JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

HUMAN STRESS DETECTION BASED ON SOCIAL INTERACTIONS

K. MALARVIZHI M.Sc.,M.Phil.,SET.,

Assistant Professor, Department of Computer Science, Velumanoharan Arts and Science College for Women, Marappalam, Peravoor (Post), Ramanathapuram-Dist, Tamilnadu, India.

ABSTRACT

Psychological stress is putting people's health at risk. Finding stress in time for proactive care is non-trivial. With the popularity of social media, people are accustomed to sharing their daily activities and interacting with friends on social media platforms, making it possible to leverage online social network data to detect stress. In this paper, we find that user stress state is closely related to his or her friends in social media, and we use large-scale datasets from real-world social platforms to systematically study the correlation of users' stress states and social Let's use Conversation. We first define a set of text, visual and social features related to stress from different aspects, and then propose a novel hybrid model - to leverage tweet content and social interaction information for stress detection. A factor graph model combined with a convolution neural network. Experimental results show that the proposed model can improve the recognition performance in F1-score by 6-9%. By further analyzing the social connection data, we also find a number of intriguing phenomena, i.e. the number of social structures with sparse connections (i.e. those with no delta connections) of stressed users is approximately 14% higher than that of non-stressed users, this shows that the social structure of friends of stressed users is less connected and less complex than that of non-stressed users.

I.INTRODUCTION

Today we all agree that social networking has completely changed the way we live, communicate, work or even our relationships with our loved ones. Simply put: Social networking has changed the way we live.

However some questions remain that how these social networking has changed our lives. Thus the question will not be asked whether social networking has changed our lives for better or for worse because many people think that social networking has good effects in our society, as many people think that social networking has only bad effects. Effects are networking. Like everything in life, we have advantages and disadvantages to social networking.

Thus we can define social networking as "the practice of increasing the number of one's business and/or social contacts by making connections through individuals often through social media sites such as Face book, Twitter, LinkedIn and Google+ etc".

So we will restrict our reflection to exploring both the good and bad effects, in order to see how social networking changed our world and how it is changing the way we live.

For that, we will first look into the history of social networking to see how social networking originated and how it got introduced in our daily lives. Next, we will study the effects (both negative and positive) of social networking in our society.

METHODOLOGY OF THE RESEARCH

To achieve the thesis objectives, a design and creation research methodology has been applied in an incremental process, where each contribution have been sequentially proposed and validated. For each contribution of this thesis, the five steps that this research methodology involves have been followed: awareness, suggestion, development, evaluation and conclusion. The first stage of the research process included identifying the research problems.

II. EXISTING SYSTEM

- Many studies on social media based sentiment analysis are at the tweet level using text-based linguistic features and classic classification approaches. A system called Mood Lens, for analyzing emotions on the Chinese micro-blog platform Weibo, classifies emotion categories into four types, i.e. angry, disgusting, joyful and sad.
- An existing system studied the problem of emotion propagation in social networks, and found that anger has a stronger relationship between different users than pleasure, indicating that negative emotions in the network can spread more quickly and widely. Since stress is mostly perceived as a negative emotion, this finding can help us combine users' social influence to detect stress.

DISADVANTAGES OF EXISTING SYSTEM

- Traditional psychological stress detection is mainly based on face-to-face interviews, self-report questionnaires or wearable sensors. However, traditional methods are actually reactive, which are usually labor-consuming, time-costly, and hysteretic.
- These functions mainly take advantage of textual content in social networks. In fact, data in social networks is usually composed of sequential and interconnected items from different sources and modalities, making it truly cross-media data.
- Although there have been few user-level emotion detection studies, the role social interactions play in one's psychological stress state, and how we can incorporate such information into stress detection, is yet to be explored. Has not yet been investigated.

III. PROPOSED SYSTEM

- Inspired by psycholinguistic principles, we first define a set of features for stress detection from tweet-level and user-level aspects, respectively: 1) tweet-level features from the content of a user's single tweet, and 2) User-level features from user's weekly tweets.
- Tweet-level attributes are primarily composed of linguistic, visual, and social attention (i.e., liked, retweeted, or commented on). Attributes are extracted from single-tweet text, images, and attention lists. However, user-level features include: (a) posting behavioral characteristics summarized from the user's weekly Tweet postings; and (b) social interaction attributes derived from User's social interactions with friends.
- Specifically, social interaction features can be further divided into: (i) social interaction content features extracted from the content of users' social interactions with friends; and (ii) social interaction structure features extracted from the structures of users' social interactions with friends.

ADVANTAGES OF PROPOSED SYSTEM

- Experimental results show that by exploiting users' social interaction features, the proposed model can improve recognition performance (F1-score) by 6–9% compared to state-of-the-art methods. This indicates that the proposed features can serve as good pointers to deal with the problem of data paucity and ambiguity. Furthermore, the proposed model can efficiently combine tweet content and social interaction to enhance the stress detection performance.
- Beyond user tweeting content, we analyze the correlation of users' stress status and their social interaction on the network, and address the problem from the point of view of: (1) social interaction content, stressed and non-stressed; - Social interaction of users by checking the difference of stressful content; and (2) social interaction structure by examining structure differences in terms of structural heterogeneity, social influence and strong/weak ties.
- We generate multiple stressed-twitter-posting datasets from several popular social media platforms by various ground-truth labeling methods and thoroughly evaluate our proposed method on several aspects.
- We conduct in-depth studies on real-world large-scale datasets and gain insights on social interactions and stress, as well as the relationship between stressed social structures.

IV. PROPOSED ALOGORITHM

DBSCAN

Clustering analysis is an unsupervised learning method that separates the data points into several specific bunches or groups, such that the data points in the same groups have similar properties and data points in different groups have different properties in some sense.

It comprises of many different methods based on different distance measures. E.g. K-Means (distance between points), Affinity propagation (graph distance), Mean-shift (distance between points), DBSCAN (distance between nearest points), Gaussian mixtures (Mahalanobis distance to centers), Spectral clustering (graph distance), etc.

Centrally, all clustering methods use the same approach i.e. first we calculate similarities and then we use it to cluster the data points into groups or batches. Here we will focus on the Density-based spatial clustering of applications with noise (DBSCAN) clustering method.

Density-Based Clustering refers to unsupervised learning methods that identify distinctive groups/clusters in the data, based on the idea that a cluster in data space is a contiguous region of high point density, separated from other such clusters by contiguous regions of low point density.

Density-Based Spatial Clustering of Applications with Noise (DBSCAN) is a base algorithm for density-based clustering. It can discover clusters of different shapes and sizes from a large amount of data, which is containing noise and outliers.

The DBSCAN algorithm uses two parameters:

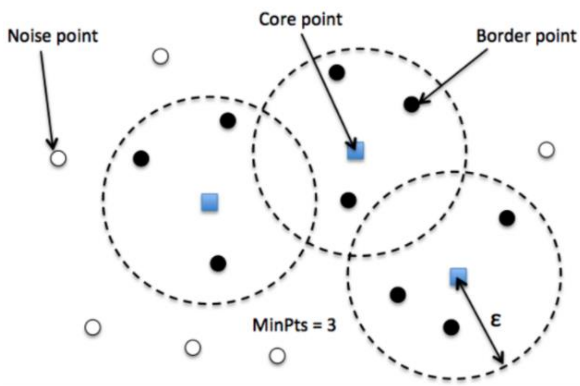
- MinPts: The minimum number of points (a threshold) clustered together for a region to be considered dense.
- Eps (ε): A distance measure that will be used to locate the points in the neighborhood of any point.

These parameters can be understood if we explore two concepts called Density Reach ability and Density Connectivity.

Reach ability in terms of density establishes a point to be reachable from another if it lies within a particular distance (eps) from it.

Connectivity, on the other hand, involves a transitivity based chaining-approach to determine whether points are located in a particular cluster. For example, p and q points could be connected if p->r->s->t->q, where a->b means b is in the neighborhood of a.

There are three types of points after the DBSCAN clustering is complete:



- Core — this is a point that has at least m points within distance n from itself.
- Border — this is a point that has at least one Core point at a distance n.
- Noise — this is a point that is neither a Core nor a Border. And it has less than m points within distance n from itself.

Algorithmic steps for DBSCAN clustering

- The algorithm proceeds by arbitrarily picking up a point in the dataset (until all points have been visited).
- If there are at least 'minPoint' points within a radius of 'ε' to the point then we consider all these points to be part of the same cluster.
- The clusters are then expanded by recursively repeating the neighborhood calculation for each neighboring point

PARAMETER ESTIMATION

Every data mining task has the problem of parameters. Every parameter influences the algorithm in specific ways. For DBSCAN, the parameters ε and minPts are needed.

- MinPts: As a rule of thumb, a minimum minPts can be derived from the number of dimensions D in the data set, as minPts ≥ D + 1. The low value minPts = 1 does not make sense, as then every point on its own will already be a cluster. With minPts ≤ 2, the result will be the same as of hierarchical clustering with the single link metric, with the dendrogram cut at height ε. Therefore, minPts must be chosen at least 3. However, larger values are usually better for data sets with noise and will yield more significant clusters. As a rule of thumb, minPts = 2·dim can be used, but it may be necessary to choose larger values for very large data, for noisy data or for data that contains many duplicates.
- ε: The value for ε can then be chosen by using a k-distance graph, plotting the distance to the k = minPts-1 nearest neighbor ordered from the largest to the smallest value. Good values of ε are where this plot shows an "elbow": if ε is chosen much too small, a large part of the data will not be clustered; whereas for a too high value of ε, clusters will merge and the majority of objects will be in the same cluster. In general, small values of ε are preferable, and as a rule of thumb, only a small fraction of points should be within this distance of each other.
- Distance function: The choice of distance function is tightly linked to the choice of ε, and has a major impact on the outcomes. In general, it will be necessary to first identify a reasonable measure of similarity for the data set, before the parameter ε can be chosen. There is no estimation for this parameter, but the distance functions need to be chosen appropriately for the data set.

V.ARCHITECTURE DIAGRAM

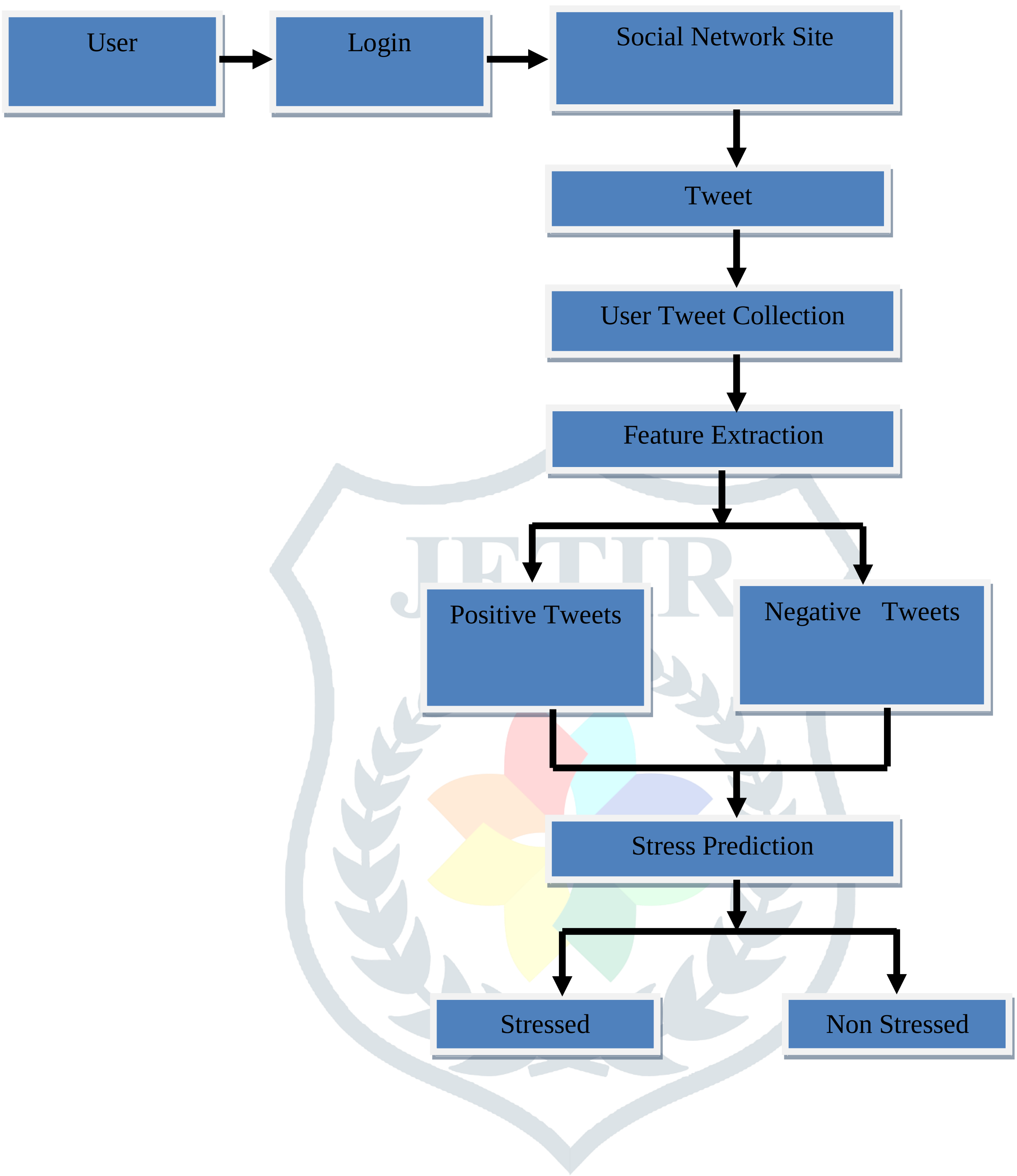


Figure 8: Architecture Diagram of Human Stress

VI. IMPLEMENTATION

Implementation is the most curial stage in archiving a successful system and giving the user confidence that the new system is workable and effective. Implementation of a modified application to replace an existing one this type of conversion is relatively easy to handle, provide there are no major changes in the system.

Each program is tested individually at the time of development using the data and has verified that this program linked together in the way specified in the program specification the computer system and its environment is tested to the satisfaction of the user. The system that has been developed is accepted and proved to be satisfactory for the user. A simple operating procedure is included so that the user can understand the different function clearly and quickly.

Initially as a first step the executable form of the application is to be created and loaded in the common server machine which is accessible to the entire user and the server is to be connected to a network. The final stage is to document the entire system which is provides components and the operating procedure of the system.

SOCIAL INTERACTIONS

We analyze 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. Our investigation unveils some intriguing social phenomena. For example, we find that the number of social structures of sparse connection (i.e. with no deltaconnections4) 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.

ATTRIBUTES CATEGORIZATION

We first define two sets of attributes to measure the differences of 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

Tweet-level attributes describe the linguistic and visual content, as well as social attention factors (being liked, commented, and retreated) of a single tweet. For linguistic attributes, we take the most commonly used linguistic features in sentiment analysis research. Specifically, we first adopt LTP , to map the words into positive/negative emotions. LIWC2007 is a dictionary which categorizes words based on their linguistic or psychological meanings, so we can classify words into different categories, e.g. positive/negative emotion words, degree adverbs. We have also tested other linguistic resources including NRC5 and HowNet6, and found that the performances were relatively the same. Furthermore, we extract linguistic attributes of emoticons (e.g., and) and punctuation marks ('!', '?', '...', '.'). Weibodefines every emoticon in square brackets (e.g., they use [ha-ha] for "laugh"), so we can 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. We use one week 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 of time also contain useful information for stress detection. Moreover, as aforementioned, the information in tweets is limited and sparse we need to integrate more complementary information around tweets, e.g., users' social interactions with friends.

VII.CONCLUSION

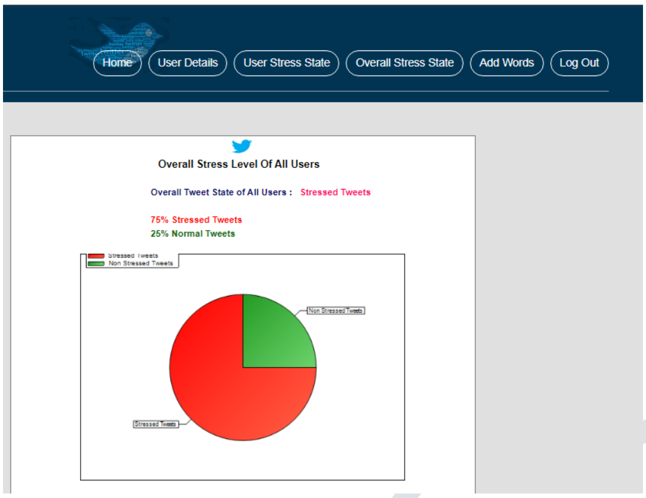
In this paper, we presented a framework for detecting users 'psychological stress states from users' weekly social media data, leveraging tweets' content as well as users' social interactions. Employing real-world social media data as the basis, we studied the correlation between user' psychological 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 convolution neural network (CNN).

VIII.FUTURE WORK:

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.

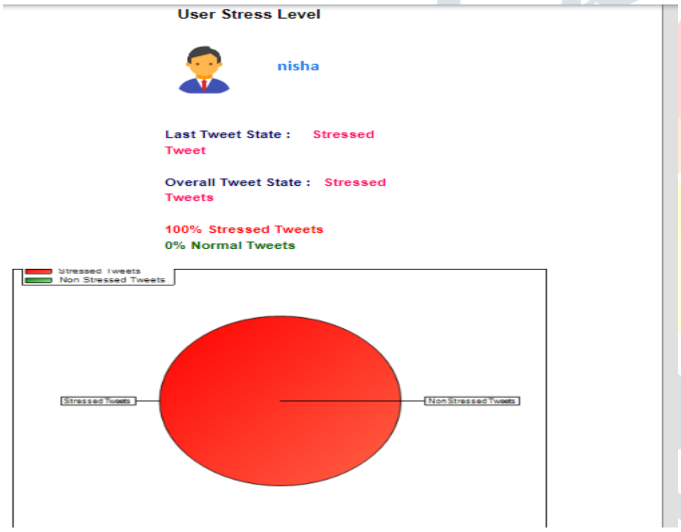
IX.RESULTS AND DISCUSSION

I conducted the simulation using dot net to verify the DB scan algorithm. The overall and stress state and the individual stress state was found.



OVERALL STRESS STATE

USER STRESS STATE



REFERENCES

[1] 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.

[2] Chris Buckley and EllenM 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.

[3] 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.

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

[5] Chih chung Chang and Chih-Jen Lin. Libsvm: a library for supportvector machines. ACM TRANSACTIONS ON 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 the buffering hypothesis. Psychological Bulletin, 98(2):310–357, 1985.

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

[9] Rui Fan, Jichang Zhao, Yan Chen, and Ke Xu. Anger is more influential 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 Proceedings of the Twenty-Third Conference on Information and Knowledge Management (CIKM'14), pages 411–420, 2014.