ISSN:2348-2079

Volume-6 Issue-2

International Journal of Intellectual Advancements and Research in Engineering Computations

Social Network Data Analysis for User Stress Discovery and Recovery

[1] R. Ragavi, [2]G.Maheswari [3] S.Geeitha,
Department of Information Technology,
Mahendra Engineering College for Women. Kumaramangalam, Tiruchengode, Tamilnadu, India
[1-2]UG Students, [3]Assistant Professor

Mail ID: [1]ragavibtechit22@gmail.com, [2]mahesh.gdana30@gmail.com

Abstract

The social networks are build to exchange images, videos and reviews between the members. The members can connect from anywhere and anytime. The checkin locations and time information are maintained under the Location Based Social Networks (LBSN). Location-Based Social Networks (LBSNs) provides the facility to access people's locations, profiles and online social connections. High scalable, high volume and high velocity data items are processed using big data mining models.
The social network data values are analyzed to estimate the user behavior and stress levels. The correlations of user interactions and stress are analyzed for the stress detection process. The social network data values are parsed and three types of attributes are extracted for the stress detection process. Stress-related textual, visual and social attributes are extracted for the stress detection process. The hybrid model combines the Factor Graph and the Convolutional Neural Network (CNN) for stress detection. Tweet content and social interaction are analyzed in the hybrid model.

The automated stress discovery process is build with hybrid model, context and connection information. The location and time information are represented in the context information. The stress detection process is improved with spatio temporal features. The hybrid model is enhanced to detect multi level stress categories. The sparse user connation parameters are integrated with the stress detection process. The hybrid model is enhanced to suggest treatment levels. Automated stress release messages and suggestions are provided by the system.

Index Terms: Location Based Social Networks (LBSN), Factor graph, Convolutional Neural Network (CNN), Social network data analysis and Stress discovery.

1. Introduction

Research in psychology has suggested that behavior and preferences of individuals can be explained to a great extent by underlying psychological constructs: personality traits. Knowledge of an individual's personality allows us to make predictions about preferences across contexts and environments, and to enhance recommendation systems. Personality can affect the decision making process and has been shown to affect preferences for websites, products, brands and services and for content such as movies, TV shows, and books.

The most widely accepted model of personality, Big Five or Five Factor Model embraces five traits: Openness, Conscientiousness, Extroversion, Agreeableness, and Emotional Stability. Further explanations of each trait are summarized. A traditional approach to measure personality requires participants to answer a series of questions evaluating their behavior and preferences. This approach is time consuming and impractical, especially in the context of on-line services. On-line users might be unwilling to spend a considerable amount of time filling-in a questionnaire, in order to personalize their search results or product recommendations.

It has been recently shown that the digital footprint of users can be used to automatically infer their personality. For example, Kosinski et al. and Youyou et al. showed that automated personality judgments based on Facebook Likes are more accurate than those made by users' friends or even their spouses. Also, Park et al. showed that similar predictions can be based on language used in social media. A variety of other approaches have been proposed using different prediction mechanisms, feature spaces, and focusing on different on-line environments.

2. Related Work

Research into micro-blog contents mainly focuses on short text mining, topic detection, and sentiment analysis. Among them, sentiment analysis, a.k.a. opinion mining, is the most related work of this study. Previous work on sentiment analysis
take blogs, reviews, and tweets as research objects, aiming to extract people's opinions towards some subjects or products. Usually, people's opinions are classified into two polarities: positive and negative, and sometimes including neutral, while peoples' emotions are classified into six categories, i.e., joy, fear, sadness, surprise, anger, and disgust.

Techniques of natural language processing and machine learning are popularly used in sentiment analysis employed the lexicon-based strategy to obtain the overall polarity of a document by computing the number of positive words and negative words in blogs or reviews also developed a syntactic parser and sentiment lexicon to discover the semantic relationship between target and expression detected the sentiment orientations to products according to the natural language expression conventions. [4] collected 40 emoticon classes found in Yahoo! blog articles, and utilized sentences containing these emoticons to automatically detect user's emotions from messenger logs.

To overcome the limitation that the performance of natural language processing techniques in sentiment analysis relies on the domain of corpus, and thus cannot easily be extended to other application fields, the performance of three machine learning methods over movie reviews, including Naive Bayes, Maximum Entropy, and Support Vector Machines. Their experimental result showed that the performance of SVM classifier with unigram presence features is superior to others improved the precision of sentiment analysis based on unsupervised learning by analyzing features of textual contents and iteration mechanisms effective techniques of feature selection and feature combination utilized both text features and text theme such as description, comment, background, etc. in sentiment analysis. Multi-classifier fusion techniques were employed improve the performance of sentiment classification estimated the emotional tendency of a document by combining the prior lexicon-based emotional tendency and the posterior training-based emotional tendency.

Most of sentiment analysis on micro-blog also makes use of [3] used emoticons as labels to reduce dependency in machine learning techniques for sentiment classification. Twitter hashtags and smileys to enhance sentiment learning. [6] improved target-dependent Twitter sentiment classification by taking target-dependent features as well as related tweets into consideration. [7] built a graph model for sentiment classification at the hashtag level in Twitter, where three approximate classification algorithms were investigated. [8] presented a two-staged SVM classifier for robust sentiment detection from biased and noisy data on Twitter.

In terms of application areas, most research in sentiment analysis aims at offering techniques to business domains by detecting users' opinions towards a product or proposal sampled and analyzed users' reviews to specific brands of products on micro-blog, and extracted users' emotion polarities using an automatic classification method. [9] predicted the direction of stock markets by analyzing the micro-blog data. [5], [2] provided a lexicon-based natural language processing method to investigate the debate performance of candidates in the 2008 US president election.

To our knowledge, this paper is the first to detect and analyze adolescents' psychological pressures from micro-blog, aiming to combine traditional adolescent mental education with micro-blog media and turn micro-blog into a new kind of adolescent mental education mode and platform.

3. Detecting Stress Based on Social Interactions

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, over half of the population has 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 has been found to be related to many diseases, e.g., clinical depressions, insomnia etc.. Moreover, according to Chinese Center for Disease Control and Prevention, 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 life quality.

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 actually reactive, which are usually labor-consuming, time-costing and hysteretic. Are there any timely and proactive methods for stress detection?

The rise of social media is changing people's life, as well as research in healthcare and wellness with the development of social networks like Twitter and Sina Weibo, more and more people are willing to share their daily events and moods, and interact with friends through the 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, and such social information can find its theoretical basis in psychology research. For example, found that stressed users are more likely to be socially less active, and more recently, there have been research efforts on harnessing social media data for developing mental and physical healthcare tools. For example, proposed to leverage Twitter data for real-time disease surveillance; while tried to bridge the vocabulary gaps between health seekers and providers using the community generated health data. There are also some research works using user tweeting contents on social media platforms to detect users' psychological stress. Existing works [1] demonstrated that leverage social media for healthcare, and in particular stress detection, is feasible.

Firstly, tweets are limited to a maximum of 140 characters on social platforms like Twitter and Sina Weibo, 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, as reported by a recent study in Pew Research Center. These phenomena incur the inherent data sparsity and ambiguity problem, which may hurt the performance of tweeting content based stress detection performance. The tweet contains only 13 characters, saying that the user wished to go home for the Spring Festival holiday. Although no stress is revealed from the tweet itself, from the follow-up interactive comments made by the user and her friends, we can find that the user is actually stressed from work. Thus, simply relying on a user's tweeting content for stress detection is insufficient.

Users' social interactions on social networks contain useful cues for stress detection. Social psychological studies have made two interesting observations. The first is mood contagion: a bad mood can be transferred from one person to another during social interaction. The second is linguistic echoes: people are known to mimic the style and affect of another person. These observations motivate us to expand the scope of tweet-wise investigation by incorporating follow-up social interactions like comments and retweeting activities in user's stress detection. This may actually help to mitigate the single user's data sparsity problem. Another reason for considering social interactions in stress detection is based on our empirical findings on a large-scale dataset crawled from Sina Weibo that the social structures of stressed users are less connected and thus less complicated than those of non-stressed users. This is consistent with the Pew Research Center's finding that stressed users are less active than non-stressed ones.

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.

To maximally leverage the user-level information as well as tweet-level content information, we propose a novel hybrid model of factor graph model combined with a convolutional neural network (CNN). This is because CNN is capable of learning unified latent features from multiple modalities, and factor graph model is good at modeling the correlations. The overall steps are as follows: 1) we first design a convolutional neural network (CNN) with cross autoencoders (CAE) to generate user-level content attributes from tweet-level attributes; and 2) we define a partially labeled factor graph (PFG) to combine user-level social interaction attributes, user-level posting behavior attributes and the learnt user-level content attributes for stress detection.

We evaluate the proposed model as well as the contributions of different attributes on a real-world dataset from Sina Weibo. Experimental results show that by exploiting the users' social interaction attributes, the proposed model can improve the detection performance (F1-score) by 6-9% over that of the state-of-art methods. This indicates that the proposed attributes can serve as good cues in tackling the data sparsity and ambiguity problem. Moreover, the proposed model can also efficiently combine tweet content and social interaction to enhance the stress detection performance. We further conduct in-depth studies on a large-scale dataset from Sina Weibo. Beyond user's tweeting contents, 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 of stressed users is around 14% higher than that of non-stressed users, indicating that the social structure of stressed users' friends tends to be less connected and complicated, compared to that of non-stressed users. The contributions of this paper are as following. We propose a unified hybrid model integrating CNN with FGM to leverage both tweet content attributes and social interactions to enhance stress detection. We build several stressed-twitter-posting datasets by different ground-truth labeling methods from several popular social media platforms and thoroughly evaluate our proposed method on multiple aspects. We carry out in-depth studies on a real-world large scale dataset and gain insights on correlations between social interactions and stress, as well as social structures of stressed users.

4. Problem Statement

The correlations of user interactions and stress are analyzed for the stress detection process. Stress-related textual, visual and social attributes are extracted for the stress detection process. The hybrid model combines the Factor Graph and the Convolutional Neural Network for stress detection. Tweet content and social interaction are analyzed in the hybrid model. The following problems are identified from the existing system. Sparse connection factors are not focused. Spatio temporal behaviors are not considered. Binary level stress analysis model. Limited detection accuracy levels.

5. User Stress Discovery and Recovery

Social network user behaviors are analyzed using the user reviews and profiles. Binary and multi level stress classes are discovered using the social network data. The tweeter data values are used in the stress discovery process. The system is divided into six major modules. They are Social Network Data Analysis, Attribute Extraction, Factor Graph Construction, Binary Level Stress Discovery, Multi Level Stress Discovery and Treatment Advisory.

The user profile and reviews are analyzed in the social network data analysis. The attribute extraction process fetches the parameters from the social network data. The factor graph is constructed user connection details. The binary level stress discovery model identifies the user behaviors. Different stress categories are discovered under the multi level stress discovery process. The treatment suggestions are provided from the treatment advisory.

5.1. Social Network Data Analysis

User profile, check in details and reviews are collected from social networks. User check-in venue and time details are maintained in location data values. User submitted messages are maintained under tweets data collection. The session identification process is carried out with the user activity information.

5.2. Attribute Extraction

Stress detection attributes are defined with tweet level and user level aspects.

Tweet level attributes are fetched from user's single tweet. User level attributes are extracted with user's weekly tweets. The tweet level attributes are extracted from the text, image and attention list. The social attention list indicates the being liked, retweeted or commented status. The user level attributes are composed with posting behavior and social interaction attributes. User's weekly tweets postings are summarized as posting behavior attributes. Social interaction attributes extracted from a user's social interactions with friends.

5.3. Factor Graph Construction

The factor graph is build to support the learning process in the convolutional neural network (CNN). Convolutional neural network (CNN) is build with cross auto encoders (CAE) to generate user level attributes from tweet level attributes. The partially labeled factor graph (PFG) is defined with user level social, posting and learnt content attributes. The factor graph is passed to the stress detection process.

5.4. Binary Level Stress Discovery

The binary level stress discovery process identifies two different classes. Stressed user and normal user classes are discovered under the binary level stress detection process. The stress discovery process is carried out on the factor graphs. User interaction and posting relationships are analyzed in the stress discovery process.

5.5. Multi Level Stress Discovery

The multi level stress discovery process is applied to different levels of stress. Initial, medium, high and critical stress levels are identified from the multi level stress discovery process. Stress categorizes are identified with attribute relationships. User stress results are ranked with stress categories.

5.6. Treatment Advisory

The treatment advisory is build to suggest treatment for the users. Online counseling, interaction sessions and medical care are suggested for the stressed users. The treatment suggestions are passed to the users and medical service providers. Automated stress release messages are passed to the users based on the stress levels.

6. Conclusion

Timely detection of psychological stress is very essential for proactive care. Online social network data are used to detect the people stress. The tweets are analyzed with the hybrid model for stress detection process. The sparse user interactions and spatio temporal factors are integrated with the hybrid model to improve the stress detection process. User reviews, connections, location and time parameters are analyzed in the stress detection process. The system detects different stress categories. The stress reduction mechanisms are also suggested by the system. The accuracy level is improved with minimum computational overhead.

References

[1] Yuanyuan Xue and et al. Detecting adolescent psychological pressures from micro-blog. Health Information Science. Springer International Publishing, pages 83–94, 2014.

[2] B. O’Connor, R. Balasubramanyan, B. Routledge, and N. Smith, “From tweets to polls: Linking text sentiment to public opinion time series,” in Proc. of the Int’l AAAI Conference on Weblogs and Social Media, 2010.

[3] D. Davidov, O. Tsur, and A. Rappoport, “Enhanced sentiment learning using twitter hashtags and smileys,” in Proc. of the 23rd International Conference on Computational Linguistics, 2010, pp. 241–249.

[4] L. Ku and C. Sun, “Calculating emotional score of words for user emotion detection in messenger logs,” in Proc. of the 13th IEEE Intl. Conf. on Information Reuse and Integration: Workshop on Empirical Methods for Recognizing Inference in Text II (EM-RIT), 2012, pp. 138–143.

[5] N. A. Diakopoulos and D. A. Shamma, “Characterizing debate performance via aggregated twitter sentiment,” in Proc, of the 28th Intl. Conf. on Human Factors in Computing Systems (CHI), 2010, pp. 1195–1198.

[6] L. Jiang, M. Yu, M. Zhou, X. Liu, and T. Zhao, “Target-dependent twitter sentiment classification,” in Proc. of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, 2011, pp. 151–160.

[7] X. Wang, F. Wei, X. Liu, M. Zhou, and M. Zhang, “Topic sentiment analysis in twitter: A graph-based hashtag sentiment classification approach,” in Proc. of the 20th ACM Conf. on Information and Knowledge Management (CIKM), 2011, pp. 1031–1040.

[8] L. Barbosa and J. Feng, “Robust sentiment detection on twitter from biased and noisy data,” in Proc. of the 23rd Intl Conf. on Computational Linguistics (COLING), 2010, pp. 36–44.

[9] J. Bollen, H. Mao, and X. Zeng, “Twitter mood predicts the stock market,” Journal of Computational Science, vol. 2, no. 1, pp. 1–8, 2011.