A Behavior Sequence Based Psychological Feature Extraction Model for Students in English Teaching

https://doi.org/10.3991/ijet.v18i02.35503

Xiaoyu Cai[1](✉), Jarrytt R. McDonald[2]
[1]Foundation Department, Shangqiu Polytechnic, Shangqiu, China
[2]Shenzhen Jinbi Experimental School and Meten English, New Jersey, USA
18403151@masu.edu.cn

Abstract—With the rapid development of science and technology, intelligent teaching has been gradually popularized in the campus, and the change of relevant teaching modes has also been carried out. In the new English teaching, it is necessary to analyze the students' psychological behavior. In order to get accurate results, the research captures the students' behavior as a behavior sequence, and uses the improved combined kernel support vector machine to extract features on the basis of approximate entropy and local mean decomposition. At the same time, it is compared with the traditional support vector machine, the traditional least square vector machine and the back-propagation neural network algorithm. The experimental results show that the improved algorithm has higher accuracy, recall and accuracy than the other three algorithms. The overall accuracy of the improved algorithm is 82.53%, the highest accuracy is 90.12%, the overall recall is 58.64% and the overall precision is 86.92%. The experimental results show that the improved algorithm has a significant improvement in performance, and can be applied to the extraction of students' psychological characteristics. Therefore, the algorithm can be used in conventional blended English teaching, which is convenient for teachers to carry out English teaching more efficiently.

Keywords—feature extraction, approximate entropy, local mean decomposition, support vector machine

1 Introduction

In the context of the current smart city, relevant teaching work is also gradually becoming intelligent. In the case of a large number of students, the traditional teaching model is easy to ignore one thing and lose the other, resulting in some students being left out and unable to learn more knowledge [1]. English teaching has always been an important part of smart teaching. English is not only a compulsory subject from primary school to university, but also an important bridge for international communication. Therefore, the importance of English teaching cannot be ignored [2]. In the current English teaching, the way of intelligent teaching is usually mixed teaching, and this teaching mode needs to teach students according to their aptitude to achieve its best effect [3]. In order to successfully carry out smart English teaching, it is necessary to form corresponding sequences according to students' behaviors and extract psychological features [4]. This is because identifying and extracting students' psychological characteristics is conducive to teachers' more detailed and accurate understanding of students, so that teachers can develop more appropriate English teaching programs for each different student. Under the current research background, the technology of feature extraction using SVM has become increasingly mature, and has its feasibility in various improvements [5]. In view of this, the experiment improves the support vector machine (SVM) on the basis of approximate entropy and local mean decomposition to form a least square vector machine with combined kernel function, and analyzes and explores the effect of the improved machine algorithm on feature extraction, so as to apply the algorithm to the actual psychological feature extraction. SVM is a machine learning method model. Based on statistical theory, it classifies data into binary categories, and uses hinge loss function to calculate empirical risk. The advantages of SVM can balance the ability to find complex models and learn even if there is no sufficient large amount of sample data.

The article is mainly divided into five parts, the first part is to introduce the background of the article, as well as other scholars related research review; The second part is about the theoretical basis of the feature extraction method used in this paper, that is, the approximate entropy and LMD are described in detail; The third part introduces the method itself, that is, the combined kernel function is used and combined with LSSVM. With the support of this function, machine learning can extract accurate features through continuous training iterations. The fourth part is to show the results of the simulation experiment and the analysis of the results. The fifth part is to summarize the article and propose the shortcomings of the study.

2 Related work

Many scholars at home and abroad have done a lot of research on feature extraction for classification learning machine algorithms such as support vector machine. Sun J and others artificially combined the dynamic financial distress prediction with time weight, using a time weighted SVM method, using dual expert voting integration, and using error based expert voting and time-based expert voting and external combination. The experimental results of sample tests show that this method is more accurate for the dynamic financial distress prediction when time changes [5]. Simian and other scholars used the mixed method of multi-core support vector regression, compared and analyzed it with the single core support vector regression method, and used gene expression programming and extreme learning machine to fit and predict the sample data. The results showed that the multi-core support vector regression method has higher accuracy and efficiency [6]. Zhao's team used the least squares support vector machine to compare the grid search and genetic algorithm on the basis of self-excitation threshold autoregression and other methods, so as to carry out ant colony optimization on the least squares support vector machine. The experimental results show that the improved least squares support vector machine is significantly better than all kinds of traditional data prediction, and the accuracy is significantly improved [7]. Ftunio's team studied different kinds of classifiers and used them to fit and classify the sample data. The final results show that the adaptive improved classifier can effectively classify the data to obtain more accurate results [8]. In order to accurately extract the characteristics of the future financial situation of stock exchange trading companies, scholars such as Vu l t used a combination of SVM and autoregressive methods to make predictions based on factor analysis and F-score analysis. Finally, the accuracy of the prediction results is high and can be applied to practical operations [9]. Shetty SH and other scholars have developed two binary logistic regression models M1 and M2. At the same time, combined with the least square vector machine, the predictive diagnostic ability of the model is tested with the help of the subjects' operating curve, area under the curve, sensitivity, specificity and annual accuracy. The results show that the predictive diagnostic accuracy of the new model is high [10]. Niu's team proposed a method based on Improved TOPSIS method and improved ant colony algorithm to optimize LSSVM and built a model for feature extraction of retail enterprise operation efficiency based on it. The entropy weight method was used to objectively weight the indicators, and then based on the improved TOPSIS method, the reverse problem in the evaluation process was eliminated. The experimental results show that the model can provide scientific and effective evaluation results [11].

Xue and other scholars combined the improved particle swarm optimization algorithm with least squares support vector machine to improve the accuracy of predicting the compressive strength potential of concrete. Using the experimental data of concrete compressive strength in the literature, the performance of the proposed impso-lssvm model is verified and evaluated. At the same time, five other algorithms are used to compare each algorithm. The experimental results show that the improved algorithm has higher accuracy, and the accuracy can be applied to predict the compressive strength of concrete [12]. Kim et al. combined support vector machine with recursive features to form a support vector machine recursive feature elimination and selection technology. This technology uses the ratio of logarithmic differential expressions to improve scalability based on the normalized T statistics of intensity correlation. At the same time, each iteration will exclude the feature set that is less dominant in correlation and redundancy from the corresponding set of features. The experimental results show that the accuracy of RNA SEQ feature extraction using this technique is very high [13]. Liao proposed a robust face feature extraction method based on dynamic depth sparse representation in order to use face extraction to estimate facial age and psychological related features. This method combines the characteristics of active appearance model, local binary pattern, Gabor and bionic features, fully considers the thinking mode of human target recognition, the similarity of adjacent ages and the classification principle of sparse signal representation, and proposes a two-factor analysis method to separate human face recognition factors to reduce the interference of face recognition factors. The experimental results show that this method has strong discrimination and robustness, and is superior to the existing age estimation methods [14]. Shukla's team, based on the characteristics of skin electrical activity, tracked skin electrical activity to obtain relevant emotional indicators of the goal. Three feature selection methods, joint mutual information, conditional mutual information maximization and dual input symmetric correlation, and machine learning techniques were used to systematically compare the characteristics of these skin electrical activities. Experimental results show that this feature extraction method is superior to all other feature groups, including the most commonly used skin conductance response related features [15]. In order to detect the fatigue state of athletes and prevent athletes from excessive fatigue and sports injury, Zhang et al. Established a fatigue sports detection model based on support vector machine according to the acceptance and rejection criteria of sequential forward floating selection algorithm and different combinations of characteristic parameters. Taking the classification performance of the established SVM detection model as the evaluation standard, and the sequence floating forward selection algorithm as the search strategy, the experiment established an optimal selection algorithm for fatigue characteristic parameters. Based on paired sample t-test and variance analysis, the comprehensive effects of individual differences and fatigue exercise on sports behavior and eye movement characteristics were analyzed and quantified. The experimental results show that the algorithm can apply PERCLOS criterion to the training of convolutional neural network, so as to recognize the fatigue state of face according to the comprehensive features of face, and improve the robustness of the algorithm [16].

From the above research, it can be seen that many scholars have been widely used in feature extraction and related data prediction of SVM and other algorithms, but the research on applying SVM feature extraction to students' psychology in English teaching is less. On the basis of approximate entropy and local mean decomposition, SVM is improved to form a new feature extraction model, and compared with the other three algorithms. It is expected that the proposed algorithm can accurately and efficiently extract features for the psychological characteristics of English teaching students.

3 Feature extraction of approximate entropy and LMD

In smart English teaching, the extraction of students' psychological characteristics is more important. Only by extracting their psychological characteristics can we carry out corresponding teaching methods and teaching recommendations. The extraction of psychological features needs to study the target behavior sequence, so it needs to adopt a feature extraction algorithm that can recognize and measure the sequence.

Approximate entropy is an algorithm for measuring the complexity of time series. Approximate entropy has low requirements for sequence length, good anti-interference and anti-noise capabilities, and fast operation speed. It has strong applicability for both deterministic signals and random signals [17]. Approximate entropy usually uses a non-negative number to represent the complexity of time series. The larger the entropy, the more complex the series is. Conversely, the smaller the entropy, the higher the self-similarity of the series.

The approximate entropy has three parameters, namely (m, r, N), where m is the embedding dimension, r is the similarity tolerance, and N is the data length. Generally, m is taken as 2, and r is taken as 0.1–0.25 times the standard deviation of the original data sequence [18]. The research needs to select the value of r according to the actual situation of the simulation experiment. The approximate entropy-based algorithm is shown in Figure 1.

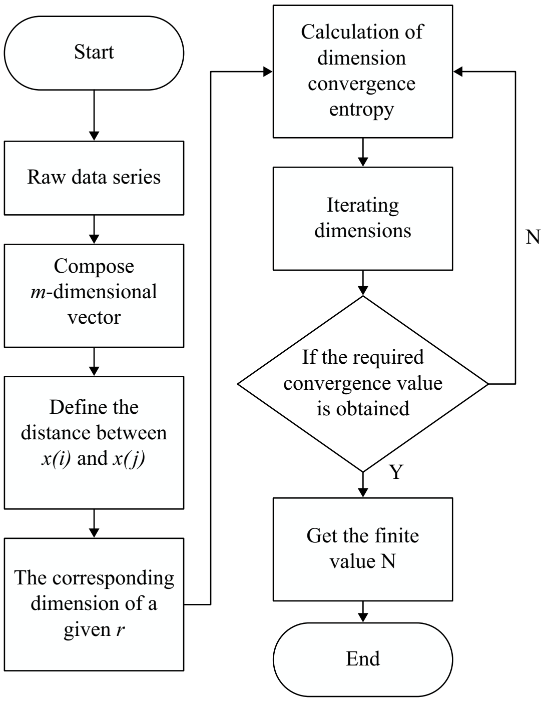


Fig. 1. Basic algorithms of approximate entropy

Set the time series with length as $\{u(i),\ i = 1, 2, \ldots N\}$, then reconstruct the m dimension vector as shown in equation (1). The value in the new set X_i in equation (1) is 1, 2, … n, $n = m + N-1$

$$X_i = \{u(i), u(i+1), \ldots, u(i+m-1)\} \tag{1}$$

Calculate the distance between vector x_i and another vector x_j, as shown in equation (2). In formula (2), k is a constant determined according to the required conditions of simulation.

$$d = \max \left| u(i+j) - u(j+k) \right|, k = 0,1,2 \ldots m-1 \tag{2}$$

For each vector, count the number of $d \leq r$ and the ratio of this number to the total distance, and record the ratio as $Cm(r)$. Take the logarithm of $Cm(r)$, and then calculate the average value of all, which is recorded as $\Lambda^m(r)$, as shown in formula (3).

$$\Lambda^m(r) = \frac{1}{m+N-1} \sum_{i=1}^{m+N-1} \ln Cm(r) \tag{3}$$

Add 1 to the value of *m* and repeat the calculation of formula (1) (2) (3) to obtain the corresponding $Cm + 1(r)$ and $\Lambda^{m+1}(r)$ at this time, the approximate entropy is obtained as shown in formula (4).

$$ApEn = \sum_{N\to\infty} \Lambda^{m}(r) - \Lambda^{m+1}(r) \tag{4}$$

Local mean decomposition (LMD) is a commonly used feature extraction algorithm. LMD was originally used to simulate and check power grid faults. With the development of science and technology, LMD has been widely used in many fields. LMD can transform a complex original signal into a product function component with instantaneous significance through a series of operations, and the resulting component frequencies are arranged from high to low [19].

After LMD decomposition, the original signal will be decomposed into envelope signal and pure FM signal, and the two signals will be multiplied to obtain a physically meaningful product function component. The obtained signal is calculated circularly until all the product function components are separated or reach the termination condition. After the original signal is obtained, all extreme points of the original signal $x(t)$ are determined, which are represented by m_1, m_2, … m_i in turn. At this time, the average value a_i and the envelope estimation value b_i are shown in equations (5) and (6) respectively.

$$a_i = \frac{m_i + m_{i+1}}{2} \tag{5}$$

$$b_i = \left| \frac{m_i - m_{i+1}}{2} \right| \tag{6}$$

The moving average method is used to smooth all a_i and b_i to obtain the local mean function $af(t)$ and the local envelope function $bf(t)$. The local mean function $af(t)$ is separated from the original signal, and formula (7) is obtained after operation. In equation (7), $hf(t)$ is the new function obtained by separation.

$$hf(t) = x(t) - af(t) \tag{7}$$

Use the new function $hf(t)$ to demodulate the envelope function, as shown in equation (8). Equation (8) $sf(t)$ is a new demodulation function.

$$sf(t) = \frac{hf(t)}{bf(t)} \tag{8}$$

Perform cyclic processing on equations (7) and (8) until $sf(t)$ is converted into pure FM signals, and the envelope estimation function $sf(t)$ value is 1 at this time. Then the corresponding envelope signal is obtained as shown in equation (9).

$$bf`(t) = \sum_{i=1}^{bf(t)} bf(t_i) \cdot \prod_{j=1}^{n} bf(t_j) \tag{9}$$

At this time, the product function component is a single component AM FM signal. At this time, its instantaneous frequency $F(t)$ can be calculated according to the envelope signal, as shown in equation (10).

$$F(t) = \frac{1}{2\pi}\frac{d[ar\cos(sf`(t))]}{dt} \tag{10}$$

The product function component is separated from the original signal and a new signal is obtained. As the high-frequency oscillation is gradually separated, the original signal is gradually smoother. The new signal is processed circularly until the new signal is monotonic. The basic flow of the whole algorithm is shown in Figure 2.

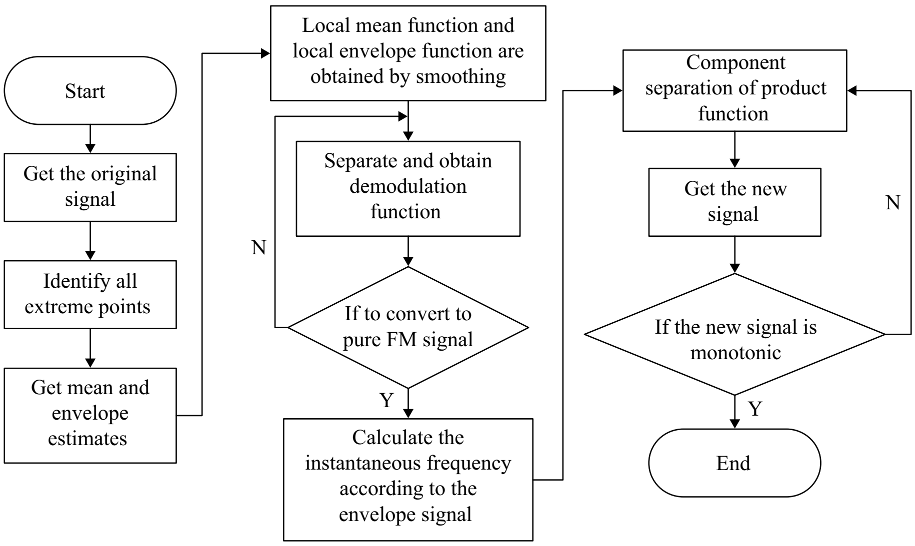


Fig. 2. Basic algorithms of LMD

For feature extraction, the more complex a behavior sequence is, the greater the probability it will produce new patterns, which will make the extraction of eigenvalues more complex and difficult. Psychological feature extraction of individual behavior is usually more difficult than simple physical signal extraction, so the experiment combines approximate entropy and LMD to increase the accuracy and efficiency of feature extraction.

4 Research on feature extraction methodology based on SVM and combined kernel function

When constructing feature extraction model, it is usually necessary to use appropriate machine learning algorithm. Because the approximate entropy and LMD used in the experiment contain vectors, support vector machine is a better choice. SVM is a machine learning method model, which classifies the data on the basis of statistical theory, and uses the hinge loss function to calculate the empirical risk. The advantage of SVM is that it can balance the ability to find complex models and learn even if there is no sufficient large amount of sample data. [20]. For SVM classification, sigmoid function is usually used, as shown in equation (11), θ is the mapping of sigmoid function.

$$h_{\theta}(x) = \frac{1}{1+e^{-\theta^{t}}} \tag{11}$$

The three key points of SVM are classification interval, duality and kernel, which determine the practicality of SVM, the principle of minimizing structural risk and the separation of dimension space from it [21]. If there is an initial training set as $\{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n), y \in [-1, 1]\}$, two different samples can be distinguished by a linear function. At this time, it is assumed that the initial training set can be linearly divided by the hyperplane, and the interval between two different categories of samples should be the maximum when the optimal hyperplane is established. The classification principle is shown in Figure 3.

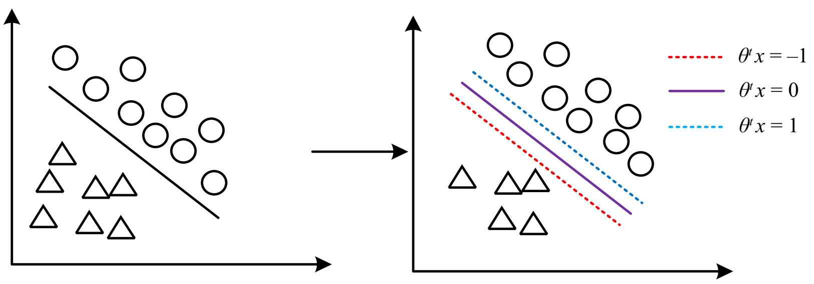


Fig. 3. Classification principle of SVM

When the samples are classified, if these data are non-linear, they will be transformed into a linear problem in high-dimensional space through spatial transformation, and an optimal classification plane will be selected to solve it. At the same time, the dual method will reduce the complexity of spatial dimension upgrading and make it simpler [22]. The dual characteristic makes it possible to obtain the corresponding classification function and operate it in the high-dimensional feature space only by performing the sample inner product operation. The principle of obtaining the optimal hyperplane is shown in Figure 4.

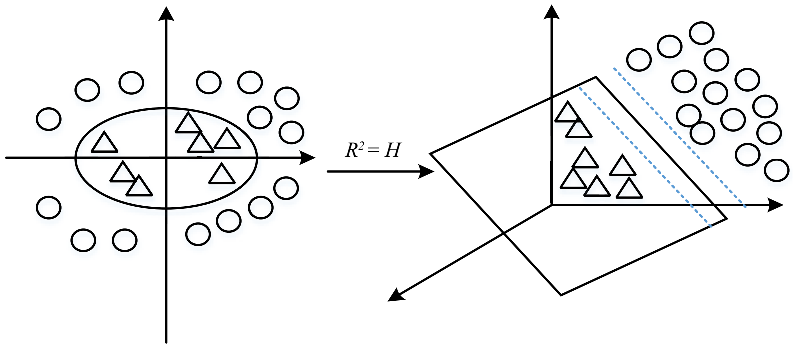


Fig. 4. Optimal classification hyperplane

Based on the idea of kernel method, SVM can map the data that is difficult to be separated by plane into high-dimensional space to obtain the classification plane through the method of feature transformation. Find the kernel function $K(x_i. x_j)$ in the input space. The dot product operations $\lambda(x_i)$ and $\lambda(x_j)$ of the high-dimensional feature space are replaced by the low-dimensional space. The decision function obtained by the dual transformation is shown in equation (12), and the dual transformation result is shown in equation (13). In equations (12) and (13), $i = 1, 2, ..., n$, $\sum_{i=1}^{n} \beta_i y_i = 0$.

$$f(x) = sign\left\{\sum_{i=1}^{n} \beta_i y_i K(x_i, x) + g\right\} \tag{12}$$

$$\max W(\beta) = \sum_{i=1}^{n} \beta_i - \frac{1}{2}\sum_{i,j=1}^{n} \beta_i \beta_j y_i y_j K(x_i, x_j) \tag{13}$$

In order to obtain the maximum classification interval, the decision function and β are transformed into a new objective function Ψ, as shown in equation (14). In equation (14), C is the penalty parameter used to control the degree of sample misclassification.

$$\Psi(\beta) = \frac{1}{2}\left\|\beta\right\|^2 + C\sum_{i=1}^{n} \beta_i \tag{14}$$

However, SVM has the advantage of fast learning, but for more complex feature extraction models, conventional SVM often has the disadvantages of slow computing speed and low accuracy of feature extraction results. It has been difficult to meet the needs of complex psychological feature extraction for teaching, so it needs to be improved [23]. The experiment combines approximate entropy, LMD and SVM. In addition, SVM is transformed into the same kind of extended least square support vector machine (LSSVM), and a new improved algorithm, ck-LSSVM, is formed based on the combined kernel (CK).

Combined kernel function is one of the common functions used by classifiers to select kernel functions. Given a coefficient of each kernel function, two or more kernel functions are combined to replace the previous single kernel function. Therefore, the combined kernel can usually avoid falling into the local optimal solution, so as to speed up the global convergence and obtain the optimal coefficient [24].

Compared with the conventional SVM, the main advantage of LSSVM is that it can better solve the complex quadratic programming problems, follow the principle of structural minimization, and only need to solve the linear equations to simplify the solution process. Suppose given a training set point, input a real number and output a scalar. The nonlinear mapping eucalyptus tree is used to establish the regression model, and the input data is mapped to the high-dimensional feature space, in which the linear regression model is constructed. The linear equation is constrained, and the Lagrange function is used to construct a binary problem. The obtained equations are transformed into matrices, and the decision function is finally obtained by using Mercer theorem. According to the basic method of SVM, the decision function will transform the kernel function from a high-dimensional space to a plane. It only needs to find the optimal classification surface and transform the classification problem into a linear problem [25].

Composite kernel combines at least two single kernel functions, and uses their respective advantages to avoid defects as much as possible. The main types of kernel functions are local kernel functions and global kernel functions. Among the local kernel functions, d-order polynomial kernel function is typical. The advantage of this kind of function is strong generalization ability, and the disadvantage is that the classification accuracy is difficult to reach a high level. Among the global kernel functions, Gaussian radial kernel function is the typical one [26]. The advantage of this kind of function is high classification accuracy, and the disadvantage is weak generalization ability. Because the advantages and disadvantages of these two kinds of functions are complementary, the experiment combines these two kinds of kernel functions and gives a weight coefficient to each of the two kernel functions to form a combined kernel function. The definition formula of composite kernel function is shown in formula (15).

$$K_m(X, X_i) = \mu K_r + (1-\mu) K_d \tag{15}$$

In equation (15), μ represents the weight coefficient, whose value is (0, 1). When μ takes 0 or 1, it is a single kernel function, K_m represents the composite function, K_r represents the Gaussian kernel function, K_d represents the d-order polynomial kernel function, and X and X_i are the two main independent variables of the kernel function respectively. When the D value of the d-order polynomial kernel function is 1 and the σ^2 value of the Gaussian kernel function is 1, both of them can maximize their advantages. Therefore, both D and j of the combined kernel are taken as 1. The test points of the combined kernel function are 0.2 and μ are 0.5, 0.6, 0.7 and 0.8 respectively to determine the best value of μ. The basic flow of the whole algorithm is shown in Figure 5.

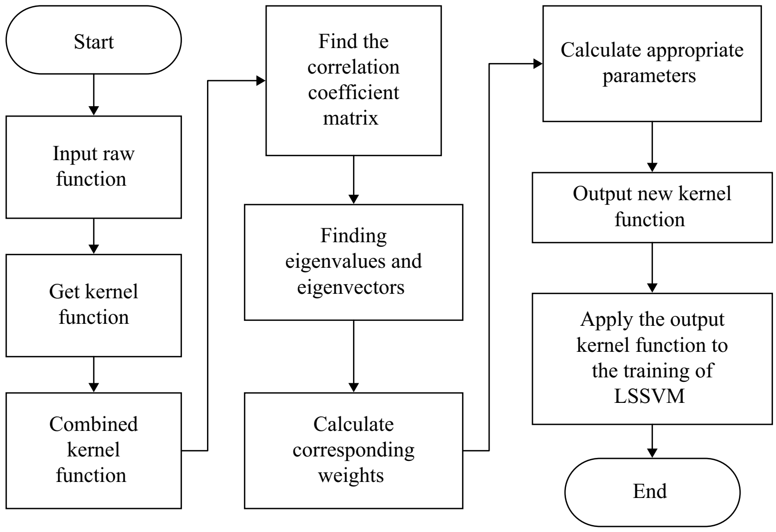


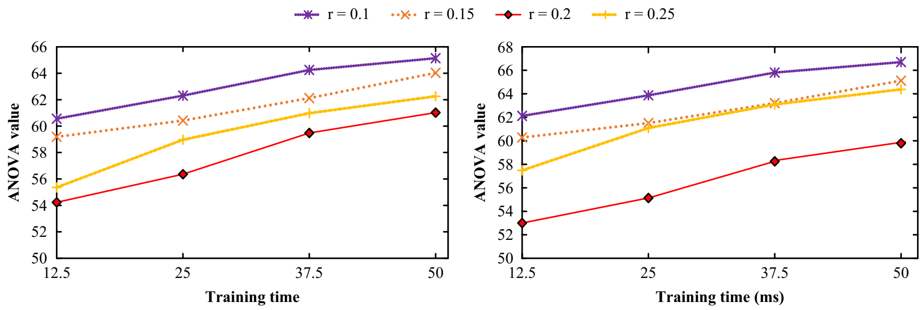
Fig. 5. Basic flow of combined kernel function

In terms of data processing, first build a sample set, then build a feature extraction model and train it. The training time is divided into two equal data sets, data set 1 and data set 2, respectively, to increase the representativeness of the experimental results, and then obtain the relevant parameters through simulation experiments. After the parameters are selected, the improved algorithm used in the experiment, namely ck-LSSVM, and several other algorithms are simulated. Among other algorithms, conventional SVM and conventional LSSVM are selected for comparison, and back propagation neural network (BPNN) commonly used for feature extraction is selected for comparison at the same time, so as to more intuitively and accurately show the performance of the improved algorithm used in the experiment. The other three algorithms are basically the same as ck-LSSVM for data processing. The reason for choosing these three algorithms is whether the result after least squares has advantages compared with SVM, whether the combined kernel function has advantages compared with LSSVM, and whether BPNN is a classical and commonly used neural network method in current depth learning, so as to find out whether it has advantages compared with conventional depth learning.

5 Experimental results and analysis of ck-LSSVM feature extraction model

5.1 Main parameter settings

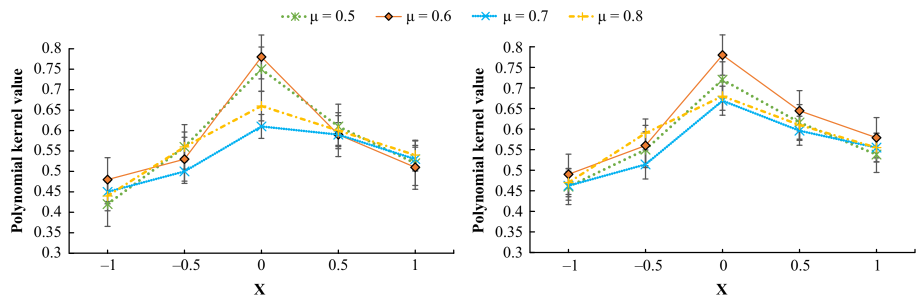
The experimental environment is ECS, with 2G running memory and 49G hard disk capacity. The algorithm is implemented based on Python. In the parameter processing of experimental data, the approximate entropy value needs to be one of 0.1–0.25. The experiment is divided into four orders, namely 0.1, 0.15, 0.2, 0.25, which is the best value for analysis. Under the same other conditions, the two groups of results of variance of approximate entropy varying with experimental time under different values are shown in Figure 6. As can be seen from Figure 6, in data set 1 and data set 2, the comprehensive value of analysis variance is the lowest when the R value is 0.2 as the experimental time increases. Since the lower the comprehensive value of variance analysis, the more stable the representative operation is and the simpler the processing is, the value selected is 0.2.



(a) Under the dataset 1 **(b) Under the dataset 2**

Fig. 6. Analysis variance value of two sets of data sets

The possible values of μ of the combined kernel are 0.5, 0.6, 0.7 and 0.8 respectively. The numerical results of two groups of polynomial kernel functions with test points of 0.2 and 0.4 under the four values are shown in Figure 7. It can be seen from Figure 7 that when the test points are 0.2 and 0.4 respectively, they are polynomial kernel function values with μ value of 0.6, and the polynomial kernel function values can reflect the overall operation speed. Therefore, μ value with the highest polynomial function value should be selected. According to the significance analysis, the value selected through comprehensive consideration is 0.6.



(a) Test point value = 0.2 **(b) Test point value = 0.4**

Fig. 7. Polynomial kernel function values at two different test points

5.2 Comparison and analysis of various algorithm results

When the training has been completed, each group of test samples is predicted and compared with the actual situation. The accuracy of the four algorithms according to the training times is shown in Figure 8. As can be seen from Figure 8, the accuracy of each algorithm shows an obvious upward trend with the increase of training times. When the training times are closer to the maximum times of each experiment, i.e. 100 times, the increase of accuracy will slow down. It can be seen that the accuracy rate of ck-LSSVM is obviously the highest among the four algorithms, and it is the highest at any training times. According to the significance analysis results, the accuracy rate of ck-LSSVM is significantly higher than that of the other three algorithms, indicating that ck-LSSVM has a significant advantage in the accuracy of operation results. The average accuracy of ck-LSSVM algorithm in data set 1 is 80.83%, in data set 2 is 84.23%, the overall accuracy is 82.53%, and the highest accuracy is 90.12%.

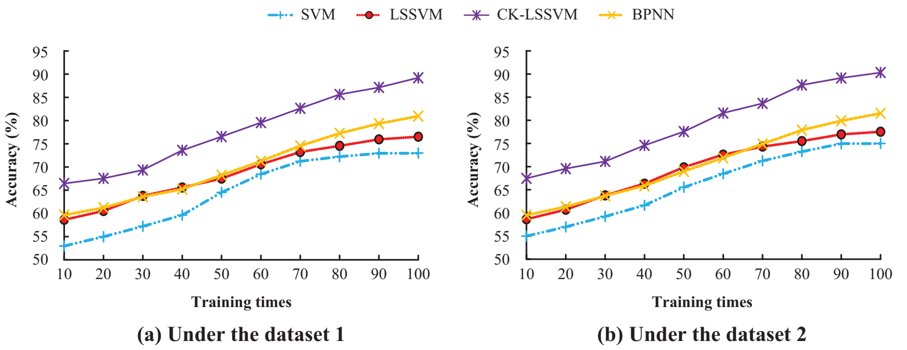
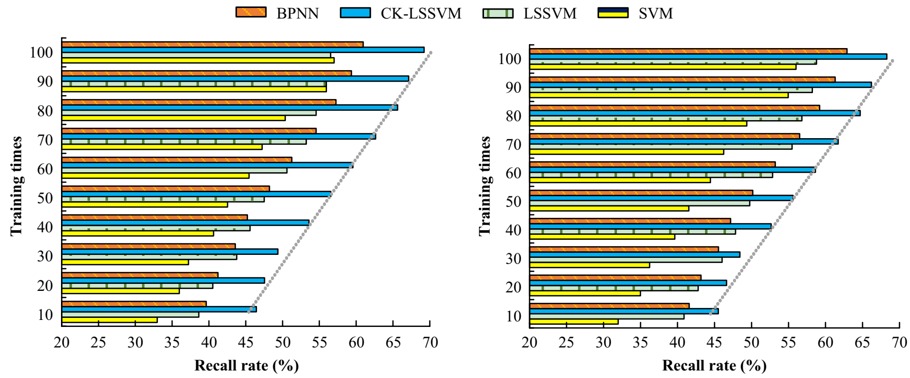


Fig. 8. Accuracy with training times of two sets of data sets

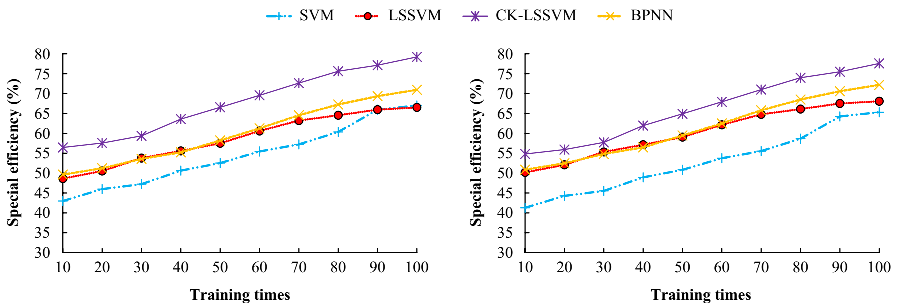
In the simulation experiment, the two groups of recall rates of the four algorithms with the number of training times are shown in Figure 9. As can be seen from Figure 9, the recall rates of the four algorithms show an obvious growth trend with the increase of training times. The recall rate of ck-LSSVM algorithm is significantly higher than that of the other two algorithms, and each training time of the two groups of results is the highest recall rate of ck-LSSVM. According to the results of significance analysis, the recall rate of ck-LSSVM in the two data sets is significantly higher than that of the other three algorithms, indicating that the measurement ability of ck-LSSVM algorithm for positive cases is significantly better than that of the other three algorithms. The overall recall rate of the two sets of data sets of ck-LSSVM algorithm is 58.64%.



(a) Under the dataset 1 **(b) Under the dataset 2**

Fig. 9. Recall rate with training times of two sets of data sets

The two groups of results of the effectiveness of the four algorithms with the number of training times are shown in Figure 10. As can be seen from Figure 10, the specific effect degree of each algorithm shows an obvious upward trend with the increase of training times. In the results of the two sets of data sets, ck-LSSVM has the highest specific effect under each training times, which is generally higher than the other three algorithms. According to the significance analysis results, the specific effect degree of ck-LSSVM in data set 1 is significantly higher than that of the other three algorithms, and that of ck-LSSVM in data set 2 is significantly higher than that of SVM and LSSVM, but not significantly with BPNN. In the overall results of the two data sets, the specific effect degree of ck-LSSVM is still higher than that of the other three algorithms, indicating that the negative case recognition and measurement ability of ck-LSSVM is significantly better than that of the other three algorithms.



(a) Under the dataset 1 **(b) Under the dataset 2**

Fig. 10. Special efficiency with training times of two sets of data sets

The two sets of results of the accuracy rate of the four algorithms with the number of training times are shown in Figure 11. As can be seen from Figure 11, the accuracy rates of the four algorithms show an obvious upward trend with the increase of training times, and the overall increase is relatively slow. The accuracy of ck-LSSVM is significantly higher than that of the other three algorithms. According to the significance analysis results, ck-LSSVM algorithm is significantly different from other algorithms, indicating that ck-LSSVM algorithm is significantly better than the other three algorithms in the performance of optimal solution. The overall accuracy of ck-LSSVM algorithm is 86.92%.

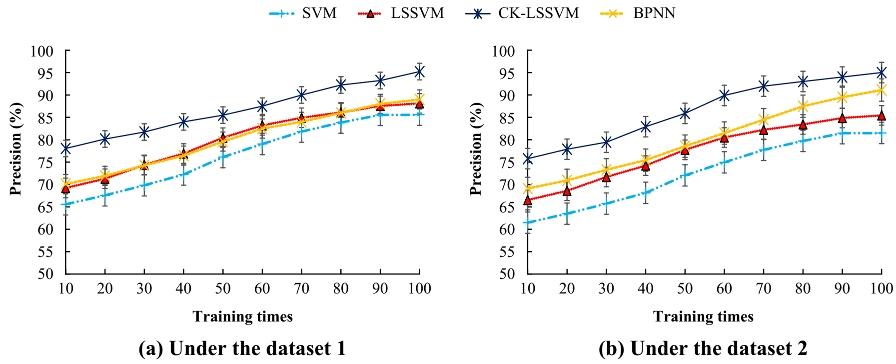


Fig. 11. Precision with training times of two sets of data sets

The average operation time of the four algorithms with the training times is shown in Figure 12. As can be seen from Figure 12, the average operation time of ck-LSSVM in data set 1 is significantly less than that of the other three algorithms, while that in data set 2 is significantly lower than that of SVM and LSSVM, but there is little difference with BPNN. According to the results of significance analysis, the average operation time of ck-LSSVM is significantly different from that of SVM and LSSVM, but it is not significantly different from that of BPNN, indicating that ck-LSSVM is significantly better than SVM and LSSVM in operation efficiency.

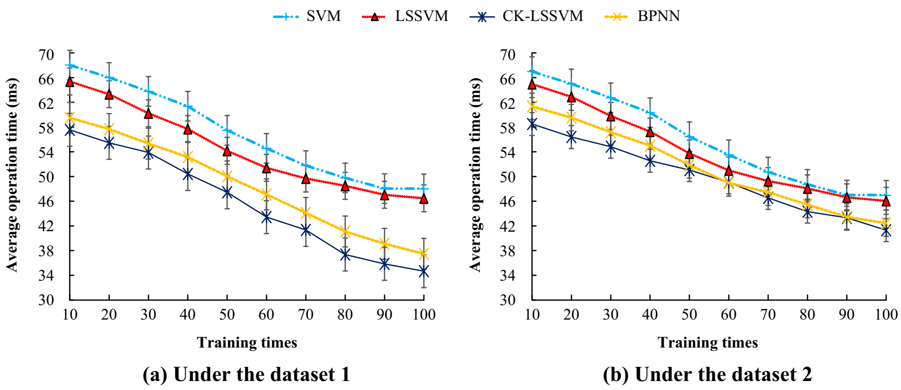


Fig. 12. Average operation time with training times of two sets of data sets

6 Conclusion

Aiming at the need of understanding students' psychological characteristics in smart English teaching, an improved ck-LSSVM algorithm based on approximate entropy and LMD is proposed to extract the characteristics of the corresponding behavior sequences. The experimental results show that the ck-LSSVM algorithm is significantly better than the conventional SVM, LSSVM and BPNN algorithm in many performances. The overall recall rate of the ck-LSSVM algorithm is 58.64% and the overall accuracy rate is 86.92%, which is not only significantly higher than the other three algorithms selected in the experiment, but also fully meets the needs of conventional feature extraction, indicating that the improvement of the ck-LSSVM algorithm is effective. Although the research has made some achievements, the training times of the research are less, and the selection of training samples is lack of universality. It is necessary to further select more representative sample data in the follow-up, which is also a problem that needs to be improved in further research in the future.

7 References

[1] Y. Zhang, et al. "Interactive Intelligent Teaching and Automatic Composition Scoring System Based on Linear Regression Machine Learning Algorithm", *Journal of Intelligent and Fuzzy Systems*, vol. 40, pp. 2069–2081, 2021. https://doi.org/10.3233/JIFS-189208

[2] J. Zhou, et al. "Application Investigation and Technological Development of Intelligent Product in College English Teaching", *Journal of Physics: Conference Series*, vol. 1237, pp. 22069–22077, 2019. https://doi.org/10.1088/1742-6596/1237/2/022069

[3] Z. Sun, T. Shangguan, "Application of Intelligent Terminal Mobile Learning Platform in English Teaching", *Journal of Physics: Conference Series*, vol. 1915, pp. 42092–42097, 2021. https://doi.org/10.1088/1742-6596/1915/4/042092

[4] M. S. Alharbi, H. Ahmad, "Saudi EFL Teacher Professional Identity and Emotions: Emotions Regulation Perspective", *Arab World English Journal*, vol. 11, pp. 385–397, 2020. https://doi.org/10.24093/awej/vol11no3.24

[5] X. Zou, R. Lv, X. Li, et al. "Intelligent Electrical Fault Detection and Recognition Based on Gray Wolf Optimization and Support Vector Machine", *Journal of Physics: Conference Series*, vol. 2181, pp. 12058–12071, 2022. https://doi.org/10.1088/1742-6596/2181/1/012058

[6] J. Sun, H. Fujita, P. Chen, et al. "Dynamic Financial Distress Prediction with Concept Drift Based on Time Weighting Combined with Adaboost Support Vector Machine Ensemble", *Knowledge-Based Systems*, vol. 120, pp. 4–14, 2017. https://doi.org/10.1016/j.knosys.2016.12.019

[7] D. Simian, F. Stoica, A. Brbulescu, "Automatic Optimized Support Vector Regression for Financial Data Prediction", *Neural Computing and Applications*, vol. 32, pp. 1–14, 2020. https://doi.org/10.1007/s00521-019-04216-7

[8] D. Zhao, J. Ding, S. Chai, "Systemic Financial Risk Prediction Using Least Squares Support Vector Machines", *Modern Physics Letters B*, vol. 32, pp. 1850183–1850192, 2018. https://doi.org/10.1142/S021798491850183X

[9] F. Tunio, Y. Ding, A. Nabi, et al. "Financial Distress Prediction Using Adaboost and Bagging in Pakistan Stock Exchange", *Journal of Asian Finance Economics and Business*, vol. 8, pp. 665–673, 2021.

[10] L. T. Vu, N. T. Nguyen, et al. "Feature Selection Methods and Sampling Techniques to Financial Distress Prediction for Vietnamese Listed Companies", *Investment Management and Financial Innovations*, vol. 16, pp. 276–290, 2019. https://doi.org/10.21511/imfi.16(1).2019.22
[11] S. H. Shetty, T. N. Vincent, "The Role of Board Independence and Ownership Structure in Improving the Efficacy of Corporate Financial Distress Prediction Model Evidence from India", *Journal of Risk and Financial Management*, vol. 14, pp. 333–345, 2021. https://doi.org/10.3390/jrfm14070333
[12] D. Niu, L. Si, S. Dai, "Comprehensive Evaluation for Operating Efficiency of Electricity Retail Companies Based on the Improved TOPSIS Method and LSSVM Optimized by Modified Ant Colony Algorithm from the View of Sustainable Development", *Sustainability*, vol. 10, pp. 860–868, 2018. https://doi.org/10.3390/su10030860
[13] X. Xue, et al. "Evaluation of Concrete Compressive Strength Based on an Improved PSO-LSSVM Model", *Computers & Concrete*, vol. 21, pp. 505–511, 2018.
[14] K. J. Kim, H. Kim, N. Baek, "An Improved SVM-T-RFE Based on Intensity-Dependent Normalization for Feature Selection in Gene Expression of Big-Data", *Lecture Notes in Electrical Engineering*, vol. 10, pp. 44–51, 2018. https://doi.org/10.1007/978-981-10-6451-7_6
[15] H. Liao, "Facial age feature extraction based on deep sparse representation", *Multimedia Tools and Applications*, vol. 78, pp. 2181–2197, 2019. https://doi.org/10.1007/s11042-018-6342-5
[16] J. Shukla, M. Barreda-Angeles, J. Oliver, et al. "Feature Extraction and Selection for Emotion Recognition from Electrodermal Activity", *IEEE Transactions on Affective Computing*, vol. 10, pp. 105–112, 2019.
[17] F. Zhang, F. Wang, "Exercise Fatigue Detection Algorithm Based on Video Image Information Extraction", *IEEE Access*, vol. 3, pp. 199696–199709, 2020. https://doi.org/10.1109/ACCESS.2020.3023648
[18] M. X. Xiao, H. C. Wei, Y. J. Xu, et al. "Combination of R-R Interval and Crest Time in Assessing Complexity Using Multiscale Cross-Approximate Entropy in Normal and Diabetic Subjects", *Entropy*, vol. 20, pp. 497–503, 2018. https://doi.org/10.3390/e20070497
[19] S. K. Rout, M. Sahani, P. K. Dash, et al. "Multifuse Multilayer Multikernel RVFLN+ of Process Modes Decomposition and Approximate Entropy Data from iEEG/sEEG Signals for Epileptic Seizure Recognition", *Computers in Biology and Medicine*, vol. 132, pp. 104299, 2021. https://doi.org/10.1016/j.compbiomed.2021.104299
[20] W. Lei, Z. Liu, M. Qiang, et al. "Time–Frequency Analysis Based on Ensemble Local Mean Decomposition and Fast Kurtogram for Rotating Machinery Fault Diagnosis", *Mechanical Systems and Signal Processing*, vol. 103, pp. 60–75, 2018. https://doi.org/10.1016/j.ymssp.2017.09.042
[21] C. Sukawattanavijit, C. Jie, H. Zhang, "GA-SVM Algorithm for Improving Land-Cover Classification Using SAR and Optical Remote Sensing Data", *IEEE Geoscience & Remote Sensing Letters*, vol. 14, pp. 284–288, 2017. https://doi.org/10.1109/LGRS.2016.2628406
[22] S. H. Hojjati, A. Ebrahimzadeh, A. Khazaee, et al. "Predicting Conversion from MCI to AD Using Resting-State fMRI, Graph Theoretical Approach and SVM", *Journal of Neuroscience Methods*, vol. 282, pp. 69–80, 2017. https://doi.org/10.1016/j.jneumeth.2017.03.006
[23] S. Raj, K. C. Ray, "ECG Signal Analysis Using DCT-Based DOST and PSO Optimized SVM". *IEEE Transactions on Instrumentation & Measurement*, vol. 66, pp. 470–478, 2017. https://doi.org/10.1109/TIM.2016.2642758
[24] B. Samy, "An SVM Framework for Malignant Melanoma Detection Based on Optimized HOG Features", *Computation*, vol. 5, pp. 14–18, 2017. https://doi.org/10.3390/computation5010004

[25] Z. Zhang, Z. Li, T. Ma, et al. "EEG Signal Classification Method Based on Improved Empirical Mode Decomposition and SVM", *Journal of Physics: Conference Series*, vol. 1846, pp. 1205–1214, 2021. https://doi.org/10.1088/1742-6596/1846/1/012054
[26] X. Jia, Y. Guo, B. Zhao, et al. "Fractional-Integral-Operator-Based Improved SVM for Filtering Salt-and-Pepper Noise", *Image Processing, IET*, vol. 13, pp. 2346–2357, 2019. https://doi.org/10.1049/iet-ipr.2019.0624

8 Authors

Xiaoyu Cai received her master's degree in education from He'nan University (2015), in curriculum and pedagogy. Currently, she works as an English teacher, Japanese teacher, and psychological counselor at Shangqiu Polytechnic, He'nan, China. Since 2011, she has been working as a school clerk, psychological counselor, and concurrently Japanese and English teachers for 10 years. she has published many papers such as "Research on Ashe Participation of Counselors in College Students' Mental Health Education". Engaged in college students' learning psychology, interpersonal communication and college students' career planning psychological problem counseling. During her tenure as a psychological counselor, she participated in several career guidance lectures for freshmen in Ashe college, participated in she compilation of 3 textbooks such as "College English", "Mechanical and Electrical English", "Career Planning under The New College Entrance Examination" – based on IMAP career empowerment system, published 5 national journal papers, 16 scientific research, and won three national patents. As a psychological counselor, she held lectures to provide employment guidance and career planning for all graduates, and has received hundreds of career planning psychological consulting services. She has been invited as a consultant for college student mental health and career guidance lectures. She also serves as the school's English and Japanese foreign translators. She has published in more than 10 internationally renowned peer-reviewed journals and conference proceedings. Her areas of interest include psychology, pedagogy, foreign and international Chinese language education, business Chinese and online education.

Jarrytt R. McDonald Received an associate of arts degree from Rowan College in Burlington County, Mount Laurel, New Jersey (2011). B.A. (Media Studies and Production) from Temple University, Philadelphia, Pennsylvania. He works in Maple leaf International Academy, Shenzhen, China as an ESL Teacher, taught Chinese students English in grades 1–9 while using an English environment. Planned many classes to suit different learners of all ages and grade levels. Assisted other foreign teachers with their planning to further improve their English teaching. His areas of interest are audio production, classroom management, multimedia expertise, adaptable skill sets, communication excellence and technical proficiency. E-mail: sqzycxy@163.com

Article submitted 2022-09-22. Resubmitted 2022-11-07. Final acceptance 2022-11-15. Final version published as submitted by the authors.