INFORMATION STUDY

Data mining-based study on sub-mentally healthy state among residents in eight provinces and cities in China

Hongmei Ni, Xuming Yang, Chengquan Fang, Yingying Guo, Mingyue Xu, Yumin He

METHODS: The strategic tree algorithm was used to identify the main manifestations of the state of sub-mental health. The backpropagation artificial neural network was used to analyze the main manifestations of sub-healthy mental states of three different degrees. A sub-mental health evaluation model was then established to achieve predictive evaluation results.

RESULTS: Using classifications from the Scale of Chinese Sub-healthy State, the main manifestations of sub-mental health selected using the strategic tree were F1101 (Do you lack peace of mind?), F1102 (Are you easily nervous when something comes up?), and F1002 (Do you often sigh?). The relative intensity of manifestations of sub-mental health was highest for F1101, followed by F1102, and then F1002. Through study of the neural network, better differentiation could be made between moderate and severe and between mild and severe states of sub-mental health. The differentiation between mild and moderate sub-mental health states was less apparent. Additionally, the sub-mental health state evaluation model, which could be used to predict states of sub-mental health of different individuals, was established using F1101, F1102, F1002, and the mental self-assessment total score.

CONCLUSION: The main manifestations of the state of sub-mental health can be discovered using data mining methods to research and analyze the latent laws and knowledge hidden in research evidence on the state of sub-mental health. The state of sub-mental health of different individuals can be rapidly predicted using the model established.
here. This can provide a basis for assessment and intervention for sub-mental health. It can also replace the relatively outdated approaches to research on sub-health in the technical era of information and digitization by combining the study of states of sub-mental health with information techniques and by further quantifying the relevant information.

© 2014 JTCM. All rights reserved.

**Key words:** Questionnaires; Mental health; Data mining; Strategic tree; Artificial neural network

**INTRODUCTION**

"Sub-health" comes from a Chinese word coined by a Chinese scholar after "the third state", which was first put forward by N. Buheman in the former Soviet Union in the 1980s. The third state manifests as many discomforts, including those of the body and mind, and social adaptability. Assessment of states of sub-mental health is a key component of research on sub-health. Because states of sub-mental health are involved in a wide range of subjects, the assessment method needs multiple dimensions and layers. At present, the combination of scale assessment with biological indexes has become the main method for the assessment of states of sub-mental health. However, there are problems in preparation of scales for states of sub-mental health. For example, the latent variable is not directly measurable, measurement error is an issue, and relations between the latent variables need to be determined. In this context, traditional statistical methods are powerless. Data mining methods can "mine" knowledge from a large amount of data. Data mining is the process of extracting previously unknown and latent useful information and knowledge from a vast amount of incomplete, noisy, vague, and random data. At present, fields concerned with the state of sub-mental health have mainly involved depression, anxiety, and learning/memory. The present study focuses on mining the most important manifestations and establishing an assessment model of the state of sub-mental health from the various manifestations of the state of sub-mental health to achieve predictive assessment results.

**METHODS**

**Study subjects**

From January 2007 to December 2008, the study group used the Scale of Chinese Sub-healthy State (CSHS-2) to survey populations aged 20-60 in 8 regions in China: Shanghai, Xian, Hebei, Henan, Yun- nan, Lu Zhou, Yang Zhou, and Anhui. Self-assessments of health were elicited, and 13 385 completed questionnaires were retrieved. Because sub-mentally healthy persons were the subjects for analysis and study, healthy persons with total assessment of the general health scores (serial number: G1) ≥85 and patients with various diseases were excluded from analysis, leaving 5643 remaining questionnaires of persons with sub-health. The population with a mental healthy state (Serial number: G2 ≥85) were then excluded, leaving the final working sample of 3970 cases with sub-mentally healthy state.

**Source of scales**

CSHS-2 is a part of the key project of the Eleventh Five-Year Science and Technology Support Program of the State Science and Technology Department’s "Study on Concept Category and Measure Standards" (No. 2006BAI13B01). The CSHS-2 is an assessment scale for states of sub-mental health. The scale was developed by Shanghai University of Chinese Medicine, combining relevant knowledge, following the principles and methods of scale establishment and incorporating experiences of scale development at home and abroad. The development of the scale integrated literature research, expert consultation, and other methods. After pre-testing, item selection, and tests of reliability and validity, this scale included four fields (somatic manifestations, mental manifestations, social adaptation, and sexual life), 20 aspects (fatigue, digestion, sleep, dysfunction, immune disorder, allergy, aging, pain, constipation, depression, anxiety, learning/memory, pressure, sense of contentment, adaptability, sense of security, self-confidence, self-realization, social support, and sexual life), and 78 items. Responses to each item were classified using a five-point Likert scale: "never" =1 point, "rarely" =2, "general" =3, "often" =4, and "always" =5. Higher scores indicate more severely sub-mental health. In addition, the questionnaire also includes 3 self-assessment items (1-100 points), assessing general health state, mental health state, and quality of life. On these three items, higher scores indicate healthier states.

In line with the research objective of the present study, the CSHS-2 sub-scale on sub-mental health was selected. This sub-scale is composed of 3 parts: depression, anxiety, and learning/memory. The relevant items include: (a) depression: do you feel lonely? (F1001), do you often sigh? (F1002), do you wish to reveal your innermost feelings? (F1003), do you blame yourself? (F1004), do you feel that living is insignificant and have no interest in life? (F1005), and have you had feelings of oppression and unhappiness in the last month? (F1006); (b) anxiety: do you lack peace of mind? (F1101), do you easily become nervous when something comes up? (F1102), do you feel worried or afraid without any cause? (F1103), have you had peace of mind in the last month? (F1104), and have you often
been terribly upset in the last month? (F1105); and (c) learning/memory: have you experienced a situation where you were unable to think through the things that were happening recently? (F1201), do you find your attention involuntarily distracted when you do something? (F1202), and how do you think your work efficiency? (F1203). In addition, a subjective assessment item reflecting a respondent’s own mental state was selected. For the mental self-assessment total score (S-G2), respondents were asked to make a self-assessment of their own overall mental state, with the best possible state being 100 points.

Data processing

In data mining, algorithms are used to mine and assimilate relevant knowledge in the survey results in the scale to establish possible models. In the present study, according to characteristics and results observed in preliminary experiments, the strategic tree and backpropagation artificial neural network algorithms were selected in Microsoft SQL Server 2008 (Microsoft Company, Redmond, WA, USA).

First, according to the mental self-assessment total score (S-G2), the 3970 sub-mentally healthy state cases were divided into 3 groups: 0=S-G2<70, (the severe sub-mental health group); 1=S-G2 70-80 (moderate sub-mental health group); and 2=S-G2 81-84 (mild sub-mental health group). The classification algorithm is a data analysis method of predictive data mining. Its purpose is to identify the model best able to accurately describe and differentiate the model of data class or concept to be divided into a certain data class according to the attributes of the entity and other constraint conditions using important sample datasets. Here, we used the strategic tree classification, and based on a dependent network, selected the item playing the key role. The artificial neural network is a highly nonlinear system able to learn, and backpropagation is a learning algorithm of the neural network. Backpropagation can determine the parameters in the network though training with a large amount of historical data to determine the structure of the whole network, and it can lay a foundation for prediction of data using the network. Consequently, this model is useful in the search for the differences of manifestations among the three groups with different degrees of sub-mentally healthy states and for the prediction of sub-mentally healthy state.

RESULTS

Main manifestations of sub-mentally healthy state mined by strategic tree method

F1001, F1002, F1003, F1004, F1005, F1006, F1101, F1102, F1103, F1104, F1105, F1201, F1202, and F1203 were used as input values, and S-G2 was used as the predictable value. The strategic tree was produced by the vTargetMail datasets (Figure 1). Figure 2 displays the weakest link in the dependent relation, and Figure 3 displays the strongest.

The dependant relation figure of the strategic tree revealed that the strongest dependent relation with S-G2 was with F1101, followed by F1102, and then F1002. The three correlative factors were only included in the models for depression and anxiety; they were not in the model for learning/memory. The strongest dependent relations with S-G2 included in the anxiety model were F1101 and F1102, but the depression model only included F1002, which had the third strongest dependent relation with S-G2.

Analysis of sub-mentally healthy states of different degrees by the artificial neural network

The three degrees of sub-mentally healthy states were determined based on S-G2 scores: 0=severe sub-mentally healthy state, 1=moderate sub-mentally healthy state, and 2=mild sub-mentally healthy state. Because there were three populations with the sub-mentally healthy states of three different degrees, F1101, F1102, and F1002 were each analyzed in the three different sub-mentally healthy states.
Comparison between severe sub-mentally healthy state and moderate sub-mentally healthy state

F1101, F1102, and F1002 were used as the input values, and the 0 and 1 groups of S-G2 were used as the predictable values. The results are shown in Figure 4. After the information in Figure 4 was further sorted, it was found that, when values for F1002 of 3.212-5 and values for F1101 of 3.149-5 were input, the influencing score on the value of S-G2=0 reached 88.49. The next highest influence was observed when values for F1102 of 3.431-5 were input, resulting in the influencing score on S-G2=0 reaching 70.69. These results suggest that often sighing, often feeling a lack of peace of mind, and often being easily nervous when something comes up are significantly correlated with severe sub-mentally healthy state. In addition, the influencing scores of F1101=1-1.898, F1002=1-1.909, and F1102=1-2.164 on the output value of S-G2=1 reached 45.01, 40.03, 39.41, respectively, suggesting that never feeling a lack of peace of mind, never sighing, and being rarely nervous when something comes up are correlated with moderate sub-mentally healthy state.

Comparison between severe sub-mentally healthy state and mild sub-mentally healthy state

F1101, F1102, and F1002 were used as input values, and the 0 and 2 groups of S-G2 were used as the predictable values. The results are shown in Figure 5. After the information in Figure 5 was further explored, it was found that, when F1101=3.149-5, F1102=3.431-5, and F1002=3.212-5 were input, the influencing scores on the output values of S-G2=0 were 23.04, 22.20, and 16.10, respectively. These results indicate that often feeling a lack of peace of mind, often being easily nervous when something comes up, and often sighing are significantly correlated with severe sub-mentally healthy state. In addition, the influencing scores of F1102=1-2.164, F1002=1-1.909, and F1101=...
1-1.898 on S-G2=2 are 100, 85.66, and 70.37, respectively. Mild sub-mentally healthy state manifests as, in order of strength, not being easily nervous when something comes up, never sighing, and never feeling a lack of peace of mind.

**Comparison between moderate sub-mentally healthy state and mild sub-mentally healthy state**

F1101, F1102, and F1002 were used as input values, and S-G2=1 and S-G2=2 were used as the predictable values. The results are shown in Figure 6.

It was found after the information in Figure 6 was examined further that, when F1002=1.909-2.561 and F1101=1.898-2.561 were input, the scores of influencing the output S-G2=1 were 4.76 and 3.89, respectively, suggesting that rarely sighing and feeling rarely lack peace of mind have a certain effect on being in the group with moderate sub-mentally healthy state, but with very small scores. The differentiation between mild and moderate degrees of sub-mentally healthy state is poorer than the differentiation of mild or moderate from severe sub-mentally healthy states.

**Prediction of sub-mentally healthy state of subjects by the artificial neural network**

To predict the sub-mentally healthy state of an identified subject with F1002=5, F1101=5, and F1102=3, the above parameters and learned system were used for prediction from input to output (Figures 7, 8).

With F1002=5, F1101=5, and F1102=3.434 as the input values, the predictive value of S-G2=0 was obtained. This indicated that always sighing, always feeling a lack of peace of mind, and often being easily nervous when something comes up are correlated more strongly with severe sub-mentally healthy state than with other degrees of sub-mentally healthy states. The predictive result is identical to the real situation.

**DISCUSSION**

In this study, the main manifestations of sub-mentally healthy state screened by the strategy tree were F1101 (do you lack peace of mind?), F1102 (are you easily...
Among these items, F1002 was involved in the tendency toward depression, and F1101 and F1102 were involved in the tendency toward anxiety. The relative strength of sub-mentally healthy manifestations was, in order from strongest to weakest, F1101, F1102, and F1002. The model between F1101, F1102, and F1002 and SG-2 in sub-mentally healthy state manifestations could be established by artificial neural network training, which could be used to predict the sub-mentally healthy state of different individuals. A special software platform for assessment of sub-mentally healthy state could be further developed on the basis of these findings to provide strategic support for the objective assessment of sub-mentally healthy state.

Neural network learning could best differentiate between moderate and severe and between mild and severe sub-mentally healthy states. The differentiation between mild and moderate sub-mentally healthy states was less apparent. However, concentrating on clinical practice, the populations of focus, and those in need of intervention, are mainly in the moderate and severe sub-healthy groups, so this limitation does not have a large effect on the application of our study. Nevertheless, states of sub-mental health need multidimensional and multi-layered assessment. In the clinical setting, sub-mentally healthy state should be comprehensively assessed in combination with somatic manifestations, manifestations of social adaptability, and certain biological indexes. However, prediction using this model offers one means of rapid assessment that can provide a basis for clinical assessment and intervention for those with a sub-mentally healthy state.
Along with the transformation of the medical model towards the biological-mental-social model of medicine, many sociological and psychological indexes are found in medical research fields. These indexes, however, often cannot directly measure the variable of interest. Sub-mentally healthy state has complicated manifestations, the relations between the variables are not very clear or definite, and there are a great number of latent variables that cannot be directly measured. Further, the investigative results are less accurate than results measured by physical equipment, so measurement errors often occur. Traditional statistical methods mostly use the theory verifying approach. In this approach, specific relationships within the data are first supposed, and then these hypotheses are supported or overturned by analysis. Of course, the information obtained through this approach is also superficial and can be identified by actual human senses or techniques near to actual human senses. The data mining method is based on discovery. This approach uses model matching and relative algorithms to determine the important relations in the data without clear and definite hypotheses. This allows for the discovery of knowledge that is unknown, difficult to predict, and even completely conflicting with a person’s actual senses.

In brief, research and analysis on latent laws and knowledge hidden in research evidence on sub-mentally healthy state using data mining methods can uncover the main manifestations of states of sub-mental health. Sub-mentally healthy states among individuals can be rapidly predicted through the established model. This can provide a basis for the assessment and intervention of sub-mentally healthy state. This approach also has the potential to update the relatively outdated methods used in research on sub-health. By combining research on sub-mentally healthy state with information techniques and further quantifying the relevant information, research on states of sub-mental health can be brought up to date in the present era of information and digitization. Moreover, because this model offers a rapid and effective method for prediction, different assessment software can be developed to serve clinical assessment needs and enable interventions for those with a state of sub-mental health.

REFERENCES

2. Xu L, Xia JL, He YM. Rationality of using structure equation model in research of sub-healthy state. Yi Xue Yu Zhe Xue (Ren Wen She Hui Yi Xue Ban) 2008; 29(4): 36-56.