Mental Health Detection from Social Media Data

Jia Jia
jjia@tsinghua.edu.cn
Tsinghua University
26/09/2017
Outline

- **Stage 1: Online detection of mental health problems**
  - Stress Detection via Harvesting Social Media
    - Detecting Stress Based on Social Interactions
    - Beyond Binary Classification: Stressor and Stress Level Detection
  - Depression Detection via Harvesting Social Media
    - A Multimodal Dictionary Learning Solution
    - Cross-Domain Depression Detection

- **Stage 2: Online Intervention of mental disorders**

- **Stage 3: Combination with offline researches**
Stress Detection via Harvesting Social Media
Introduction

• What is stress?
  • Definition: stress is the non-specific response of the body to any demand for change.
  • Stress is not clinical but cause many physical, mental, and social problems, even lead to suicide.

• Traditional stress detection methods
  • Face-to-Face interview by professional psychologists
  • Self-report questionnaires
  • Effective but reactive

• Social media is changing the ways people do with their healthcare and wellness
  • User generated contents (UGC) reflect emotions, moods and daily lives, and even influence their mental states
Research I

Detecting Stress Based on Social Interactions

- **Our work:** binary stress detection via social media

- **Challenges:**
  - Cross-media feature learning problem
  - Time-series modeling problem
  - Integration of social interactions and tweet contents

- **Solutions:**
  1. Propose a cross-media auto-encoder (CAE) to learn a joint representation from the cross-media social media data
  2. Propose a CNN with CAE to learn a unified high-level attributes from time-series data
  3. Propose a hybrid model combining Factor Graph model with CNN+CAE, to model the correlations of users’ social interactions and time-series tweet content

Detecting Stress Based on Social Interactions in Social Networks (TKDE 2017)
Framework
Dataset

- **Dataset DB1**
  - 19,000 weeks of stressed tweets and over 17,000 weeks of non-stressed tweets.
  - 492,676 tweets from 23,304 users in total.

- **Dataset DB2**
  - Small well-labeled dataset collected from the users who shared psychological stress scale PSTR via Weibo.

- **Dataset DB3 and DB4**
  - To further test our method, we collected two more datasets from Tencent Weibo (DB3) and Twitter (DB4).
Performance

- Performance:
  - 93.40% F1-Measure on Sina Weibo dataset D1
  - 87.85% F1-Measure on Sina Weibo with PSTR label dataset D2
  - 88.32% F1-Measure on Tencent Weibo dataset D3
  - 82.24% F1-Measure on Twitter dataset D4

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

<table>
<thead>
<tr>
<th>Method</th>
<th>Acc.</th>
<th>Rec.</th>
<th>Prec.</th>
<th>F1</th>
<th>CPU time</th>
</tr>
</thead>
<tbody>
<tr>
<td>LRC</td>
<td>76.18</td>
<td>87.94</td>
<td>78.58</td>
<td>83.00</td>
<td>39.43s</td>
</tr>
<tr>
<td>SVM</td>
<td>72.58</td>
<td>87.39</td>
<td>75.16</td>
<td>80.82</td>
<td>≈10min</td>
</tr>
<tr>
<td>RF</td>
<td>77.73</td>
<td>89.63</td>
<td>79.35</td>
<td>84.18</td>
<td>67.71s</td>
</tr>
<tr>
<td>GBDT</td>
<td>79.75</td>
<td>82.99</td>
<td>85.90</td>
<td>84.43</td>
<td>262.86s</td>
</tr>
<tr>
<td>FGM</td>
<td>91.55</td>
<td>96.56</td>
<td>90.44</td>
<td>93.40</td>
<td>≈20min</td>
</tr>
</tbody>
</table>
Studies of Social Interaction

1. Interaction contents of stressed users' tweets contain much more words regarding topics like death, sadness, anxiety, anger and negative emotion, while non-stressed users' tweets contain more words related to topics like friends, family, affection, leisure, and positive emotion.

2. Being stressed is a mutually correlated behavior.

3. The social structure of stressed users' friends tends to be less connected and less complicated.

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.
Research II

Beyond Binary Classification: Stressor and Stress Level Detection

- **Our work:** stressor and stress level detection via social media

- **Challenges:**
  - Stressor subject identification
  - Stressor event detection
  - Data and representation

- **Solutions:**
  - Extract a set of discriminant features
  - 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
  - Lookup a standard psychological stress scale to measure the precise stressor and stress level (Social Readjustment Rating Scale)
Framework
Collection Method

1. Categorize the stressor events into 43 categories based on the professional life events stress scale.

2. Manually define a set of keyword patterns collected from the LIWC 6 dictionary for each stressor event category.

3. Filter matched tweets from the aforementioned one billion Weibo dataset.

4. Collect the top 12 stressor event categories and invited 30 volunteers to manually label the stressor events and stressor subjects of the tweets.

Dataset

- Collection Method
  - Nearly 2,000 posts
  - Randomly selected 600 posts that are labeled as non-stress related to be the negative samples
Performance

(a) Stressor Event Detection

(b) Stressor Subject Detection

(c) Stress Level Measurement

F1-Measure

- Stressor Event Detection: 90%
- Stressor Subject Detection: 81%
- Stress Level Measurement: 200
Depression Detection via Harvesting Social Media
Motivation

• Depression is a leading cause of disability worldwide
  • 350 million people of all ages suffer from depression

• Clinical diagnosis is effective
  • Make face-to-face interviews referring to DSM criteria
  • Nine classes of depression symptoms are defined in the criteria, describing the distinguishing behaviors on daily lives

• Clinical diagnosis is not proactive
  • More than 70% of people in the early stages of depression would not consult the psychological doctors

• People are increasingly relying on social media platforms
  ▪ User generated contents (UGC) reflect daily lives and moods
  ▪ Online behaviors may not be covered in previous depression criteria
Research I

A Multimodal Dictionary Learning Solution

- **Our work:** depression detection via harvesting multimodal social media

- **Challenges:**
  - No public available large-scale benchmark datasets
  - Users’ behaviors on social media are multi-faceted
  - Few users’ behaviors are symptoms of depression

- **Solutions:**
  - Construct well-labeled datasets by rule-based heuristic methods
  - Extract six groups of depression-oriented features
  - Leverage multimodal dictionary learning method
Framework

Feature Extraction
- Online Behavior
- Social Media
- Clinical Depression Criteria

Multimodal Features
- Social Network
- User Profile
- Visual
- Emotional
- Topic-level
- Domain-specific

Dictionaries

Multimodal Depressive Dictionary Learning
- Joint Sparse Representation
- Latent Features of Depression and Non-depression Dataset
- Multimodal Classifier

Negative Words Count per Tweet
- Mean: 0.29
- Mean: 0.52

Posting Ratio at Late Night
- Mean: 0.27
- Mean: 0.39

Behaviors Analysis
- Detect depressed users

Update \( \{D^s, w^s\} s \in \{1, ..., 6\} \)
Methods & Experiments

- **Method**: multimodal dictionary learning

- **General idea**:
  - Why dictionary learning: learn the latent and sparse representation
  - Why multimodal learning: jointly model cross-modality relatedness to capture the common patterns and learn the joint sparse representations
  - Train a classifier to detect depressed users with the learned features specifically

- **Datasets**:
  - D1 & D2 for model learning: 1400 depressed users & 1400 non-depressed users, selected by strict sentence pattern (e.g. “I’m diagnosed depression”).
  - D3 for depression behaviors discovery: 37000 depression-candidate users, selected by just loosely contained the character string “depress”.

- **Performance**: 85% F1-Measure on D1 & D2
Case Study: Depression Behaviors Discovery in Twitter

• **Posting time.** Depressed users post more tweets between 23:00 and 6:00, indicating that they are susceptible to insomnia.

• **Emotion catharsis.** All users say more about their bad moods, but depressed users express more emotions, especially negative emotions.

• **Self-awareness.** Depressed users use more first personal pronouns, which may reflect their suppressed monologues and strong senses of self-awareness.

• **Live sharing.** Depressed users post more antidepressant and depression symptom words, indicating that they are willing to share what they encountered in the real life.
Motivation:

- Success of depression detection in Twitter (Source Domain)
- Probably insufficient labeled (self-reported depression) samples for model training in other platforms (Target Domain, e.g. Weibo), due to cultural differences

Match the pattern “I am diagnosed with depression” in 100 million randomly crawled tweets:

- 481 users accessed in Twitter
- 142 users accessed in Weibo

Utilizing the source domain dataset to improve depression detection performance for a target domain
Challenges:
- How to bridge the gap between the heterogeneous feature spaces of different domains?
- How to design a model that exploits the source domain data to enhance detection for the target domain?

Solutions:
- Reveal the different feature patterns across domains, as well as the possibly disparate contribution to detection of the same feature
- Propose a cross-domain Deep Neural Network model with Feature Adaptive Transformation & Combination strategy (DNN-FATC)
Source domain: Twitter  Target domain: Weibo
Datasets and Features

- **Datasets:**
  - **Weibo dataset** $\mathcal{D}_T$: 580 depressed users & 580 non-depressed users selected by strict sentence pattern (e.g. I’m diagnosed depression).
  - **Twitter dataset** $\mathcal{D}_S$: 1400 depressed users & 1400 non-depressed users, as shown in Research 1

- **Feature patterns across domains**
  - **Universal features**: domain-independent detection indicators
  - **Singular features**: have distinctive, or even opposite implications on depression detection in different domains
  - **Isomerous features**: follow distinctive integral distributions across domains
Methods and Experiments

- **Method**: a cross-domain Deep Neural Network model with Feature Adaptive Transformation & Combination strategy (DNN-FATC)
  - **Feature Adaptive Transformation**: Feature Normalization & Alignment (FNA) & Singular Feature Conversion (SFC)
  - **Feature Combination (FC)**: combine the exclusive features in $\mathcal{D}_T$ into the deep model framework

- **Performance**: 78.5% F1-Measure on Weibo dataset, outperforming other transfer-learning methods and deep learning methods

Table 2: Performance of all compared approaches. For each method, precision, recall and F1-measure (%) are orderly shown in the corresponding cell.

<table>
<thead>
<tr>
<th></th>
<th>DL</th>
<th>ARC-t</th>
<th>MMDT</th>
<th>HFA</th>
<th>DNN-FATC</th>
</tr>
</thead>
<tbody>
<tr>
<td>71.3</td>
<td>72.2</td>
<td>71.6</td>
<td>73.8</td>
<td>73.6</td>
<td></td>
</tr>
<tr>
<td>76.2</td>
<td>75.2</td>
<td>77.0</td>
<td>76.6</td>
<td>83.8</td>
<td></td>
</tr>
<tr>
<td>73.3</td>
<td>73.7</td>
<td>73.9</td>
<td>75.1</td>
<td>78.5</td>
<td></td>
</tr>
</tbody>
</table>
Case Study: Depression Behaviors Discovery in Weibo

- **Tweeting time:** The depressed users are more likely to post tweets between 22:00 and 6:00.

- **Gender:** Women users are more likely to suffer from depression.

- **Linguistic characteristics:** Depressed users use more biology-related words and first-person singulars, manifesting more concerns about health issues, as well as personal affairs.

- **Retweets count:** Depressed users are retweeted less, which may reflect their lack in social engagement and attention from others.
Comparison of depression indicators between Twitter and Weibo

- **Universal features:**
  - Negative words count
  - First-person singular pronouns count
  - Retweets count
  - Post time distribution

- **Singular features:**
  - Tweets count
  - Positive words count
  - Saturation of image
Conclusion

- **Current work:**
  - **Stage 1:** Online detection of mental health problems

- **Future work:**
  - **Stage 2:** Online Intervention of mental disorders
  - **Stage 3:** Combination with offline researches
Thanks