Data Mining of Psychological Tendency and Health of Ideological and Political Students in Higher Vocational Tourism English Courses

Jin Zhou[1(✉)] and Wenjuan Xie[2]

[1] English Department, Sanya Aviation and Tourism Vocational College, Sanya 572000, Hainan, China
zhoujin55400@163.com
[2] Shaoyang University, Shaoyang 422000, Hunan, China

Abstract. In order to improve the quality of students' psychological tendency health data, this paper puts forward the research of mental tendency health data mining based on SVM. By cleaning the students' mental tendency health data and optimizing the SVM with particle swarm optimization algorithm, the classification of students' mental tendency health data is completed. According to the weighted results of mental propensity health data, the primary classification and feature recognition of large-scale mental propensity health data were carried out. Based on the optimal modeling of students' psychological tendency health data, the objective function of maximum difference of students' psychological tendency health data is defined. Maximum likelihood estimation is used to obtain the frequency distribution of the data in the Tourism English Course of higher vocational colleges. Experimental results show that the proposed method can not only improve the accuracy and efficiency of mental health data classification, but also control the integrity of mental health data over 80%.

Keywords: Support Vector Machine · Higher Vocational Colleges · Tourism English · Course Ideology and Politics · Psychological Tendency · Health Data Mining

1 Introduction

In recent years, the development of education in our country is progressing, and the importance of ideological and political education of tourism English courses in higher vocational colleges is also increasing. According to the development of national economic situation, society needs a large number of high-level skilled talents. Higher vocational education is an important part of higher education in our country, aiming at cultivating a large number of highly skilled talents with reasonable structure and good quality. The proportion of professional talents exported from higher vocational colleges will account for a large proportion under the requirement of economic structure. College students leave their parents and hometown to face the brand-new world as a social person, which is called "the second weaning period of life". At this particular stage, it is very easy to experience a wide range of maladjustment [1].

S. Wang (Ed.): IoTCare 2022, LNICST 501, pp. 178–190, 2023.
https://doi.org/10.1007/978-3-031-33545-7_13

In this period, the immature development of human nature, the ability of self-regulation and self-control, and the burden of employment and interpersonal relationship caused by the fierce competitive environment make most students' psychological pressure gradually increase. At the same time, most of the higher vocational students have poor academic performance and have not formed good study habits. Higher vocational students will face more psychological pressure. College students usually shoulder the family and even the country's hope, generally received a good education, should become the pillars of the country, contribute to the country's construction of their own talents. But the emergence of psychological problems led to a variety of tragic events, how to timely detect and intervene in the mental health problems of college students, is particularly important.

With the development of science and technology, the application of computer has gradually entered the operation and management of all walks of life. Through the psychological assessment system, college psychological counselors can quickly collect students' psychological conditions and make judgments, improve work efficiency and reduce work intensity. To a certain extent, it provides a great help, while the database also accumulated a large number of psychological data. But at present most of the psychological assessment conducted by colleges and universities are based on the basic psychological information of students access, query, statistics and backup operations. A large amount of psychological data is not analyzed in depth and not fully utilized. From it, we can mine some hidden information knowledge and grasp the trend of students' psychological development. Therefore, it can not effectively provide decision-making help for psychological counseling.

Wang Feng et al. [2] Considering that the traditional psychological stress testing method is mainly based on questionnaire survey and evaluation with the aid of professional equipment, which has high cost and great intrusion into the evaluated object, etc. Based on the analysis of perceptual data of smartphones, this paper studies the methods of assessing college students' psychological stress, extracts reasonable features from perceptual data, and puts forward a more efficient method of assessing psychological stress. Secondly, this paper introduces how to transform psychological stress assessment into classification problem and construct classification model by semi-supervised learning. Finally, the model is tested on the open data set StudentLife. The experimental results show that this method is superior to the baseline method in the aspects of psychological stress detection accuracy and recall. Zhou Xianyu et al. [3] Aiming at the problems of the traditional evaluation methods, such as low real-time evaluation, poor evaluation effect of single modal data, and biased social desirability response, an automatic evaluation method of college students' mental health based on multi-modal data fusion is proposed in this paper. The model can accurately assess the mental health status of college students. It has a good application prospect in the intelligent learning environment, and can provide decision-making basis and technical support for improving students' mental files and optimizing mental health services.

Data mining is the mining of information and knowledge from massive data. Data mining technology in the retail industry, financial industry, telecommunications and other industries has been fruitful. If data mining technology is applied to the development of mental health management system, a large number of fuzzy and random data information

in students' mental health archives can be processed by relevant algorithms. To a certain extent, it will help counselors to make judge and prevent students' psychology more scientifically and quickly, to guide and intervene psychology in time, to improve work efficiency and to reduce psychological events.

Based on the above research background, this paper uses support vector machine to design a mental health data mining method to ensure the healthy development of students' psychology.

2 The Design of Data Mining Method for Students' Psychological Tendency Health

2.1 Cleaning Student Mental Health Data

The data mining method of students' mental health is to clean up the data by two parts: data processing and output. The above two steps are implemented by concurrent [4, 5]. In the part of data processing, we complete the redundant processing, decontamination and determine the expiration time of the data. Using the timer to collect the data of students' mental tendency health in the buffer queue, complete the detection of students' mental tendency health data, and output the real data. When the existing data has expired, using comparative judge to judge the authenticity of students' mental health data, and transmit the output to the online platform of vocational colleges.

Through reading the cache queue, the conflict detection mechanism, the converter and the calculator, the mental propensity health data are processed.

The mental propensity health data processed by calculator were stored in EPC format. Among them, *EPC* represents the EPC code value of students' mental health data in Tourism English Course of higher vocational colleges. *RSSI* represents the signal strength of the mental health data. timestamp represents the time to gather data on students' mental health preferences. What expiretime represents is the expiration time of students' mental health data in the Tourism English course of higher vocational colleges. Z represents the value of mental health data of students' psychological tendency in tourism English curriculum of higher vocational colleges.

The data with the same characteristics were detected by conflict detection mechanism. First, the processor is used to collect a characteristic data C from the mental health data of students. According to whether there is the same data in the reading buffer queue, we can judge whether there is conflict in the thinking and politics of tourism English courses in higher vocational colleges. If there is data conflict, get the students' mental health conflict data T in higher vocational colleges tourism English curriculum ideological and political Z value. When the value of Z is 1, C. The expiretime and Z values of the expiration time of the mental propensity health data in the reading cache were modified, where:

$$T(\text{expiretime}) = C(\text{expiretime}) \tag{1}$$

If the value of Z is greater than 1, modify the value of Z in the read buffer of the student mental health conflict data so that:

$$T'(Z) = T(Z) + 1 \tag{2}$$

If there is no conflict, use the converter to process the students' mental health data q, get the caching requirements of the students' mental health data. Convert it to the appropriate format to insert at the end of the queue.

The function of the converter is to change the format of students' mental health data. Uniform data format for easy handling of data.

Using the signal-distance strength propagation model, the washing process time t and the washing process speed v of students' mental health data are calculated as follows:

The data of mental health tendency of students selected for the first time in two consecutive times were recorded as $RSSI_1$ and $RSSI_2$. Based on the signal-distance strength propagation model. The distance between the computing processor and the mental propensity health data at t_1 and t_2 time D_1, D_2:

$$D_1 = 10^{\frac{RSSI_1}{10-N}} \tag{3}$$

$$D_2 = 10^{\frac{RSSI_2}{10-N}} \tag{4}$$

N stands for empirical value.

The comparator is used to compare the given threshold γ and the mental propensity health data Z. According to the comparison results, the authenticity of label data is judged. When the data of students' mental propensity health is real, the data are exported and transmitted to the online platform of vocational colleges.

2.2 Designing an Algorithm of Students' Psychological Tendency Health Data Classification Based on SVM

Based on the cleaning of students' mental health data, the SVM training set is preprocessed to improve the training speed, and the SVM is trained to obtain the decision function. It can be seen that not all training samples work in the SVM classification process. Instead, only the training samples corresponding to the non-zero solution c_i of the dual problem, that is, the support vector, act on the decision function [6]. In other words, the complexity of the SVM decision for an unknown sample is $G(|l|)$. When $|l|$, the number of support vectors, is large, the estimator is large, resulting in a slow classification speed. Based on this, particle swarm optimization algorithm is used to reduce the support vector. After training SVM, the fuzzy membership vector of the set of support vectors is used as the particle in the particle swarm, and the average classification error of the test set is used as fitness function. The optimal set of support vectors is selected to reduce support vectors, so as to improve the classification speed of students' mental health data.

Particle swarm optimization algorithm searches for the solution by adjusting the position of particles. Assuming that the population $X = \{x_1, x_2, \ldots, x_n\}$ is composed of n particles in the D dimension search space, the current position of the i particle is $X_i = \{x_{i1}, x_{i2}, \ldots, x_{in}\}$. The current velocity of the particle is $V_i = \{v_{i1}, v_{i2}, \ldots, v_{in}\}$. The best position for particle i is $P_i = \{p_{i1}, p_{i2}, \ldots, p_{in}\}$. The best place for all particles to pass is $P_g = \{p_{g1}, p_{g2}, \ldots, p_{gn}\}$. The i-particle at time $t + 1$ is:

$$v_{id}^{t+1} = v_{id}^{t} + \zeta_1 \psi_1 \left(p_{id}^{t} + x_{id}^{t}\right) + \zeta_2 \psi_2 \left(p_{gd}^{t} - x_{id}^{t}\right) \tag{5}$$

$$x_{id}^{t+1} = x_{id}^{t} + v_{id}^{t+1} \tag{6}$$

Among them, $1 \leq d \leq D$; $1 \leq i \leq n$; ψ_1 and ψ_2 are random numbers uniformly distributed on the interval of (0,1). ζ_1 and ζ_2 are called learning factors.

In particle swarm optimization algorithm, each particle represents a solution, and particle swarm optimization algorithm is applied to support vector machine to reduce support vector. The number of support vectors obtained by trained SVM is the dimension of particles. Assume that the range of membership of these samples is $[R_{\min}, R_{\max}]$. This range is chosen as the position range of the initialized particle, and the weight vector of the sample calculated is regarded as a particle in the initialized particle swarm space. Each particle has its position and velocity. The position indicates the membership degree of the sample. The velocity changes the membership degree and sets a threshold. Particle output, when the sample membership is greater than the threshold value, so that its membership to maintain the original value, that the sample was selected. Otherwise, its membership is assigned a value of 0, indicating that the sample is not selected. Therefore, the problem of selecting support vectors is transformed into the PSO optimization problem of selecting optimal particles.

The training set is divided into two parts: one for the training set and the other for the testing set. The training sample set is processed and the training SVM gets the support vector set. The average classification error of the test set is used as the fitness value of the particles. Define the fitness function as:

$$Fitness = \frac{1}{H} \sum_{i=1}^{H} (y_i - g_i)^2 \tag{7}$$

H is the number of samples in the test set, y_i is the predicted value, and g_i is the actual value. The smaller the fitness of the particles, the better.

According to the above process, the classification algorithm of students' mental health data is designed using SVM.

2.3 Calculate the Correlation Between the Characteristics of Students' Psychological Tendency and Health Data

In the process of mining students' psychological tendency health data in the ideological and political ideological and political courses of tourism English courses in higher vocational colleges, the preliminary classification and feature identification of the psychological tendency health data are carried out. The degree of association between various types of features can be calculated. The specific steps are detailed as follows:

Suppose that F_{ij} represents the frequency of the appearance of i in Tourism English Course Thought and Politics D_j. N_i represents the number of times a mental health data feature appears. F_{ik} is the probability of the appearance of i in Tourism English Course Ideological and Political k.Then use the following formula to express the weighted results of the psychological predisposition health data:

$$\chi_{ij} = \frac{U(w) \times F_{ij}}{N_i \otimes D_j} \otimes F_{ik} \tag{8}$$

In the formula, $U(w)$ represents the constant coefficient.

Suppose, the maximum value of information entropy of each feature class is represented by $\sigma(j)$. W represents the number of characteristics of mental disposition health data. Φ represents the total dimensionality of the candidate feature set for mental dispositional health data. f_e represents the th feature of the e-th type of data. Then use the following formula to combine the weighting method of inter-class and intra-class information entropy distribution of feature items [7] to carry out preliminary classification and feature identification of large-scale psychological tendency health data:

$$\lambda(S_{ac}) = 1 - \frac{\max(S_{ac}) \times l}{\phi \otimes coff} \cdot f_e \frac{\Phi(\wp(I) * W)}{\sigma(j) \otimes W} \tag{9}$$

In the formula, $\max(S_{ac})$ represents the correlation between the psychological disposition health data and the maximum correlation coefficient. l represents the feature selection pros and cons evaluation function. ϕ stands for the feature recognition rate of mental tendency health data. *coff* represents the weight that balances the maximum recognition rate and feature dimension. $\wp(I)$ represents the number of categories.

Suppose that any two mental disposition health data are represented by a_i and a_j. ς represents the threshold between the given mental disposition health data distances. The distance from a_i to a_j in the mental health data is smaller than that of ς. However, the distance between a_i and the mental health data other than a_j is greater than ς. Then use the following formula to calculate the correlation function between the psychological tendency health data:

$$A = \sum \frac{V \times (a_i, a_j)}{\varsigma \times Q(\varsigma_i)} \tag{10}$$

In the formula, V represents the probability of the same feature appearing in the health data of different psychological tendencies. $Q(\varsigma_i)$ stands for independence of mental predisposition health data.

Suppose, a one-dimensional mental orientation health data sequence is represented by n. N''' represents the number of vector points in the phase space reconstruction of mental orientation health data. The similarity between any two psychological tendency health data is defined as the maximum difference between the two vectors, which is expressed by the following formula:

$$K_m(r) = \frac{(x_i - x_j)}{N''' \otimes n} \otimes X(E) \tag{11}$$

In the formula, $X(E)$ represents the similarity between the health data of various psychological tendencies.

To sum up, it can be shown that in the process of mining students' psychological inclination health data in the ideological and political ideological and political courses of tourism English courses in higher vocational colleges, a preliminary classification and feature identification of the large-scale students' psychological inclination health data is carried out. Calculating the correlation between various types of features lays the foundation for the realization of the data mining of students' psychological tendency health.

2.4 Optimal Modeling of Students' Psychological Tendency Health Data

According to the correlation between the characteristics of students' psychological tendency health data, the optimal modeling is carried out on the students' psychological tendency health data. If the mental time series set of different stages in the student's learning process is defined as $\{X(t), t = 1, 2, \cdots, n\}$. Because the rebellious psychology of college students has obvious differences in the form of expression, it is necessary to reconstruct the phase space of the formation process of students' psychology. So get the matrix:

$$\begin{bmatrix} x(1)\ x(1+\tau)\cdots x(1+(m-1)\Omega) \\ x(2)\ x(2+\tau)\cdots x(2+(m-1)\Omega) \\ \cdots \qquad\qquad \cdots \\ x(k)\ x(k+\tau)\cdots x(k+(m-1)\Omega) \end{bmatrix} \tag{12}$$

In the formula, m represents the complexity of students' psychology. Ω represents the duration of the student's mentality, satisfying:

$$k = n - (m-1)\Omega \tag{13}$$

Consider each row in the matrix as a variety of influencing factors of students' psychological changes, the total number is k. According to the degree of conflict, the index of the affected categories of each influencing factor in the students' psychological performance status is obtained, and the psychological performance status of different impact categories is obtained. For m-dimensional students, the total number of different mental performance state sequences corresponding to the degree of psychological complexity is m. Assuming that the probability of occurrence of b kinds of different mental performance state sequences is $p_1, p_2, \cdots, p_k$, then according to the form of Shannon entropy, the probability of students' mental state sequence is sorted, then the permutation entropy is:

$$H_{PE}(m) = -\sum_{j=1}^{b} p_j \mathrm{In} p_j \tag{14}$$

Integrate the state sequence given by the above formula. The following formula is used to express the emotional characteristics of behavior corresponding to the psychological state of students at this stage:

$$0 \le H_{PE} = \frac{H_{PE}}{\mathrm{In}(m)} \le 1 \tag{15}$$

Suppose $\{x(t), t = 1, 2, \cdots, N\}$ represents a time series set of behavioral emotional characteristics of students' mental states. Due to the complexity of the behavioral tendency of college students dominated by rebellious psychology, it is necessary to establish a relational expression with the duration t_i of rebellious psychology. Select the duration time series $X(t_i + \tau)$ to form a new duration phase point column $Y(t)$. The mental duration t_i is determined by calculating the correlation of $X(\tau)$ and $Y(\tau)$.

According to the behavioral emotional characteristics corresponding to the psychological state of students, the ant theory is used to model the health data of students' psychological tendency. It is assumed that t_{ij} represents the behavioral tendency strength of college students' psychology in the τ period. $\Delta t_{i,k}(\tau)$ represents the pheromone that the psychology of college students searched by ant k has hindered the tendency behavior. $\chi(0 \leq \chi \leq 1)$ represents the influence of student psychology on tendentious behavior. Use the following formula to calculate the intensity of behavioral tendency to search for the next period, the formula is:

$$t_{ij}(\tau + 1) = \chi \cdot t_{ij}(\tau) + \sum \Delta t_{i,k}(\tau) \tag{16}$$

Assuming that l_k represents the length of the path that the k th ant traveled in this cycle, there are:

$$\Delta t_{ii,k}(\tau) = \frac{R}{L_k} \tag{17}$$

Among them, R is a constant, assuming that ε_{ij} represents the visibility of the path (i, j), which is usually taken as $\frac{1}{d_{ij}}$. Set d_{ij} to represent the length of path (i, j). The corresponding importance of path visibility is $(\beta \geq 0)$. The relative importance of the path trajectories is $\alpha(\alpha \geq 0)$. U represents the feasible point set. The migration probability of ant k in the τ time domain is $p_{ij,k}(\tau)$. Then use the following formula to define $p_{ij,k}(\tau)$:

$$p_{ij,k}(\tau) = \begin{cases} \frac{[t_{ij}(\tau)]^{\alpha}[\varepsilon_{ij}]^{\beta}}{\sum\limits_{l \in U}[t_{ij}(\tau)]^{\alpha}[\varepsilon_{ij}]^{\beta}}; & j \in U, \\ 0 & ; j \notin U. \end{cases} \tag{18}$$

Based on the above elaboration, the objective function of the influence of students' psychology on behavior is calculated by the following formula:

$$\min B = g(x) \cdot x \in [a, b] \tag{19}$$

Among them, $g(x)$ is the optimization function.

The optimization model of college students' mental health data is constructed using the following formula, namely:

$$\begin{cases} t_j(\tau + 1) = \varphi \cdot t_j(\tau) + \sum\limits_{k} \Delta t_j \\ \Delta t_j = R/L_j \end{cases} \tag{20}$$

In the above formula, Δt_j represents the increase in the regional attraction strength of the j th ant in this cycle. L_j represents the amount of change in $g(x)$ in this cycle, which is defined as $g(x + r) - g(x)$. The optimization of the function is carried out with the help of the continuous movement of m ants. When $\varepsilon_{ij} \geq 0$ is satisfied, ant i transfers from its nearest neighbor I to the neighbor of ant j according to probability p_j. When $\varepsilon_{ij} \leq 0$ is satisfied, the ant i searches the nearest neighbor I, and the search radius is r, that is, each ant or the place where the other ants are located. Or perform a near-neighbor search and gradually converge to the global optimal solution of the problem.

By solving the problem of students' psychological predisposition behavior disorder, the optimal modeling of students' psychological predisposition health data was carried out.

2.5 Realization of Students' Psychological Tendency Health Data Mining

Suppose that the frequency of occurrence of the i-th feature is represented by $\lambda_i^{(c)}$. ∂_i represents the frequency vector of the mental orientation health data i. $V_m(r)$ represents the maximum difference between the two psychological orientation health data. Its objective function is defined by the following formula:

$$\arg\min A = \frac{\Omega(y(c) - y(d))}{V_m(r) \otimes \lambda_i^{(c)}} \otimes \partial_i \tag{21}$$

In the formula, $y(c)$ represents the amount of information including c data. $y(d)$ represents the frequency of d features. Ω stands for the best classification scheme.

Suppose that $p(d|\hat{c})$ represents the probability of generating students' psychological tendency health data d in the ideological and political course $\hat{c}$ of tourism English in higher vocational colleges. $p(c|\hat{c})$ represents the probability of generating the health data $\hat{c}$ of all students' psychological inclinations in the students' psychological inclination health database in the ideological and political c of the tourism English course in vocational colleges. The frequency distribution of each data in the ideological and political $\hat{c}$ of the tourism English course in higher vocational colleges is obtained by using maximum likelihood estimation, which is expressed as:

$$\hat{c}(q) = \frac{p(d|\hat{c}) \otimes p(c|\hat{c})}{LLR \otimes \ell_{(W)} * t_m^{(d)}} \times \frac{\mu(S)}{\sigma(A)} \tag{22}$$

In the formula, *LLR* represents the log-likelihood ratio obtained by taking the logarithm. $\ell_{(W)}$ represents the current feature correction set. $t_m^{(d)}$ represents the distribution of the $i+1$-th important feature symbol after searching for d second-important feature symbols. $\mu(S)$ stands for calculating the objective function value from the classification result and the dimension of the feature subset. $\sigma(A)$ represents the minimum feature dimension.

Assuming that $\arg\max D_E$ represents the optimization objective function under a certain number of categories, it can be expressed by the following formula:

$$\arg\max D_E = \Omega(d) * \mathrm{X}(Z)\frac{\delta(u)}{\beta(q)} * K(N) \tag{23}$$

In the formula, $\Omega(d)$ represents the similarity data execution value between the cluster centers. $\mathrm{X}(Z)$ represents the similarity between two close cluster centers. $\delta(u)$ represents the probability of symbolic distribution constructed from the feature space u of students' psychological tendency health data. $\beta(q)$ represents the relationship between students' psychological predisposition health data. $K(N)$ represents the optimized feature subset from the original student mental orientation health data features.

According to the above elaboration, the following formula is used to complete the mining of students' psychological tendency health data:

$$\partial(\hat{c})\frac{\arg\max D_E * \hat{c}}{\delta(u)} \tag{24}$$

To sum up, by defining the objective function of the maximum difference of students' psychological tendency health data. The frequency distribution of each data in the ideological and political courses of tourism English courses in higher vocational colleges is obtained by using maximum likelihood estimation, and the data mining of students' psychological tendency and health is realized.

3 Experimental Comparative Analysis

3.1 Experimental Data Source

The data used in this paper come from 941 records in a university student psychological survey database. The SCL-90 scale is used to measure the mental health of college students.

The mental health indicators recorded in the database included somatization, interpersonal sensitivity, obsessive-compulsive symptoms, anxiety, depression, hostility, paranoia, phobia, psychosis and positive average scores.

The psychological data used in this paper have the following characteristics:

1. Being scientific: The SCL-90 Psychological Measurement Scale used in this school is widely used in the professional field of psychology and has high popularity. Moreover, it can quantify the students' behavior and psychological symptoms, and the data obtained is very professional and scientific.
2. High reliability: The measurement data of this part are stored in the database of relevant parts of the school. Data collected under the unified leadership of the school student department, distributed through questionnaires, or scattered online are more reliable.
3. Data imbalance: 941 student records used in this article measured the results of "abnormal" data only 76, measuring the results of "normal" data only 865. That is, the original data results are "abnormal" and results are "normal" data serious imbalance. This will affect the feasibility of data mining and the accuracy of test results.

3.2 Data Purification

Data cleansing is to delete the untrue and incomplete data from the original data, which can not meet the quality requirements of decision tree algorithm. Thus, the quality of data analysis can be ensured and more accurate data mining results can be obtained.

1. Purge False Records: Because the SCL-90 test has 90 multiple choice questions. After statistics that the test time is less than 4 min of the record is not earnest answer left unreal data. This part of the record is therefore culled.
2. Purifying unrelated attribute factors: Because the scope of the census was all college students, the data of students' name, sex, age, specialty, grade and time consumption had no effect on the model. It will increase the time and cost of data analysis and reduce the efficiency of data. Therefore, these factors are not considered in constructing the decision. After data purification, 886 relatively reliable records were obtained. After the adjustment of the database, the score table of each factor in the sample set

of decision tree training is obtained. The scores of each field in the sample set of decision tree training correspond to the scores measured by SCL-90, that is, the score of attribute factor of mental health index.

3. Purify the status of students: Because the SCL-90 is not always accurate, it can not be directly used as a basis for the diagnosis of psychological crisis, only as a reference. Therefore, a status attribute is attached to mark whether the students have psychological crisis. State refers to the mental health education center after the relevant workers screened to confirm the existence of psychological crisis.

After data purification, 886 relatively reliable records were obtained. Among them, 100 were selected as test samples and 786 as training samples.

3.3 Result Analysis

Data mining method based on smartphone perception, Ecological Instantaneous Assessment Theory and this paper's method are used to carry out the mental health data mining experiment. The results of the accuracy and efficiency tests for classifying student mental health data are shown in Figs. 1 and 2.

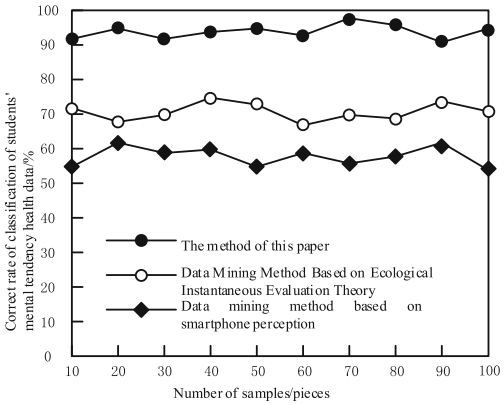

Fig. 1. Correct rate of classification of students' mental tendency health data

It can be concluded from Fig. 1 and Fig. 2 that the accuracy and efficiency of this method are higher than those of traditional data mining methods based on smartphone perception and ecological instantaneous assessment. The main reason is that SVM is used to classify and recognize the data features of mental health of students, and to calculate the degree of association among the features. Thus greatly improved the accuracy and efficiency of this method in the classification of mental health data.

On the basis of ensuring the accuracy and efficiency of the classification of students' psychological inclination health data. When mining students' mental health data, it is also necessary to ensure the integrity of the data, as shown in Fig. 3.

According to the results of Fig. 3, the data integrity of mental health data mined by this method is over 80%, and the data integrity of mental health data mined by smartphone is

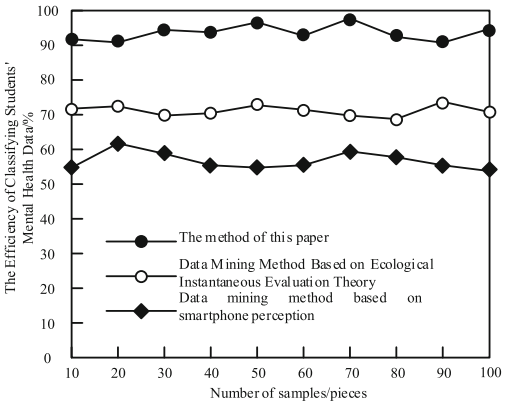

Fig. 2. The Efficiency of Classifying Students' Mental Health Data

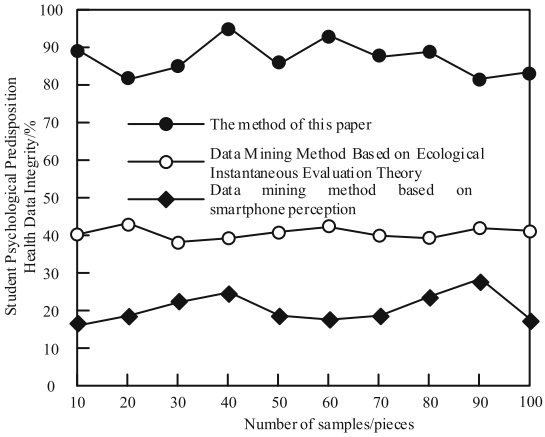

Fig. 3. Completeness of Student Psychological Predisposition Health Data

under 30%. When the data mining method based on ecological instantaneous assessment theory is used to mine students' mental health data, the data integrity fluctuates at 40%. Compared with the test results of three different methods, this method can be used to mine students' mental health data with higher integrity.

4 Conclusion

This paper puts forward a data mining research on mental health of students in Tourism English Course based on SVM. Experimental results prove that this method can improve the classification efficiency of mental health data. It can guarantee the integrity of mental health data to more than 80%. But there are still many deficiencies in this study. In the future research, we hope to introduce principal component analysis to extract the characteristics of students' mental health data and ensure the quality of data mining.

Fund Project. 2022 Hainan Higher Education and teaching reform research project: Research on the generation path of integrating curriculum ideology and politics into tourism professional English Curriculum -- a case study of Higher Vocational Colleges in Hainan Province (Project No.: Hnjgzc2022–111).

References

1. Wei, Z., Yong, W., Ning, Z.: Improved student psychology based optimization algorithm using hybrid strategy. Appl. Res. Comput. **39**(6), 1718–1724 (2022)
2. Rahman, M.M., Singh, M.K.M.: English Medium university STEM teachers' and students' ideologies in constructing content knowledge through translanguaging. Int. J. Biling. Educ. Biling. **25**(7), 2435–2453 (2022)
3. Shi, Q., Cai, N., Jiao, W.: Monitoring and evaluating college students' mental health based on big data analysis. Am. J. Health Behav. **46**(2), 164–176 (2022)
4. Zhou, X., Liu, L., Chen, Y., et al.: Research on design and application of an automatic assessment model for college students' mental health based on multimodal data fusion. E-educ. Res. **42**(8), 72–78 (2021)
5. Mou, K., Xu, B.: Research on emotional analysis platform of college students' public opinion information based on big data. In: 2022 IEEE 6th Information Technology and Mechatronics Engineering Conference (ITOEC), vol. 6, pp. 2105–2110. IEEE (2022)
6. Yu, Y., He, A., Zheng, S., et al.: Association between school bullying and mental health of adolescent students in low-income and middle-income countries. Chin. J. School Health **06**, 842–875 (2021)
7. Liu, Y., Chen, Q.: Personalized information mining of mobile terminal users based on tag mapping. Comput. Simul. **39**(1), 177–180, 208 (2022)