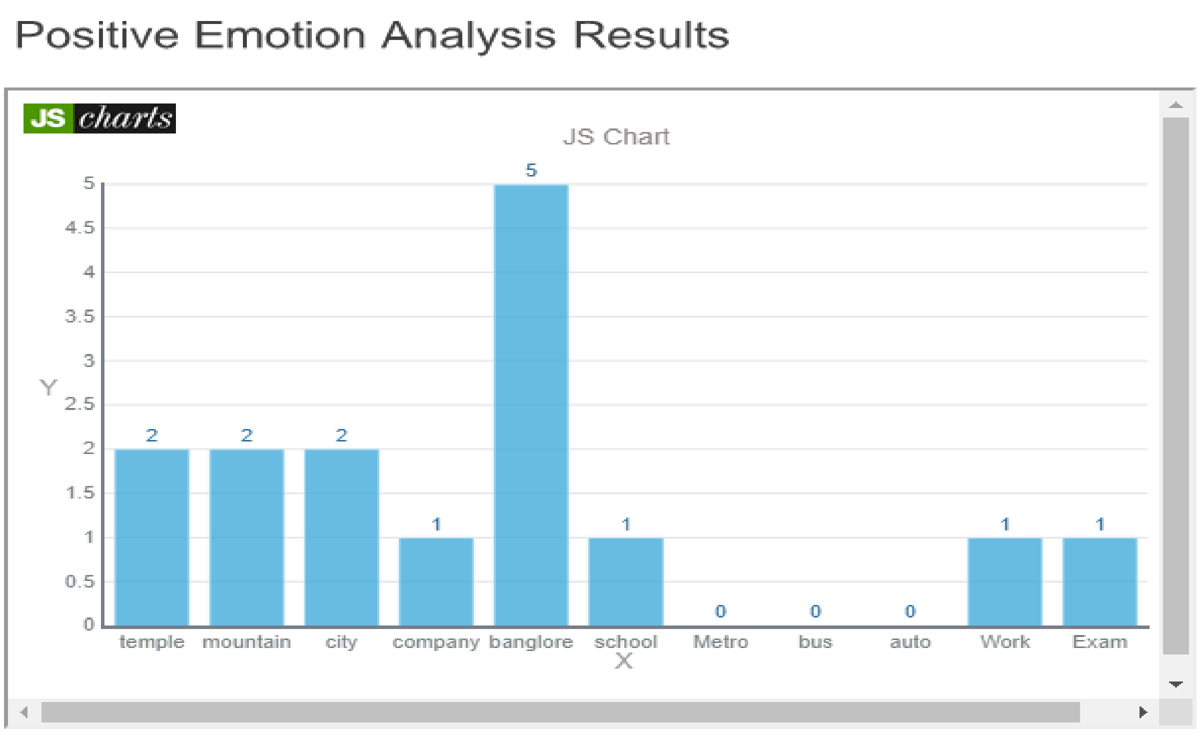


Fig 2. Positive Emotion Analysis Results

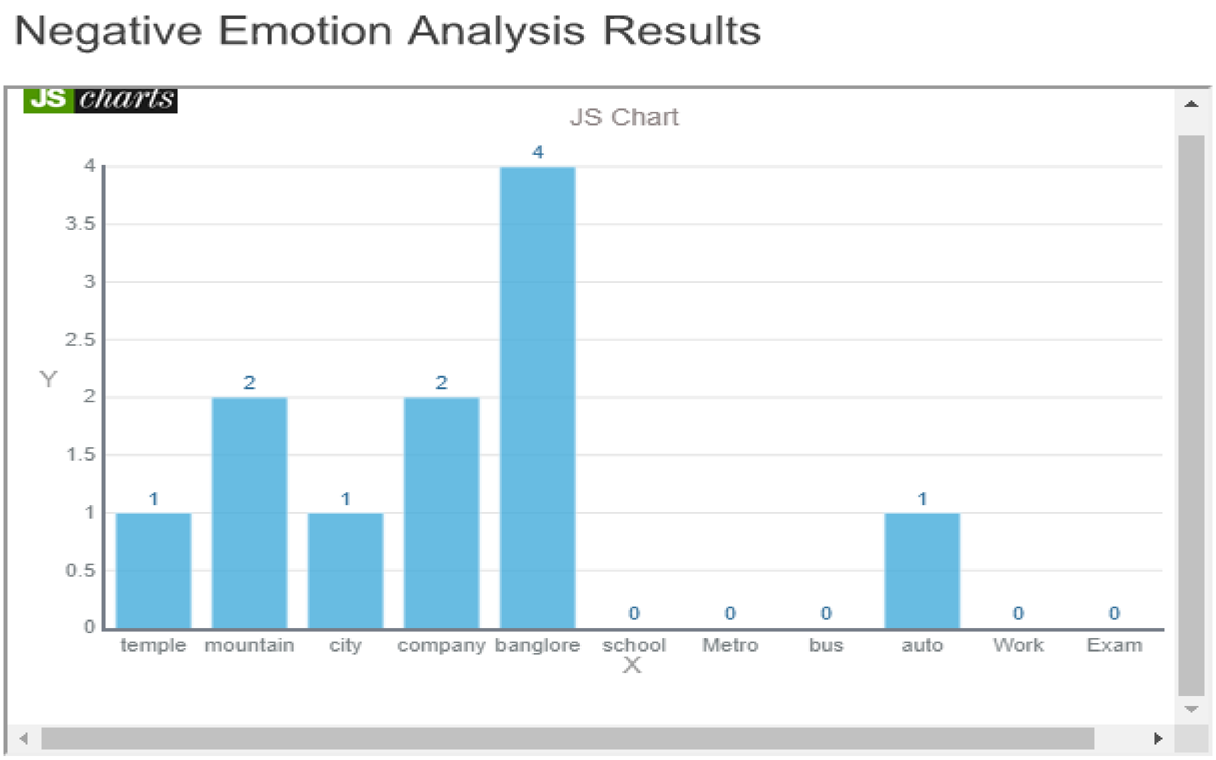


Fig 3. Negative Emotion Analysis Results

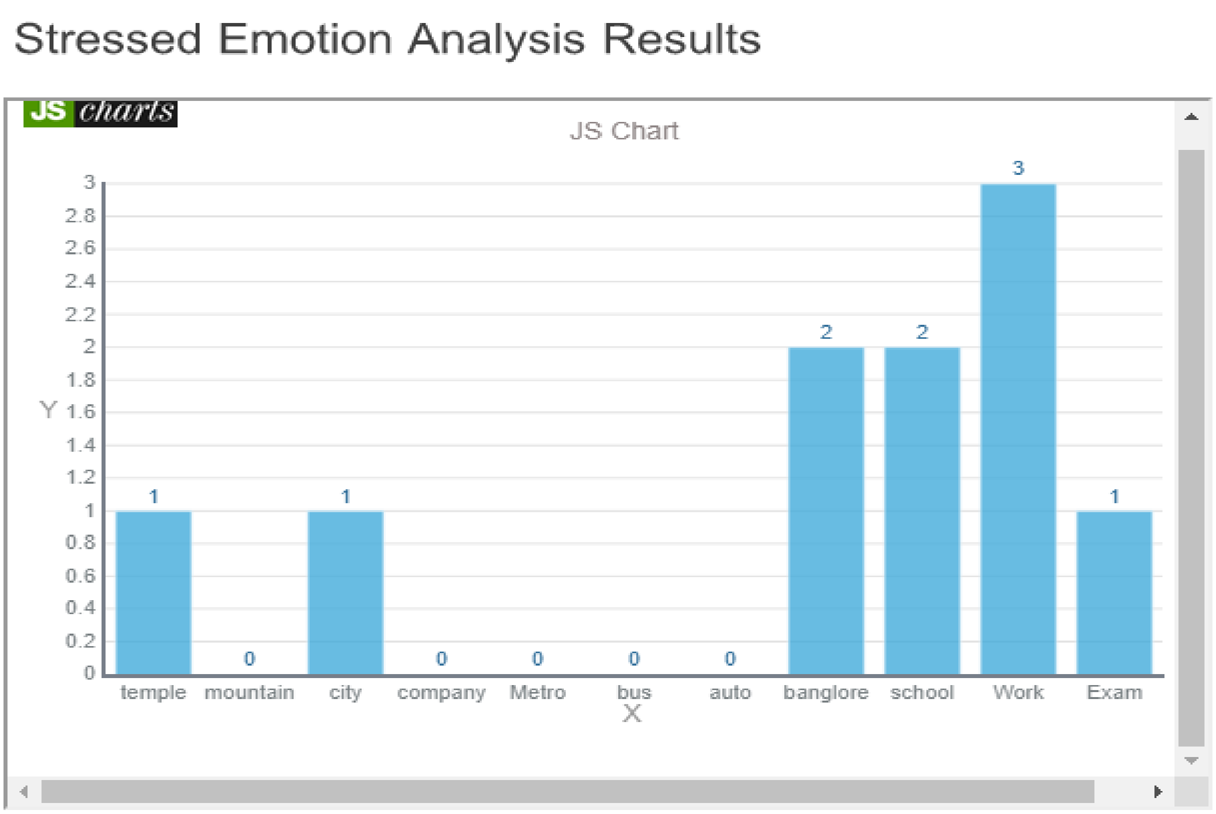


Fig 4. Stressed Emotion Analysis Results

VII. CONCLUSION

In this paper, we presented a framework for detecting users' psychological stress states from users' weekly social media data, leveraging tweets' content as well as users' social interactions. Employing real-world social media data as the basis, we studied the correlation between user' psychological stress states and their social interaction behaviors. To fully leverage both content and social interaction information of users' tweets, we proposed a hybrid model which combines the factor graph model (FGM) with a convolutional neural network (CNN). In this work, we also discovered several intriguing phenomena of stress. We found 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 less complicated than that of non-stressed users. These phenomena could be useful references for future related studies.

REFERENCES

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

[2]Chris Buckley and EllenM Voorhees. Retrieval evaluation with incompleteinformation. In Proceedings of the 27th annual internationalACM SIGIR conference on Research and development in informationretrieval, pages 25–32, 2004.

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

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

[5]Chih chung Chang and Chih-Jen Lin. Libsvm: a library for supportvector machines. ACM TRANSACTIONSON INTELLIGENTSYSTEMS AND TECHNOLOGY, 2(3):389–396, 2001.

[6]Dan C Ciresan, Ueli Meier, Jonathan Masci, Luca Maria Gambardella,and J ¨ urgen Schmidhuber. Flexible, high performanceconvolutional neural networks for image classification. In Proceedingsof International Joint Conference on Artificial Intelligence, pages1237–1242, 2011.

[7]Sheldon Cohen and Thomas A. W. Stress, social support, and thebuffering hypothesis. Psychological Bulletin, 98(2):310–357, 1985.

[8]Glen Coppersmith, Craig Harman, and Mark Dredze. Measuringpost traumatic stress disorder in twitter. In Proceedings of theInternational Conference on Weblogs and Social Media, pages 579–582,2014.

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

[10]Zhanpeng Fang, Xinyu Zhou, Jie Tang, Wei Shao, A.C.M. Fong,Longjun Sun, Ying Ding, Ling Zhou, , and Jarder Luo. Modelingpaying behavior in game social networks. In In Proceedings of theTwenty-Third Conference on Information and Knowledge Management(CIKM'14), pages 411–420, 2014.