Analysis and processing of misdiagnosis data for depression based on modified entropy weight method

Lu Shiying[1,a], Tang Ji'an[2,b], Liu Feng[3,c*], Qin Sishi[4,d], Chen Jie[5,e], Chen Lei[6,f]

[1]Changsha university of science and technology, Hunan China
[2]Shanghai University of International, Business and Economics, Shanghai China
[3]Institute of Artificial Intelligence and Change Management/Shanghai University of International Business and Economics
[4]Shanghai University of International, Business and Economics, Shanghai China
[5]Changshu Institute of Technology, Jiangsu China
[6]Wuxi Prithink Information Technology Co., Ltd., Jiangsu China

Abstract—Depression is always the core field of psychological research, and the analysis of misdiagnosis data of depression is also the vital content of depression research. Based on the analysis of misdiagnosis data processing, this paper adopts a order relation analysis method, to correct the problem of inconsistent entropy and entropy transfer relation (when all entropy value tend to be 1). This paper obtains multi-index comprehensive quantitative values, from various angles analysis of misdiagnosis data depression, so as to avoid subjective and one-sided evaluation results. It not only improves the rapidity and practicability of the algorithm, but also makes the analysis of misdiagnosis data more objective and accurate, which can be applied to medical field.

1 Introduction

Depression disorder has always been one of the front-burner issue in the field of psychological research, the analysis of misdiagnosis data of depression is also the core content of depression research. Due to the extensive attention of misdiagnosis, how to scientifically and reasonably evaluate the misdiagnosis data of depression is of great significance for the validation, promotion and development of the psychological research. There are many comprehensive evaluation and analysis methods which are mostly studying on depression on risk factors and prediction models based on statistical, including entropy weight method, artificial neural network method, analytic hierarchy process, logistic regression model design, fuzzy mathematics method and its combination method (cross-sectional and longitudinal studies). However, these multi-objective fuzzy comprehensive analysis methods have the limitation of artificial assignment of weight of traits, making evaluation results less scientific.

To this end, the modify entropy weight method was introduced to the comprehensive evaluation to overcome the limitation of artificial weighting of each target trait, thereby achieving a more scientific and effective comprehensive evaluation of misdiagnosis data.

2 principal and method of comprehensive evaluation of clinical symptoms study by modify entropy wight method.

2.1 data processing of clinical symptoms by modified entropy weight method and sequence relation analysis method:

Standardization of membership matrix of evaluation indexes:

The membership evaluation matrix is constructed by constructing the index values of n evaluated objects corresponding to m evaluation indexes R:

$$R=\begin{pmatrix} r_{11} & \cdots & r_{1m} \\ \vdots & \ddots & \vdots \\ r_{n1} & \cdots & r_{nm} \end{pmatrix}$$

2.1.1 Calculate the index entropyH_j and weight w_j

$$w_j=\begin{cases} (1-H^{35.35})w_{0j}+H^{35.35}w_{3j} & H_j<1 \\ 0 & H_j=1 \end{cases}$$

$$w_j=\frac{1-H_j}{n-\sum_{j=1}^{n} H_j}$$

$$w_j=\frac{1-H_j}{\sum_{j=1}^{n}(1+H_j-H_k)}(2)$$

[a]529455452@qq.com, [b]2513217525@qq.com, [c]lsttoy@163.com, [d]1034109212@qq.com , [e]natsusu@qq.com, [f]757076251@qq.com

Note: the average value of all entropy values not equal to 1 in H

2.1.2 Calculate the combined weight

By calculating vector resemblance-degree of index vector and the system integrative vector then standardize it the index weight is obtained. The original data were preprocessed by the extremum processing method to eliminate the dimensional differences among different indicators v, the objective weight of the second index is obtained by the modified entropy weight method w, the combined weight is calculated a, the calculation formula is as follows [2]:

$$a_1 = \frac{w_1 v_1}{v_1 \sum_{k=1}^{n} w_k + w_3 v_2 + w_4 v_3 + v_4 \sum_{k=5}^{n} w_k + \sum_{k=8}^{n} w_k} \quad (3)$$

2.2 clinical symptom case analysis by modified entropy weight method and order relation analysis method

Cases in misdiagnosed cases analysis of data, different analysis show different clinical symptoms. On this basis, we divided the different clinical symptoms into cognitive psychological symptoms and physiological neurological symptoms, and then subdivided them into ten secondary indexes, respectively, anxiety, emotional fluctuation, neural inhibition, retardation of thinking, slow movements, loss of interest, the circulatory system, the digestive system, the respiratory system, the nervous system.[4]According to the data of symptoms, the modified entropy weight method was substituted into the formula to calculate the relevant weights, so as to observe which symptoms accounted for the greater weight in clinical manifestations, so as to facilitate the next analysis. (specific weights are shown in table 1)

TABLE 1: WEIGHT OF CLINICAL SYMPTOMS IN MISDIAGNOSIS DATA

Index classification	wight	Index name	wight	Index name	weight
Cognitive psychological symptom	0.4735	anxiety	0.4122	Anxiety and insomnia	0.6612
				Strong motive of suicide	0.2718
				Hypochondriac delusion	0.0670
		Emotional fluctuation	0.1350	Mood swings	0.1358
				fussiness	0.8642
		Neural inhibition	0.0495	torpor	0.4358
				hypobulia	0.5642
		Retardation of thinking	0.0472	Loosening of association	0.5269
				Thought disturbance	0.4731
		Slowed movements	0.0308	Slowed movements	0.4876
				bradyphrasia	0.5124
		Loss of interest	0.3253	inappetence	0.7160
				Reduce or lack pursuits within the scope of work	0.2840
Physical and neurological symptoms	0.5265	Circulatory system	0.1067	Chest congestion and palpitation	0.7612
				Coronary heart disease	0.2388
		Digestive system	0.2923	Disorders of digestion	0.6044
				Enteritis with diarrhea	0.3255
				Nausea and vomiting	0.0701
		Respiratory system	0.0573	Asphyxia with hard breathing	0.5343
				dyspnea	0.4657
		Nervous system	0.5437	Dizziness and headache	0.7014
				Nervous debility	0.2986

Multiple regression analysis is an important method of data processing, the multicollinearity is a quite important link in analysis. By explaining variables linear correlation between phenomena, variable selection explained variables had no significant influence, and it guaranteed the significance of regression equation fitting and simplicity to deal with problems.

At the same time, we calculated and analyzed the improvement rate of depression in the data, which was recorded as 1 for recovery, 0.7 for significant improvement, 0.3 for improvement, and 0 for invalid treatment. The improvement rate and weighted improvement rate were calculated respectively for data comparison and further research (as shown in table 2).

TABLE 2. THE IMPROVEMENT RATE AND WEIGHTED IMPROVEMENT RATE

recovery	Clearly improved	improved	Futile	Improvement rate	Weighted improvement rate
53	22	6	2	0.975904	0.860241
0	8	2	0	1	0.66
11	36	6	4	0.929825	0.687719
17	0	2	0	1	0.947368
17	6	2	0	1	0.888
14	12	4	0	1	0.813333
0	91	46	7	0.951389	0.602083
0	5	1	1	0.857143	0.571429
53	18	12	4	0.954023	0.822989
30	0	0	0	1	1
0	16	0	0	1	0.7
0	186	0	0	1	0.7
7	43	8	5	0.920635	0.652381
55	28	9	5	0.948454	0.815464
0	0	0	31	0	0
13	0	0	0	1	1
42	0	31	23	0.760417	0.598958
0	15	2	0	1	0.676471
30	24	0	0	1	0.866667
0	15	3	3	0.857143	0.571429
0	8	0	10	0.444444	0.311111

3 Discussion

Early depression diagnosis mainly depends on the doctor to the patient interviews, only to confirm a few depression risk factors, in order to a common medium for large depression risk factor determination. Beck, a professor at the university of Pennsylvania in 1961 with its 25 years clinical experience depression, the patient of the common problems of paying the statistical and pioneering depression self-rating scale (beck 'inventory, BDI).[5] Since depression diagnosis to comprehensive various statistical composition, with scale standards constantly improve, and adopted by constantly updating optimization and risk factor, doctor's inquiry with scale survey, as the necessary option of depression diagnosis, has been used up to now. Nevertheless, there are many problems with this diagnostic method.[6] Firstly, the design criterion of the scale is that depressive disorder is not diagnosed until the score reaches the diagnostic threshold, which often leads to overdiagnosis or underdiagnosis. Secondly, the subjective concealment and other interference factors of the subjects cannot be avoided in the process of inquiry. Meanwhile, the extreme lack of mental health professionals, the difference of doctors' experience, the neglect of patients and privacy concerns also make the overall recognition rate of depression low.[7]

In the last two years, some research teams have begun to model and reason about depression risk data based on statistical risk factors. Okamoto constructed a depression prediction model based on step linear discriminant analysis, which could reach the prediction accuracy of 78.2%. Benjamin predicted the likelihood of depression in adolescents based on regression tree and ascension classifier model. Based on the stepwise logistic regression model, King designed the risk prediction algorithm predict A, which can be used to accurately predict the incidence risk of anxiety disorders. Victor constructed a fuzzy reasoning system based on questionnaire score to predict individual depression risk.

Nevertheless, the above models are basically based on social statistics, lack of hobbies, chronic disease status, self-efficacy and other risk factors obtained from the self-filling scale. Although the disadvantages of the scale threshold judgment are solved, the subjective influence of the subjects cannot be eliminated.

Depression is a complex psychological disorders, and its inducing factors and symptoms are diverse, and dynamic change over time. So the stability prediction model and generalization ability, still need to consider the correlation between different time points, understand the evolution of the risk of depression over time development characteristics, to achieve more accurate evaluation and forecast.

4 Conclusion

An improved formula based on the theory of information entropy is proposed. The objective weights in this formula can be derived from the data included in alternatives. The coefficients of weight in this model are derived from the avail value of data reflecting the information entropy, by this method the problem of weight allocation can be avoided.

In the study, the method of order relation analysis is adopted, which does not construct judgment matrix, but requires strong consistency condition among evaluation indexes. In the entropy weight method, the greater the difference of an index between the evaluation objects, the more information it contains, the lower the entropy. But, in the traditional entropy weight method, when all the entropy values are close to 1, even if the difference between the indexes is small, it will cause the change of entropy weight times, which will result in some indexes being given weights that are not in accordance with the actual situation, and affect the final judgment result.[1] Therefore, the improved entropy weight method overcomes the shortcomings of the traditional algorithm to maintain the ability of the gap, while combining the sequential relationship analysis method for the weight calculation of clinical symptoms. Through multiple regression analysis of misdiagnosis data and actual data, the practicability and rapidity of the algorithm can be improved from different perspectives of the data.

In this paper, a method of misdiagnosis data processing based on modified entropy weight method and multiple regression analysis is proposed. It makes the data analysis more practical and improves the performance of the shortcut of the whole method.

However, this paper does not carry out in-depth analysis on the correlation between symptoms. In the next research, we will optimize the algorithm design and use BP neural network to construct the relationship diagram, so as to provide help for doctors to diagnose patients with depression.

Acknowledgment

This paper was supported by National Natural Science Foundation of China(72042004).

References

[1]Junwei YANG;Jianjun WANG;Linsheng SHAO;Yanliang LI;Bianping ZHAO;Xin JIA;Qi LUO.Comprehensive Evaluation of New Fresh Edible Sweet-glutinous Maize Varieties Based on Entropy Weight Method[J].Asian Agricultural Research,2018,v.10,96-98.

[2] Entropy weight coefficient model and its application in evaluation of groundwater vulnerability of the Sanjiang Plain[J].Journal of Northeast Agricultural University(English Edition),2007,83-88.

[3] Merijn Eikelenboom,Johannes H. Smit,Aartjan T.F. Beekman,Brenda W.J.H. Penninx.Do depression and anxiety converge or diverge in their association with suicidality?[J].Journal of Psychiatric Research,2012,46(5),608-615.

[4] Zhang SuoYuan;Chen ling.JiZhenPeng;Da-fang Chen;Fu-de Yang;Day. The misdiagnosis for depression in patients with bipolar disorder Ⅰ type a risk factor for suicide risk analysis [J]. International journal of psychiatry, 2015, v. 42, 24-27.

[5] Emmanuel Jesulola,Christopher F. Sharpley,Vicki Bitsika,Linda L. Agnew,Peter Wilson.Frontal alpha asymmetry as a pathway to behavioural withdrawal in depression: Research findings and issues[J].Behavioural Brain Research,2015,292,56-67.

[6]Saime Akdemir Akar,Sadık Kara,Sümeyra Agambayev,Vedat Bilgiç.Nonlinear analysis of EEGs of patients with major depression during different emotional states[J].Computers in Biology and Medicine,2015,67,49-60.

[7] Q. Zhao,Y. Tang,S. Chen,Y. Lyu,A. Curtin,J. Wang,J. Sun,S. Tong.Early perceptual anomaly of negative facial expression in depression: An event-related potential study[J]. Neurophysiologie Clinique / Clinical Neurophysiology,2015,45(6),435-443.

[8] Baolin Liu,Zhixing Jin,Zhao Qing,Zhongning Wang.The processing of phonological, orthographical, and lexical information of Chinese characters in sentence contexts: An ERP study[J].Brain Research,2011,1372,81-91.

[9] David J Kupfer,Ellen Frank,Mary L Phillips.Major depressive disorder: new clinical, neurobiological, and treatment perspectives[J].The Lancet,2012,379(9820),1045-1055.

[10] Bistricky,Atchley,Ingram,O'Hare.Biased processing of sad faces: An ERP marker candidate for depression susceptibility[J].Cognition & Emotion,2014,28(3),470-492.

[11] Philip van Eijndhoven,Guido van Wingen,Koen van Oijen,Mark Rijpkema,Bozena Goraj,Robbert Jan Verkes,Richard Oude Voshaar,Guillén Fernández,Jan Buitelaar,Indira Tendolkar.Amygdala Volume Marks the Acute State in the Early Course of Depression[J].Biological Psychiatry,2008,65(9),812-818.

[12] Cornelia Stoeckel,Patricia M. Gough,Kate E. Watkins,Joseph T. Devlin.Supramarginal gyrus involvement in visual word recognition[J].Cortex,2009,45(9),1091-1096.

[13]Osamu Abe,Hidenori Yamasue,Kiyoto Kasai,Haruyasu Yamada,Shigeki Aoki,Hideyuki Inoue,Kunio Takei,Motomu Suga,Koji Matsuo,Tadafumi Kato,Yoshitaka Masutani,Kuni Ohtomo.Voxel-based analyses of gray/white matter volume and diffusion tensor data in major depression[J].Psychiatry Research: Neuroimaging,2009,181(1),64-70.

 https://doi.org/10.1051/e3sconf/202021403033

[14] Qiyong Gong,Qizhu Wu,Cristina Scarpazza,Su Lui,Zhiyun Jia,Andre Marquand,Xiaoqi Huang,Philip McGuire,Andrea Mechelli.Prognostic prediction of therapeutic response in depression using high-field MR imaging[J].NeuroImage,2011,55(4),1497-1503.

[15] Norman A.S. Farb,Adam K. Anderson,Richard T. Bloch,Zindel V. Segal.Mood-Linked Responses in Medial Prefrontal Cortex Predict Relapse in Patients with Recurrent Unipolar Depression[J].Biological Psychiatry,2011,70(4),366-372.