What Does Social Media Say about Your Stress?

Huijie Lin,[1] **Jia Jia,**[1*] **Liqiang Nie,**[2] **Guangyao Shen,**[1] and **Tat-Seng Chua**[3]
[1]Department of Computer Science and Technology, Tsinghua University, Beijing, China
Tsinghua National Laboratory for Information Science and Technology (TNList)
Key Laboratory of Pervasive Computing, Ministry of Education
[2]School of Computer Science and Technology, Shandong University
[3] School of Computing, National University of Singapore
{linhuijie, nieliqiang, thusgy2012}@gmail.com, jjia@mail.tsinghua.edu.cn, dcscts@nus.edu.sg

Abstract

With the rise of social media such as Twitter, people are more willing to convey their stressful life events via these platforms. In a sense, it is feasible to detect stress from social media data for proactive health care. In psychology, stress is composed of stressor and stress level, where stressor further comprises of stressor event and subject. By far, little attention has been paid to estimate exact stressor and stress level from social media data, due to the following challenges: 1) stressor subject identification, 2) stressor event detection, and 3) data collection and representation. To address these problems, we devise a comprehensive scheme to measure a user's stress level from his/her social media data. In particular, we first build a benchmark dataset and extract a rich set of stress-oriented features. We then propose a novel hybrid multi-task model to detect the stressor event and subject, which is capable of modeling the relatedness among stressor events as well as stressor subjects. At last, we lookup an expert-defined stress table with the detected subject and event to estimate the stressor and stress level. Extensive experiments on real-world datasets well verify the effectiveness of our scheme.

1 Introduction

Aided by the convenience and constant access provided by mobile devices, about 65% of American adults and 92% of teens use social networking sites, a nearly tenfold jump in the past decade, reported by Pew Research Center[1]. We rely on social networks to share our or others' daily activities with a wide audience. Besides, we use social networks to disclose emotions and moods, happiness and unhappiness, since disclosure is intrinsically rewarding and can improve interpersonal intimacy. Inevitably, the language used in social media postings may signal feelings of exhaustion, sleeping problems, sweating, loss of appetite and difficulty concentrating that characterize the major stress. It is thus feasible to detect users' psychological stress from social media data.

In psychology, stress is our bodies' response to any kind of frustrations, demand or threat [Selye and others, 1936], which fill up our modern life. Thereby, stress is so commonplace that it has become a way of life. Some stress is normal and even useful. For example, it can help us win a race or finish an important job on time. But when stress becomes overwhelming or lasts too long, it can damage our health, mood, relationships, and quality of life. Also, it can increase the risk of strokes, ulcers, and mental disorders. It is hence of importance to recognize the signs of stress in advance, which enables us to take proactive care to reduce the harmful effects of stress.

It is worth mentioning that few efforts thus far have been dedicated to stress detection by harvesting social media, except the followings. Lin et al. and Xue et al. proposed novel methods to detect tweet-level stress [2014a; 2013; 2014] and user-level stress [2014b] over a short time window respectively. Despite great success, they only classified samples into stressed or non-stressed categories, which is incapable of measuring the exact stressor and stress level. Stress is indeed much more complicated, which is composed of two key factors: stressor and stress level. Quite literally, stressor, comprising of stressor event and stressor subject, triggers stress; meanwhile, different stressors incur different stress levels. For instance, a layoff usually makes people more stressed in comparison to a project deadline. In addition, stressor events happening to other subjects can also be someone's stress trigger, but may have different affects. For instance, "my friend's father just passed away". Moreover, stress level depends on the stressor, and is measurable using various psychological stress scales, despite that stress is often thought of as a subjective experience. As claimed by psychologists, detecting stressor and measuring the stress levels are of essential importance to proactive care.

In this paper, we work towards measuring stress via social media data. It is, however, non-trivial due to the following challenges: 1) Stressor subject identification. As known, social media postings are usually informally and ambiguously written with grammatical problems, while the current subject Lexicon tools are not well suited for such scenarios. 2)

*Corresponding author: J. Jia (jjia@mail.tsinghua.edu.cn)
[1]http://t.cn/RyRCFem

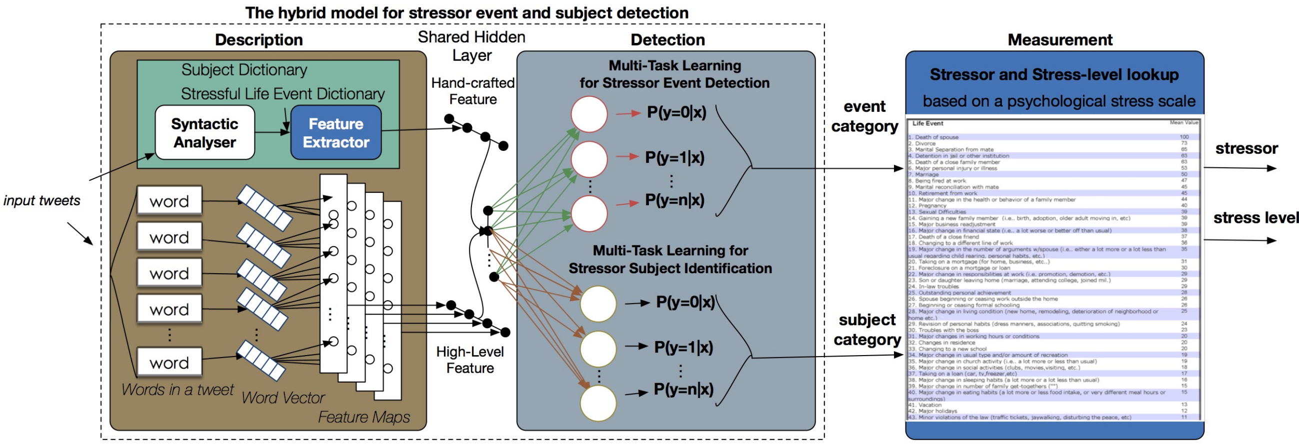


Figure 1: Illustration of our scheme. It comprises of three components: 1) description component that extracts features from input tweets; 2) detection component that integrates handcrafted and CNN features with Multi-Task Learning (MTL) for stressor event and subject detection, where the CNN is fine-tuned by MTL to compose the hybrid model, and 3) measurement component leveraging a psychological stress scale to estimate the stressor and the corresponding stress level.

Stressor event detection. Psychologists manually define a set of stress-related events, which are usually not independent but correlated in a nonuniform way. Meanwhile, for some specific event categories, such as death, there are insufficient training samples. How to model the event relatedness and borrow training samples for the sample-lacking categories are largely untapped. And 3) data and representation. There is no publicly available benchmark datasets thus far that can be leveraged to conduct stress study and verify the effectiveness of models. In addition, it is of importance to extract a rich set of stress-oriented features to describe the data.

To address the aforementioned challenges, we devise a comprehensive scheme that is able to automatically detect stressor subjects and events, and further estimate the stressors and stress levels. As illustrated in Figure 1, our scheme consists of three components. In particular, we first extract a set of discriminant features, including the hand-crafted statistical features and the high-level semantic features. We then propose a hybrid model combining multi-task learning with convolutional neural network (CNN) to respectively identify the stressor subjects and stressor events of the given social posts. It is worth mentioning that the relatedness among stressor events as well as stressor subjects is well captured and modeled in our model. Once we obtain the stressor subject and event category, we can lookup a standard psychological stress scale to measure the precise stressor and stress level. To verify our scheme, we construct a representative dataset from Weibo by an extendable set of seed words, and invite 30 volunteers to manually build the ground truth. Extensive experiments well validate our proposed scheme.

The main contributions are as follows:

- We extracted a set of stress-related features and proposed a hybrid model to boost the identification performance of stressor events and subjects.
- We constructed a benchmark dataset in the field of stress detection. Meanwhile, we have released the data together with codes and parameters to facilitate the research community[2] .

2 Related Work

Literature on mental health care via social media and personal event detection is related to our work.

Mental Health Care via Social Media: In recent years, there exist some studies on leveraging social media data for mental health care. De Choudhury et al. [2013] were the first to explore social media data for depression prediction. Tsugawa et al. [2015] further leveraged Twitter postings to predict depression for Japanese-speaking users. Xue et al. demonstrated the promising performance via using social media data for stress detection[2013; 2014]. In summary, these efforts have shown the feasibility of leveraging social media data for mental health care. They, however, mainly focused on the binary classification results, while the trigger event and level of mental problems are more important for successive cares.

Personal Event Detection: As compared to general event detection, personal event detection in social media is a quite new topic. Li and Cardie [2014] attempted to construct the history of users' life events based on their Twitter stream and proposed an unsupervised framework to create a chronological list for personal events. Li and Ritter [2014] learned a supervised classifier with manually defined textual features, which is able to classify users' tweets into predefined life event categories. However, each step of the system induces errors and negatively affects the following parts. Moreover, they assume that the detected personal event is related to the tweet owner, which may be violated in the real-world scenarios.

---
[2]http://stressmeasure.droppages.com/

3 Problem Formulation

To formulate our problem, we declare some notations in advance. In particular, we use bold capital letters (e.g., $\mathbf{X}$) and bold lowercase letters (e.g., $\mathbf{x}$) to denote matrices and vectors, respectively. We employ non-bold letters (e.g., x) to represent scalars, and Greek letters (e.g., λ) as parameters. If not clarified, all vectors are in column form.

Suppose that we have K stressor events and M stressor subjects. Let us denote $\mathbf{e_i} \in \mathbb{R}^K$ as the event label vector, and $\mathbf{s_i} \in \mathbb{R}^M$ as the subject label vector, for the i-th tweet. Given a set of tweets $\mathcal{T} = \{t_1, t_2, \cdots, t_N\}$, it consists of N distinct training samples. Let $\mathbf{x_i} \in \mathbb{R}^D$ be the feature vector of the i-th tweet. Each training sample $t_i = (\mathbf{x_i}, \mathbf{e_i}, \mathbf{s_i})$ consists of a feature vector denoted by $\mathbf{x_i}$, with the corresponding stressor event label $\mathbf{e_i}$ and the stressor subject label $\mathbf{s_i}$. Let $\mathbf{X} = [\mathbf{x_1}, \mathbf{x_2}, \cdots, \mathbf{x_N}]^T \in \mathbb{R}^{N\times D}$ be the feature matrix, $\mathbf{E} = [\mathbf{e_1}, \mathbf{e_2}, \cdots, \mathbf{e_N}]^T \in \mathbb{R}^{N\times K}$ be the stressor event label matrix, and $\mathbf{S} = [\mathbf{s_1}, \mathbf{s_2}, \cdots, \mathbf{s_N}]^T \in \mathbb{R}^{N\times M}$ be the stressor subject label matrix, respectively.

With the notations above, we can formally define our problem as: *Given a set of tweets $\mathcal{T}$ with their feature matrix $\mathbf{X}$, the corresponding stressor event label matrix $\mathbf{E}$ and the stressor subject label matrix $\mathbf{S}$, we aim to learn a model that is able to automatically detect stressor event and subject labels for unlabeled tweets.*

4 Data Description

4.1 Stressor Event and Subject Dictionaries

The word embeddings [Mikolov *et al.*, 2013] have been found effective in estimating the semantic similarities among different words. In the light of this, it makes sense to gather a set of semantic similar words and hence to construct dictionaries. We learned the word embeddings with a 200-dimensional vector on a one billion Weibo dataset, which was crawled from Weibo between 2009.6 and 2012.12 using Weibo's open APIs. The word segmentation method used to segment the tweets into words is the word segmentation module of LTP [Che *et al.*, 2010]. Based on the trained word embeddings, we constructed two dictionaries:

- **Stressor Event Dictionary.** Based on the event categories in the professional life events stress scale [Holmes and Rahe, 1967], we first selected the top 12 categories of events that occur frequently in Weibo. We then defined a set of seed words for each selected category using LIWC [Tausczik and Pennebaker, 2010], and further calculated the similarities between each seed word and the words from the vocabulary of the learned word embeddings. Thereafter, we selected the top 100 most similar words for expansion. We ultimately got a stressor event dictionary with more than $1,500$ words. The samples of the dictionary is shown in Table 1.
- **Stressor Subject Dictionary.** Stressor subject is defined as the social relation between a user and the one that a stressor event is related to, which plays an important role in stress measurement. There are some existing dictionaries, e.g., LIWC, containing some words of different social relations. However, stressor subjects expressed in social media data are usually informal, making it difficult to directly use the formally constructed dictionaries. Inspired by the work in [Haslam, 1994], we categorized the subjects into six categories. We then collected a set of seed words for each category using LIWC. Thereafter, similar to the stressor event seed words expansion, we again employed the learned word embeddings to select the top 40 most similar words for each category. We finally collected a stressor subject dictionary containing about 200 words. The samples of the dictionary is shown in Table 2.

Table 1: Summary of the manually labeled tweets for each stressor event category and the sampling words of the constructed stressor event dictionary.

Events	Labeled	Sampling words
Marriage	227	marry wedding bride
Financial	114	income salary rent
Illness	424	hospital sick pain
School	171	school holiday finals
Birth	133	born life baby
Fired	102	fired job lose
Argue	107	cold war quarrel argue
Blamed	199	question blame afraid
Pregnancy	132	baby pregnant mother-to-be
Habits	102	revise habits smoke drink
Death	127	pass away R.I.P
Divorce	112	divorce ex-wife cry

Table 2: Summary of the manually labeled tweets for each stressor subject category and the sampling words of the constructed stressor subject dictionary.

Category	Labeled	Sampling words
I	647	I my our we
family	277	mother daughter
friend	327	friend teammate
spouse	207	wife husband dear
boss	161	boss teacher tutor
relative	123	aunt uncle cousin

4.2 Hand-Crafted Feature

We defined some hand-crafted features based on the two predefined dictionaries and the grammatical analysis of tweets using the LTP tool [Che *et al.*, 2010]:

- **Stressor Event Word Distribution**. It is represented by the word histogram over stressor events, extracted from a tweet using the stressor event dictionary. It indicates the occurrence of the possible stressor events.
- **Stressor Subject Word Distribution**. It is represented by the word histogram over stressor subjects, extracted from a tweet using the stressor subject dictionary. It can help locate the possible stressor subjects in a tweet.

4.3 High-Level Feature

To aggregate the word-level vector representations to learn tweet-level features, we applied convolutional neural networks (CNN) [LeCun and Bengio, 1995]. We hence utilized the CNN to learn tweet-level features from a set of low-level word vectors extracted from pre-trained word embeddings. The CNN features derived will be combined with the aforementioned hand-crafted features in subsequent steps to detect the stressor event and subject.

5 Stressor Event and Subject Detection

The stressor events and subjects are not independent. Some of them share certain similar patterns. For example, "pregnancy" shares many similar description patterns with "birth", but greatly differs from "death". Another example, subjects of "family" category share more similar patterns with "relative" compared to "friend". Moreover, the training samples of stressor event and subject categories are unbalanced. Further, the dimension of the learnt tweet features is usually very high, which requires sufficient training samples. To capture the relatedness among different event and subject categories and leverage the shared patterns to strengthen the learning performance, we regard each event or subject detection as a task and propose to use a multi-task learning (MTL) [Caruana, 1997] model. The MTL model is able to adaptively capture the relatedness among tasks, as well as to learn the task sharing and task-specific features.

Objective Function. We align each event category with a task. Regarding stressor subject detection, each subject is treated as a task. To save the space, we only detail the learning process of stressor event. The prediction function of the i-th task is defined as $f_i(\mathbf{x}) = \mathbf{w_i^T x}$, where $\mathbf{w_i} \in \mathbb{R}^D$ is the weight vector for task $\mathbf{T_i}$. The weight matrix over K tasks is denoted as $\mathbf{W} = [\mathbf{w_1}, \mathbf{w_2}, \cdots, \mathbf{w_K}]^T \in \mathbb{R}^{K\times D}$. We formulate the objective function of stress event detection as,

$$\min_{\mathbf{W},\Omega} \frac{1}{2}||\mathbf{XW}-\mathbf{Y}||_F^2 + \lambda tr\left(\mathbf{W}^T\mathbf{\Omega}^{-1}\mathbf{W}\right) + \frac{1}{2}\gamma||\mathbf{W}||_F^2 \quad s.t.\ \mathbf{\Omega} \succeq 0,\ tr(\mathbf{\Omega}) = 1, \tag{1}$$

where the first term measures the empirical error; the second term encodes the relatedness among different tasks; and the third term controls the generalization error; λ and γ are the regularization parameters, and $\mathbf{\Omega} \in \mathbb{R}^{K\times K}$ is a positive semi-definite matrix that we aim to learn, with the (i,j)-th entry represents the relation between task i and task j.

Solution. We adopt the alternative optimization procedure to solve $\mathbf{W}$ and $\mathbf{\Omega}$:

- **Optimizing W with $\mathbf{\Omega}$ fixed**. In this condition, the derivative of Eqn.(1) with respect to $\mathbf{W}$ is:

$$\nabla\mathbf{W} = \mathbf{X^T}(\mathbf{XW}-\mathbf{Y}) + \lambda\left(\mathbf{\Omega}^{-1}+\mathbf{\Omega^{-T}}\right)\mathbf{W} + \gamma\mathbf{W}. \tag{2}$$

Let $\nabla\mathbf{W} = 0$, we then have,

$$\left(\mathbf{X^TX} + \lambda\left(\mathbf{\Omega}^{-1}+\mathbf{\Omega^{-T}}\right) + \gamma\mathbf{I}\right)\mathbf{W} = \mathbf{X^TY}. \tag{3}$$

Thus, we have a closed-form solution for $\mathbf{W}$:

$$\mathbf{W} = \left(\mathbf{X^TX} + \lambda\left(\mathbf{\Omega}^{-1}+\mathbf{\Omega^{-T}}\right) + \gamma\mathbf{I}\right)^{-1}\mathbf{X^TY}. \tag{4}$$

Because of its low dimension, the right part in Eqn.(4) can be easily solved by computing the pseudo-inverse based on singular value decomposition (SVD) [Ben-Israel and Greville, 2003].

- **Optimizing $\mathbf{\Omega}$ with W fixed**. When $\mathbf{W}$ is fixed, Eqn.(1) is reduced to:

$$\min_{\mathbf{\Omega}} tr\left(\mathbf{W}^T\mathbf{\Omega}^{-1}\mathbf{W}\right) = \min_{\mathbf{\Omega}} tr\left(\mathbf{\Omega}^{-1}\mathbf{WW}^T\right), \quad s.t.\ \mathbf{\Omega} \succeq 0,\ tr(\mathbf{\Omega}) = 1. \tag{5}$$

Let us denote $\mathbf{A} = \mathbf{WW}^T$. We can restate Eqn.(5) as

$$\begin{aligned} & tr\left(\mathbf{\Omega}^{-1}\mathbf{A}\right) \\ &= tr\left(\mathbf{\Omega}^{-1}\mathbf{A}\right) tr(\mathbf{\Omega}) \\ &= tr\left(\left(\mathbf{\Omega}^{-\frac{1}{2}}\mathbf{A}^{\frac{1}{2}}\right)\left(\mathbf{A}^{\frac{1}{2}}\mathbf{\Omega}^{-\frac{1}{2}}\right)\right) tr\left(\mathbf{\Omega}^{\frac{1}{2}}\mathbf{\Omega}^{\frac{1}{2}}\right) \\ &\geq \left(tr\left(\mathbf{\Omega}^{-\frac{1}{2}}\mathbf{A}^{\frac{1}{2}}\mathbf{\Omega}^{\frac{1}{2}}\right)\right)^2 = \left(tr\left(\mathbf{A}^{\frac{1}{2}}\right)\right)^2. \end{aligned} \tag{6}$$

The equality holds if and only if $\left(\mathbf{\Omega}^{-\frac{1}{2}}\mathbf{A}^{\frac{1}{2}}\right) = a$ for some constant a and $tr(\mathbf{\Omega}) = 1$. We thus can derive a closed-form solution for $\mathbf{\Omega}$, as

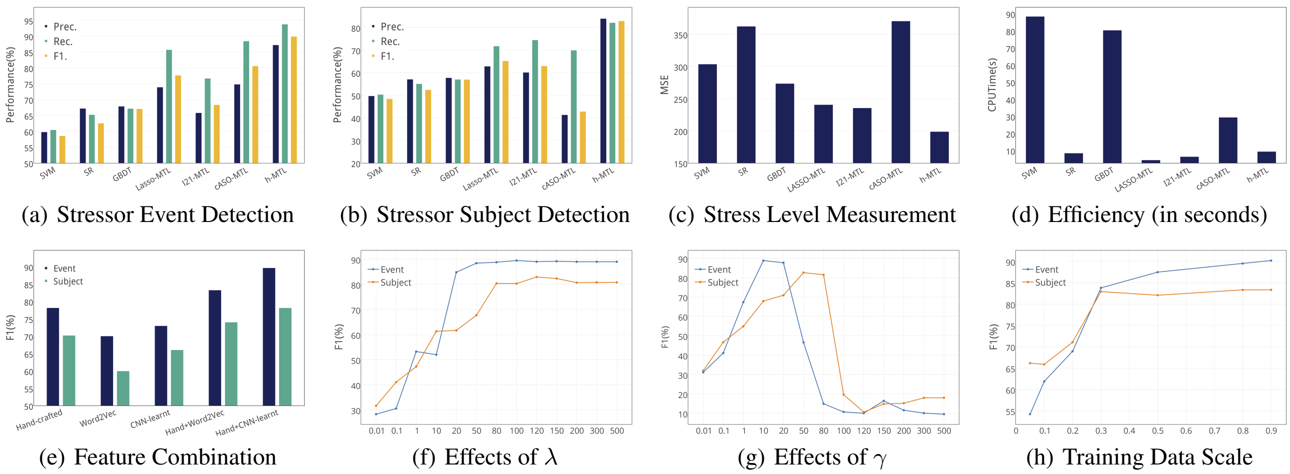
$$\mathbf{\Omega} = \frac{\left(\mathbf{WW}^T\right)^{\frac{1}{2}}}{tr\left(\left(\mathbf{WW}^T\right)^{\frac{1}{2}}\right)}. \tag{7}$$

Hybrid Model. Instead of the shallow MTL, we enhance our proposed MTL model with stacked auto-encoders by treating it as the final classifier layer on top of the CNN. The gradient of the final multi-task learning layer is propagated to fine-tune the convolution layer in the training stage using the back propagation method to further improve the learnt features of CNN. As our aim is to detect both stressor event and subject, in this work, we train two independent models for stressor event and subject detection respectively.

6 Stress Measurement

The measurement of psychological stress has been well studied by psychologists in the past decades. Most of these measurements [Rowlison and Felner, 1988; Brantley and Jones, 1993] are based on questionnaires and interviews. Among them, the Social Readjustment Rating Scale (SRRS) [Holmes and Rahe, 1967] is one of the widely-accepted metrics. It is composed of 43 predefined stressors (stressor events and subjects) and the corresponding stress scores, which are representative of major stressors that occur in people's lives.

We have thus far detected the stressor event and stressor subject from a given tweet via our proposed hybrid model, we then can leverage the psychological stress measure scale to estimate the stressor and the correponding stress level, which is valuable for insightful analysis of stress. To accomplish this, we first split the stressors of SRRS into stressor events and stressor subjects, and then group them into two separate categories, i.e., stressor event category and stressor subject category. Taking the stressor "Death of Spouse" as an example, it is splitted into the event "Death" and the subject "Spouse" separately. Hereafter, we can align the detected stressor event and stressor subject of a tweet to the stressor in the SRRS table, and get the corresponding stress level of the tweet.



(a) Stressor Event Detection (b) Stressor Subject Detection (c) Stress Level Measurement (d) Efficiency (in seconds)

(e) Feature Combination (f) Effects of λ (g) Effects of γ (h) Training Data Scale

Figure 2: Experimental results of following validations: (a) performance of stressor event detection; (b) performance of stressor subject detection; (c) performance of stressor level measurement; (d) efficiency of model training and testing; (e) impact of feature combinations; (f) effects of the λ parameter of MTL controling the weight of task relatedness; (g) effects of the γ parameter of MTL controlling the weight of the generalization error, and (h) impact of training data scales.

7 Experiments

7.1 Experimental Setup

Dataset. To construct a benchmark dataset, we first categorized the stressor events into 43 categories based on the professional life events stress scale [Holmes and Rahe, 1967]. We then manually defined a set of keyword patterns collected from the LIWC dictionary for each stressor event category. Using the collected keywords as seeds, we filtered matched tweets from the aforementioned one billion Weibo dataset. We then collected the top 12 stressor event categories and invited 30 volunteers to manually label the stressor events and stressor subjects of the tweets from the filtered Weibo data. In particular, we randomly and equally divided the tweets into 10 groups. For each group of tweets, three distinct volunteers with diverse backgrounds were invited to manually label the stressor event and subject of each tweet. Therefore, each tweet was labeled three times, and it was first labeled to be stress or non-stress related, and the stress related tweets were further categorized into one of 12 stressor event categories and six stressor subject categories, respectively. Following that, voting was performed to obtain the final ground truth. For the cases that there are two classes having the same number of ballots, a discussion was carried out among the labelers to decide the final ground truths. We finally collected a dataset containing nearly $2,000$ tweets. We also randomly selected 600 tweets that are labeled as non-stress related to be the negative samples. The detailed distribution of the labeled dataset is shown in Table 1 and 2.

Evaluation Metrics. To throughly verify our model, we adopted the following metrics:

- **Effectiveness.** We evaluated the performance of stressor event and subject detection of our model and the baselines in terms of Recall (Rec.), Precision (Prec.) and F1-Measure (F1). As for the measurement of stress level measure performance, we used the Mean Squared Error (MSE).
- **Efficiency.** We evaluated the efficiency of the methods by comparing the CPU time of training each model. All experiments were performed on an x64 machine with two 2.6GHz intel E5-2650 CPUs and 128GB RAM.

7.2 Model Evaluation

We compared the stressor event, stressor subject detection and the overall stress level measurement performance with some state-of-the-art models. All the reported results in this paper were based on 10-fold cross validation:

- **Support Vector Machine (SVM) [Chang and Lin, 2011]:** It is a popular binary classifier that is found to effective in several classification problems.
- **Softmax Regression (SR) [Böhning, 1992]:** It is a model that is used to predict the probabilities of the different possible outcomes of a categorically distributed dependent variable, given a set of independent variables.
- **Gradient Boosted Decision Tree (GBDT) [Friedman, 1999]:** It trains a gradient boosted decision tree model with features associated with each tweet.
- **Lasso-MTL [Nie *et al.*, 2010]:** It is the Lasso regularized MTL which can achieve the goal of reducing model complexity and feature learning.
- **$\ell_{2,1}$-MTL [Tibshirani, 2011]:** It is the $\ell_{2,1}$-norm regularized MTL which encodes grouped sparsity by assuming that all tasks share a common set of features.
- **cASO-MTL [Chen *et al.*, 2009]:** It is a convex relaxation of the alternating structure optimization [Ando and Zhang, 2005], which decomposes the predictive model of each task into the task-specific feature mapping and task-shared feature mapping.

- **h-MTL:** It is our proposed hybrid model.

To ensure a fair comparison, we fed all the comparison methods with the same feature settings. For the SVM, SR and GBDT methods, we use the scikit-learn [Pedregosa *et al.*, 2011] implementation. As for the comparison methods of MTL, we use the MALSAR package [Zhou *et al.*, 2011] implementations with least squares loss.

Evaluation Results. The evaluation results are shown in Figure 2. It can be seen that shallow classification methods would not gain satisfactory performance on stressor event or subject detection tasks. However, our proposed hybrid model (h-MTL) is able to improve the detection performance by around 10% on F1-measure, indicating that the relatedness among different categories is important and our method can better capture the relatedness to enhance the detection performance. For the performance on stress level measurement tasks, it can be seen from Figure 2(c) that our scheme gains much lower MSE (and thus higher prediction accuracy) compared with other methods. Meanwhile, Figure 2(d) shows that the training time of the proposed model is comparatively lower than SVM, GBDT and cASO-MTL, while achieving much higher performance. All the improvements reported are significant at 0.05 level under two-tailed t-tests.

7.3 Feature Discrimination Analysis

To evaluate the effectiveness of different features for stressor event or subject detection, we further conducted an experiment to feed our model with different types of features, and the results in terms of F1-measure are shown in Figure 2(e). We see that even using only the proposed hand-crafted features gives us promising performance, which verifies the effectiveness of our proposed features. Besides, we see that using the CNN learnt features or the word vector representations alone leads to suboptimal performances. However, by integrating the CNN features or word vector representations with our hand-crafted features, we are able to obtain further improved performance against both hand-crafted or learnt features.

7.4 Parameter Sensitivity Analysis

There are two key parameters in our model, which are λ and γ in Eqn.(1) that controls the impact of task relatedness and the generalization error, respectively.

We first fixed λ for stressor event ($\lambda = 70$) and subject ($\lambda = 100$) detection, and then varied γ from 0.001 to 500. The fixed values for λ are chosen based on the best performance of the corresponding task, and the experimental results over different γ's are shown in Figure 2(f). It can be seen that the detection performance obtains the best scores when γ is around $10 \sim 20$ for stressor event detection and $50 \sim 80$ for stressor subject detection. When γ goes too high or too low, the detection performance drops severely. This result is reasonable because the choice of γ trades off between model generalization and complexity based on regularization, and thus the best performance can only be achieved with proper settings of γ.

We then fixed γ for stressor event ($\gamma = 10$) and subject ($\gamma = 55$) detection with varying λ, where the results are illustrated in Figure 2(g). It can be seen that the performance increases with the increase of λ, and tends to be stable when λ reaches 70 for stressor event and 100 for stressor subject detection. This observation further verifies that different tasks (i.e., event/subject categories) can be inherently related to each other, and by considering such an inherent relation with the spectral component, our hybrid MTL model is able to achieve progressively enhanced performances.

7.5 Training Data Scalability Analysis

Figure 2(h) shows the trend of detection performance with different proportions of training data. We can see that when using only 5% of all training data, the detection performance experiences a severe drop for both stressor event and subject detection. When adopting approximately 30% of all training data, our model can obtain an equally competitive performance compared with using the full data for stressor subject detection, while for event detection, it achieves over 80% of the F1-score. Moreover, the performance of event detection keeps increasing given more training data, while the performance of subject detection tends to keep stable. These results verify the data scalability of our hybrid MTL model.

8 Conclusion

This paper focuses on the broadest view of digital health with a comprehensive scheme for stress level estimation based on social media information, which comprises of three components. We first extract discriminative hand-crafted statistical features and high-level semantic features, and then proposed a hybrid model combining multi-task learning with convolutional neural network (CNN) to respectively identify the stressor subjects and stressor events of the given social posts. Finally, we evaluate the precise stressors and stress levels align with a widely-accepted psychological stress scale. To verify our scheme, we constructed a publicly available dataset with crowd-sourcing labeled ground truth. Experimental results show that our scheme achieves promising performances on measuring stress. Our technique serves as an assistant tool for psychologists to help detect stress, and our model as well as the released dataset further helps to promote the research of social media oriented intelligent digital health, which is destined to grow rapidly in the years ahead.

In future work, we plan to take the personalized coping ability of stress into consideration for fine-grained stress measurement, which is of importance for personalized cares according to psychologists.

9 Acknowledgments

This work is supported by the National Basic Research Program (973 Program) of China (2012CB316401, 2013CB329304), National Natural and Science Foundation of China (61370023). This work is partially supported by the major project of the National Social Science Foundation of China(13&ZD189).

References

[Ando and Zhang, 2005] Rie Kubota Ando and Tong Zhang. A framework for learning predictive structures from multiple tasks and unlabeled data. *Journal of Machine Learning Research*, 6(3):1817–1853, 2005.

[Ben-Israel and Greville, 2003] Adi Ben-Israel and Thomas NE Greville. *Generalized inverses: theory and applications*, volume 15. Springer Science & Business Media, 2003.

[Böhning, 1992] Dankmar Böhning. Multinomial logistic regression algorithm. *Annals of the Institute of Statistical Mathematics*, 44(1):197–200, 1992.

[Brantley and Jones, 1993] Phillip J Brantley and Glenn N Jones. Daily stress and stress-related disorders. *Annals of behavioral medicine*, 15(1):17–25, 1993.

[Caruana, 1997] Rich Caruana. Multitask learning. *Machine learning*, 28(1):41–75, 1997.

[Chang and Lin, 2011] Chih-chung Chang and Chih-Jen Lin. Libsvm: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2(3):27, 2011.

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

[Chen *et al.*, 2009] Jianhui Chen, Lei Tang, Jun Liu, and Jieping Ye. A convex formulation for learning shared structures from multiple tasks. In *Proceedings of the Annual International Conference on Machine Learning*, pages 137–144. ACM, 2009.

[De Choudhury *et al.*, 2013] Munmun De Choudhury, Michael Gamon, Scott Counts, and Eric Horvitz. Predicting depression via social media. In *Proceedings of International AAAI Conference On Web and Social Media*, pages 128–137. AAAI Press, 2013.

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

[Haslam, 1994] Nick Haslam. Categories of social relationship. *Cognition*, 53(1):59–90, 1994.

[Holmes and Rahe, 1967] Thomas H Holmes and Richard H Rahe. The social readjustment rating scale. *Journal of psychosomatic research*, 11(2):213–218, 1967.

[LeCun and Bengio, 1995] Yann LeCun and Yoshua Bengio. Convolutional networks for images, speech, and time series. *The handbook of brain theory and neural networks*, 3361(10):255–258, 1995.

[Li and Cardie, 2014] Jiwei Li and Claire Cardie. Timeline generation: Tracking individuals on twitter. In *Proceedings of International Conference on World Wide Web*, pages 643–652. ACM, 2014.

[Li *et al.*, 2014] Jiwei Li, Alan Ritter, Claire Cardie, and Eduard Hovy. Major life event extraction from twitter based on congratulations/condolences speech acts. In *Proceedings of Empirical Methods in Natural Language Processing*, pages 1997–2007. ACL, 2014.

[Lin *et al.*, 2014a] Huijie Lin, Jia Jia, Quan Guo, Yuanyuan Xue, Jie Huang, Lianhong Cai, and Ling Feng. Psychological stress detection from cross-media microblog data using deep sparse neural network. In *Proceedings of International Conference on Multimedia and Expo*, pages 1–6. IEEE, 2014.

[Lin *et al.*, 2014b] Huijie Lin, Jia Jia, Quan Guo, Yuanyuan Xue, Qi Li, Jie Huang, Lianhong Cai, and Ling Feng. User-level psychological stress detection from social media using deep neural network. In *Proceedings of ACM International Conference on Multimedia*, pages 507–516. ACM, 2014.

[Mikolov *et al.*, 2013] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. Distributed representations of words and phrases and their compositionality. In *Advances in Neural Information Processing Systems*, pages 3111–3119. Curran Associates, Inc., 2013.

[Nie *et al.*, 2010] Feiping Nie, Heng Huang, Xiao Cai, and Chris H Ding. Efficient and robust feature selection via joint l21-norms minimization. In *Advances in Neural Information Processing Systems*, pages 1813–1821. Curran Associates, Inc., 2010.

[Pedregosa *et al.*, 2011] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[Rowlison and Felner, 1988] Richard T Rowlison and Robert D Felner. Major life events, hassles, and adaptation in adolescence: confounding in the conceptualization and measurement of life stress and adjustment revisited. *Journal of personality and social psychology*, 55(3):432, 1988.

[Selye and others, 1936] Hans Selye et al. A syndrome produced by diverse nocuous agents. *Nature*, 138(3479):32, 1936.

[Tausczik and Pennebaker, 2010] Yla R Tausczik and James W Pennebaker. The psychological meaning of words: Liwc and computerized text analysis methods. *Journal of language and social psychology*, 29(1):24–54, 2010.

[Tibshirani, 2011] Robert Tibshirani. Regression shrinkage and selection via the lasso: A retrospective. *Journal of the Royal Statistical Society*, 73(3):273–282, 2011.

[Tsugawa *et al.*, 2015] Sho Tsugawa, Yusuke Kikuchi, Fumio Kishino, Kosuke Nakajima, Yuichi Itoh, and Hiroyuki Ohsaki. Recognizing depression from twitter activity. In *Proceedings of the Annual ACM Conference on Human Factors in Computing Systems*, pages 3187–3196. ACM, 2015.

[Xue *et al.*, 2013] Yuanyuan Xue, Qi Li, Ling Feng, Gari D Clifford, and David A Clifton. Towards a micro-blog platform for sensing and easing adolescent psychological pressures. In *Proceedings of the ACM conference on Pervasive and ubiquitous computing adjunct publication*, pages 215–218. ACM, 2013.

[Xue *et al.*, 2014] Yuanyuan Xue, Qi Li, Li Jin, Ling Feng, David A Clifton, and Gari D Clifford. Detecting adolescent psychological pressures from micro-blog. In *Health Information Science*, pages 83–94. Springer, 2014.

[Zhou *et al.*, 2011] J. Zhou, J. Chen, and J. Ye. *MALSAR: Multi-tAsk Learning via StructurAl Regularization*. Arizona State University, 2011.