Int. J. Mach. Learn. & Cyber.
DOI 10.1007/s13042-017-0697-1

ORIGINAL ARTICLE

Multi-kernel SVM based depression recognition using social media data

Zhichao Peng[1] · Qinghua Hu[1] · Jianwu Dang[1]

Received: 23 August 2016 / Accepted: 19 May 2017

Abstract Depression has become the world's fourth major disease. Compared with the high incidence, however, the rate of depression medical treatment is very low because of the difficulty of diagnosis of mental problems. The social media opens one window to evaluate the users' mental status. With the rapid development of Internet, people are accustomed to express their thoughts and feelings through social media. Thus social media provides a new way to find out the potential depressed people. In this paper, we propose a multi-kernel SVM based model to recognize the depressed people. Three categories of features, user microblog text, user profile and user behaviors, are extracted from their social media to describe users' situations. According to the new characteristics of social media language, we build a special emotional dictionary consisted of text emotional dictionary and emoticon dictionary to extract microblog text features for word frequency statistics. Considering the heterogeneity between text feature and another two features, we employ multi-kernel SVM methods to adaptively select the optimal kernel for different features to find out users who may suffer from depression. Compared with Naive Bayes, Decision Trees, KNN, single-kernel SVM and ensemble method (libD3C), whose error reduction rates are 38, 43, 22, 21 and 11% respectively, the error rate of multi-kernel SVM method for identifying the depressed people is reduced to 16.54%. This indicates that the multi-kernel SVM method is the most appropriate way to find out depressed people based on social media data.

Keywords Chinese microblog · Depression recognition · Multi-kernel · Social media · SVM

✉ Jianwu Dang
dangjianwu@tju.edu.cn

Zhichao Peng
zcpeng@csu.edu.cn

Qinghua Hu
huqinghua@tju.edu.cn

1 School of Computer Science and Technology, Tianjin University, Tianjin 300350, China

1 Introduction

The World Health Organization (WHO) describes depression as a common mental disorder, characterized by sadness, loss of interest or pleasure, feelings of guilt or low self-worth, disturbed sleep or appetite, feelings of tiredness, and poor concentration. According to a report issued by the WHO in 2012, more than 350 million people in the world were suffering from depression. At its worst, depression can lead to suicide. A survey shows that almost one million people with depression end their lives each year.[1] Depression has become the fourth major disease in the world and is possible to become the second major disease following heart disease by 2020 according to the WHO.

If depressed people can cooperate with doctors for targeted treatment, depression can be effectively treated. But only less than half of the global patients receive effective treatment. Additionally, the situations of depression in different countries are very different. In November 13, 2014, *Nature* published a special issue on depression, reporting the burden of depression, the present research situation, the research progress and research challenges [25, 26]. As shown in Table 1, the country

[1] World Health Organization, http://www.who.int/topics/depression/en/.

Published online: 02 June 2017

Table 1 Prevalence of depression in typical country

Country	Prevalence of depression (%)	Psychiatrists per 100,000 people
Afghanistan	22.05	0.16
Switzerland	6.16	41.42
United States	4.45	7.79
China	3.02	1.53

with the highest prevalence of depression in the world is Afghanistan, which has the least medical resources and the lowest health care level because of the civil war. In contrast, Switzerland has the best health care for mental health, with more than 40 psychiatrists per 100,000 people. Even for the countries that possess the sound medical security, the burden of the disability from depression only is reduced just by 10–30%.

The relatively low prevalence of depression in China could be the result of the way in which it is diagnosed, rather than lower actual rates. Because of the relative lack of psychiatrists, people with depression here often are misdiagnosed as stomach pain or headache. As a result, they may not adopt the standard diagnostic methods that focus on mood, motivation and fatigue to calculate the incidence of the disease [26]. Additionally, because of social discrimination on mental disorders, many people with depression don't like to see the doctor. As a result, they missed the diagnosis.

According to the data of Chinese Association of Mental Health, more than 200,000 people commit suicide because of depression each year in china. The three main reasons that trigger this phenomenon are social pressure, family and marriage. In addition, the depressed patient trends to be younger, and 25–50 years old people are the main part of depressed people, accounting for 80%. China's educational pattern and one-child policy which has been enforced for a long term to some extent make many young people egoistical and not independent to overcome difficulties, which also induce such depression [48].

With the rapid development of the internet, people often express what they see, hear and think on the social media sites. The characters and emotion icons (emoticons) published on social media show their true and real-time thoughts and feelings about a variety of happenings in daily life. Of course, the potential depressed people including those who are misdiagnosed or not diagnosed also publish postings on social media. These postings can describe the mental status of the writers who feel sad, tired, guilty, helpless, or self-hatred. By analyzing the postings, we can identify the mental changes of the writers, and find out the potential symptom complex, even though we are not specialists in mental medical field. For this reason, we attempt to exploit a new way to find out the people with the potential depression symptom from social media sites.

Sina Weibo (a Chinese Microblog site which is similar to Twitter) is the most popular Chinese social media site. On March 18th, 2012, a posting by Zoufan caused huge shock in Sina Weibo, and then many microblog users forwarded the posting, that is "I suffer from depression, so I'll go to die. No special reason, and please do not care about my death. Bye-bye, everyone". In the next day, police confirmed that Zoufan had committed suicide. Zoufan was a college student in Nanjing and had passed the psychological health testing in school successfully. Perhaps she didn't want teachers and classmates to know that she had suffered from depression, so she did not truthfully answer questions in the students' psychological test. However, if we looked carefully through all of her 1896 postings, it was clear that she lived alone, and showed a tendency of suicide to some extent. Based on those postings, as a non-specialist we could realize that she had depressive tendency obviously. The majority of people with depression do not receive any treatment. However, utilizing the postings would be an efficient way to find out them. Accordingly, we timely conduct the intervention and prevention to avoid the occurrence of such tragedies like Zoufan. To do it, applying the machine learning on the social media data is a powerful way to identify potential depressed people.

Many machine learning and statistical methods have been used to discover the potential knowledge from social media data, such as the sentiment analysis, the personalized recommendation system, the public opinion monitoring, special people's discovery and so on [40]. Recently, researchers have started to conduct studies by using social media data to investigate people's mental state such as depression [12, 31, 43, 44]. In these papers, researchers focus on analyzing the sentiment polarity of each posting, and the content of each posting do not play an important role comparing to the behavior features.

Accordingly, this study attempts to find out the potential depressed people by considering simultaneously the behavior and text features. We begin with the analysis of the typical characteristics of the depressed people, and their features expressed in social media data. Three categories of features, user microblog text, user profile and user behaviors, are extracted from their social media to describe the users' situations. To extract the text features, we build up an emotional dictionary. Considering the heterogeneity between word frequency and another two features, we employ multi-kernel SVM methods to adaptively select the optimal kernel for different features to find out users who may suffer from depression. We construct a depression recognition model on the basis of multi-kernel SVM. At last, by comparing a number of different machine learning methods (Naïve Bayes, SVM, KNN, Decision tree and libD3C), we evaluate the efficiency of the method using the social media to find out the potential depressed people.

The rest of the paper is organized as follows. Section 2 reviews the main related works, and put forward our research on depression recognition problem from Chinese microblog. To solve this problem, framework of depression recognition is proposed in Sect. 3 firstly and then construct a special emotional dictionary in Sect. 4. Additionally, depression recognition method based on multi-kernel is formulated in Sect. 5 and experimental results are presented in Sect. 6. Section 7 concludes the paper with future perspectives.

2 Related works

Depression is a serious illness, which is deeply related with emotions. Emotion is extensively researched for a long time [10]. Emotion recognition is an important research field of pattern recognition and human–computer interaction for years. Emotions can be divided into the basic emotions and complex emotions, but the detailed classification is different in different psychiatrists' opinion, and the results also show some differences [8]. According to the widely used model by researchers which is proposed by Ekman [17], there are six basic emotions: surprise, fear, disgust, anger, happiness and sadness. By combining those six emotions, we can get the rest of a variety of complex emotional descriptions, such as depression, tension, anxiety, etc. Nowadays the explosion of UGC (User Generated Content) has generated abundant multimedia information in social media sites, and such information not only contains user's different viewpoints and ideas, but also contains their emotional information. The latter component can be used for mental health investigation. Accordingly, emotional analysis based on social media has become a research hotspot [22]. At present, most of the emotional analysis researches on Chinese microblog are still in the stage of investigating whether the postings are objective or subjective, and/or it has positive or negative emotion. Existing studies on microblog emotion recognition mainly focus on identifying the basic emotions. Because complex emotion analysis has different strategies in different fields, and the study of investigating further development of people's emotion (such as depression) is very limited.

Depression is a mental disease which has a very complex etiology. Experts in psychiatry, psychology, medicine, and sociology, etc. have conducted a lot of related researches. Psychologists use different depression measurement scales such as SDS (Self-Rating Depression Scale) and CES-D (Center for Epidemiologic Studies Depression Scale) to identify people's levels of depression. These scales mainly test people's thoughts and feelings which are related to depression.

2.1 Study of depression based on physiological status

Medical researchers investigate a number of behavioral signals to detect people's mental state, such as brain signals, heart rate, blood pressure, voice prosody, and facial expression to get psychophysiological information [16]. They use many physiological devices such as EEGs, heart rate trackers, and skin conductors to acquire abundant data; however, these devices are not convenient to carry, and often difficult to use, so typically limited to be used in clinics. In order to find some convenient ways to monitor depressed people, researchers try to install app software in the smartphone to get different types of sensor data in order to record the activities of people's daily life. Burns et al. [3] developed a smartphone app Mobilyze that collects sensor data to detect the user's cognitive state. Mobilyze built machine learning models to predict mood, emotions, cognitive or motivational states, activities, environmental context, and social context. This system has been used for intervention and not for depression detection. Doryab et al. [16] developed a system called Big Black Dog (BBD, an Android app that uploads captured data every day to server) to detect the onset of major depression, considering the earlier diagnosis and treatment of first episodes, relapses, and recurrences. BBD collects sensor data that can reveal behavioral and environmental factors, including noise amplitude (from microphone), location, WiFi SSIDs, light intensity (from ambient light sensor), and movement (from accelerometer).

2.2 Study of depression based on social media

With rapid development of mobile network technology and wide use of smart phones, Social network [18, 19] also develops rapidly, and many people use one or more social networking services (such as Facebook, Twitter, Microblog, Wechat, QQ, etc.) to express their views and emotions. At the same time a lot of personal information (postings sending time, postings sending frequency, user activity state, etc.) are exposed. Such information provides a possible way for us to find out the potential depressed people who express their inner emotion on social media sites. The main features of depressed individuals differing from normal person are linguistic content and styles, sentiment state, behavior habits etc. Some researchers conducted emotion analysis on the basis of online community or forum on depression [9, 24]. Nguyen et al. [31] studied the characteristics of online depression communities in comparison with those joining other online communities based on the website of depression.livejournal.com (the largest community in Live Journal interested in "depression"). He analyzed the effectiveness and content based on mental health communities using sentiment information, topics of interest and language styles.

More researchers detect and predict depression by using the online social networking sites [1, 46, 30]. Park et al. [32–34] made an effort to examine social network determinants of depressive symptoms. Facebook contains a wide range of information about users, including demographic features such as age and gender, as well as social features such as friends list, like, interest and location tagging. Together, these features could represent how a user maintains relationships online as well as offline. Facebook activities have predictive power in distinguishing depressed and non-depressed individuals by the wide set of features available on Facebook. De Choudhury et al. [12] used a crowdsourcing methodology to build a large corpus of postings on Twitter that have been shared by individuals who are diagnosed with clinical depression. He developed a probabilistic model trained on this corpus to determine whether postings could indicate depression, and created a "social media depression index" with potential for detecting depression in populations. He found that it was possible to predict episodes of depression by using features such as changes in postings frequency and increased concern over health issues.

De Choudhury et al. [11, 13] also used online social networking sites to detect and predict special depressive populations, such as postpartum depression. They leveraged Facebook data shared voluntarily by 165 new mothers as streams of evidence for characterizing their postnatal experiences. They leveraged multiple measures including activity, social capital, emotion, and linguistic style in participants' Facebook data in pre- and postnatal periods.

Wang et al. [43, 44] built an association model to predict depression. The model was established on basis of sentiment analysis algorithm, and also combined sufferer behavior characteristic with the principle symptom which affects depression detection significantly. He put forward an improved model with additional factors (like linkage features and propagation coefficient) based on the above works [43, 44]. In his papers, all postings were used to analyze sentiment polarity which is positive or negative, and sentiment analysis was an unimportant factor in the model. In fact, however, postings are an important factor to judge whether the writer has depressive tendency or not.

By means of digging the emotional information and the behavior features from users' published postings, a new way using social media data is put forward to find out the potential depressed people.

3 Framework of depression recognition and data processing

3.1 Framework of depression recognition

Social media sites have become an integral part of people's daily social lives and emerged as one of the main forms of communication among certain social groups. And people can publish their thoughts and feelings on social media sites every day. Therefore, utilizing the postings published on the social media sites becomes a brand new way to discover the special emotion tendency of the writers whether they suffer from depression or not.

In this study, we attempt to discriminate whether there is a depressive tendency through user's postings and Internet behaviors, so as to make further prevention and intervention. Accordingly, we take advantages of social media data to detect potential depressed people, and provide useful clues to psychiatrist, and thus intervention and treatment can be conducted instantly. Figure 1 shows the framework proposed in this study for identifying potential depressed people using social media data. In this framework, we divide microblog features into three major categories: user microblog text, user profile and user behaviors. The user microblog text plays an indispensable role in depression recognition. Therefore, we construct a depression sentiment dictionary based on the text first, and combine with other features to identify the mental state of the users. By utilizing the depression emotional dictionary, we calculate the term frequency within the text, and form a user text feature based on the feature selection. User profile features and user behavior features are extracted as other different features. Considering the dissimilarity of different kinds of features and the nonlinearity of the data, the multi-kernel method for the user's different characteristics is employed and the SVM classifier is used to find the depressed people.

3.2 Data acquisition

Our data comes from Sina Weibo which was established in August 2009, and by 2016. The number of its MAU (Monthly Active User) is reported to reach 198 million worldwide. It works in the same way as Twitter, with similar features like the use of the @ and hash tag (#) and users can choose the objective they follow. Like Twitter, Sina Weibo has the 140-character limit but as these generally correspond to syllables rather than single letters like in alphabet characters. Postings in Chinese microblog generally contain more information than that in Twitter. Apart from text, users can also insert other issues such as emoticons, web links and photos into their postings.

We can get the microblog data through an open platform API interface, however, due to the consideration of Sina

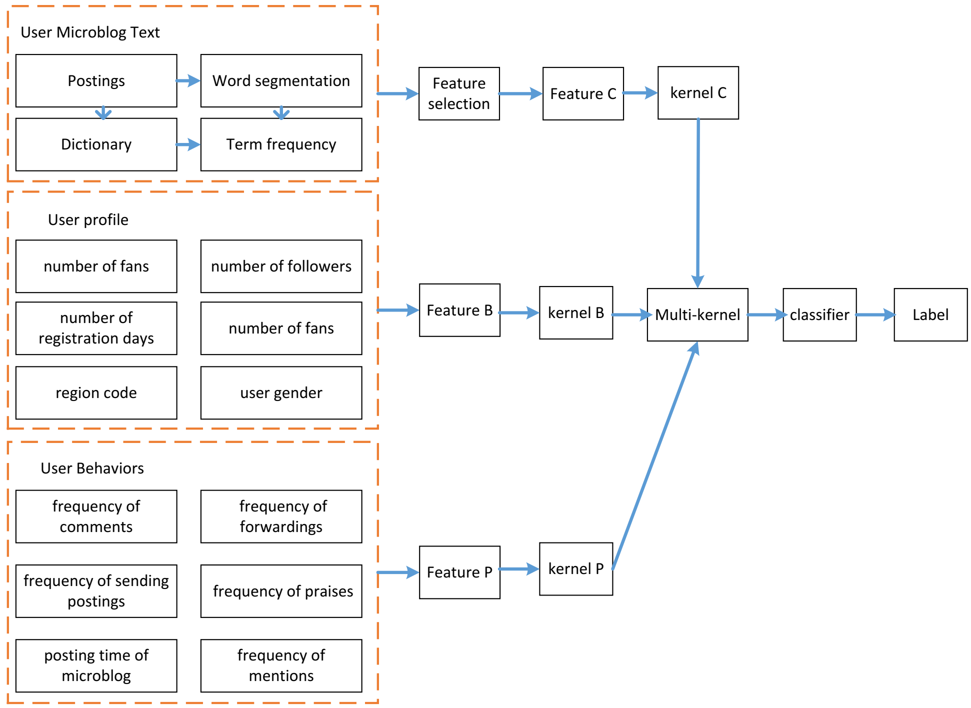

Fig. 1 Framework of depression recognition

Weibo for protecting user data, the microblog API interface allows the user to crawl the specified data with maximum of 30%, such as the number of followers and fans etc. In order to get the complete data for more accurate experiment results, we use the data provided by a data mining company. The acquired data include the user basic information and all the microblog data posted by the users.

3.3 Data preprocessing

We attempt to find out the potential depressed people from the postings, mainly based on the user's published microblog postings, their behavior characteristics and their profile. Of course, the postings are the most obvious feature, for example, "I still feel like life is boring. Everything is boring. What I do is to make myself more and more sad. I feel inferior every day. Why do people live such a life like me? If I die, it will not be so bad". As non-specialists, we can find out depressive tendency obviously from these type of postings. According to intuitive experience, therefore, we manually label the microblog data, and send the result to psychiatrists for second screening.

Our data are collected from August 14, 2009 to August 15, 2013. We mainly analyze the original postings which contain their own viewpoints. The forward postings which don't express their views are filtered out. Meanwhile, we also eliminate the official microblog accounts for those postings do not express any personal emotions.

We extract two kinds of information from the acquired data. One is the user's microblog text, and the other is the user's profile and network behavior information such as the sending time, the number of forwarding and comments, etc. For microblog text, we usually process the punctuation marks and stop words, so that it is convenient for text segmentation. We also remove all the HTTP links and redundant spaces etc.

Because users registered their accounts in different time and different people have different microblog sending frequency, some users only published several postings, others probably published postings up to tens of thousands. We need to screen the microblog data obtained, in accordance with the order of time to obtain a maximum of 1000 users' postings starting from the user registration time, and remove users who published totally less than 100 valid postings.

Because a posting is relatively short (usually contains only a dozen or even several words or emoticons), there will be serious data sparse problem, which will seriously restrict further improvement of accuracy. To avoid high dimensional and sparse feature vectors, we get users' postings together and build an emotional lexicon as well as behavior information to carry out emotional analysis.

As shown in Table 2, according to the requirement, we select 135 eligible users from 141 depressive users and 252 users from 263 non-depressed users. Among them, depressive tendency people published a total of 62,759 postings,

Table 2 Data for preprocessing

Type	Primary user number	User number	Microblog numbers after filtering	Average number of each user
Depression	141	135	62,759	465
Non-depression	263	252	170,114	675

465 postings for each person on average. In contrast, non-depressed people published a total of 170,114 postings, 675 postings for each on average. 83 depressed people are female out of 135 depressed people, and 52 depressed people are male. The ratio of female to male is 60:40%. So gender should be taken into consider as an important feature to find out the depressed people. In additional each depressed user published 465 postings on average, and each non-depressed user published 675 postings on average. This implies that non-depressed people are more active than depressed people on social media sites. Therefore, user activity frequency also should be taken into consider as an important feature.

3.4 User behavior analysis

At present, DSM-IV (The Fourth Edition of the Diagnostic and Statistical Manual of Mental Disorders) is the most perfect diagnostic systems of mental disorders used widespread. In the DSM-IV, six out of nine core symptoms of depression are primarily experiential and thus not directly observable (e.g., depressed mood, feelings of worthlessness and guilt, thoughts of death, fatigue) and only three are major behavioral in nature and thus potentially directly observable (e.g., weight loss or gain, in- or hypersomnia, psychomotor agitation or retardation). People usually express their emotions consciously by words or emoticons on social media sites. Moreover, people expose their own behavior via their unconscious activities such as the frequency and time of sending postings.

As shown in Fig. 2a, b, the moods, interests and emotions of the non-depressed people are mainly involved with happiness, enjoyment, expectation, music and game etc. while involved with depression, loneliness, painfulness and even death for the depressed people.

We analyze all the postings made by each user during the course of a day (24-h cycle and local time). Figure 3 shows the diurnal status of postings sending between depressed people and normal people, measured as the percentage of postings according to every two hours and included the entire history of microblog data of the two class users. We find out that depressed people have different patterns in sending time compared with normal people. Depressed people averagely send less postings from 6:00 to 18:00 than

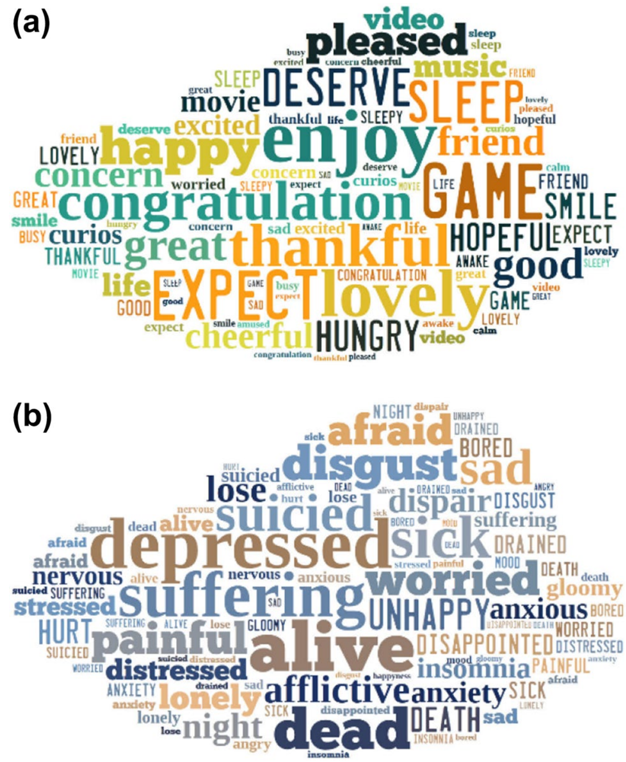

Fig. 2 The moods, interests and emotions for the two classes. **a** The moods, interests and emotions of the non-depressed people. **b** The moods, interests and emotions of the depressed people

the normal people. But, the depressed people averagely send more postings from 18:00 to 6:00 than normal people. It implies that posting time is another important indicator of depressed people. So according to posting time of microblog, we divide a day into two parts, one is a "night" window as "6:00 p.m.–5:59 a.m.", and another is the "day" window for the same user as "6:00 a.m.–5:59 p.m.".

4 Emotional dictionary construction

4.1 Feature selection

Feature selection is the key operation of data preprocessing. Filter methods as a common feature selection method always employ special statistics measurement to filter out the useless features. For example, some feature ranking methods are used in image classification and protein–protein interaction prediction [50]. For the text classification, word frequency is used in feature ranking method. In this study, we use TF-IDF (Term Frequency-Inverse Document Frequency) [39] method to build a special emotional dictionary, and use TF method to statistics word frequency feature from microblog text.

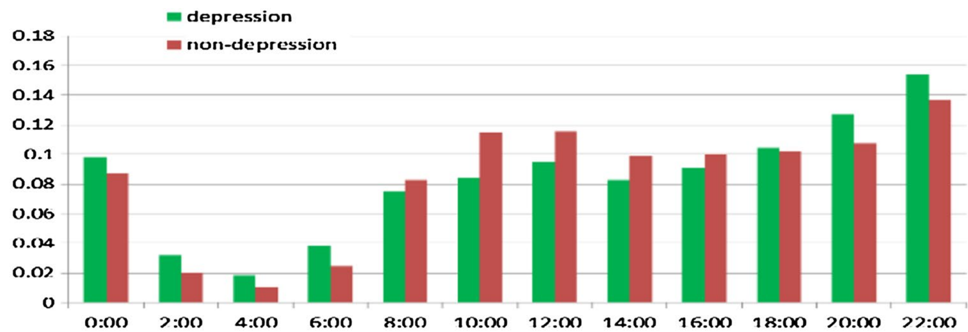

Fig. 3 Diurnal status of postings sending time for the two classes

According to the characteristics of social media websites, three categories of features, user profile features, user behavior features and microblog text features, are selected for special sentiment classification. The details of feature selection method are as follows to extract these three categories of features.

1. User profile features

We utilize a set of profile of the microblog user to characterize the general features associated with them. We select six user profile features: number of postings, number of followers, number of fans, number of registration days, user gender and region code. The first four features are the measures of overall profile of the microblog users in social media. The feature of registration days means total number of days (total-days) from the date of registration. The first three features inputted into the classifier are normalized by number/total-days respectively. We also take other features into account such as user gender and region code. Sina Weibo provides a list of region code in accordance with the order of <province, city>, for examples, <11, 1> that indicates to the Dongcheng District city of Beijing.

2. User behavior features

Interaction frequency is the most distinct feature in the interaction behavior of social media. We can find out the behavior features from user's original posting by the way of interaction activities with other users. We select six user behavior features: frequency of forwardings, frequency of comments, frequency of praises, frequency of sending postings, frequency of mentions and posting time of microblog.

The first five features are the behavior characteristics of the interaction between the other users. These features are normalized by number/online-days respectively. The online-days are the total number of days crawled to analyze the user's postings started from the register time to the time that meet the experiment requirement. Because the depressed people are much easier to insomnia at night than normal people, the feature of posting time of microblog should be taken into consideration. In our experiment, we label posting time of microblog as '1' when the posting time belongs to "night" window, and correspondingly label '0' for "day" window.

3. Microblog text features

The extraction of text features is performed by counting emotional words of microblog text from emotional dictionary for depression recognition. First of all, emotional dictionary is a very important tool (or measure) in the sentiment analysis of microblog. But there are different methods to build emotional dictionary for sentimental analysis in different languages. For example, as an English emotional dictionary, LIWC lexicon [36] has been well validated and widely used in emotion analysis, which contains a dictionary of several thousand words with emotion-indicative categories such as positive emotions, negative emotions and others emotions. There are several Chinese emotional dictionaries which are available to us such as HowNet [15] and NTUSD emotional dictionary [27], Chinese Affective Lexicon Ontology (CALO) [47]. HowNet and NTUSD are mainly used for the coarse-grained sentiment analysis, such as positive or negative emotions, while CALO is mainly used for fine-grained sentiment analysis. All of these dictionaries are not proper for special sentiment recognition such as depression recognition.

Based on the six kinds of basic emotions, CALO adds a kind of emotion lexical ontology, which is positive emotions "good". Finally, CALO consists of seven kinds of emotions with the 21 subclasses in the lexical ontology. The seven kinds of emotions are surprise, fear, disgust, anger, happiness, sadness, good.

We build the emotion feature dictionary for depression analysis according to CALO classifier method. Because depression is a complex emotion which mainly contains the pain associated with anger, sadness, sorrow, self-guilt, shame and other emotions under different circumstances. Different from CALO, as the "surprise" emotion is not associated with "depression" emotion recognition, we adjust the emotional vocabulary ontology, and add

Table 3 Basic emotional dictionary

Class	Examples
Depression	煎熬,安眠药,复发,绝望的,忧郁,懦弱,心情不好,头晕,哭泣,心痛,脆弱,解脱,焦虑,折磨,睡不着,泪流满面,失眠,崩溃
Good	喜欢,恭敬,敬爱,英俊,优秀,信任,信赖,可靠,渴望,保佑,倾慕,宝贝
Happiness	喜悦,欢喜,笑眯眯,欢天喜地,踏实,宽心,安心
Fear	害怕,奇怪,奇迹,大吃一惊,瞠目结舌,不安
Anger	发作,生气,发火,气愤,恼火,大发雷霆,七窍生烟
Sadness	绝望,忧伤,无所谓,难受,去死,自杀,悲伤,痛苦
Disgust	他妈的,怨恨,狂妄,该死,崩溃,不甘心,将就,不耐烦,苦恼,王八蛋,令人作呕,轻蔑,厌恶,恨不得,恶心小人

"depression" as a new class, and exclude the "surprise" class. Finally, the emotion feature dictionary is built with seven kinds of emotion, including depression, good, happiness, fear, sadness, disgust and anger.

4.2 Emotional dictionary ontology

Characters with cyberspeak style and emoticons are the main ways to express people's feelings on social media sites. Therefore, we build the text emotional dictionary, cyberspeak emotional dictionary and emoticon dictionary for the emotion recognition.

1. Text emotional dictionary

As shown in Table 3, the basic emotional dictionary consists of emotional words that are commonly used in social media sites, but the low frequency words in CALO are not included in this dictionary. However, emotional dictionary breaks through the limitations to words, and the phrases which have explicit sentiment are also included in this dictionary. In addition, considering the difference between the special domain emotion recognition and the general emotion recognition, for example, depression sufferers tend to consider the topic about "death" or "die", some words should be paid special attention in building up the feature dictionary. At last, negative words totally reverse the meaning of the expression, such as "not", "never". The negative words such as "not good" are imported as a new item into the dictionary.

Additionally, cyberspeaks with a certain emotional tendency are very prevalent in postings and obvious to display user's current emotional state. Therefore, the emotional dictionary needs to be included the cyberspeaks and colloquial words. As shown in Table 4, the word like "SB", which is word showing anger emotion, is also added into the dictionary. Moreover, postings are often written in a colloquial style, and then modal particles often occur in postings to express feelings directly, such as "Ha-Ha". Such words are added into the vocabulary too.

Table 4 Cyberspeak emotional dictionary

Class	Examples
Depression	稀饭,果酱
Good	给力,牛B,高富帅,白富美,顶,大虾,8错,GX,NB,牛X
Happiness	得瑟,嘚瑟,嘻嘻,哈哈(Ha-Ha)
Fear	悲剧,悲催,菜鸟,damn
Anger	靠,TMD,TNND,MD,SB,Kao,傻逼,狗逼,操蛋,犯贱,傻B,欠揍,找死
Sadness	屌丝,晕,压力山大
Disgust	垃圾,他妈的,该死

Table 5 Emoticon dictionary

Class	Examples
Depression	[囧] [困]
Good	[给力] [威武] [鼓掌] [good] [赞] [OK] [耶] [可爱]
Happiness	[微笑] [太开心] [哈哈] [嘻嘻] [偷笑] [偷乐] [亲亲] [爱你] 呲牙
Fear	[汗] [害羞] [抓狂]
Anger	[怒] [闭嘴] [怒骂] [草泥马]
Sadness	[泪] [失望] [蜡烛] [悲伤] [衰] [泪流满面] [生病] [委屈] [可怜]
Disgust	[鄙视] [挖鼻] [白眼] [阴险] [猪头] [弱]

2. Emoticon dictionary

Emoticon is another way of expressing emotions in social media sites, which is different from the normal text. A lot of users in the microblog will more or less use the emoticons. Sina Weibo provides some of the default expression symbols, such as "", and so on. As shown in Table 5, the emoticons crawled down from website are formed by the text that is contained in brackets.

4.3 Procedure of building emotional dictionary

CALO is a good emotional lexicon for classification of basic emotion, but it is not a good classification for complex emotion recognition such as depression recognition. In order to overcome the deficiency of CALO for depression recognition, we focus on the following four aspects of the issues in building emotional dictionary.

1. High-dimensional and small-sample data will lead to the "over fitting" problems. CALO includes a total of 27,466 emotional words. However, there are only 387 users in our database. So if we use CALO as emotional dictionary in this experiment, it will lead to over-fitting.
2. A large part of the idioms in CALO are not commonly used in postings, however, cyberspeaks often occur in social media sites. Thus some explicit expressions of

cyberspeaks should be included in the emotional dictionary.

3. CALO is a universal emotional dictionary, and it doesn't take the emotional features in specific domain into consideration. Thus it is not suitable for depression recognition.
4. We also consider some of the emotional words which are related with the CES-D depression scale, but not included in CALO about depression in, such as bad appetite, feel weak, etc.

To solve the above issues, we employ a new method to build emotional dictionary for depression recognition. The procedure for building emotional dictionary as follow:

Step 1 We aggregate the microblog corpus together, and then segment them into phases or words. In order to build up the emotional dictionary, we use a corpus of 5 million postings. A word segmentation tool, ICTCLAS [49], is used to segment the postings. The conjunctions and stop words are removed after the segmentation.

Step 2 We use TF-IDF to analyze the emotional words or phrases statistically. The main idea of TF-IDF method is that if the TF of a word or phrase is high in one article but very low in other articles [39]. Hence we should think the word or phrase as an effective classification feature which can distinguish the two articles from each other. The formula of TF-IDF is shown as the following:

$$TF-IDF(wi) = freq(wi) \times \left(\frac{logN}{df(wi)}\right). \tag{1}$$

The $freq(wi)$ is the frequency of word wi appears in the corpus, N is the number of documents in the corpus, $df(wi)$ is the document frequency of corpus that contain word wi. In the corpus, the TF-IDF value of each candidate word is first calculated, and then sorted according to the value of TF-IDF(wi). The candidate word would be chose for further selection if the value of TF-IDF(wi) is beyond the threshold value.

Step 3 We manually remove the objective word, and then add the subjective words to build a basic emotional dictionary according to similar classification method of CALO. Additionally, ICTCLAS is also a good tool which includes many cyberspeaks. We also build a cyberspeaks emotional dictionary according to the similar method mentioned above.

Step 4 For emoticons, we classify all the emoticons that may have an emotional meaning to build the emoticon dictionary by the way of artificial labeling.

As a result, a new emotional dictionary is constructed for depression recognition and this dictionary holds the capacity of 1381 emotional words or phrases.

5 Depression recognition and prediction

SVM based on structural risk minimization has tremendous advantages in small samples and better generalization ability. Structural risk minimization considers the empirical risk and structural risk simultaneously. If we can ensure the accuracy of classification (empirical risk), while reducing the VC dimension of the learning machine, the expected risk of learning machines on the entire set of samples can be controlled. We use SVM method to achieve the task of depression recognition and prediction, while comparing with other classifiers such as traditional machine learning methods, Naive Bayes, KNN and Decision Tree [4].

Additionally, the latest ensemble learning method is also employed for comparing with SVM method. Ensemble methods can obtain better predictive performance using multiple learning algorithms than the traditional learning algorithms alone, and also are used in imbalanced classification problem [29]. LibD3C is an ensemble method that is based on k-means clustering and the framework of dynamic selection and circulating in combination with a sequential search method [28].

To better understand SVM, we introduce below the principle of SVM and multi-kernel SVM method.

5.1 SVM

SVM is a supervised learning model with associated learning algorithms that analyze data used for classification and regression analysis. SVM as a parametrically kernel-based method to deal with supervised classification problems, is proven to be successful. Its theory and method are stated detailed in [2, 5–7, 20, 41]. Single kernel SVM has been widely used in data analysis domains, including social media, mostly owing to its great classification ability [41]. SVM achieves high performance in text categorization since they accept high dimensional feature spaces and sparse feature vectors. Also text classification using SVMs is very robust to outliers and does not require any parameter tuning. We only give a brief introduction of SVM for a typical binary classification problem.

We are given a training dataset of n points of the form $G = (x_i,y_i), i=1, \ldots, n, x_i \in R^d$, where the $y_i \in \{+1, -1\}$, and y_i is a class label of x_i, each indicating the class to which the point x_i belongs. Each x_i is a d-dimensional real vector. We want to find the maximum-margin hyperplane that divides the group of points x_i with $y_i = +1$ from the group of points with $y_i = -1$, so that the distance between the hyperplane and the nearest point x_i from either group is maximized. The decision function of SVM is $g(x) =< w, \varphi(x) > +b$, where $\varphi(x)$ is a mapping of sample x_i from space to a high dimensional feature space. A decision hyperplane can be defined by an intercept term $b$and a decision hyperplane normal vector w which is perpendicular to the hyperplane. This vector is commonly referred to in the machine learning literature as the weight vector. $< \cdot, \cdot >$ denotes the dot product in the feature space. The optimal values of w and b can be obtained by solving the following regularized optimization problem:

$$\min \mathrm{J}(w, \varepsilon) = \frac{1}{2}\|w\|^2 + C\left(\sum_{i=1}^{n} \varepsilon_i\right) \tag{2}$$

$$\begin{aligned} \text{Subject to:} & y_i[> w, \varphi(x_i) > +b] \geqslant 1 - \varepsilon_i, i = 1, 2, ..., n \\ & \varepsilon_i \geqslant 0, \end{aligned} \tag{3}$$

where ε_i is the i-th slack variable and C is the regularization parameter. Instead of solving this optimization problem, we use the Lagrangian dual function to obtain a dual formula:

$$\max \ \mathrm{Q}(\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i,j=1}^{n} \alpha_i \alpha_j y_i y_j < x_i, x_j > \tag{4}$$

$$\begin{aligned} \text{Subject to } & \sum_{i=1}^{n} y_i \alpha_i = 0 \\ & \alpha_i \geqslant 0, i = 1, 2, ..., n, \end{aligned} \tag{5}$$

where α_i is a Lagrange multiplier which corresponds to the sample x_i. And if α^* is the optimal solution, then we get:

$$w^* = \sum_{i=1}^{n} \alpha^* y \alpha_i. \tag{6}$$

The weight coefficient vector of the optimal hyperplane is a linear combination of the training sample vectors. This is a two order function extremum problem with inequality constraints, and there is a unique solution. The process of optimization is actually making the classification margin maximum.

The Karush–Kuhn–Tucker (KKT) conditions are satisfied at the solution of any constrained optimization problem (convex or not), with any kind of constraints, provided that the intersection of the set of feasible directions with the set of descent directions coincides with the intersection of the set of feasible directions for linearized constraints with the set of descent directions. This rather technical regularity assumption holds for all support vector machines, since the constraints are always linear. Furthermore, the problem for SVMs is convex (a convex objective function, with constraints which give a convex feasible region), and for convex problems (if the regularity condition holds), the KKT conditions are necessary and sufficient for w, b and α to be a solution [21]. Thus solving the SVM problem is equivalent to finding a solution to the KKT conditions. This fact results in several approaches to find the solution [42].

There is only a part of the nonzero solution. The corresponding samples of these solutions are the support vectors. The optimal classification function is obtained after solving the above problem:

$$f(x) = \mathrm{sgn}\{< w^*, x > +b^*\} = \mathrm{sgn}\left\{\sum_{i=1}^{n} \alpha_i^* y_i < x_i, x > +b^*\right\}. \tag{7}$$

In the case of inseparability, some training samples cannot meet these conditions. A relaxation parameter is introduced in this condition.

For non-linear tasks, we can transform them into high-dimensional space by a kernel function, where a linear classification hyperplane exists.

The dual problem, whether it is optimal or classification function involving only the training samples between the inner product operations $< x, x_i >$.

The kernel function is defined as follows:

$$K(x_i, x_j) = \varphi(x_i) \cdot \varphi(x_j). \tag{8}$$

x_i, x_j are input vectors, φ is a map to transform source data from input space to feature space. In (8), a scalar value obtained from kernel function is equal to the dot product of transformed vectors in feature space. That means kernel function can determine the distance or describe the similarity of the two feature vectors of high dimensions to some degree.

The resulting algorithm is formally similar except that the dot product is replaced by a nonlinear kernel function.

And the objective function (4) becomes:

$$Q(\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i,j=11}^{n} \alpha_i \alpha_j y_i y_j K(x_i, x_j). \tag{9}$$

And the corresponding classification function (7) becomes:

$$f(x) = \mathrm{sgn}\left\{\sum_{i=1}^{n} \alpha_i^* y_i K(x_i, x_j) + b^*\right\}. \tag{10}$$

5.2 Multi-kernel SVM

In different problem domains, the kernel function should have different forms and parameters. At present, the selection of the kernel function and its parameters is based on experience, with a certain randomness. How to select the appropriate kernel function according to the actual data model in the SVM algorithm is very important problem. In this paper, we put forward multi-kernel SVM methods to adaptively select the optimal kernel for different features to get the best experimental results.

Kernel functions must be satisfied the conditions of continuous, symmetric, and most preferably, and should have a positive (semi-) definite Gram matrix for the convergence of iterative method used in solving the constrained optimization problem. Kernels need to satisfy the Mercer's theorem, sequentially, they are called positive (semi-) definite matrices, meaning their kernel matrices have only non-negative values. The use of a positive definite kernel insures that the optimization problem will be convex and solution will be unique. There are different forms of kernel functions, such as linear, polynomial, Gaussian, Laplacian, exponential, ANOVA and sigmoid kernel function etc.

Choosing the most appropriate kernel highly depends on the problem at hand because it depends on what we are trying to model. There are three kinds of kernel functions widely used which are linear, polynomial and Gaussian kernel functions.

1. Linear kernel function

$$K_{lin}(x_i, x_j) = < x_i, x_j > +c. \tag{11}$$

The linear kernel which is the simplest kernel function is given by the inner product $< x_i, x_j >$ plus an optional constant c. Kernel algorithms using a linear kernel are often equivalent to their non-kernel counterparts.

2. Polynomial kernel function

$$K_{pol}(x_i, x_j) = [< x_i, x_j > +c]^d. \tag{12}$$

The polynomial kernel is a non-stationary kernel and is well suited for problems where all the training data are normalized. Adjustable parameters are the constant term c and the polynomial degree d.

3. Gaussian kernel function

$$K_{Gau}(x_i, x_j) = \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\sigma^2}\right). \tag{13}$$

The Gaussian kernel is an example of radial basis function (RBF) kernel. $\left\|x_i - x_j\right\|^2$ may be recognized as the squared Euclidean distance between the two feature vectors. σ is a free parameter.

An equivalent, but simpler, it could also be implemented by defining a parameter $\gamma = \frac{1}{2\sigma^2}$.

$$K_{RBF}(x_i, x_j) = \exp(-\gamma \left\|x_i - x_j\right\|^2). \tag{14}$$

The adjustable parameter sigma γ plays a major role in the performance of the kernel, and should be carefully tuned to the problem at hand.

The choice of the kernel function is very important since it implicitly defines the metric properties of the feature space. Different features spaces obtained through different kernel functions have different classification performances. The polynomial kernel is a typical global kernel function and allows us to model feature conjunctions up to the order of the polynomial. Radial basis kernel functions allow to pick out hyperspheres in contrast with the linear kernel, which allows only to pick out hyperplanes.

Multi-kernel SVM is a powerful machine learning tool which allows automatically to obtain suitable combinations of several kernels over several features (therefore selecting the features as well). Multi-kernel SVM is introduced to learn with multi-modality attributes or adaptively select the optimal kernel if multiple candidates exist [14, 45]. Multi-kernel SVM adequately utilizes particular characteristic of each source and provides more possibility to choose suitable kernels or their weighted combination especially for data from multiple heterogeneous sources [35].

Multi-kernel SVM aims to construct a kernel model where the kernel is a linear combination of fixed basic kernels. Learning the kernel then consists of learning the weighting coefficients for each base kernel, rather than optimizing the kernel parameters of a single kernel. A multi-kernel function can be written as the linear combination of basic kernels:

$$K(x_i, x_j) = \sum_d \beta_d K_d(x_i, x_j), \beta_d \geqslant 0 \sum_d^{\beta_d} = 1, \tag{15}$$

where β_d is the weight on the corresponding basic kernel $K_d(x_i, x_j)$. Usually, heterogeneous kernels are selected as basic kernels. A basic kernel is computed for each kinds of data. Multi-kernel method is used to determine the weights for the basic kernels and then is used to select the relevant features with non-zero weights of the corresponding basic kernels.

The decision function in the classification is written as:

$$f(x) = \operatorname{sgn}\left\{ \sum_{i=1}^{n} \alpha_i^* y_i \sum_d \beta_d K_d(x_i, x_j) + b^* \right\}. \tag{16}$$

Basically, in the classification problems, the kernel type can be different or the same (distinguished by some parameters with different values in this case). And each kind of feature has a relevant kernel to obtain a corresponding result. Fusion of all results leads to the final one. Hence, multi-kernel SVM will well broaden the kinds of kernel used and develop the result estimation.

In our application, the text feature has 1381 dimensions which is sparse, however, the behavior features and the profile features only have 16 dimensions, thus the best kernel

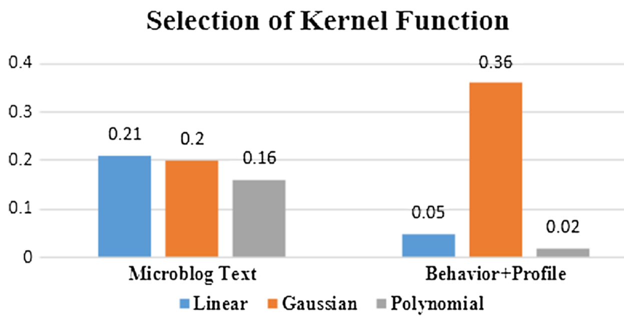

Fig. 4 Selection of kernel function

Table 6 Confusion matrix

	Positive (actual) (P = TP + FN)	Negative (actual) (N = TN + FP)
Test outcome		
Positive	True positive (TP)	False positive (FP)
Negative	False negative (FN)	True negative (TN)

functions for these types of features should be different. So we exploit multi-kernel SVM to automatically combine a set of features about the classification task. What's more, our dataset which includes 135 depressive users and 252 non-depressed users is small sample, SVM can get much better results than other algorithms in small sample dataset.

Based on the analysis above, in order to get the best mapping and the best classification results, we use multi-kernel SVM to choose the most appropriate kernel function according to different characteristics in our work.

6 Experimental analysis and comparison

After constructing emotional dictionary, we use TF method to statistic the frequency of words or phrases as microblog text features. And we get the dataset for experiment by combining the other two features in this paper. After that, we divide the dataset into two parts, one is microblog text data, another is user profile and behavior data. We employ three base kernels, linear kernel, polynomial kernel and Gaussian kernel for each part. The MK-SVM method is adopted to learn SVM classifier [23, 37], and then get each kernel coefficient from a linear combination of fixed basic kernels.

The penalty parameter C is used to determine the level of confidence interval and experimental risk in a given feature space. The value of C equals to 10 in this study. Sequential Minimal Optimization(SMO) is used as an optimizer parameter for each SVM kernel. Figure 4 shows the selection of kernel function which mainly include the value of Beta in Eqs. (15) and (16). The summation of Beta in microblog text data and the behavior and profile data is 0.57 and 0.43, respectively. the values of Beta for the linear kernel, Gaussian kernel and polynomial kernel are 0.21, 0.20 and 0.16, respectively in microblog text dataset. These three kinds of kernels are about the same. But the value of Beta for Gaussian kernel (0.36) is more than linear kernel (0.05) and polynomial kernel (0.02). As a result, Gaussian kernel plays an important role than the other two kernels in the behavior and profile dataset.

Normally, the larger portion of the data is used for training, and a smaller fraction of the data is used for testing. We adopt fivefold cross validations for these experiments. Accordingly, we divide the collected data for each class into five equal sized folds, and use fourfolds of the labeled data for training and onefold for testing. Then we get the classification model from training data using selected classification algorithms.

As shown in Table 6, we use the confusion matrix to represent the classification results. The accuracy of classifiers is measured based on precision and recall, and the F1-measure is also calculated based on the weighted harmonic means of precision and recall.

$$\text{Precision} = \frac{TP}{TP + FP} \tag{17}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{18}$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{19}$$

$$\text{F1 measure} = \frac{2 \times \text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}}. \tag{20}$$

Finally, we run various classification algorithms (Naive Bayes, KNN, Decision Tree and libD3C) using WEKA and compare the results with single and multiply kernel SVM method. As shown in Tables 7, 8, 9 and 10, the precision, recall, F1-measure and accuracy of Naïve Bayes, Decision Tree, KNN, libD3C, single-kernel SVM and multi-kernel SVM are listed by using different kinds of features, based on fivefold cross validation.

The precision is 75.56% under the conditions of using multi-kernel SVM method based on all-features, and under the same conditions we get the F-measure (76.12%) and the recall (76.69%). We obtain better precision, recall and F-measure than other methods from these experiment results, and it is also better in all-features than part features except the recall which is lower in all-features than the microblog text features. Compared to Naïve Bayes, Decision Tree, KNN and libD3C, the precision and recall of multi-kernel SVM method based on all-features are

Table 7 Precision of Naïve Bayes, KNN, decision tree, libD3C, single-kernel SVM and multi-kernel SVM using different features

Features	Naïve Bayes (%)	KNN (%)	Decision Tree (%)	LibD3C (%)	Single-kernel SVM (%)	Multi-kernel SVM (%)
User profile + user behavior	61.6	61.9	69.3	67.4	71.85	62.22
Microblog text	62.8	63.1	66.2	76.4	61.48	62.96
All-features	63.5	62	69.1	74.4	67.41	75.56

Table 8 Recall of Naïve Bayes, KNN, decision tree, libD3C, single-kernel SVM and multi-kernel SVM using different features

Features	Naïve Bayes (%)	KNN (%)	Decision Tree (%)	LibD3C (%)	Single-kernel SVM (%)	Multi-kernel SVM (%)
User profile + user behavior	68.5	57.8	68.5	71.9	67.36	67.2
Microblog text	52.6	39.3	66.7	71.9	71.55	80.19
All-features	54.1	42.2	71.1	71.1	71.1	76.69

Table 9 F1-measure of Naïve Bayes, KNN, decision tree, libD3C, single-kernel SVM and multi-kernel SVM using different features

Features	Naïve Bayes (%)	KNN (%)	Decision Tree (%)	LibD3C (%)	Single-kernel SVM (%)	Multi-kernel SVM (%)
User profile + user behavior	69.1	59.8	73.6	69.5	69.53	64.62
Microblog text	57.3	48.4	66.4	74.0	66.14	70.54
All-features	58.4	50.2	70.1	72.7	69.2	76.12

Table 10 Accuracy of Naïve Bayes, KNN, decision tree, libD3C, single-kernel SVM and multi-kernel SVM using different features

Features	Naïve Bayes (%)	KNN (%)	Decision Tree (%)	LibD3C (%)	Single-kernel SVM (%)	Multi-kernel SVM (%)
User profile + user behavior	75.45	72.87	80.36	78.04	77.93	76.23
Microblog text	72.61	70.8	76.49	82.43	78.04	81.65
All-features	73.13	70.8	78.81	81.39	79.07	83.46

Table 11 Performance comparison of error rate

Features	Naïve Bayes (%)	KNN (%)	Decision tree (%)	LibD3C (%)	Single-kernel SVM (%)	Multi-kernel SVM (%)
Error rate	26.87	29.2	21.19	18.61	20.93	16.54
Error reduction	38	43	22	11	21	N/A

improved, and it also is found that the recall improves much more than the improvement of the precision.

Meanwhile the multi-kernel SVM with all-features can achieve better accuracy of 83.46% than the other methods, and also get better accuracy than part features. We also find that the accuracy in Naïve Bayes, KNN, Decision tree and libD3C methods with user profile and behavior features is better than that in microblog text features. But the accuracy in single-kernel SVM and multi-kernel SVM methods with microblog text features is better than that in user profile and behavior features. This shows that SVM method can achieve better performance for text classification, and multi-kernel SVM method has better classification performance for multi-source and heterogeneous data in comparison with single-kernel SVM method.

As shown in Table 11, by comparing with other methods, including Naive Bayes, Decision Trees, KNN, libD3C and single-kernel SVM, it is found that the error rate of multi-kernel SVM method for identifying the depressed people was reduced respectively from 26.87, 29.2, 21.19, 18.61 and 20.93–16.54% representing a 38, 43, 22, 11 and 21% reduction in error rate.

In conclusion, the multi-kernel method can greatly reduce situations in which sample data are misclassified to other classes, and can get the best classification results.

7 Conclusions and future work

Nowadays people with depression are not diagnosed or misdiagnosed in the world, especially in China. With the development of Internet, social media provides a new approach to find out the potential depressed people. To this end, we propose a depression recognition model based on multi-kernel SVM. First, we analyze the characteristics of depressed people, and then construct the framework of depression recognition. Second, according to the peculiarity of social media language, we built a special emotional dictionary to extract microblog text features in accordance. Third, considering the heterogeneity between word frequency and another two features, we employ multi-kernel SVM methods to adaptively select the optimal kernel for different features. At last, we compared the performance of multi-kernel SVM with other several machine learning algorithms such as Naive Bayes, KNN, Decision Tree, libD3C and SVM for classifying the emotion of postings. The error reduction rate is 38, 43, 22, 11 and 21% compared with Naive Bayes, KNN decision trees, libD3C and single-kernel SVM, respectively. The error rate of multi-kernel SVM method for identifying the depressed people is reduced to 16.54%.

In additional, because there is not public dataset for depression recognition, the datasets used by researchers are very different in some aspects such as different language, features, samples and unbalance of the positive and negative samples. But compared the experimental results with other studies, our methods are very effective for depression recognition in the whole. Wang et al. and Banitaan et al. adopted Bayes, Decision trees and rules-based methods to acquire the classification model which implemented using WEKA, and then got the best accuracy of 81.4 and 83.3%, respectively. De Choudhury et al. employed SVM classifier with a RBF kernel and got the best accuracy of 74.57%. To my knowledge, it is the first paper for depression recognition using multi-kernel SVM and we get better accuracy of 83.46%. Actually although we don't know whether the data is linear or manifold pattern, multi-kernel SVM method can automatically combine a set of features about the classification task and achieve better experimental results.

We judge the user's mental status by all the sending postings for some time. But we do not consider the level of depression and dynamic status varied with the time. These should be done in the future work. In addition, we intend to integrate other pieces of information such as sleep data, exercise and physical activities, as well as food information.

Acknowledgements This work is partly supported by National Program on Key Basic Research Project under Grant 2013CB329304, National Natural Science Foundation of China under Grant 61222210 and New Century Excellent Talents in University under Grant NCET-12-0399.

References

1. Banitaan S, Daimi K (2014) Using data mining to predict possible future depression cases. Int J Public Health Sci 3(4):231–240
2. Burges CJ (1998) A tutorial on support vector machines for pattern recognition. Data Min Knowl Discov 2(2):121–167
3. Burns MN, Begale M, Duffecy J, Gergle D, Karr CJ, Giangrande E, Mohr DC (2011) Harnessing context sensing to develop a mobile intervention for depression. J Med Internet Res 13(3):e55
4. Caruana R, Niculescu-Mizil A (2006) An empirical comparison of supervised learning algorithms. In: Proceedings of ICML, pp 161–168
5. Chapelle O, Vapnik V (1999) Model selection for support vector machines. In: Proceedings of NIPS, pp 230–236
6. Chapelle O, Vapnik V, Bousquet O, Mukherjee S (2002) Choosing multiple parameters for support vector machines. Mach Learn 46(1–3):131–159
7. Cortes C, Vapnik V (1995) Support-vector networks. Machine Learn 20(3):273–297
8. Dang J, Li A, Erickson D, Suemitsu A, Akagi M, Sakuraba K (2010) Comparison of emotion perception among different cultures. Acoust Sci Technol 31(6):394–402
9. Dao B, Nguyen T, Phung D, Venkatesh S (2014) Effect of mood, social connectivity and age in online depression community via topic and linguistic analysis. In: Proceedings of International Conference on Web Information Systems Engineering, pp 398–407
10. Darwin C (1872) The expression of the emotions in man and animals, 1st edn. John Murray, London
11. De Choudhury M, Counts S, Horvitz E (2013) Social media as a measurement tool of depression in populations. In: Proceedings of the 5th Annual ACM Web Science Conference, pp 47–56
12. De Choudhury M, Counts S, Horvitz EJ, Hoff A (2014) Characterizing and predicting postpartum depression from shared Facebook data. In: Proceedings of the 17th ACM conference on Computer supported cooperative work and social computing, pp 626–638
13. De Choudhury M, Gamon M, Counts S, Horvitz E (2013) Predicting depression via social media. In: Proceedings of the ICWSM, pp 128–137
14. Do H, Kalousis A, Woznica A, Hilario M (2009) Margin and radius based multiple kernel learning. In: Joint European Conference on Machine Learning and Knowledge Discovery in Databases, pp 330–343
15. Dong Z, Dong Q (2006) HowNet and the computation of meaning. World Scientific, Singapore, pp 85–95

16. Doryab A, Min JK, Wiese J, Zimmerman J, Hong JI (2014) Detection of behavior change in people with depression. In: Proceedings of AAAI Workshop on MAIHA, pp 12–16
17. Ekman P (1992) An argument for basic emotions. Cognit Emot 6(3–4):169–200
18. Ellison NB (2007) Social network sites: definition, history, and scholarship. J Comput Mediat Commun 13(1):210–230
19. Ellison NB, Steinfield C, Lampe C (2007) The benefits of Facebook "friends:" Social capital and college students' use of online social network sites. J Comput Mediat Comm 12(4):1143–1168
20. Filippone M, Camastra F, Masulli F, Rovetta S (2008) A survey of kernel and spectral methods for clustering. Pattern Recognit 41(1):176–190
21. Fletcher R (1987) Practical methods of optimization, 2nd edn. Wiley, New Jersey
22. Gilbert E, Karahalios K (2009) Predicting tie strength with social media. In: Proceedings of the SIGCHI conference on human factors in computing systems, pp 211–220
23. Gönen M, Alpaydın E (2011) Multiple kernel learning algorithms. J Mach Learn Res 12:2211–2268
24. Haimson OL, Ringland K, Simpson S, Wolf CT (2014) Using depression analytics to reduce stigma via social media: BlueFriends. iConference (Social Media Expo)
25. Heidi L (2014) Medical research: if depression were cancer. Nature 515(7526):182–184
26. Kerri S (2014) Mental health: a world of depression. Nature 515(7526):180–181
27. Ku L, Chen H (2007) Mining opinions from the Web: Beyond relevance retrieval. J Am Soc Inf Sci Technol 58(12):1838–1850
28. Lin C, Chen W, Qiu C, Wu Y, Krishnan S, Zou Q (2014) LibD3C: ensemble classifiers with a clustering and dynamic selection strategy. Neurocomputing 123:424–435
29. Lin C, Huang Z, Yang F, Zou Q (2012) Identify content quality in online social networks. IET Commun 6(12):1618–1624
30. Neuman Y, Cohen Y, Assaf D, Kedma G (2012) Proactive screening for depression through metaphorical and automatic text analysis. Artif Intell Med 56(1):19–25
31. Nguyen T, Phung D, Dao B, Venkatesh S, Berk M (2014) Affective and content analysis of online depression communities. IEEE Trans Affect Comput 5(3):217–226
32. Park M, Cha C, Cha M (2012) Depressive moods of users portrayed in Twitter. In: Proceedings of the ACM SIGKDD Workshop on healthcare informatics, pp 1–8
33. Park M, McDonald DW, Cha M (2013) Perception differences between the depressed and non-depressed users in Twitter. In: Proceedings of ICWSM, pp 476–485
34. Park S, Lee SW, Kwak J, Cha M, Jeong B (2013) Activities on Facebook reveal the depressive state of users. J Med Internet Res 15(10):e217
35. Peng S, Hu Q, Chen Y, Dang J (2015) Improved support vector machine algorithm for heterogeneous data. Pattern Recognit 48(6):2072–2083
36. Pennebaker JW, Francis ME, Booth RJ (2007) Linguistic inquiry and word count (Computer Software). LIWC Inc
37. Rakotomamonjy A, Bach F, Canu S, Grandvalet Y (2007) More efficiency in multiple kernel learning. In: Proceedings of the 24th ICML, pp 775–782
38. Rakotomamonjy A, Bach F, Canu S, Grandvalet Y (2007) SimpleMKL. J Mach Learn Res 9:2491–2521
39. Salton G, McGill MJ (1983) Introduction to modern information retrieval. McGraw-Hill computer science series
40. Schoen H, Gayo-Avello D, Takis MP, Mustafaraj E, Strohmaier M, Gloor P (2013) The power of prediction with social media. Internet Res 23(5):528–543
41. Vapnik V (2013) The nature of statistical learning theory. Springer Science & Business Media, Berlin
42. Wang T, Rao J, Hu Q (2014) Supervised word sense disambiguation using semantic diffusion kernel. Eng Appl Artif Intell 27:167–174
43. Wang X, Zhang C, Ji Y, Sun L, Wu L, Bao Z (2013) A depression detection model based on sentiment analysis in micro-blog social network. In: Trends and Applications in Knowledge Discovery and Data Mining, pp 201–213
44. Wang X, Zhang C, Sun L (2013) An improved model for depression detection in micro-blog social network. In: Proceedings of the IEEE 13th ICDMW, pp 80–87
45. Wang Z, Chen S, Sun T (2008) MultiK-MHKS: a novel multiple kernel learning algorithm. IEEE Trans Pattern Anal Mach Intell 30(2):348–353
46. Wilson ML, Ali S, Valstar MF (2014) Finding information about mental health in microblogging platforms: a case study of depression. In: Proceedings of the 5th Information Interaction in Context Symposium, pp 8–17
47. Xu L, Lin H, Pan Y, Ren H, Chen J (2008) Constructing the affective lexicon ontology. J China Soc Sci Tech Inf 27(2):180–185
48. Yang B, Ollendick TH, Dong Q et al (1995) Only children and children with siblings in the People's Republic of China: levels of fear, anxiety, and depression. Child Dev 66(5):1301–1311
49. Zhang H, Yu H, Xiong D, Liu Q (2003) HHMM-based Chinese lexical analyzer ICTCLAS. In: Proceedings of the second SIGHAN workshop on Chinese language processing, pp 184–187
50. Zou Q, Zeng J, Cao L, Ji R (2016) A novel features ranking metric with application to scalable visual and bioinformatics data classification. Neurocomputing 173:346–354