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—Psychological stress is threatening people's health. It is non-trivial to detect stress timely for proactive care. With the popularity of social media, 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. 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, we 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.

Index Terms—Stress detection, factor graph model, micro-blog, social media, healthcare, social interaction.

✦

1 INTRODUCTION

1.1 Motivation

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 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[2], 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, [7] 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, [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 contents on social media platforms to detect users' psychological stress. Existing works [28],

• *Huijie Lin, Jia Jia*, Jiezhong Qiu and Guangyao Shen are with the Department of Computer Science and Technology, Tsinghua University; with Key Laboratory of Pervasive Computing, Ministry of Education; with Tsinghua National Laboratory for Information Science and Technology (TNList), Beijing 100084, China (email: linhuijie@gmail.com, jjia@mail.tsinghua.edu.cn, xptree@gmail.com, thusgy2012@gmail.com). Jia Jia is the corresponding author.*

• *Yongfeng Zhang is with University of Massachusetts Amherst, US (email: zhangyf07@gmail.com).*

• *Lexing Xie is with Research School of Computer Science at the Australian National University (email: lexing.xie@anu.edu.au).*

• *Jie Tang and Ling Feng are with the Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China (email: jietang@tsinghua.edu.cn, fengling@mail.tsinghua.edu.cn)*

• *Tat-Seng Chua is with the Institute of Systems Science, National University of Singapore (email: chuats@comp.nus.edu.sg).*

[1]http://tinyurl.com/htunr9g

[2]http://www.weibo.com, one of the most popular social media platforms in China

Fig. 1. Sample tweets from Sina Weibo. In each tweet, the top part is tweet content with text and an image; the bottom part shows the social interactions of tweets where there are multiple indicators of stress: mentions of '*busy*' and '*stressed*', '*working overtime*', '*failed the exam*', '*money*' and a stressed emoticon.
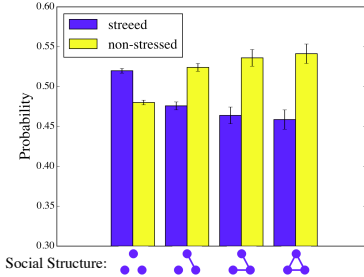

Fig. 2. The sampling test results of the diversity of users' social structures from Sina Weibo, by using the top 3 interacted friends of the users.

[47] demonstrated that leverage social media for healthcare, and in particular stress detection, is feasible.

Limitations exist in tweeting content based stress detection. 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[3]. These phenomena incur the inherent data sparsity and ambiguity problem, which may hurt the performance of tweeting content based stress detection performance. For illustration, let's see a Sina Weibo tweet example in Figure 1. 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* [37]: a bad mood can be transferred from one person to another during social interaction. The second is *linguistic echoes* [34]: 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. The bottom of Figure 2 illustrates four social interaction structure patterns. Each node in a structure pattern represents a user's interacting friend (who either commented or retweeted the tweets). If two nodes are also friends on social network, there is an edge linking both; otherwise, there is none. We examined 3000 users on Sina Weibo. For each user, we collected and merged his/her one week tweets into one and sense stress from it. Meanwhile, we captured the top-3 most active friends the user interacted with. As shown in Figure 2, stressed users' interaction structures are less connected, and thus less complicated than those of non-stressed users.

1.2 Our Work

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 (i.e. with no delta connections[4]) 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.

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.

The rest of this paper is organized as follows. Section 2 gives an overview of related works. Section 3 presents our problem statement. Then in Section 4, we introduce the definitions of the proposed attributes. Section 5 presents the hybrid model and training method for stress detection. Experimental results are shown in Section 6. Then in Section 7, we present several in-depth studies on our dataset for further insights. Finally, we make some conclusions and discuss in Section 8.

2 RELATED WORK

Psychological stress detection is related to the topics of sentiment analysis and emotion detection.

Research on tweet-level emotion detection in social networks. Computer-aided detection, analysis, and application of emotion, especially in social networks, have drawn much attention in recent years [8], [9], [28], [41], [52], [53]. Relationships between psychological stress and personality traits can be an interesting issue to consider [11], [16], [43]. For example, [1] providing evidence that daily stress can be reliably recognized based on behavioral metrics from users mobile phone activity. Many studies on social media based emotion analysis are at the tweet level, using text-based linguistic features and classic classification approaches. [53] proposed 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. [9] 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

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. However, these work mainly leverage the textual contents in social networks. In reality, data in social networks is usually composed of sequential and inter-connected items from diverse sources and modalities, making it be actually cross-media data.

Research on user-level emotion detection in social networks. While tweet-level emotion detection reflects the instant emotion expressed in a single tweet, people's emotion or psychological stress states are usually more enduring, changing over different time periods. In recent years, extensive research starts to focus on user-level emotion detection in social networks [29], [36], [38], [50]. Our recent work [29] proposed to detect users psychological stress states from social media by learning user-level presentation via a deep convolution network on sequential tweet series in a certain time period. Motivated by the principle of homophily, [38] incorporated social relationships to improve user-level sentiment analysis in Twitter. Though some user-level emotion detection studies have been done, *the role that social relationships plays in one's psychological stress states*, and *how we can incorporate such information into stress detection* have not been examined yet.

Research on leveraging social interactions for social media analysis. Social interaction is one of the most important features of social media platforms. Now many researchers are focusing on leveraging social interaction information to help improve the effectiveness of social media analysis. [12] analyzed the relationships between social interactions and users' thinking and behaviors, and found out that Twitter-based interaction can trigger effectual cognitions. [49] leveraged comments on Flickr to help predict emotions expressed by images posted on Flickr. However, these work mainly focused on the content of social interactions, e.g., textual comment content, while ignoring the inherent structural information like *how users are connected*.

3 PROBLEM FORMULATION

Before presenting our problem statement, let's first define some necessary notations.

Let V be a set of users on a social network, and let $|V|$ denote the total number of users. Each user $v_i \in V$ posts a series of tweets, with each tweet containing text, image, or video content; the series of tweets contribute to users social interactions on the social network.

Definition 1. Stress state. The stress state y of user $v_i \in V$ at time t is represented as a triple (y, v_i, t), or briefly y_i^t. In the study, a binary stress state $y_i^t \in \{0, 1\}$ is considered, where $y_i^t = 1$ indicates that user v_i is stressed at time t, and $y_i^t = 0$ indicates that the user is non-stressed at time t, which can be identified from specific expressions in user tweets or clearly identified by user himself, as explained in the experiments. Let Y^t be the set of stress states of all users at time t.

Definition 2. Time-varying user-level attribute matrix. Each user in V is associated with a set of attributes A. Let X^t be a $|V| \times |A|$ attribute matrix at time t, in which every row x_i^t corresponds to a user, each column corresponds to

[4]Meaning that three points are connected with each other

TABLE 1
Summary of tweet-level attributes. The column "#" indicates the feature vector length for each type of feature

Category	Short Name	#	Description
Linguistic	Positive & Negative Emotion Words	2	Number of positive and negative emotion words
	Positive & Negative Emoticons	2	Number of popular positive and negative emoticons, e.g., ☺ and ☹
	Punctuation Marks & Associated Emotion Words	4	To signify the intensity of emotion four typical punctuation marks ('!', '?', '...', '.') are considered.
	Degree Adverbs & Associated Emotion Words	2	In examples "{I feel a little bit sad}" and "{I feel terribly sad}" , 'sad' expresses different negative feelings. We use 1-3 to represent neutral, moderate, and severe degree of positive emotions, and the minus to represent the negative ones.
Visual	Five-color theme	15	A combination of five dominant colors in HSV color space, indicating main color distribution of images, has been revealed to be important on human emotions by psychology and art theories.
	Saturation	2	The mean value of saturation and its contrast.
	Brightness	2	The mean value of brightness and its contrast.
	Warm / Cool color	1	Ratio of cool colors with hue ([0-360]) in the HSV space in [30, 110].
	Clear / Dull color	1	Ratio of colors with brightness ([0-1]) and saturation < 0.6.
Social	Social Attention	3	Number of comments, retweets, and likes

an attribute, and an element $x_{i,j}^t$ is the j-th attribute value of user v_i at time t.

A user-level attribute matrix describes user-specific features, and can be defined in different ways. This study considers user-level content attributes, statistical attributes, and social interaction attributes. A detailed discussion of the matrix can be found in Section 4.

Definition 3. Time-varying edge set. Users are linked by edges of certain types. Let $E^t \subseteq V \times V \times C$ be a set of edges between users at time t. Three types of edges are considered in the study. For an edge $e = (v_i, v_j, c) \in E^t, c = 0$ indicates that v_i follows or is followed by v_j at time t, $c = 1$ indicates that there are positive words in comments between user v_i and v_j at time t, and $c = 2$ indicates that there are negative words in comments between them at time t.

Definition 4. Time-varying attribute-augmented network. An attribute-augmented network at time t is comprised of four elements, including 1) a user set V^t, 2) an edge set E^t, 3) a user-level attribute matrix set X^t, and 4) a stress state set for all users Y^t at time t, denoted as $G^t = (V^t, E^t, X^t, Y^t)$.

Given a sequence of labeled time-varying attribute-augmented networks at different times, our goal is to learn a model that can best fit the relationships among users' stress states, user-level attributes, and users' social linkage, and then detect users' unknown stress states with the model.

Problem 1. Psychological stress detection: Given a series of T partially labeled time-varying attribute-augmented networks $\{G^t = (V_L^t, V_U^t, E^t, Y_L^t) \mid t \in \{1, 2, \cdots, T\}\}$, V_L^t is a set of users with labeled stress states Y_L^t at time t, and V_U^t is a set of unlabeled users, the objective is to learn a function

$$f : \{G^1, G^2, \ldots G^T\} \rightarrow \{Y_U^1, Y_U^2, \ldots Y_U^T\}$$

to predict unlabeled users' stress states.

4 ATTRIBUTES CATEGORIZATION AND DEFINITION

To address the problem of stress detection, 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.

4.1 Tweet-level Attributes

Tweet-level attributes describe the linguistic and visual content, as well as social attention factors (being liked, commented, and retweeted) of a single tweet.

For linguistic attributes, we take the most commonly used linguistic features in sentiment analysis research. Specifically, we first adopt LTP [4] — A Chinese Language Technology Platform — to perform lexical analysis, e.g., tokenize and lemmatize, and then explore the use of a Chinese LIWC dictionary — LIWC2007 [14], 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 NRC[5] and HowNet[6], and found that the performances were relatively the same, so we adopted the commonly used LIWC2007 dictionary for experiments. Furthermore, we extract linguistic attributes of emoticons (e.g., ☺ and ☹) and punctuation marks ('!', '?', '...', '.'). Weibo defines every emoticon in square brackets (e.g., they use [haha] for "laugh"), so we can map the keyword in square brackets to find the emoticons. Twitter adopts Unicode as the representation for all emojis [15], [24], which can be extracted directly. The list of linguistic attributes and descriptions are shown in Table 1.

As for the visual attributes, we use API from OpenCV [7] to perform picture processing and color-related attributes computation, e.g., saturation, brightness, warm/cool color, clear/dull color in Table 1. For a special class of attributes named five-color theme, we adopt algorithm from papers on affective image classification [32] and color psychology theories [23], [45]. In this work, we did not adopt the direct emotional detection results as visual features because we need multi-dimensional visual features for deep model learning, while a direct visual emotional classification result only gives a single or very few dimensions as features. However, with the development of emotion-sensitive visual representation techniques, it would be possibility to adopt

[5]http://www.saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm
[6]http://www.keenage.com
[7]http://opencv.org

automatic visual features in the future. The details of tweet-level attributes are summarized in Table 1.

4.2 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.

Thus, appropriately designed user-level attributes can provide a macro-scope of a user's stress states, and avoid noise or missing data. Here, we define user-level attributes from two aspects to measure the differences between stressed and non-stressed states based on users' weekly tweet postings: 1) *user-level posting behavior attributes* [29] from the user's weekly tweet postings; and 2) *user-level social interaction attributes* from the user's social interactions beneath his/her weekly tweet postings. The details of user-level attributes are summarized in Table 2.

5 MODEL FRAMEWORK

Two challenges exist in psychological stress detection. 1) *How to extract user-level attributes from user's tweeting series and deal with the problem of absence of modality in the tweets?* 2) *How to fully leverage social interaction, including interaction content and structure patterns, for stress detection?* To tackle these challenges, we propose a novel hybrid model by combining a factor graph model with a convolutional neural network (CNN), since CNN is capable of learning unified latent features from multiple modalities, and factor graph model is good at modeling the correlations. In this section, we will first introduce the architecture of our model, and then describe the details of each part of the proposed model.

5.1 Architecture

Figure 3 shows the architecture of our model. There are three types of information that we can use as the initial inputs, i.e., tweet-level attributes, user-level posting behavior attributes, and user-level social interaction attributes, whose detailed computation will be described later. We address the solution through the following two key components:

- First, we design a CNN with cross autoencoders (CAE) to generate user-level interaction content attributes from tweet-level attributes. The CNN has been found to be effective in learning stationary local attributes for series like images [3], [6] and audios [30], [48].
- Then, we design a partially-labeled factor graph (PFG) to incorporate all three aspects of user-level attributes for user stress detection. Factor graph

model has been widely used in social network modeling [10], [39], [44]. It is effective in leveraging social correlations for different prediction tasks.

Take the user labeled with a red star in Figure 3 as an example. We extract attributes from each tweet of the user to form tweet-level attributes as shown in the cylinders. Different colors represent different modalities and blank (white color) represents modalities that are not available in the tweet. The tweet-level attributes in the cylinder are fed to cross autoencoders (CAEs) [28]. The CAEs are embedded in a CNN [26], [29] that will integrate attributes from CAEs into the aggregated user-level content attributes by pooling each attribute map. The user-level content attributes, user-level posting behavior attributes, and user-level social interaction attributes together form the user-level attributes. The user-level attributes of a user at time t are denoted by x_i^t (i=1,2,$\cdots$) in Figure 3. The route of the other users' attributes in Figure 3 are similar, which finally form their user-level attributes. We focus on the attribute flow of the user with red star and omit the detailed route of other users' attributes in the figure. The stress state of each user at time t is denoted by y_i^t (i=1,2,$\cdots$), respectively. The user-level attributes and the stress states are connected by an attribute factor, while stress states of different users are connected by social factors. Stress states of the same user at adjacent times are connected by dynamic factors. We define the graph as a (PFG). By calculating the factors, we can finally derive all users' stress states over different weeks.

In the following, we will describe the details of the CNN with CAE and PFG used in the architecture that tackles the tweet series with cropped modalities and leverages the social interaction information between users, respectively.

5.2 Learning Aggregated Attributes From Tweet Series

To aggregate user-level attributes, we need to face two major challenges: (1) Missing modality, e.g., tweets with only text but no picture AND (2) How to generate a distributed and modality-invariant representation for each tweets.

To solve above challenges in cross-media tweet data, we use a cross auto-encoder (CAE) [28] to learn the modality-invariant representation of each single tweet with different modalities. Denoting the text, visual, and social attributes of a tweet by v_T, v_I, and v_S, the CAE is formulated as follows:

$$\begin{cases} u = f(w_T v_T + w_I v_I + w_S v_S + b) \\ (\widetilde{v}_T, \widetilde{v}_I, \widetilde{v}_S) = f(\widetilde{w}u + \widetilde{b}) \end{cases} \quad (1)$$

where u is the modality-invariant representation. w_T, w_I, w_S, and b are parameters in the encoder, whereas $\widetilde{w}_T$, $\widetilde{w}_I$, $\widetilde{w}_S$, and $\widetilde{b}$ are parameters in the decoder. $f(\cdot)$ is the activation function. We use a sigmoid activation function $f(z) = \frac{1}{1+\exp(-z)}$ in our model. $\widetilde{v}_T$, $\widetilde{v}_I$, $\widetilde{v}_S$ are the reconstructed input modalities.

The basic idea of CAE is to force the model to reconstruct missing modalities in the training stage and to learn cross modalities correlation from the data (e.g. negative words in text correlate with cool color in pictures). [18] While training the cross auto-encoder, we use training data that contains all the three modalities. We manually disable the

TABLE 2
Summary of user-level attributes. The column "#" indicates the feature vector length for each type of feature

| Category	Short Name		#	Description
Posting Behavior	Social Engagement		3	The numbers of @-mentions, @-retweets, and @-replies in weekly tweet postings, indicating one's social interaction activeness with friends.
	Tweeting time		24	The numbers of tweets posted in hours with a 24-dimensional vector.
	Tweeting type		4	Categorize users' tweets into mainly four types based on general categories of social media platforms: (1) Image tweets (tweets containing images); (2) Original tweets (tweets that are originally authored and posted by the user); (3) Information query tweets (tweets that ask questions or ask for help); (4) Information sharing tweets (tweets that contain outside hyperlinks). We use a 4-dimensional vector of the numbers of tweets in the above 4 types respectively to quantify the tweeting type attribute.
	Tweeting linguistic style		10	Adopt 10 categories from LIWC that are related to daily life, social events, e.g., personal pronouns, home, work, money, religion, death, health, ingestion, friends, and family. We extract words from users' weekly tweet postings, and use a 10-dimensional vector of numbers of words in the 10 categories
Social Interaction	Content Style	Words	10	A 10-dimensional integer vector, with each value representing the number of words from social interaction content of users weekly tweet postings in each word category from LIWC;
		Emoticons	2	A 2-dimensional integer vector with each value representing the number of positive and negative emoticons (e.g., ☺ and ☹) in tweets.
	Social Influence	Stressed Neighbor Count	1	The number of the user's stressed neighbors.
		Strong-tie Count	1	The number of stressed neighbors with strong tie.
		Weak-tie Count	1	The number of stressed neighbors with weak tie.
		Follower Count	1	The number of the user's followers.
		Fans Count	1	The number of the user's fans.
	Social Structure		8	Representing the structure distribution of the user's interacted friends, where each element refers to the existence of the corresponding structure in Fig. 6.
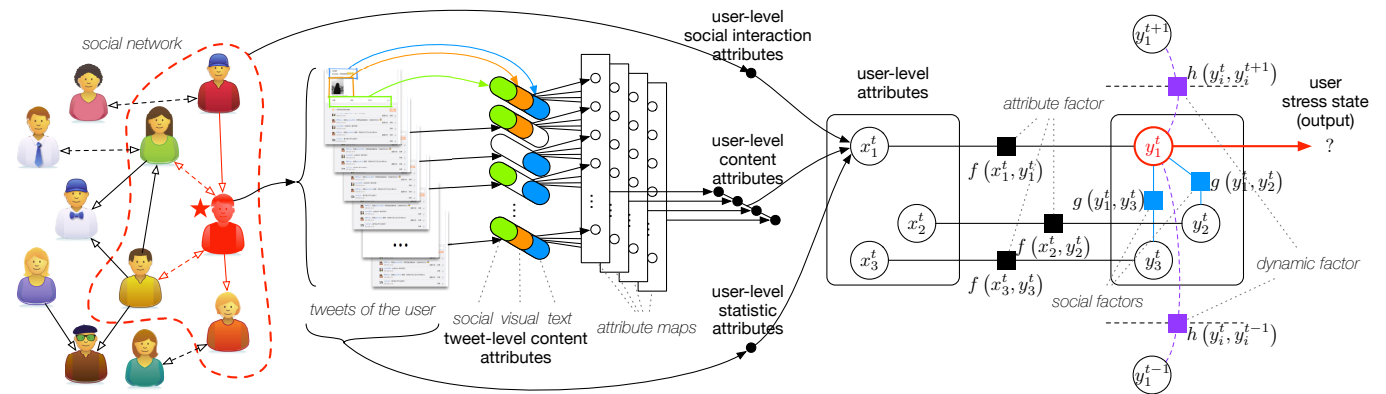
Fig. 3. Architecture of our model. The model consists of two parts. The first part is a CNN. The second part is a FGM. The CNN will generate user-level content attributes by convolution with CAE filters as input to the FGM. Take the user labeled with a red star as example. Tweet-level attributes of the user are processed through a convolution with CAE to form the user-level content attributes. The user-level attributes are denoted by x_i^t in the left box. Every x_i^t contains three aspects: user-level content attributes, user-level posting behavior attributes, and user-level social interaction attributes. Data of other users follows the same route. In the FGM, attribute factors connect user-level attributes to corresponding stress states. Social factors connect the stress state of different users. Dynamic factors connect stress state of a user over time. The output of the user's user-level stress state at time t is y_1^t as highlighted in red, which actually denotes the stress state of the user in weekly period in this paper.

visual modalities and/or social interaction [8] modality of the training data, and require it to reconstruct all three modalities. We train the CAE with a cropped set of data $\overline{v}_T, \overline{v}_I, \overline{v}_S$ that inputs from one or two modalities are absent, while requiring it to reconstruct all the three.

We use the stochastic gradient descent to train the CAE. Denoting all the parameters in the CAE as θ, the energy function is defined as follows.

$$
\begin{aligned}
J\left(\overline{v}_T, \overline{v}_T, \overline{v}_S; \theta\right) = \frac{1}{2} & \left(\sum_{M \in T, I, S} \|\widetilde{v}_M - v_M\|^2\right) \\
& + \frac{\lambda}{2} \left(\sum_{m \in T, I, S} \|w_M\|^2 + \|\widetilde{w}_M\|^2\right).
\end{aligned}
\tag{2}
$$

The first term measures the reconstruction accuracy. The second term is the weight decay regularization term that prevents parameters in the model from diverging arbitrarily. λ is the regularization weight. Using data with different modalities as input, the CAE can be trained and learn a modality-invariant representation u.

The attributes of tweets, which come from a user's weekly tweets in timeline, form a time series. To model a user as a subject of series of tweets, we apply CNN [26] which has large learning capacity, but has much fewer connections and parameters to learn than similar-size standard network layers. It focuses on learning stationary local attributes from series like images (pixel series), audio, and other time series. We can learn user-level content attributes from a series of individual tweets in a time series to describe a user's stress state over a week. All attributes of tweets in a time series form a one-Dimensional series. We use an 1-Dimension CNN in our model.

CAE units are listed in the attribute maps of the CNN. They connect to a patch of instance. CAE units take patches with missing modalities and generate modality-invariant attribute maps. The CAE units are used as filters in the 1-D CNN and convolute over the sequence of tweets to form one feature map. Thus the latent user-level content attributes can be generated from the tweet-level attributes of single tweets.

Pooling is another important step to summarize attribute maps into fewer attribute instances. Though different users have different number of tweets in different weeks, the period of time over which the tweets are sampled are the same. We simply pool each attribute map into one pooled attribute. There are two commonly used pooling operations: max-pooling and mean-pooling. When max pooling is used, the pooled attribute unit is assigned with the maximal activation among all units in the attribute map. When mean-pooling is applied, the mean of activations of all units in the attribute map is assigned to the pooled attribute unit. Since we pool over the period of time rather than a certain number of tweets, we use mean-over-time (MOT) in this paper, which can be calculated by summing up the activations, since the tweet instances are sampled in the same length of time intervals.

[8]Different from the social interaction attributes in this paper, the social interaction here is the attribute of a single tweet defined in [28]. It is simply the mean and variance of interaction numbers of a tweet.

5.3 Learning Latent Correlations Between Tweet's Content And Social Interactions

As the social correlation between users and time-dependent correlation are hard to be modeled using classic classifiers such as SVM, we use a partially-labeled factor graph model (PFG), which was first proposed in [39], to incorporate social interactions and tweets' content for learning and detecting user-level stress states.

We define an objective function by maximizing the conditional probability of users' stress states $\boldsymbol{Y}$ given a series of attribute-augmented networks

$$
\boldsymbol{G} = \{G^t = \{(V^t, E^t, X^t, Y^t)\}\}, t \in \{1, \ldots, T\}
$$

and $V = V^1 = \cdots = V^T, |V| = N$, i.e., $P_\theta(\boldsymbol{Y}|\boldsymbol{G})$. The factor graph [25] provides a way to factorize the "global" probability as a product of "local" factor functions, which makes the maximization simple, i.e.,

$$
P(\boldsymbol{Y}|\boldsymbol{G}) = \prod_{t=1}^{T}\prod_{i=1}^{N} f(\boldsymbol{x_i^t}, y_i^t)h(y_i^t, y_i^{t+1}) \prod_{e \in E^t} g(y_e)
\tag{3}
$$

The joint probability has three types of factor functions, corresponding to the intuitions we have discussed.

Attribute factor. We use this factor $f(\boldsymbol{x_i^t}, y_i^t)$ to represent the correlation between user v_i's stress state at time t and her/his attributes $\boldsymbol{x}_i^t$. More specifically, we instantiate the factor by an exponential-linear function:

$$
f(\boldsymbol{x}_i^t, y_i^t) = \frac{1}{Z_\alpha} \exp\left\{\alpha^{\mathrm{T}} \boldsymbol{x}_i^t\right\}
\tag{4}
$$

where α is a parameter of the proposed model, and Z_α is a normalization term.

Dynamic factor. We use this factor $f(y_i^t, y_i^{t+1})$ to represent the time correlation between user v_i's stress state at time t and $t + 1$. More specifically, we instantiate the factor by an exponential-linear function:

$$
h(y_i^t, y_i^{t+1}) = \frac{1}{Z_\gamma} \exp\left\{\gamma^{\mathrm{T}} h'(y_i^t, y_i^{t+1})\right\}
\tag{5}
$$

where γ is the model parameters for this type of factor, $h'(\cdot)$ is defined as a vector of indicator functions, and Z_γ is the normalization term.

Social factor. We use social factor $g(y_e)$ (where $e = (v_i^t, v_j^t, c) \in E^t$) to represent the correlation between user v_i and v_j's stress states according to c at time t:

$$
g(y_e) = \frac{1}{Z_{\beta_c}} \exp\left\{\beta_c{}^{\mathrm{T}} g'(y_i^t, y_j^t)\right\}
\tag{6}
$$

where β_c is the model parameters for this type of factor, $g'(\cdot)$ is defined as a vector of indicator functions, and Z_{β_c} is the normalization term.

Finally, by combining Eq.4, 5, 6 into Eq.3, the objective function as the log-likelihood of the proposed model is:

$$
\begin{aligned}
\mathcal{O} = & \sum_{t=1}^{T}\sum_{i=1}^{N} \alpha^{\mathrm{T}} \boldsymbol{x}_i^t + \sum_{t=1}^{T}\sum_{i=1}^{N} \gamma^{\mathrm{T}} h'(y_i^t, y_i^{t+1}) \\
& + \sum_{t=1}^{T}\sum_{e \in E^t} \beta_c^{\mathrm{T}} g'(y_i^t, y_j^t) - \log Z
\end{aligned}
\tag{7}
$$

where $Z = Z_\alpha \prod_{c \in C} Z_{\beta_c} Z_\gamma$ is the global normalization term.

Input: a series of time-varying attribute augmented network G with stress states on some of the user nodes, learning rate η;

Output: parameter value $\theta = \{\alpha, \{\beta_c\}, \gamma\}$ and full stress state vector Y;

Randomly initialize Y;
Initialize model parameters θ;
repeat
> Compute gradient $\nabla\alpha, \nabla\beta_c, \nabla_\gamma$;
> Update $\alpha \leftarrow \alpha + \eta \times \nabla\alpha$;
> Update $\beta_c \leftarrow \beta_c + \eta \times \nabla\beta_c$;
> Update $\gamma \leftarrow \gamma + \eta \times \nabla\gamma$;

until *convergence*;

Algorithm 1: Learning and inference by factor graph

Learning. Learning the predictive model is to estimate a parameters configuration $\theta = (\alpha, \{\beta_c\}, \gamma)$ from the partially-labeled dataset and to maximize the log-likelihood objective function Eq. 7, i.e., $\theta^* = \arg\max_\theta \mathcal{O}(\theta)$.

For optimization, we adopt a gradient decent method. Specifically, we derive the gradients with respect to each parameter in our objective function of Eq. 7.

$$\frac{\partial\mathcal{O}}{\partial\alpha} = E\left[\sum_{t=1}^{T}\sum_{i=1}^{N} f(\boldsymbol{x}_i^t, y_i^t)\right] - E_{P_\alpha(\boldsymbol{Y}|\boldsymbol{G})}\left[\sum_{t=1}^{T}\sum_{i=1}^{N} f(\boldsymbol{x}_i^t, y_i^t)\right]$$

$$\frac{\partial\mathcal{O}}{\partial\beta} = E\left[\sum_{t=1}^{T}\sum_{e\in E^t} g(y_e)\right] - E_{P_\beta(\boldsymbol{Y}|\boldsymbol{G})}\left[\sum_{t=1}^{T}\sum_{e\in E^t} g(y_e)\right]$$

$$\frac{\partial\mathcal{O}}{\partial\gamma} = E\left[\sum_{t=1}^{T}\sum_{i=1}^{N} h(y_i^t, y_i^{t+1})\right] - E_{P_\gamma(\boldsymbol{Y}|\boldsymbol{G})}\left[\sum_{t=1}^{T}\sum_{i=1}^{N} h(y_i^t, y_i^{t+1})\right]$$

$$(8)$$

where in the first equation, $E\left[\sum_{t=1}^{T}\sum_{i=1}^{N} f(\boldsymbol{x}_i^t, y_i^t)\right]$ is the expectation of the summation of the attribute factor functions given the data distribution over Y and G in the training set, and $E_{P_\alpha(\boldsymbol{Y}|\boldsymbol{G})}\left[\sum_{t=1}^{T}\sum_{i=1}^{N} f(\boldsymbol{x}_i^t, y_i^t)\right]$ is the expectation of the summation of the attribute factor functions given by the estimated model. The other expectation terms have similar meanings in the other equation.

As the network structure in the real world may contain cycles, it is intractable to estimate the marginal probability in the second terms of 8. In this work, we adopt Loopy Belief Propagation (LBP) [33] to calculate the marginal probability of $P(\boldsymbol{Y})$ and compute the expectation terms. The learning process can then be described as an iterative algorithm. Each iteration contains two steps. Firstly, we call LBP to calculate marginal distributions of unknown variables $P_\alpha(\boldsymbol{Y}|\boldsymbol{G})$. Secondly, we update α, β, γ with the learning rate η by Eq.9 The learning algorithm terminates when it reaches convergence.

$$\theta_{new} = \theta_{old} + \eta\frac{\partial\mathcal{O}}{\partial\theta} \qquad (9)$$

Detection. With the estimated parameter θ, we can now assign the value of unknown labels Y by looking for a label configuration that will maximize the objective function, i.e.,

$$Y^* = \arg\max \mathcal{O}(\boldsymbol{Y}|\boldsymbol{G}, \theta) \qquad (10)$$

In this paper, we use a max-sum algorithm [31] to solve this problem.

6 EXPERIMENTS

In this section, we will present the effectiveness and efficiency of our hybrid model on user-level stress detection.

6.1 Dataset Collection

To conduct observations and evaluate our succecive model, we first collect a set of datasets using different labeling methods, which are listed as following:

Dataset DB1: It is a challenge to construct a dataset with reliable ground truth labels from large-scale noisy social media data. The data crawled from social platforms is usually massive, thus manual labeling methods are not feasible due to the uncontrollable cost and quality. To solve this problem, we employed a sentence pattern labeling method to automatically extract labeled data from the crawled large-scale social media data. We first crawled 350 million tweets data via Sina Weibo's REST APIs[9] from Oct. 2009 to Oct. 2012. Sina weibo, as the biggest microblog website in China, provides users an open online platform for information sharing, communication and obtaining. Similar to Twitter and Facebook, users on Sina Weibo can post contents with multiple modalities, including text, image, social action (retweet, comment, favorite), video and etc. Despite these user generated contents, user relationship, which takes the form of following on Sina Weibo, also contains abundant information for data analysis. Utilizing above information and features extracted from multiple modalities, we are able to investigate users emotions, stresses and opinions.

We then tried to identify the weekly stressed state of users. Facing the vast scale of social images, manually labeling is powerless. Instead, we use tags and comments for automatic image labeling, which is a common method in previous work. [20], [21], [46] This is done by searching for tweets containing patterns like "*I feel stressed this week*" and "*I feel stressed so much this week*", which are used to indicate that the users are stressed. The weeks containing such sentence patterns are labeled as "stressed" weeks. Similarly, we identify "non-stressed" weeks of users by searching for tweets with patterns like "*I feel relaxed*" and "*I feel non-stressed*". These sentence patterns have been shown to have high precision against user-assigned psychological state labels validated by online surveys in weibo [29].

In this way, we collected over 19,000 weeks of tweets that are labeled as stressed, and over 17,000 weeks of non-stressed users' tweets. There are 492,676 tweets from 23,304 users in total. We use this dataset for experiments, analysis and further in-depth studies, which is represented by DB1 in this paper. Details of the dataset are shown in Table 3.

Dataset DB2: We verified the reliability of the above ground truth labeling method through dataset DB2 in Table 4. It is a small dataset collected from the users who have shared the score of a psychological stress scale PSTR[10] designed by psychologists via Weibo. Guided by the rules of the PSTR scale, a user is taken as stressed when the score is larger than 80, otherwise non-stressed. We thus crawled the scores posted by users, and used the scores as ground truth label for the set of tweets in +-3-day window.

---

[9]http://open.weibo.com
[10]http://types.yuzeli.com/survey/pstr50

TABLE 3
Overview of the Weibo-Stress dataset

	non-stressed	stressed	total
#tweets	253,638	239,038	492,676
#users	12,230	11,074	23,304
#weeks	17,861	19,136	36,997
#tweets/week*	14.2	12.5	13.3
#weeks/user*	1.46	1.73	1.59
#interacting users/week*	5.79	6.99	6.35

* means average number.

Dataset DB3 and DB4: To further test our method, we collected two more datasets from Tencent Weibo (DB3) and Twitter (DB4). They are again labeled using the sentence pattern labeling method as described above for DB1. In particular, as social platforms of different languages, Weibo and Twitter have many differences. [51]. For example, their top topics differs very much. Thus, experiments on Twitter can validate the universality of our method. The details of the two datasets are presented in Table 4.

6.2 Experimental Setup

In the following experiments, we first train and test our model on the large-scale Sina Weibo dataset DB1. We then test our model on the other 3 datasets to show effectiveness of the proposed model on different data sources or different ground truth labeling methods. For all of our analysis, we use 5-fold cross validation, with over 10 randomized experimental runs.

Comparison Methods. We compare the following classification methods for user-level psychological stress detection with our FGM+CNN model (denoted as FGM here).

- **Logistic Regression (LRC) [19]:** it trains a logistic regression classification model and then predicts users' labels in the test set.
- **Support Vector Machine (SVM) [5]:** 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) [42]:** 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) [13]:** it trains a gradient boosted decision tree model with features associated with each user.
- **Deep Neural Network (DNN) [29] 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.

We employ scikit-learn[11] for the above methods.

Evaluation Measures. For a fully investigation of the proposed methods, we consider the following aspects:

- **Effectiveness.** We evaluate the detection performance of our model and comparison methods

[11]http://scikit-learn.org

in terms of Accuracy (Acc.), Recall (Rec.), Precision (Prec.) and F1-Measure (F1) [2].
- **Efficiency.** We evaluate efficiency of the methods by comparing the CPU time of training each model. All experiments are performed on an x64 machine with 2.9GHz intel Core i7 CPU and 8GB RAM.

6.3 Experimental Results on DB1

Comparison of Detection Performance. To evaluate the effectiveness of our model, we first conduct a test using different models based on the Weibo-Stress dataset. In this experiment, we used all the three attributes described in previous section: user-level social interaction attributes, user-level posting behavior attributes and user-level content attributes generated from the tweet-level attributes by CNN+CAE. Table 5 shows the experimental results. We see that FGM gains superior results against the comparative methods, which verifies that our proposed model can effectively leverage the social interaction and social structure attributes for stress detection. Compared with the results in [29], which also aims at user-level stress detection based on social media data sources, our proposed model improves the detection performance by up to 9% on F1-score. These results demonstrate the feasibility of stress detection via the brand new information source of social interactions, and that our proposed model can significantly enhance the performance by leveraging the social interaction information. We further perform t-tests and all the p-values are ≤ 0.01, indicating that the improvements of our proposed models over the comparison methods are statistically significant.

Comparison of Model Efficiency. To evaluate the efficiency of the aforementioned methods, we compare the CPU time of training each model. The comparison results are also shown in Table 5. Overall, all methods have good efficiency performance, and the running time of different methods ranges from seconds to minutes. FGM results in a slightly lower but better performance compared to other methods.

Factor Contribution Analysis. The definition of factors is important to the performance of the Factor Graph Model. We have three types of factors in our model, i.e., attribute factor, social factor, and dynamic factor. To analyze the impact of different factors in our model, we compare the detection performance with different combinations of factors in this experiment, as shown in Figure 4(a). Specifically, we first use all the three factors, denoted as FGM, then we remove the following factors respectively: social factor, dynamic factor and both of them, denoted as FGM-S, FGM-D and FGM-S-D We see that the worst performance is achieved if we incorporate only the attribute factor. However, integrating attribute factor with social or dynamic factor both improve the performance, revealing that both of the two factors are effective for stress detection. Specifically, incorporating social factor significantly improves the detection performance to around 91% on accuracy, indicating that the social factor is extremely effective. The best detection performance is observed when using all three types of factors.

TABLE 4
Details of other Datasets

Platform	Stress label	Number of tweets	Number of users	Number of weeks	Tweets per week
DB2:Sina Weibo (2010.2-2011.9)	stressed	1,459	98	98	14.9
	non-stressed	1,845	112	112	16.5
	summary	3,304	210	210	15.7
DB3:Tencent Weibo (2011.11-2013.3)	stressed	138,570	7,845	8,974	15.4
	non-stressed	172,585	8,239	9,976	17.3
	summary	311,155	16,084	18,950	16.4
DB4:Twitter (2009.6-2009.12)	stressed	54,748	4,905	6,081	9.0
	non-stressed	75,357	4,018	6,545	11.5
	summary	130,105	8,923	12,626	10.3

(a) Factor Contribution   (b) Data Scale Analysis   (c) Convergence Analysis   (d) DNN Layer Analysis
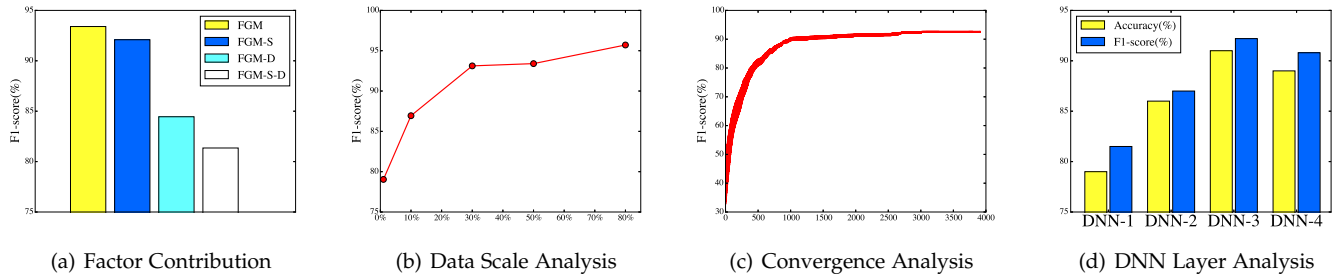
Fig. 4. Experiment results analysis from various perspectives. (a) Attribute contribution analysis; (b) Factor contribution analysis; (c) Results of detection performance with different training data scales; (d) Convergence Analysis of FGM.
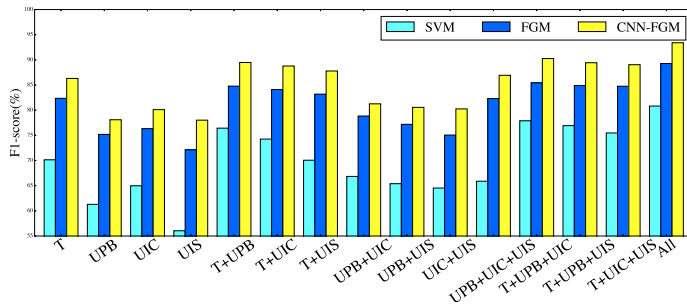

Fig. 5. Experiment results analysis of different attribute combinations on different models, with T, UPB, UIC, and UIS representing Tweet-level attributes, User-level Posting Behavior attributes, User-level social Interaction Content attributes and User-level social Interaction Structure attributes respectively. For example, 'UIC+UIS' here means a combination of User-level social Interaction Content attributes and User-level social Interaction Structure attributes.

TABLE 5
Comparison of efficiency and effectiveness using different models (%).

Method	Acc.	Rec.	Prec.	F1	CPU time
LRC	76.18	87.94	78.58	83.00	39.43s
SVM	72.58	87.39	75.16	80.82	≈10min
RF	77.73	89.63	79.35	84.18	67.71s
GBDT	79.75	82.99	85.90	84.43	262.86s
FGM	91.55	96.56	90.44	93.40	≈20min

Training Data Scale Analysis. To evaluate the data scalability of the proposed model, we try to train the model with different scale of training data, and compare the final detection performance in F1-score. In this test, we use all the three attributes as input. Figure 4(b) shows the trend of detection performance with different proportions of training data. It is clear that when using only 1% of all training data, our model fails to achieve meaningful detection performance. When adopting approximately 30% of all training data, our model can obtain an equally competitive performance of around 93% compared with that when using 50% of training

data. Moreover, the performance keeps increasing given more training data. These results verify the scalability of our model on large-scale real-world social media datasets.

Convergence Analysis. We further investigate the convergence of the learning algorithm for FGM, and Figure 4(c) presents the F1-score with increasing number of iterations. We see that the algorithm converges within around 2000 iterations, which is rapid enough for us to conduct efficient model training on large scale datasets in practice.

Impact of size of network. Size of network is a critical issue in setting up DNN model. Shallow networks result in trivial model that cannot catch any underlying correlation in data, whereas too deep networks lead to over-complex model which is difficult to tune and may suffer from problems like over-fitting. To choose an appropriate DNN model for classification, we test DNN with different number of layers. Figure 4(d) summarizes the experiment results. It is clear that 2-layer is not sufficient for the model to achieve a satisfactory result. 3-layer model improve significantly while 4-layer model reaches the peak. 5-layer model does not get better result. This is mainly because at 5-layer the network may be too large that it cannot be tuned well with the available data and within a feasible training time.

Attribute Contribution Analysis. As described in Section 4, we have defined several set of tweet-level and user-level attributes from a single tweet's content as well as users' posting behaviors and social interactions in a weekly period. To evaluate the contribution of different attributes and compare the effectiveness of our model of leveraging different attributes, we compared the proposed model with other existing models by using different combinations of attributes as input. As described in Section 4, the proposed attributes are categorized into four groups: tweet-level attributes, user-level posting behavior attributes, user-level

social interaction content attributes and user-level social interaction structure attributes, denoted as T, UPB, UIC, and UIS respectively. We compare the detection performance of the proposed CNN+FGM model with SVM and CNN with traditional autoencoder, with all the possible combinations of these four set of attributes. For the SVM with the tweet-level attributes, we simply take the average of the feature vectors from a user's weekly tweets.

The results of this experiment are shown in Figure 5. We see that all the models achieve the best detection performance when utilizing all the three set of attributes. When using only the tweet-level features, the detection performance of the proposed model and the DNN model drops to around 86% and 82% respectively in F1-score, which is acceptable. While for SVM, the detection performance drops to around only 70%, which is poor for a binary classification. This result demonstrates the effectiveness of the feature aggregation of CNN, which is much better than simply summarizing the feature vectors manually.

Figure 5 also shows the effectiveness of different attributes. We can see that by using only user-level attributes, the detection performance of all the models drops drastically compared to that using only tweet-level attributes, which shows the importance of the tweet-level attributes. By combing different types of user-level attributes, the detection performance improves by around 3-8% in F1-score, showing that the user-level attributes are supplementary to each other. Meanwhile, by combining the user-level attributes with tweet-level attributes, the detection performance improves up to 10-20% in F1-score. This result indicates that the user-level attributes are great supplements to tweet-level attributes.

When using only two set of attributes, the detection performance drops to around 91% in F1-score. In case of using sole attributes, we see that using solely user-level social interaction attributes gets the best detection performance of around 90% in F1-score, as compared to the other attributes. This reveals that the proposed user-level social interaction attributes are quite effective for stress detection.

Impact of different modalities in content attributes: Tweets content come with multiple modalities. To evaluate the contribution of each data modality, we conduct experiments with different combination of attributes. Since text is the necessary part of a tweet, we test using solely text attributes, and the two combinations of text and visual attributes, and text and social attributes, as well as using all attributes. The results are shown in Table 6. It is interesting to note that using only text attribute could achieve rather high performance. Simply combining visual or social attributes with text attributes may even reduce the performance, especially the social attributes. This trend is even more obvious when both types of attributes (content and posting behavior) are used. Nevertheless, using all attributes together outperforms that using only the text attributes; and the highest performance is observed when using all attribute and working with both types of attributes.

6.4 Results on Other Datasets

We further evaluate our model on other datasets, DB2-DB4, as shown in Table 4, to show that our model is universally

TABLE 6
Comparison of results using different modalities.

	Text	Text + visual	Text + Social	All
Accuracy	0.8713	0.8761	0.8628	0.9155
F1-score	0.8794	0.8865	0.8711	0.9340

applicable. For these experiments, we use all the proposed attributes with MOT pooling, and a 4-layer DNN model.

DB2 from Sina Weibo with PSTR label. We use a matured model trained with large scale Sina Weibo dataset, and then test it against another set of subject independently sampled from Sina Weibo. For the test set, we collect weekly tweets from the users that have shared the score of a psychological stress scale with 50 items via Sina Weibo. Detection result shows that the test accuracy is 84.26% and F1-score is 0.8785, which demonstrates that the overall model is consistent and the sentence pattern based ground truth labeling method is reliable.

DB3 from Tencent Weibo. We test on data collected from another major Chinese social media platform. For this test, we use the attribute extractor trained with large scale Sina Weibo dataset and only finetune the network with Twitter dataset in 5-fold. The accuracy is 86.18% and F1-score is 0.8832 which demonstrate the capability of the model.

DB4 from Twitter. We also test against the Twitter dataset. We still use the attribute extractor trained with large scale Sina Weibo dataset and only finetune the network with Twitter dataset in 5-fold. The accuracy is 77.43% and F1-score is 0.8224. One reason for this modest result is that users in Twitter dataset and Sina Weibo dataset come from different language and culture background, so that the language patterns and sentimental signals from these two different language environments can be different, thus the attribute extractor trained with large scale Sina Weibo dataset may not be fully functional for Twitter datasets. Nevertheless, we still achieved acceptable performance in Twitter dataset, which implies that the basic stress patterns between social relations can be transferred in between different language environments. Another factor could be that the scale of this dataset is rather small. Subjects in the Twitter dataset are on the order of 10% than that in large-scale Sina Weibo dataset. We look into the collected data and find that, by coincidence, all tweets in this dataset have no social activity. We conjecture this is also one of the causes of the unsatisfactory result.

7 STUDIES OF SOCIAL INTERACTION

We have presented the experimental results on stress detection in the previous section, while in the setting of social networks, it would be helpful to further analyze how a user's stress status is developed and how they affect each other. To do so, we try to conduct several studies on DB1 to offer insights on how social interactions contribute to user stress and the task of stress detection from the following aspects:

(1) Content. *How are users' social interaction contents (e.g., language used) related to users' stress states?*

(2) Structure. *Compared to non-stressed users, do stressed users show different structural diversity patterns when they behave in social networks? Do differences of social influence and strong/weak ties exist between stressed and non-stressed users?*
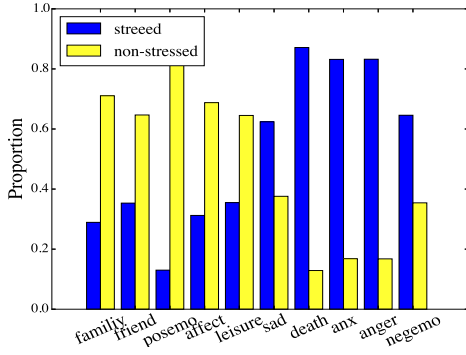
Fig. 6. Distribution of stress states (stressed and non-stressed) over different word categories from tweets' comments and retweets. Here we show 10 most widely used word categories in our data set.

7.1 Content

Content of social interaction refers to the content of tweets' comments and retweets, including text, emoticons, and punctuation marks. Based on a widely used psychological dictionary LIWC2007 [40], we extract emotional words from the interaction content of tweets, and categorize the extracted words into corresponding groups defined in LIWC2007. We compare the frequencies of different word categories between stressed and non-stressed users.

Figure 6 shows the comparison results of the most widely used word categories in our data set, we observe that there is an obvious difference in interaction contents between stressed and non-stressed users. That is, interaction contents of stressed users' tweets contains much more words from categories like death, sadness, anxiety, anger, and negative emotion, while non-stressed users' tweets contain more words from categories like friends, family, affection, leisure, and positive emotion.

7.2 Structure

To examine structure properties (i.e., social influence and strong/weak tie) of (non)stressed users, we use risk ratio (RR) to measure the correlation between users' stress states and different structural attributes. Risk ratio is an effective measurement widely used in the statistical analysis and relevant fields. The risk ratio of a stressed state, associated with a structural attribute a, is calculated as follows:

$$RR(a) = \frac{P(stressed\ user\ has\ attribute\ a)}{P(stressed\ user\ does\ not\ have\ attribute\ a)}. \quad (11)$$

A larger risk ratio implies that users with attribute a are more likely to be stressed. In this section, we investigate representative sociology theories, and quantitatively analyze the correlations between users' stress states and fundamental social concepts, so as to examine how and why a user's stress state is developed and affected by other users.

7.2.1 Structural Diversity

We are interested in whether stressed and non-stressed users have any structural difference in respective friends' connection. In sociology, social structure refers to a society's framework, consisting of various relationships among people, as well as groups that direct and set limits on human behaviors. In social networks, direct connections (following or followed) of users that interact with each other via comments and retweets also form a kind of social structure. For

this in-depth study, we select top four users with the most frequent interactions from users' weekly tweet postings, where four is adopted because this is the minimum number of nodes required to produce structural combinations (10 combinations), so as to calculate the probability of each combination, and incorporating more nodes would make the calculation combinatorial expensive. We measure the connection of the interacting users by the following link, that is, if A is following or followed by B, then A and B are connected, and cliques made up of different nodes are treated the same. We compare the proportion of different social structures of interacting users to measure the structural diversity. The results in Figure 7 clearly show us that structural differences do exist between stressed and non-stressed users. 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 tends to be less connected and less complicated, compared to that of non-stressed users. This phenomenon has also been reported by the current psychological research result that stressed users are more likely to be socially less active [7].
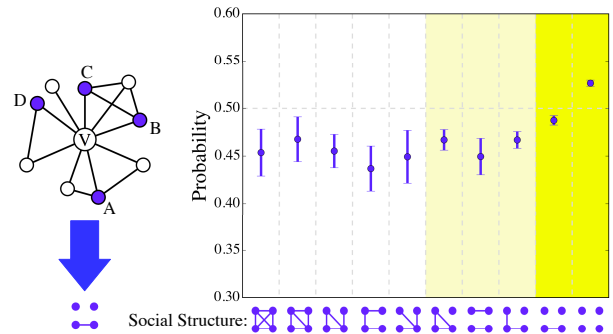

Fig. 7. Distribution of stress states (stressed and non-stressed) over different social structures. The dot represents a friend of the user, and the line represents the connection of friends.

7.2.2 Social Influence

Social influence is an important factor that governs the dynamics of social networks. The principle of social influence [22] suggests that users tend to change their behaviors to match their friends' behaviors. In this study, we try to examine whether users' stress states will be influenced by their neighbors' states by looking at the probability of a user's stress state when he/she has different types of relationships with other stressed users. As for the stress state labeling, all users including friends are labeled using the sentence pattern method described in previous section.

Figure 8(a) shows the probability that a user being stressed, conditioned on the number of stressed neighbors that the user has in the social network. We can see that being stressed is a mutually correlated behavior. In particular, the chance that a none-stressed user becoming stressed increases to three times higher for those with stressed neighbors than for those without. Another trend observed from Figure 8(a) is that the likelihood of a user becoming stressed increases with the number of stressed neighbors.

7.2.3 Strong/Weak Tie

Strong/Weak Tie [17] is one of the most basic principles in social network theories. We classify the constructed social
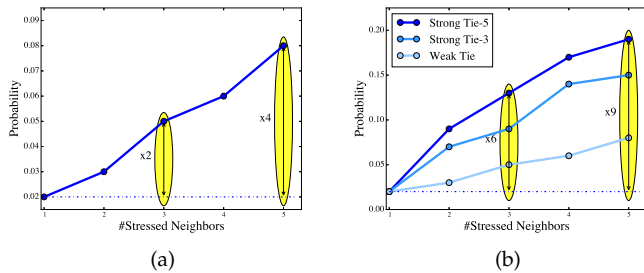
Fig. 8. Social influence and Social tie analysis. (a) Variation trend of probability of a user being stressed when she/he has different number of stressed neighbors; (b) Variation trend of probability of a user being stressed when she/he has different number of stressed neighbors with strong/weak ties.

relationships into strong or weak ties by the number of times that two users interact with each other via *comment*, *@-mention*, *retweet*, or *like* in a week. In our work, we tried different values for the threshold and finally chose three by cross-validation. If two users interact with each other more than three times, we call the relationship a *strong* tie, and otherwise a *weak* tie. This definition of user ties is adopted as the standard treatment in the research of social network analysis [17], so as to capture the most recent user relationships in a shifting environment. Figure 8(b) illustrates the results. We can see that strong ties indeed have strong influence on users' stress states, and the influence of weak ties is relatively weak. For example, when a user has three stressed strong-tie connections, the probability that the user will become stressed increases to 13%, more than twice as high as for a user with three stressed weak-tie connections.

Summary. Based on the experimental results and analyses we know that: 1) users' stress states are not only revealed in their own tweets, but also affected by the contents of their social interactions, including commenting on and re-tweeting others' tweets; and 2) users' stress states are revealed by the structure of their social interactions, including structural diversity, social influence, and strong/weak ties. These insights quantitatively prove the necessity and effectiveness of combining social interactions for stress detection.

8 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 phone data, weather conditions and individual traits. In *ACM International Conference on Multimedia*, pages 477–486, 2014.

[2] 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.

[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 chinese language 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 support vector machines. *ACM TRANSACTIONS ON INTELLIGENT SYSTEMS AND TECHNOLOGY*, 2(3):389–396, 2001.

[6] Dan C Ciresan, Ueli Meier, Jonathan Masci, Luca Maria Gambardella, and Jürgen Schmidhuber. Flexible, high performance convolutional neural networks for image classification. In *Proceedings of International Joint Conference on Artificial Intelligence*, pages 1237–1242, 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. 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.

[11] Golnoosh Farnadi, Geetha Sitaraman, Shanu Sushmita, Fabio Celli, Michal Kosinski, David Stillwell, Sergio Davalos, Marie Francine Moens, and Martine De Cock. Computational personality recognition in social media. *User Modeling and User-Adapted Interaction*, pages 1–34, 2016.

[12] Eileen Fischer and A. Rebecca Reuber. Social interaction via new social media: (how) can interactions on twitter affect effectual thinking and behavior? *Journal of Business Venturing*, 26(1):1–18, 2011.

[13] Jerome H. Friedman. Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 29(5):1189–1232, 1999.

[14] Rui Gao, Bibo Hao, He Li, Yusong Gao, and Tingshao Zhu. Developing simplified chinese psychological linguistic analysis dictionary for microblog. pages 359–368, 2013.

[15] Johannes Gettinger and Sabine T. Koeszegi. *More Than Words: The Effect of Emoticons in Electronic Negotiations*.

[16] 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.

[17] Mark S. Granovetter. The strength of weak ties. *American Journal of Sociology*, 1973.

[18] Quan Guo, Jia Jia, Guangyao Shen, Lei Zhang, Lianhong Cai, and Zhang Yi. Learning robust uniform features for cross-media social data by using cross autoencoders. *Knowledge Based System*, 102:64–75, 2016.

[19] David W. Hosmer, Stanley Lemeshow, and Rodney X. Sturdivant. Applied logistic regression. *Wiley series in probability and mathematical statistics*, 2013.

[20] Sung Ju Hwang. Discriminative object categorization with external semantic knowledge. 2013.

[21] Sepandar D. Kamvar. We feel fine and searching the emotional web. In *In Proceedings of WSDM*, pages 117–126, 2011.

[22] Herbert C. Kelman. Compliance, identification, and internalization: Three processes of attitude change. *general information*, 1(1):51–60, 1958.

[23] Shigenobu Kobayashi. The aim and method of the color image scale. *Color research & application*, 6(2):93–107, 1981.

[24] Novak P Kralj, J Smailovi, B Sluban, and I Mozeti. Sentiment of emojis. *Plos One*, 10(12), 2015.

[25] Frank R Kschischang, Brendan J Frey, and H-A Loeliger. Factor graphs and the sum-product algorithm. *Information Theory, IEEE Transactions on*, 47(2):498–519, 2001.

[26] Yann LeCun and Yoshua Bengio. Convolutional networks for images, speech, and time series. *The handbook of brain theory and neural networks*, 3361, 1995.

[27] Kathy Lee, Ankit Agrawal, and Alok Choudhary. Real-time disease surveillance using twitter data: demonstration on flu and cancer. In *Proceedings of ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1474–1477, 2013.

[28] H. 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.

[29] H. Lin, J. Jia, Q. Guo, Y. Xue, Q. Li, J Huang, L. Cai, and L. Feng. User-level psychological stress detection from social media using deep neural network. In *Proceedings of ACM Int. Conference on Multimedia*, 2014.

[30] Li Liu and Ling Shao. Learning discriminative representations from rgb-d video data. In *Proceedings of International Joint Conference on Artificial Intelligence*, pages 1493–1500, 2013.

[31] H-A Loeliger. An introduction to factor graphs. *Signal Processing Magazine, IEEE*, 21(1):28–41, 2004.

[32] Jana Machajdik and Allan Hanbury. Affective image classification using features inspired by psychology and art theory. In *Proceedings of the international conference on Multimedia*, pages 83–92, 2010.

[33] Kevin P Murphy, Yair Weiss, and Michael I Jordan. Loopy belief propagation for approximate inference: An empirical study. In *Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence*, pages 467–475, 1999.

[34] Cristian Danescu niculescu mizil, Lillian Lee, Bo Pang, and Jon Kleinberg. Echoes of power: Language effects and power differences in social interaction. *eprint arXiv:1112.3670*, 2011.

[35] Liqiang Nie, Yi-Liang Zhao, Mohammad Akbari, Jialie Shen, and Tat-Seng Chua. Bridging the vocabulary gap between health seekers and healthcare knowledge. *Knowledge and Data Engineering, IEEE Transactions on*, 27(2):396–409, 2015.

[36] Federico Alberto Pozzi, Daniele Maccagnola, Elisabetta Fersini, and Enza Messina. Enhance user-level sentiment analysis on microblogs with approval relations. In *AI* IA 2013: Advances in Artificial Intelligence*, pages 133–144. 2013.

[37] Neumann R and Strack F. "mood contagion": the automatic transfer of mood between persons. *Journal of Personality and Social Psychology*, pages 211–223, 2000.

[38] Chenhao Tan, Lillian Lee, Jie Tang, Long Jiang, Ming Zhou, and Ping Li. User-level sentiment analysis incorporating social networks. In *Proceedings of the SIGKDD international conference on Knowledge discovery and data mining*, pages 1397–1405, 2011.

[39] Wenbin Tang, Honglei Zhuang, and Jie Tang. Learning to infer social ties in large networks. In *Machine Learning and Knowledge Discovery in Databases*, pages 381–397. 2011.

[40] Y. R. Tausczik and J. W. Pennebaker. The psychological meaning of words: Liwc and computerized text analysis methods. *Journal of Language and Social Psychology*, 29(1):24–54, 2010.

[41] Mike Thelwall, Kevan Buckley, Georgios Paltoglou, Di Cai, and Arvid Kappas. Sentiment strength detection in short informal text. *Journal of the American Society for Information Science and Technology*, 61(12):2544–2558, 2010.

[42] Svetnik V. Random forest: a classification and regression tool for compound classification and qsar modeling. *Journal of Chemical Information and Computer Sciences*, 43(6):1947–1958, 2003.

[43] Ben Verhoeven, Walter Daelemans, and Barbara Plank. Twisty: A multilingual twitter stylometry corpus for gender and personality profiling. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation LREC*, pages 1632–1637, 2016.

[44] Chi Wang, Jie Tang, Jimeng Sun, and Jiawei Han. Dynamic social influence analysis through time-dependent factor graphs. *Advances in Social Networks Analysis and Mining (ASONAM), 2011 International Conference on*, pages 239 – 246, 2011.

[45] Xiaohui Wang, Jia Jia, Peiyun Hu, Sen Wu, Jie Tang, and Lianhong Cai. Understanding the emotional impact of images. In *Proceedings of the 20th ACM International Conference on Multimedia*, MM '12, pages 1369–1370, 2012.

[46] Lexing Xie and Xuming He. Picture tags and world knowledge: learning tag relations from visual semantic sources. In *ACM Multimedia Conference*, pages 967–976, 2013.

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

[48] Jian Bo Yang, Minh Nhut Nguyen, Phyo Phyo San, Xiao Li Li, and Shonali Krishnaswamy. Deep convolutional neural networks on multichannel time series for human activity recognition. In *Proceedings of International Joint Conference on Artificial Intelligence*, pages 3995–4001, 2015.

[49] Yang Yang, Jia Jia, Shumei Zhang, Boya Wu, Qicong Chen, Juanzi Li, Chunxiao Xing, and Jie Tang. How do your friends on social media disclose your emotions? In *Proceedings of the 28th AAAI Conference on Artificial Intelligence (AAAI'14)*, pages 306–312, 2011.

[50] Shuo Zeng, Mingfeng Lin, and Hsinchun Chen. Dynamic user-level affect analysis in social media: Modeling violence in the dark web. In *IEEE International Conference on Intelligence and Security Informatics (ISI)*, pages 1–6, 2011.

[51] Qian Zhang and Bruno Goncalves. Topical differences between chinese language twitter and sina weibo. *Computer Science*, 2015.

[52] Yuan Zhang, Jie Tang, Jimeng Sun, Yiran Chen, and Jinghai Rao. Moodcast: Emotion prediction via dynamic continuous factor graph model. *2013 IEEE 13th International Conference on Data Mining*, pages 1193–1198, 2010.

[53] Jichang Zhao, Li Dong, Junjie Wu, and Ke Xu. Moodlens: an emoticon-based sentiment analysis system for chinese tweets. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1528–1531, 2012.

Huijie Lin is currently a Ph.D. candidate of Tsinghua University. His research interests include social media analysis and affective computing.

Jia Jia is an associate professor in Department of Computer Science and Technology, Tsinghua University. She got bachelor degree at Tsinghua University in 2003, and received her Ph.D. degree from Tsinghua University in 2008. Her main research interest is social affective computing and human computer speech interaction. She is serving as the secretary-general of Professional Committee of Speech in Chinese Information Processing Society, and also a committee member of Multimedia Federation in China Society of Image and Graphic. She has been awarded the ACM Multimedia Grand Challenge Prize (2012) and the Scientific Progress Prizes from the National Ministry of Education (2009, 2016).

Jiezhong Qiu is currently an Ph.D student of Tsinghua University.

Yongfeng Zhang is a Postdoc Research Associate in University of Massachusetts Amherst, US.

Guangyao Shen is currently a Master student of Tsinghua University.

Lexing Xie received the B.S. degree from Tsinghua University, Beijing, China, in 2000, and the M.S. and Ph.D. degrees from Columbia University, in 2002 and 2005, respectively, all in electrical engineering. She is a Lecturer in the Research School of Computer Science at the Australian National University; she was with the IBM T.J. Watson Research Center, Hawthorne, NY from 2005 to 2010. Her recent research interests are in multimedia mining, machine learning and social media analysis. Dr. Xie has won several awards: the best conference paper award in IEEE SOLI 2011, the best student paper awards at JCDL 2007, ICIP 2004, ACM Multimedia 2005 and ACM Multimedia 2002.

Jie Tang is an associate professor at the Department of Computer Science and Technology, Tsinghua University. His main research interests include data mining algorithms and social network theories. He has been visiting scholar at Cornell University, Chinese University of Hong Kong, Hong Kong University of Science and Technology, and Leuven University. He has published over 100 research papers in major international journals and conferences including: KDD, IJCAI, AAAI, ICML, WWW, SIGIR, SIGMOD, ACL, Machine Learning Journal, TKDD, and TKDE.

Ling Feng is a professor of computer science and technology at Tsinghua University, Beijing. Her research interests include context-aware data management toward ambient intelligence, knowledge-based information systems, data mining and warehousing, and distributed object-oriented database management systems. She has published more than 150 scientific articles in high-quality international conferences or journals, and received the 2004 Innovational VIDI Award by the Netherlands Organization for Scientific Research, the 2006 Chinese ChangJiang Professorship Award by the Ministry of Education, and the 2006 Tsinghua Hundred-Talents Award.

Tat-Seng Chua joined the National University of Singapore, Singapore, in 1983, and spent three years as a Research Staff Member with the Institute of Systems Science, National University of Singapore. He was the Acting and Founding Dean of the School of Computing, National University of Singapore, from 1998 to 2000. He is currently the KITHCT Chair Professor with the School of Computing, National University of Singapore. His research interests include multimedia information retrieval, multimedia question answering, and the analysis and structuring of user-generated contents.

Dr. Chua has organized and served as a Program Committee Member of numerous international conferences in the areas of computer graphics, multimedia, and text processing. He was the Conference Co-Chair of ACM Multimedia in 2005, the Conference on Image and Video Retrieval in 2005 and ACM SIGIR in 2008, and the Technical PC Co-Chair of SIGIR in 2010. He serves on the editorial boards of the ACM Transactions of Information Systems, Foundation and Trends in Information Retrieval, The Visual Computer, and Multimedia Tools and Applications. He is on the Steering Committees of the International Conference on Multimedia Retrieval, Computer Graphics International, and Multimedia Modeling Conference Series. He serves as a member of international review panels of two large-scale research projects in Europe. He is the Independent Director of two listed companies in Singapore.