2016 IEEE/WIC/ACM International Conference on Web Intelligence

Predicting Depression from Internet Behaviors by Time-frequency Features

Changye Zhu and Baobin Li
School of Computer and Control
University of Chinese Academy of Sciences
Beijing,100190, China.
Email: zhuchangye14@mails.ucas.ac.cn
libb@ucas.ac.cn

Ang Li
Department of psychology
Beijing Forestry University
Beijing, 100083, China
Email: angli@bjfu.edu.cn

Tingshao Zhu
Institute of Psychology
Chinese Academy of Sciences
Beijing, 100101, China.
Email:tszhu@psych.ac.cn

***Abstract*—Early detection of depression is important to improve human well-being. This paper proposes a new method to detect depression through time-frequency analysis of Internet behaviors. We recruited 728 postgraduate students and obtained their scores on a depression questionnaire (Zung Self-rating Depression Scale, SDS) and digital records of Internet behaviors. By time-frequency analysis, we built classification models for differentiating higher SDS group from lower group and prediction models for identifying mental status of depressed group more precisely. Experimental results show classification and prediction models work well, and time-frequency features are effective in capturing the changes of mental health status. Results of this paper might be useful to improve the performance of public mental health services.**

***Index Terms*—Internet Behaviors, Time-frequency Analysis, Depression Questionnaire,Self-rating Depression Scale.**

1. Introduction

Mental health is an essential component of the human health. Currently, like depression, mental health problems have been one of the leading causes for global burden of disease [1], [2]. Therefore, preventing mental health problems can be helpful to improve human well-being.

Early detection is a basis of the prevention of mental health problems [3]. However, because of several reasons such as lacking mental health knowledge, and stigmatizing attitudes towards mental patients, people with mental health problems are not motivated to seek professional help [4], [5]. More importantly, traditional methods for detecting individual mental health problems (e.g. self-report techniques and clinical diagnosis) cannot identify individual mental health status in real-time, which may lead to delayed reporting and can have negative impacts on personal mental health.

The development of Internet and information technology gives us an opportunity to find new method for detecting mental health problems. First, based on information technology, Internet behaviors can be collected and processed in a non-intrusive, accurate and efficient manner. Given that the relationship between Internet behaviors and psychological features (e.g. personality) has been confirmed in previous studies, which implies the possibility of detecting mental health problems through Internet behaviors analysis. Amichai-Hamburger and Ben-Artzi (2000) found that there exists relationship between Internet behaviors and personality [6]. Gosling et al. (2011) collected digital records of human behaviors on social media and proved the accuracy of predicting personality by perceiving Internet behaviors [7]. Furthermore, Kosinski and colleagues (2013) established computational models for predicting Facebook user's psychological profile and personal preference [8]. Wu et al. (2015) argued that computer-based personality judgments are more accurate than those made by humans [9].

In view of these advantages, some studies have been conducted to identify individual mental health problems based on Internet behaviors. Park et al. (2012) found that it is possible to identify one user's depressive emotion by analyzing his/her Twitter [10]. Besides, they found that Facebook activities can be used to distinguish depressed users from healthy ones [11]. X. Wang et al. (2013) analysed sentiments in Micro-blog social network to build up a depression detection model [12], and the model improved by linkage features reached an accuracy of 95% [13].

Recently, a few researches established depression detection models based on social media data [14], [15]. Results of these studies confirmed that Internet behaviors are sensitive to the variation of mental health status. However, these studies were limited to data on one single website (e.g. Facebook and Twitter), which only represented a small part of individual Internet behaviors. On the other hand, Internet behaviors are time series in essentially, but these studies ignored this important feature–time.

Our goal is to recognize depressed users from their web-behaviors. If this goal can be realized, the depressed persons can get effective treatments earlier. To reach this goal, we build up several models by the aid of the time-frequency analysis method to predict whether an user is depressed and the depressed ones' extent of depression. In this paper, we mainly discuss and explore the features related to mental status obtained in the time and frequency domain, respectively.

The whole procedure of our research mainly includes

978-1-5090-4470-2/16 $31.00 © 2016 IEEE
DOI 10.1109/WI.2016.59

four stages: collecting data, preprocessing data, extracting features and building computational models. In the first stage, digital records of Internet behaviors from the web server of LAN gateway are collected, which includes individual browsing histories on different web sites. All these records are of signals varying with the time, so in the stage of extracting features, we employ Fourier transform to analysis collected data, and obtain features in the time and frequency domain, which are closely related web users' mental status. Signals or data can be understood better by studying the time and frequency jointly [18]. Some properties can be present clearly in the frequency domain while we may find them in the domain of time scarcely. Finally, these important features will be used to build classification models and prediction models in the forth stage. Classification models are built for differentiating between the mentally healthy group and the depressed group. Prediction models are built to evaluate mental status of depressed persons more precisely.

The remainder of this paper is organized as follows. Section 2 introduces the process of experiment for collecting data, and the data are also described in this section. Methods for analyzing data are described in Section 3, including ways of data preprocessing and feature extraction with DFT and clustering. Section 4 describes the procedure of building model and results of depression recognition with these models. The experimental results show that our computational models work well, and time-frequency features are very important for recognising the mental status of depression, which can also improve the performance of models distinctly. This paper ends with conclusions in Section 5.

2. Database

We have recruited 728 postgraduate students in the University of Chinese Academy of Sciences (UCAS) to participate this research from 2012-4-1 to 2012-6-30. In this period all participants have completed a depression questionnaire – Zung Self-rating Depression Scale (SDS), and allowed us to download their digital records of Internet behaviors.

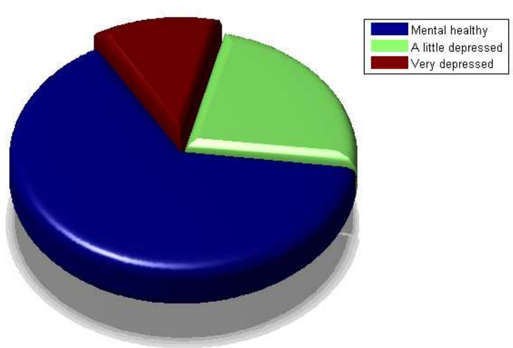


Figure 1. Percentage of users' mental states. The blue part denotes mental healthy persons(about 61.7%). The green and red part are for a little and very depress persons, which occupy 34.1% and 4.2, %respectively

TABLE 1. AVERAGE AND STANDARD DEVIATION OF THREE GROUP USERS' SCORES

Group	Mental healthy ($SDS < 40$)	A little depressed ($40 \leq SDS \leq 60$)	Very depressed ($SDS > 60$)
Number of people	449	248	31
Average SDS score	29.7	47.6	65.6
SD. of SDS score	5.4	5.8	3.9

SDS is one of the most commonly used tools designed for measuring the presence and severity of depressive disorder in patients [16]. It consists of 20 self-rating items, such as "I feel down-hearted and blue". Participants rated themselves on each item by a 4-point Likert-type scale (1 = A little of the time to 4 = Most of the time). A SDS score is computed by summing up scores of all 20 items. People at a high risk of depressive disorder are those scoring higher than 50. A Chinese version of SDS was used in this study [17]. Figure 1 and Table 1 show the distribution of all users' SDS scores. In our database, mental healthy persons account for 61.7% of all participants (the blue part in Figure 1), and slightly depressed and seriously depressed persons occupy 34.1% and 4.2% of them, which are denoted by the green and red parts in Figure 1. The majority of people are healthy, so our object is to distinguish the depressed ones from them.

To surf the Internet in UCAS, everyone needs to log in the LAN gateway beforehand, so users browsing histories and Internet behaviors are all recorded by the gateway server. From 2012-4-1 to 2012-6-30, we had collected data recorded in the web server of UCAS LAN gateway (see Figure 2). The obtained information includes mainly five aspects: browsing webs, usage of search engine, sending and receiving emails, chatting online and posting on forums. And the raw data are in the form of web logs, in which every browsing history is a record.

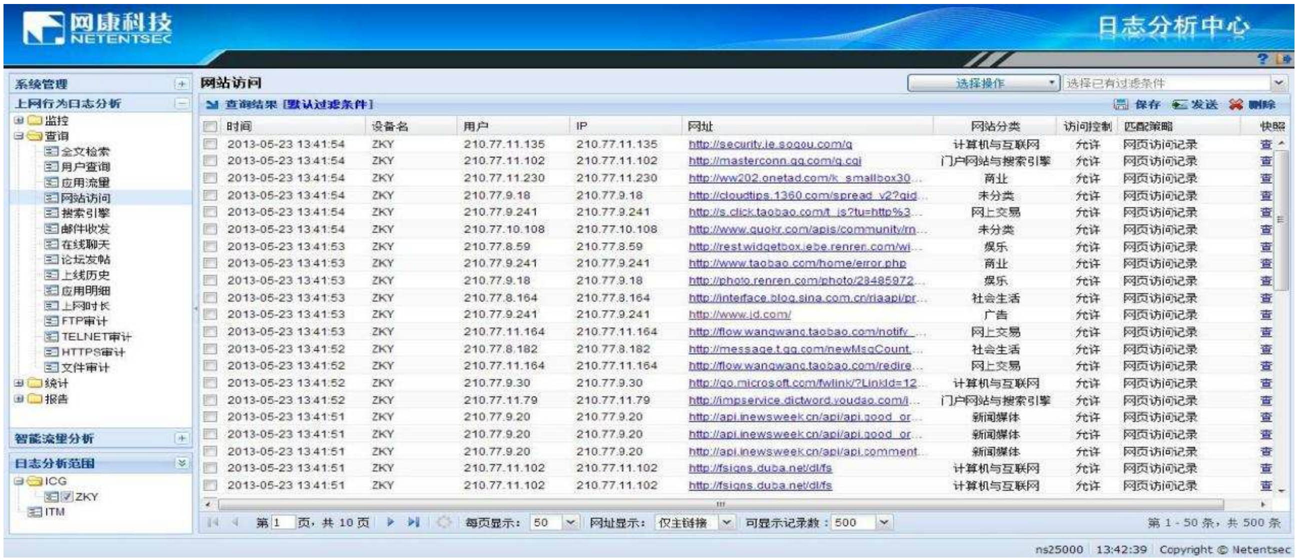
From these data, we acquired 19 different attributes of Internet behaviors which have been shown in Table 2, including statistics for all words, pronouns and punctuations. Each entry comprises several attributes. And a 90*19 matrix was recorded during 90 days for every person.

3. Methods

Figure 3 shows details of three stages: preprocessing data, feature extraction and building models, which are denoted by the red, green and blue blocks, respectively. The data preprocessing consists of selecting attributes, Z-score normalization and reducing dimension. The feature extraction is conducted with the discrete fourier transform (DFT) and K-means clustering, and classification models and prediction models are built up with three methods respectively. Let us begin with preprocessing row data.

3.1. Preprocessing data

The original raw data may be noisy, high-dimensional and redundant. To make them noiseless and simple, we



(a) Program interface

ipspy-2010-03-**.log

Sat Mar ** 11:20:06 20092801051**** 2 210.77.6.127 220.181.38.89 TCP 80

Sat Mar ** 11:20:12 20092801051**** 2 210.77.6.127 219.133.60.3 TCP 80

...

(b) Web logs

Figure 2. The example of records in the LAN gateway. The information we can get from a web user includes mainly five aspects: browsing webs, usage of search engine, sending and receiving emails, chatting online and posting on forums. The columns in the interface from left to right are time of recording, names of users' computers, users' IP addresses, classes of web pages, authority of access, etc. Each row in the interface is an access record, whose form is a web log in (b).

TABLE 2. DETAILS OF 19 USEFUL ATTRIBUTES OF INTERNET BEHAVIORS

Attribute type	Attribute name	Meaning
Words statistics	WC	The total number of words
	WPS	Words in each sentence
	Sixltr	Number of words of more than 6 characters
	Dic	Capture rate
	Numerals	Arabic numerals
	Funct	Function words
Pronoun statistics	Pronoun	Pronoun
	Ppron	Personal pronoun
	I	I
	We	We
	You	You
	Shehe	She or he
	They	They
Punctuation statistics	Dash	Dash
	Quote	Quote
	Apostro	Apostrophe
	Parenth	Parentheses
	OtherP	Other punctuation
	AllPct	All punctuation

need data preprocessing techniques to make them easy to be analyzed. Red blocks in Figure 3 describe the operations we take before extracting features. In this paper, the procedure of data preprocessing consists of three steps: selecting attributes, Z-score normalization and reducing dimension.

3.1.1. Selecting attributes. The raw data may be redundant, and this redundancy will bring more complicated for analyzing data. To reduce data redundancy, we only need to retain those closely correlated attributes. Here, we use the correlation coefficient to measure dependency . The correlation coefficient is an important method for measuring the linear correlation between attributes, ranging from -1 to 1. The greater the absolute value of a correlation coefficient is, the more correlated the attributes are. Positive values represent the positive correlation, while negative ones means the negative correlation. For those greater positive attributes, we need to delete some ones for reducing the redundancy of records. Among all attributes, correlation coefficients of 'Dic' and 'Funct', 'WC' and 'WPS' are more than 0.95, which means they are highly positively correlated, so 'Dic' and 'WC' are removed.

3.1.2. Z-score normalization. From Table 2, we find that there many types of attributes, and different attributes of Internet behaviors data have different scales. If their units of measure are different, their influences on recognition results would be different. So we need to make their influence alike and take the method of Z-score normalization.

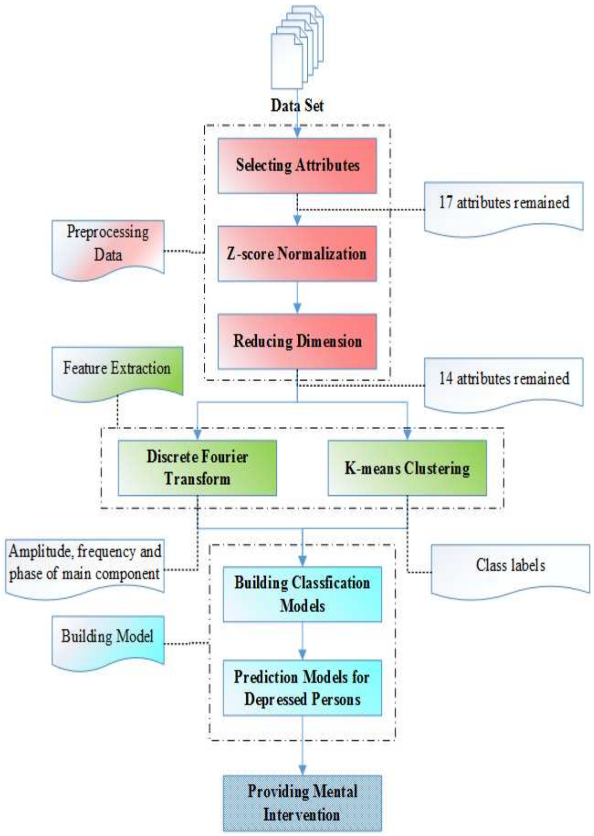


Figure 3. Data flow diagram. Our methods includes three stages: Data preprocessing, feature extraction and model building. Data preprocessing consists of selecting attributes, Z-score normalization and reducing dimension; feature extraction is conducted with Discrete Fourier Transform (DFT) and K-means clustering; classification models and prediction models are built up with three methods respectively.

Z-score normalization is also known as standard score normalization, whose formula is expressed as follows

$$z = \frac{x-\mu}{\sigma},$$

where x is a variable, μ stands for the average value of x and σ stands for standard deviation of x. z represents the distance between x and μ, measured by standard deviation σ. x is lower than μ when z is negative. All attributes of each person are normalized by Z-score.

3.1.3. Reducing Dimension. After two steps above, the number of attributes is seventeen. To compress implicitly related attributes further, the principal component analysis (PCA) is used in this paper. PCA is a kind of linear transformation which transforms multiple variables to some important variables. It can retain original information and reduce the number of variables. The number of attributes to be retained depends on the eigenvalues of PCA and determines the reconstruction error.

In this paper, attributes of pronoun and punctuation are reduced by the method of PCA. According to the eigenvalues, the top five attributes of the result of PCA remain, making the reconstruction error less than 5%. To make it clear, Table 3 shows the eigenvalues of PCA on pronoun and punctuation attributes respectively. In this table, eigenvalues drop quickly at last and three attributes in red could be ignored.

TABLE 3. PCA EIGENVALUES FOR ATTRIBUTES ABOUT PRONOUN AND PUNCTUATION

Components	1	2	3	4	5	6	7
Eigenvalues of Pronoun	3.06	1.03	0.968	0.957	0.823	0.156	0.011
Eigenvalues of Punctuation	2.10	1.02	1.00	0.985	0.747	0.145	

3.2. Extracting features

After preprocessing data, what we should do is to select and extract important features closely related to subjects' mental status. These features are the key factors for identifying web users' status. Two methods will be used to extract features in this stage - Discrete Fourier Transform (DFT) and K-means clustering. DFT helps us get the initial time-frequency features, while the method of clustering clearly shows us important and helpful ones which are more closely related mental status.

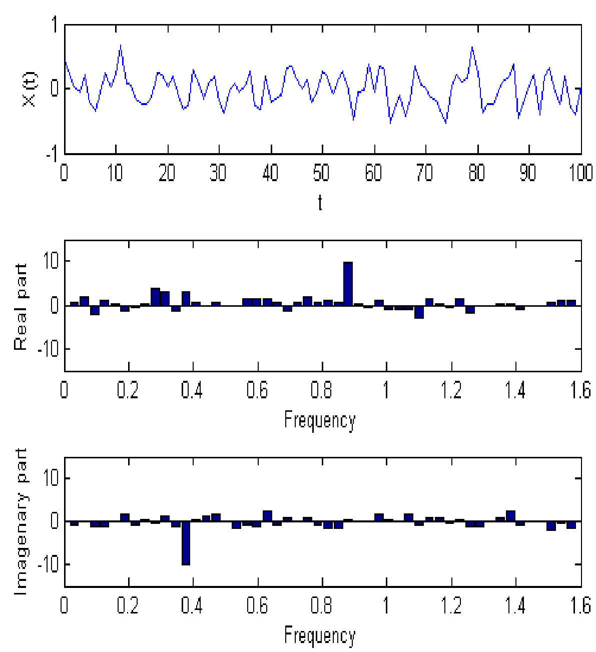


Figure 4. An example of DFT. The first graph is $x(t) = sin(0.88t) + cos(0.37t)$ mixed by Gaussian noise. Next two graphs are results of DFT. Its major component signal and noise are very clear in transform results.

3.2.1. Discrete Fourier Transform (DFT). Fourier transform [18] is one of the most useful methods for mathematical physics and engineering. It has extensive applications for signal processing, including filtering, correlation, coding, synthesis and feature extraction for pattern identification. Discrete Fourier transform is the discrete form of the Fourier transform, which transforms the signal sampling of time domain into the sampling of frequency domain.

In the discrete Fourier transform, a series of N points $\{x[n]_{0 \leq n < N}\}$ is intuitively decomposed into a linear combination of cosinoidal vibrations as follows

$$x(t) = \sum_{n=1}^{N/2} X(n) cos(\frac{2\pi n}{N} t + \varphi_n),$$

where N is length of the series, $X(n)$ is the amplitude of the nth component and φ_n is the phase of it.

The DFT of the series is formally defined as follows

$$\hat{x}[k] = \sum_{n=0}^{N-1} e^{-\frac{2\pi i}{N}} x[n], \quad k = 0, 1, \ldots, N-1,$$

where e stands for the base of natural logarithms and i is the imaginary unit. In fact, DFT can be thought of as a transformation on periodic signals derived by periodic extension. In practice, the fast Fourier transform (shorted as FFT) is used to calculate DFT efficiently.

Moreover, by the aid of the discrete Fourier transform, some properties can be present clearly in the frequency domain while we may find them in the domain of time scarcely. As the example in Figure 4, the original signal is $x(t) = sin(0.88t) + cos(0.37t)$ mixed by Gaussian noise. After DFT, we can get the main frequency of a time series X(t) at 0.88 and 0.37 easily and distinguish noise from it.

This paper performs DFT on each column of every 90*14 matrix and retains the first 22 components of transformation results. The features extracted from transformation results are frequency (substituted by n here), amplitude $X(n)$ and phase φ_n of the component with the largest amplitude. For example, Table 4 shows features derived from DFT on the attribute 'WPS'.

TABLE 4. SOME FEATURES EXTRACTED BY DFT FOR 'WPS'

User number	Frequency n	Amplitude $X(n)$	Phase φ_n
1	41	6.10	-2.56
2	19	15.47	-1.64
3	13	11.45	-1.72
...	...	...	...

3.2.2. Clustering. An attribute is a vector consisting of 90 elements. If one plots a vector as a curve, the curve can show its rise and fall, maximum and minimum easily. Clustering is utilized to find similarity and diversity among all users. It can help us distinguish curves with different fluctuations. The K-means algorithm will be used as the clustering methods in this paper because of its simplicity and efficiency.

The standard algorithm for clustering was first proposed by Stuart Lloyd in 1957 as a technique for pulse-code modulation, though it wasn't published outside of Bell Labs until 1982 [19]. Clustering is a process dividing all objects into multiple classes of similar objects. The objects in the same class are alike, while objects from different classes are unlike. In the K-means algorithm, a data set of N samples is partitioned into $K(K < N)$ classes with the following two conditions:

1. Each class contains at least one sample.
2. Each sample only belongs to one class.

For a given K, the algorithm gets an initial partition, and changes the classes of samples by iteration later on, making the classes better than before.

Using K-means, one should set the number of clusters K at the beginning. The preferable K comes from trial and error. Firstly, we set a relatively large K, such as 20. If there were some clusters which contain few samples, K would be decreased until all clusters have enough samples (more than 20 samples for instance).

In our experiment, each attribute is clustered into 16 clusters firstly. The clustering result comprises only three clusters including enough samples, so the number of clusters is adjusted to three. The curves of samples in the same cluster are similar and vice versa. Figure 5 shows curves of three clusters of attribute 'WPS', where curves in the same row are of the same cluster. The features extracted from clustering are class labels of attributes. They represent fluctuations of attribute curves, similarity and diversity among all users.

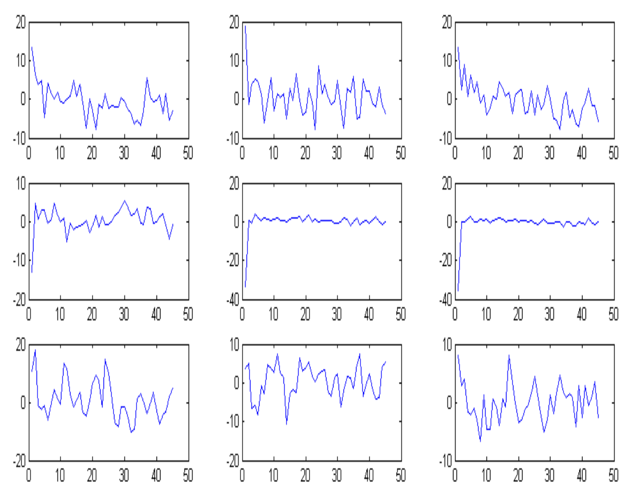


Figure 5. Curves of 3 clusters for the attribute 'WPS'. Curves in the same row are from the same cluster. By comparison, curves from the same class are similar and curves from different classes are dissimilar.

4. Results

Time-frequency features extracted in the above section will be used to establish the computational model to predict uses' depressed status. Firstly, users are classified into two groups: one is mentally healthy and the other is depressed. Then, classification models are built up to classify these two groups with methods of the naive bayes([20]), BP neural network([21]) and C4.5 decision tree([22]). Finally, for depressed group, prediction models are built up to predict their depressed status more precisely by means of logistic regression([20]), BP neural network, and CART decision tree([22]). We will sketch out these methods in the following.

4.1. Modeling Approaches

The naive bayes is a classifier constructed by estimating probabilities of classes and conditional probabilities of attributes. It is robust to irrelevant features, but if features are not conditionally independent or are not Gaussian distributed, its accuracy will decrease. The BP neural network is a kind of an artificial neural network trained by the back-propagation algorithm. It can match categorical or continuous outputs, so it is used to build up classification or prediction models.

A decision tree is a flowchart-like structure in which each internal node represents a ”test” on an attribute (e.g. whether a coin flip comes up heads or tails), each branch represents the outcome of the test and each leaf node represents a class label. The paths from roots to leaves represent classification rules. It is trained by measuring decrease of entropy. Like the BP neural network, it can also be used to predict discrete or continuous values.

The logistic regression can be seen as an analogue to the linear regression. It assumes a standard logistic distribution of errors, and measures the relationship between the categorical dependent variable and independent variables by estimating probabilities using a logistic function, which is the cumulative logistic distribution.

In the next two subsections, these approaches will be used to building models to classify mentally healthy and unhealthy persons.

4.2. Classification Results

According to the discussion in the section 2, web users are divided into three groups – a healthy group, a slightly depressed one and seriously depressed one by their SDS scores. To identify whether a user is mental healthy or not, the healthy group is labeled as a healthy class while the slightly depressed group and seriously depressed group are combined into the unhealthy class or a depressed class. Numbers of samples corresponding to these two different classes are 449 and 279 respectively. In particular, the classification model are built up to classify these two group with methods of the naive bayes, BP neural network and C4.5 decision tree.

Features to be used in this model include 14 clustering features (class labels from K-means clustering) and 42 frequency features (amplitudes, frequencies and phases from DFT). The number of total features are 56. Here, for illustrating the significance of features in the time and frequency domain, we will use three different types feature in the classification model. They are

1. 14 clustering feature, i.e. features in the time domain;

TABLE 5. AVERAGE PRECISION (P) AND RECALL (R) OF DEPRESSED CLASS IN PERCENT

Features \ Methods		Naive Bayes	BP Neural Network	Decision Tree
Clustering Labels	P	76.8	74.3	75.0
	R	55.7	53.1	56.6
DFT Parameters	P	73.7	71.7	67.4
	R	53.1	50.4	57.2
All Features	P	75.6	73.4	72.3
	R	62.3	55.0	56.7

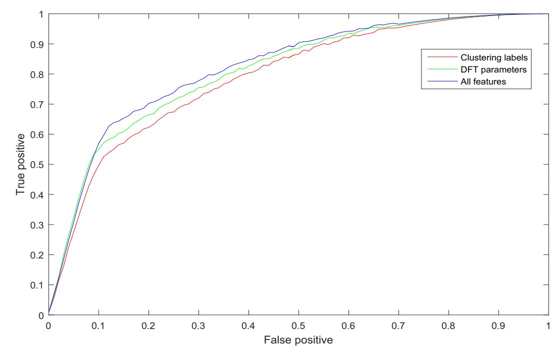


Figure 6. Receiver operating characteristic (ROC) curves of three naive bayes models built up with different types of features. The areas under the curve of three models are 0.79, 0.81, and 0.82, respectively.

2. 42 frequency features;

3. 56 features including all feature in the time and frequency domain.

Table 5 shows the average precision (shorted as ’P’in the table) and recall rate (shorted as ’R’in the table) of the depressed group in these models built with different features, which are calculated by the ten-fold cross-validation. And Figure 6 shows the receiver operating characteristic (ROC) curves of three naive bayes models built up with different types of features. The X-axis and Y-axis denote false positive rate (1-specificity) and true positive rate (sensitivity), respectively. If the area under the curve (AUC) is larger, the model is better. The AUC of three models are 0.79, 0.81, and 0.82, respectively.

According to Table 5, the best classification approach is the naive bayes. For clustering labels, the naive bayes’precision is the highest while the decision tree has the best recall rate. And models constructed with clustering labels performs better than those only with DFT features as a whole. For features combining cluster labels with DFT features, the naive bayes’ recall rates increase significantly by more than 6% and its F-measure (harmonic average of precision and recall) is the highest. That is to say, compared with clustering features and frequency features, time-frequency ones work very well.

For this classification, we should explain more in details as follows. Firstly, the samples of the healthy group are more than those of the depressed group. It leads to the classifiers are partial to the healthy group, so as to get higher precision of depression compared to its recall. Secondly, the naive bayes considers features to be independent, while neural network and decision tree treat the features as a whole. Maybe the features extracted by DFT and clustering are independent essentially, so the naive bayes can get better performance. At last, the combination of all features mainly improves the recall of depression, which means the combined features in the time and frequency domain are more suitable to reflect the nature of depression.

Because identifying depression is more important, the following subsection will introduce the procedure of building prediction models for depressed persons.

4.3. Prediction Results

The mental unhealthy group includes two types of persons–slightly depressed ones and seriously depressed ones. As mentioned in Section 2, both of their SDS scores are higher than 40 while seriously depressed web users scores higher than 50.

People scoring higher than 50 are at a high risk of depression, and they need to pay more attention to keep mentally healthy. To focus on depressed persons' mental status and predict more precisely, we build up prediction models whose outputs are continuous for the depressed class.

Three types of features to be used are the same as those employed in the above classification models. Prediction models are built up to predict depressed users' SDS scores with methods of the logistic regression, BP neural network, and CART decision tree. Table 6 shows the mean absolute error (MAE) in ten-fold cross-validation between predicted scores and real scores in these models built with different features.

TABLE 6. MEAN ABSOLUTE ERROR (MAE) BETWEEN PREDICTED SCORES AND REAL SCORES OF DEPRESSED CLASS

Features \ Methods	Logistic Regression	BP Neural Network	Decision Tree
Clustering Labels	4.23	4.17	7.68
DFT Parameters	4.10	3.85	9.74
All Features	4.33	3.76	7.55

The standard deviation of scores of depressed group is 7.96. The MAE of prediction results of logistic regression and BP neural network are much lower than this value. Obviously decision tree is not suitable for this task. Compared with other methods, BP neural network performs the best in this prediction task. Moreover, for three types of features, frequency features are better than clustering features, while a combination of them works best. As a whole, the BP neural network built up with all features performs well, producing the lowest MAE of 3.76.

Both logistic regression and BP neural network can be considered as functions which output continuous values, and their training process is to adjust corresponding parameters to training samples. For decision tree, it is more difficult to produce continuous values, and needs more training samples. So, both logistic regression and BP neural network outperform decision tree. And the function of BP neural network fits the depression samples better than that of logistic regression.

5. Conclusions

In this paper, we mainly discussed some methods for feature extraction in web behavior data. We use clustering and time-frequency analysis (such as DFT) to extract features and build up classification models to recognise depression. Time-frequency features has improved the performance of models obviously. Compared with traditional psychological methods, our models can detect the users' mental status automatically and predict depression more effectively for early mental intervention.

Though the precision and recall of our models are not very high, we could still see the validity of the features from time-frequency analysis. In following research, we will try other time-frequency analysis tools such as wavelets to improve performance of models.

Acknowledgments

The authors thanks anonymous reviewers for their valuable suggestions and comments that have helped improve the presentation of this paper.

This work was supported in part by NSFC under Grant 11301504, in part by National Basic Research Program of China (973 Program2014CB744600), and in part by Strategic Priority Research Program (XDA06030800).

The authors would like to thank all participants in this experiment.

References

[1] M. Prince, V. Patel, S. Saxena, "No health without mental health", *The lancet*, vol. 370, pp. 859-877, 2007.

[2] T. Vos, AD. Flaxman, M. Naghavi, "Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990C2010: a systematic analysis for the Global Burden of Disease Study 2010", *The lancet*, vol. 380, pp. 2163-2196, 2012.

[3] PY. Collins, V. Patel, SS. Joestl, "Grand challenges in global mental health", *Nature*, vol. 475, pp. 27-30, 2011.

[4] AF. Jorm, "Mental health literacy: empowering the community to take action for better mental health", *American Psychologist*, vol. 67, pp. 231-243, 2012.

[5] S. Clement, O. Schauman, T. Graham, "What is the impact of mental health-related stigma on help-seeking? A systematic review of quantitative and qualitative studies", *Psychological Medicine*, vol. 45, pp. 11-27, 2015.

[6] YA. Hamburger, E. Ben-Artzi, "The relationship between extraversion and neuroticism and the different uses of the Internet", *Computers in Human Behavior*, vol. 16, pp. 441-449, 2000.

[7] SD. Gosling, AA. Augustine, S. Vazire, "Manifestations of personality in online social networks: self-reported Facebook-related behaviors and observable profile information", *Cyberpsychology, Behavior, and Social Networking*, vol. 14, pp. 483-488, 2011.

[8] M. Kosinski, D. Stillwell, T. Graepel, ”Private traits and attributes are predictable from digital records of human behavior”, *Proceedings of the National Academy of Sciences of the United States of America*, vol. 110, pp. 5802-5805, 2013.

[9] Y. Wu, M. Kosinski, D. Stillwell, ”Computer-based personality judgments are more accurate than those made by humans”, *Proceedings of the National Academy of Sciences*, vol. 112, pp. 1036-1040, 2015.

[10] M. Park, C. Cha, M. Cha, ”Depressive moods of users portrayed in Twitter”, in Proceedings of the ACM SIGKDD Workshop on Healthcare Informatics (HI-KDD), 2012, pp. 1-8.

[11] S. Park, SW. Lee, J. Kwak, ”Activities on Facebook reveal the depressive state of users”, *Journal of medical Internet research*, vol. 15, pp. 163-177, 2013.

[12] X. Wang, C. Zhang, Y. Ji, ”A Depression Detection Model Based on Sentiment Analysis in Micro-blog Social Network”, Trends and Applications in Knowledge Discovery and Data Mining, Springer Berlin Heidelberg, 2013, pp. 201-213.

[13] X. Wang, C. Zhang, L. Sun, ”An Improved Model for Depression Detection in Micro-Blog Social Network”, IEEE International Conference on Data Mining Workshops 2013, pp. 80-87.

[14] M. Choudhury, S. Counts, EJ. Horvitz, ”Characterizing and predicting postpartum depression from shared Facebook data”, in Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW 2014), pp. 626-638.

[15] M. Gamon, MD. Choudhury, S. Counts, ”Predicting depression via social media”, in Proceedings of the 7th International AAAI Conference on Weblogs and Social Media (ICWSM 2013), pp. 128-137.

[16] WW. Zung, ”A self-rating depression scale”, *Archives of general psychiatry*, vol. 12, pp. 63-70, 1965.

[17] L. Shu, ”Self-rating depression scale and depression status inventory”, *Chinese mental health journal*, pp. 194-197, 1999.

[18] L. Cohen, TimeCFrequency Analysis. New York: Prentice-Hall, 1995.

[19] SP. Lloyd, ”Least squares quantization in PCM”, *IEEE Transactions on Information Theory*, vol. 28, pp. 129-137, 1982.

[20] AY. Ng and MI. Jordon, ”On discriminative vs. generative classifiers: A comparison of logistic regression and naive bayes,” *Advances in Neural Information Processing Systems*, vol. 2, pp. 169-187, 2002.

[21] H. William and S. A. Teukolsky, Numerical recipes: The art of scientific computing, 3rd ed. New York: Cambridge University Press, 2007.

[22] JR. Quinlan, C4.5: Programs for machine learning, Morgan Kaufmann Publishers Inc., 2014.