Predicting Mental Health Status in the Context of Web Browsing

Dong Nie, Yue Ning, Tingshao Zhu

Institute of Psychology, University of Chinese Academy of Sciences
CAS, Beijing, 100190, China
*Corresponding Author: tszhu@psych.ac.cn

Abstract

Currently, people around the world are suffering from mental disorders. Given the wide-spread use of the Internet, we propose to predict users’ mental health status based on browsing behavior, and further recommend suggestions for adjustment. To identify mental health status, we extract the user’s web browsing behavior, and train a Support Vector Machine (SVM) model for prediction. Based on the predicted status, our recommender system generates suggestions for adjusting mental disorders. We have implemented a system named WebMind as the experimental platform integrated with the predicting model and recommendation engine. We have conducted user study to test the effectiveness of the predicting model, and the result demonstrates that the recommender system performs fairly well.

Keywords: Mental Health, Web Browsing Behavior, Prediction and Recommendation.

I. Introduction

Nowadays, many people around the world are suffering from mental disorders [15], such as, insomnia, depression, anxiety, tension, etc. It is reported that there are 26 million depression patients in China, and more than 30 million adolescents have various mental disorders [9]. In China, since psychotherapy is quite expensive and time-consuming, a large number of mental patients have to withstand illnesses without treatment. Therefore, mental disorder becomes increasingly serious. It is very important and urgent to find an efficient way to help people cope with mental disorders.

It is now very convenient to access the Internet. As reported, people spend more and more time on the web, and some of them have Internet addiction disorder [21]. Since most users take web browsers as their main entrance to the Internet on PC and mobile devices, this motivates us to help web users with mental disorders using browsers.

Much research indicate that it is feasible to cure mental patients through Internet psychological services. Online Cognitive Behavioral Therapy (CBT) has been developed to help people with depression [13], and it is proved to be effective. In this paper, we propose to predict users’ mental health status based on browsing behavior, and then recommend appropriate suggestions to help people alleviate mental disorders. We integrate the predicting model and recommendation engine into a recommender system — WebMind. To predict users’ mental health status, we train a browsing behavior model, then generate personalized adjustment recommendations based on the current status. The main focus of the prediction process is to collect users’ browsing history, and train a predicting model. In our work, we use WebMind to conduct user studies, collecting browsing log and evaluation results.

The rest of this paper is organized as follows. In Section II, we introduce some related works. Section III describes the whole process to build mental health prediction model including the data collection, preprocessing, training, and testing result. We present the method of generating recommendation in Section IV, then discuss the conclusion and future work in section V.

II. Related Work

Many researchers have already found that web behavior relates to mental health disorders [12], [14]. However, they just report the correlation between some web behavior and certain mental health problems, and this correlation does not necessarily mean these web behaviors can be used for predication, not even to mention how to alleviate mental disorders in practice, as there are too many factors that may affect mental health [19]. According to our previous research [22], mental health can be predicted based on web usage behavior. For example, frequently using instant message (IM) tools in a short time denotes that the user is in anxiety to some extent. Web user’s mental health
might relate to her web usage behavior, just the same as in the real life, the behavior will reflect one’s recent mental health [22]. Since web behavior is a part of individual’s behavior nowadays, it could also be used to identify the user’s mental health status. In this paper, we intend to train a machine learning model for predicting.

As quite a lot of recommendation algorithms have been developed so far, of which, content-based method [18] has been popular in early times, and collaboration filtering [16], [10] is the most used recommendation algorithm, while others are for special purpose. Much research have also been done to build hybrid algorithms (e.g. [20]) to build recommender system with better performance. With tremendous development of Internet, people rely more and more on the web, mental health should also be taken into consideration in recommender systems, especially those who use Internet quite often. Currently, people may have mental disorders due to huge working pressure. However, it is a little difficult to consult with a psychologist face-to-face, which is very expensive and time consuming, and can not be done for huge population. It becomes more and more critical to provide mental health intervention in time for these web users.

In this paper, we focus on web browsing behavior, and based on the mental health status predicted, provide suggestions for intervention. We have conducted an user study, using a customerized web browser, WebMind, to collect users’ all browsing behavior data, then make prediction about users’ mental health status, present suggestions, and ask the participant to give evaluation as well. We have trained nine models, each for a psychological dimensions respectively. For recommendation, we construct a psychological expert library containing suggestions, and use a hybrid of collaborating filtering and content-based algorithm, to choose personalized suggestions.

### III. Predicting Model

Research indicates that web behaviors can be used to identify mental health status instead of questionnaires [11], [17]. In this paper, we try to learn a model to predict mental health status by using browsing behavior.

Let $U$ be the matrix of users’ behaviors (e.g., behavior features in Tab. I), and $P$ the matrix of users’ mental health status. More specially, we encode one person’s behavior into a feature vector with $b$ dimensions. At the same time, each person is required to complete SCL-90 test to assess his/her current mental health status in 9 dimensions. The Symptom Checklist-90-R (SCL-90-R) is a relatively brief self-report psychometric instrument (questionnaire) published by the Clinical Assessment division of the Pearson Assessment and Information group. SCL-90 is designed to evaluate a broad range of psychological problems and symptoms of psychopathology, and it can also be used to measure the performance of psychiatric and psychological treatments or just for research purposes. [7] These nine dimensions of SCL-90 are somatization ($som$), obsessive-compulsive ($o-c$), interpersonal sensitivity ($i-s$), depression ($dep$), anxiety ($anx$), hostility ($hos$), phobic anxiety ($phob$), paranoid ideation ($par$), and psychoticism ($psy$).

Thus, we label each user by a SCL-90 vector to describe his/her mental health status,

\[(som, o-c, i-s, dep, anx, hos, phob, par, psy)\]

The predicting model can be defined as a mapping $F$:

\[U_{u \times b} \times R_{b \times 9} \rightarrow P_{u \times 9}\]

where $R$ is a projection matrix which may reveal the essential mapping from browsing behaviors to mental health status. In order to get an optimal $R$, we define the object function:

\[f(u, r) = F(u, r) - P_0\]

where $P_0$ is the psychological status resulted from a questionnaire as the ground truth and $r$ is the projection matrix. Then our task is to find such a matrix $R$ to minimize $f$,

\[f_r = \arg \min f(u, r)\]

If we can collect $n$ users’ data, i.e., $U$ and $P$, we are able to estimate the key matrix $R$, which is the predicting model for identifying mental health status.

### A. Collecting Data

To train the predicting model, the first step is to collect browsing history. We have implemented a web browser WebMind to collect users’ browsing history, as shown in Fig. 1

![Figure 1. WebMind Browser](image)
WebMind has been developed on an open source IE-based browser, and we have integrated some new features. It records the user’s browsing history in details, and cache the web pages that visited. The user can check his/her mental health status by clicking the button in the upper right corner, and a small pop-up window presents mental health information which is predicted by our model. WebMind also provides relative psychological intervention recommendations on a popping up window.

We have conducted an user study in February 2012, and 47 graduate students were recruited as volunteer to use WebMind instead of their own browsers. The experiment is arranged in 4 weeks. Subjects are required to use WebMind as their default browser, for example, use WebMind to surf Internet. WebMind automatically records users’ behavior when they are browsing the web, and their web behavior data is stored in several files. During the study, we also asked the subjects to complete SCL-90 [8] with 90 multiple-choice questions, and recorded the nine dimensions assessment for each subject.

We finally collect 47 copies of effective data in the form of xml. We take the complete browsing history of each subject. After preprocessing, we extract 53 browsing features for each subject, and some of them are in Table I:

### Table I. Browsing Behavior Features

<table>
<thead>
<tr>
<th>Feature</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>surfUrls</td>
<td>The number of surfed urls</td>
</tr>
<tr>
<td>maxContentClass</td>
<td>The most frequent surfed content class</td>
</tr>
<tr>
<td>minContentClass</td>
<td>The least frequent surfed content class</td>
</tr>
<tr>
<td>maxEmotionClass</td>
<td>The most frequent surfed emotion class</td>
</tr>
<tr>
<td>minEmotionClass</td>
<td>The least frequent surfed emotion class</td>
</tr>
<tr>
<td>startSurfWay</td>
<td>User’s habitual start surfing way</td>
</tr>
<tr>
<td>startSurfTime</td>
<td>Time start to surf internet</td>
</tr>
<tr>
<td>maxTabs</td>
<td>Max tabs open at a time</td>
</tr>
<tr>
<td>searchEngineUse</td>
<td>Total times user uses search engine at a time</td>
</tr>
<tr>
<td>socialNetUse</td>
<td>Total times user uses social networks at a time</td>
</tr>
<tr>
<td>emailUsePerDay</td>
<td>Total times user uses email at a time</td>
</tr>
<tr>
<td>unhealthSurfs</td>
<td>Total times user visits unhealthy pages</td>
</tr>
</tbody>
</table>

### B. Training Model

To learn such predicting models, we propose to train SVM classifier based on browsing behavior features. SVM uses kernel function to minimize the structural risk [5], [2], which greatly enables the ability to solve cases with high dimension feature [3]. After decades’ development, SVM is now thought to be one of the best classification algorithms [1]. In this paper, we estimate $R$ using SVM and the training set acquired from the user study.

We compute each subject’s 9 dimensions of mental health status according to SCL-90 [6], and take them as the mental health label for each subject. Here, each subject’s browsing vector (behavior features) is taken as training data, while the vector of mental health status is taken as class label. Since the original value of mental health status are continuous, we run the discretization on each dimension. Technically, people in high score group indicate a high potential in having corresponding mental disorder, and vice versa. Thus, we only focus on the two extreme ends: high-score and low-score, where high-score group is from the top 27%, and low-score group is the bottom 27% ones. Tab. II shows some examples of training data on depression dimension. The label ‘-1’ means low-score group while ‘+1’ means high-score group.

### Table II. part of training data for depression

<table>
<thead>
<tr>
<th>label</th>
<th>m_surfUrls</th>
<th>m_maxTabs</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1</td>
<td>57</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>-1</td>
<td>17.8</td>
<td>3.8</td>
<td>3.8</td>
</tr>
<tr>
<td>-1</td>
<td>78</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>+1</td>
<td>42.7</td>
<td>2</td>
<td>11.3</td>
</tr>
<tr>
<td>+1</td>
<td>91.9</td>
<td>5.6</td>
<td>40.8</td>
</tr>
<tr>
<td>+1</td>
<td>57</td>
<td>2</td>
<td>5</td>
</tr>
</tbody>
</table>

After preprocessing, we obtain training data which can be used to train SVM classifier — libsvm [4]. SVM classifier is trained for each of 9 dimensions, and they are trained independently.

### C. Experimental results

To test the performance of predicting model, we randomly pick 20% of the whole data as testing set, and all remaining for training. Nine classification models are trained, for somatization, obsessive-compulsive, interpersonal sensitivity, depression, anxiety, hostility, phobic anxiety, paranoid ideation, and psychoticism respectively. The results are shown in Tab. III

### Table III. classification accuracies

<table>
<thead>
<tr>
<th>Mental State</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>somatization</td>
<td>77.78%</td>
</tr>
<tr>
<td>obsessive-compulsive</td>
<td>100%</td>
</tr>
<tr>
<td>Interpersonal Sensitivity</td>
<td>88.90%</td>
</tr>
<tr>
<td>depression</td>
<td>88.90%</td>
</tr>
<tr>
<td>anxiety</td>
<td>77.78%</td>
</tr>
<tr>
<td>hostility</td>
<td>77.78%</td>
</tr>
<tr>
<td>phobic anxiety</td>
<td>88.90%</td>
</tr>
<tr>
<td>paranoid ideation</td>
<td>88.90%</td>
</tr>
<tr>
<td>psychoticism</td>
<td>77.78%</td>
</tr>
</tbody>
</table>

Although the number of our data is limited, the result is encouraging. It explicitly indicates that people’s mental
status correlates with their browsing behavior, so we can identify people’s mental health in time if we can track people’s browsing behavior. Taking depression as an example, the prediction precision is 88.90%, which is fairly well. In fact, the result also demonstrates web behavior can be used to do psychological prediction to some extent. Furthermore, if we can predict all nine dimensions of SCL-90, we are able to know people’s mental health in every aspect, which is really useful to help people.

IV. Recommendation

Based on the prediction in Section III, we try to produce recommendations to help people improve mental health. On the latest version of WebMind, we have integrated a mental health recommender system. We first manually produced a mental intervention recommendation library by consulting with several psychologists, for example, “You must learn to rationally and objectively analyze the difficulties encountered by yourself in daily lives, and do not make personal goals too high”. The recommender system will pick up the appropriate suggestions based on the current mental health status predicted.

To recommend appropriate suggestions, we propose the prediction-based recommendation method as follows. We predict the status of each dimension at first, then compare it with the standard score of this dimension, and we can find the dimension with the highest difference. This process can be shown in the following equation:

$$\text{Dim} = \arg \max_d \{\text{predictScore}(d) - \text{standardScore}(d)\}$$

where $\text{Dim}$ is the dimension that the user needs suggestions to the most, $D$ is a set of 9 SCL-90 dimensions. After the user has been identified to a specific dimension, we then randomly pick up suggestions as recommendation. By this, we can provide corresponding intervention suggestions according to $\text{Dim}$. The recommendation is achieved by a projection from psychological dimension to suggestion, we can describe it as follows:

$$S = f(\text{Dim})$$

where $S$ is the suggestions we want to recommend, and $\text{Dim}$ is the dimension identified above.

The baseline model is just a random pick-up from all suggestions, and we want to test whether the recommendation based on mental health status works or not.

We have implemented the recommendation in WebMind, and to test the performance we conducted another user study by recruiting 52 participants.

Each participant is required to use WebMind as their default browser for 3 weeks. They do not need do anything special in the first week, from the second week on, the intervention recommendation is provided. Each time of generating recommendation, we just randomly use one of these two methods (i.e., prediction-based and baseline model). The subject is instructed to look into the suggestions and rate the recommendation. WebMind records their feedback and store the information in the form of xml files.

The participants are the first grade graduate students and the sample size is 52, with average age of 23.15 (S.D.=1.28). In such research sample, the male population is 41, accounting for 78.85%, while the female population is 11, accounting for 21.15%. The Han ethnic population is 47, accounting for 90.38%, and other ethnic’s population is 5, accounting for 9.62%. The population from whether one-child family or not-one-child family haven’t been counted.

The users’ feedback is shown in Fig. 2:

![Figure 2. Ranks of two different recommender systems](image)

Rank 4 represents “Best” for the recommendation, and rank 1 represents “Worst” for the recommendation. We have totally collected 326 ratings for our recommendation model through 52 subjects’ using log, meanwhile, 352 ratings for random recommendation model is collected. As shown in Fig. 2, the experiment result shows our recommendation method performs better than random pick-up, which is optimistic. The high score rating is 55.2%, which is fairly higher than random recommendation model(38.06%). We are now working on new algorithms to improve the effectiveness of recommendation, to supply more effective suggestions to help mental health.

V. Conclusions

Based on the 9 classification models and the corresponding prediction accuracies (77.78%-100%), we can conclude that web user’s mental health status can be predicted by browsing behavior. The experiment result
indicates that it is an effective way to predict people’s mental health by analyzing browsing behavior. Moreover, the result also shows that machine learning algorithm might be useful for mental health intervention.

The recommendation experiment we have conducted in this paper demonstrates that it is effective to adjust people’s mental health by providing recommendations. Furthermore, our research also shows that recommender system performs fairly well with high accuracy on mental prediction.

Apparently, there exists vast space we can do to either predict mental health status or help people alleviate mental disorders. We are now exploring more efficient learning algorithms to train better user model to predict mental health. Meanwhile, we will try different methods to produce recommendations more effectively.

Acknowledgments

The authors gratefully acknowledges the generous support from NSFC (61070115), Institute of Psychology (113000C037), Strategic Priority Research Program (XDA06030800) and 100-Talent Project (Y2CX093006) from Chinese Academy of Sciences.

References


