# PAD Dimension Prediction Based on Clustering PSO-LSSVM Model (Postprint)

**Authors:** Hu Yanxiang, Sun Ying, Zhang Xueying, Duan Shufei

**Date:** 2019-01-28T00:00:00+00:00

## Abstract

To address the prediction accuracy issue of PAD (pleasure, arousal, dominance), a clustered PSO-LSSVM model is proposed that integrates particle swarm optimization (PSO) algorithm-optimized least squares support vector machine (LSSVM) with emotion clustering analysis. Emotional features are extracted from three types of emotional speech in the TYUT2.0 and Berlin speech databases. Based on these features and annotated P, A, D values, separate emotion-dimension PSO-LSSVM models are constructed for each of the three single emotions, and a mixed-emotion-dimension PSO-LSSVM model is constructed for the three emotions collectively. Subsequently, the mixed-emotion-dimension PSO-LSSVM model is employed to predict P, A, D values, and the distances to the basic emotion PAD values are computed. Finally, emotions with distances exceeding the threshold are clustered as mixed emotions, while those with distances below the threshold are assigned to the nearest emotion, and the corresponding emotion' s regression model is utilized to predict their P, A, D values. Results demonstrate that this model achieves smaller prediction errors for P, A, D compared to LSSVM and PSO-LSSVM models, with stronger correlation between predicted and annotated values, indicating that the clustered PSO-LSSVM model provides more reliable and accurate PAD predictions.

## Full Text

## Forecast of PAD Dimensions Using Clustering PSO-LSSVM Model

**Hu Yanxiang, Sun Ying†, Zhang Xueying, Duan Shufei**
(College of Information & Computer, Taiyuan University of Technology, Taiyuan 030024, China)

**Abstract**

To address the accuracy problem in PAD (Pleasure, Arousal, Dominance) prediction, this paper proposes a clustering PSO-LSSVM model that combines Least Squares Support Vector Machine (LSSVM) optimized by Particle Swarm Optimization (PSO) with affective clustering analysis. Emotional features were extracted from three types of emotional speech in the TYUT2.0 and Berlin voice libraries. Based on these features and annotated P, A, D values, separate PSO-LSSVM models were established for each individual emotion type, along with a mixed-emotion PSO-LSSVM model for all three emotions. The mixed-emotion model was then used to predict P, A, D values, and the distances between these predictions and the PAD values of basic emotions were calculated. Finally, emotions with distances greater than the threshold were clustered as mixed emotions, while those with distances smaller than the threshold were clustered into the nearest emotion category, and their P, A, D values were predicted using the corresponding emotional regression model. The study demonstrates that this model yields smaller prediction errors for P, A, D compared to LSSVM and PSO-LSSVM models, with stronger correlation between predicted and annotated values, indicating that the clustering PSO-LSSVM model provides more reliable and accurate PAD predictions.

**Keywords:** emotional dimensions PAD; least squares support vector machine; particle swarm optimization algorithm; affective clustering analysis

## 0 Introduction

Affective computing is indispensable for the development of mature artificial intelligence. Emotional dimensions describe emotions from psychological attributes and hold significant importance in affective computing. With the advancement of intelligent human-computer interaction, discrete emotions can no longer satisfy the recognition of natural emotions in daily life, making the study of continuous emotional dimensions containing emotional information particularly crucial. Given the difficulty of real-time monitoring of emotional dimensions, mathematical modeling becomes necessary for dimension prediction, with data simulation and fitting serving as the primary means for effective forecasting.

Artificial intelligence prediction methods mainly include neural networks and support vector machines. However, neural networks require large data samples and cannot guarantee prediction accuracy with small samples of emotional speech. Support Vector Machines (SVM) can effectively solve complex problems involving small samples and nonlinearity, but require solving quadratic programming equations with high computational complexity. Least Squares Support Vector Machine (LSSVM) improves upon SVM through equality constraints, converting the traditional quadratic programming problem into solving a system of linear equations, thereby reducing computational complexity. The introduction of parameter selection optimization methods such as Particle

Swarm Optimization (PSO) makes parameter selection in the modeling process more objective. The PSO-LSSVM algorithm has been applied in solar greenhouse temperature prediction and throttling liquid velocity prediction, but has not yet been reported for emotional dimension prediction. However, the complex interactions between emotional features of different emotions can affect the prediction accuracy of the PSO-LSSVM regression model for emotional dimensions. Therefore, introducing affective clustering analysis into this regression model is necessary to reduce the influence between emotional features of different emotions.

This paper proposes performing clustering processing based on the distance between the initial prediction results of the PSO-LSSVM model and the PAD values of basic emotion centers, then using the PSO-LSSVM regression model for secondary dimension prediction. Based on the PAD three-dimensional emotion model, we first extract emotional features from emotional speech; then establish three groups of regression models—LSSVM regression model, PSO-LSSVM regression model, and clustering PSO-LSSVM regression model—to construct mapping relationships between speech emotional features (prosodic features, MFCC features, and nonlinear features) and manually annotated P, A, D values, achieving P, A, D value prediction for emotional utterances; finally, we compare and analyze the prediction results of the three regression models. Experimental results demonstrate that the clustering PSO-LSSVM regression model achieves higher prediction accuracy for P, A, D values.

## 1 PAD Three-Dimensional Emotion Model

Human emotions in daily life are subtle and complex, such as mixed feelings of sorrow and joy or tears of joy, which do not completely belong to a single basic emotion category. Therefore, the continuous space theory of emotion was proposed to address this issue. This theory suggests that human emotions consist of several dimensions in space that can cover almost all human emotions, enabling continuous and smooth transitions between different emotions. A typical continuous emotion model is the PAD three-dimensional emotion model, developed by Mehrabian at UCLA, which uses semantic differential evaluation to divide emotions into three dimensions: P representing Pleasure (pleasure-displeasure, indicating the positive or negative characteristic of an individual' s emotional state), A representing Arousal (arousal-nonarousal, indicating the neurophysiological activation level of an individual), and D representing Dominance (dominance-submissiveness, indicating the individual' s control state over the situation and others). The PAD three-dimensional emotion model describes emotions from three perspectives, facilitating the quantification of continuous and diverse emotions in daily life. The PAD three-dimensional emotion model is shown in Figure 1 [Figure 1: see original paper].

## 2.1 Least Squares Support Vector Machine Theory

The LSSVM algorithm is based on Support Vector Machine (SVM). SVM maps input data to a high-dimensional feature space through nonlinear mapping, converting the problem into a quadratic programming problem with inequality constraints, but its computational complexity is high. Therefore, least squares linear theory is introduced into SVM for improvement, transforming the traditional quadratic programming problem into solving a system of linear equations, reducing computational complexity.

The LSSVM algorithm basic setup: For a sample set with n samples, where $x_i \in \mathbb{R}^d$ is the input variable, $y_i \in \mathbb{R}$ is the corresponding output, and $\{(x_i, y_i) | i = 1, 2, ..., n\}$, the LSSVM optimization problem is:

$$\min_{w,b,e} J(w,e) = \frac{1}{2} w^T w + \frac{1}{2} C \sum_{i=1}^{n} e_i^2$$
$$\text{s.t.} \quad y_i = w^T \Phi(x_i) + b + e_i, \quad (i = 1, 2, ..., n)$$

where $C$ is the regularization parameter and $e_i \in \mathbb{R}$ is the error variable.

To solve this optimization problem, it is converted into the following linear problem:

$$\begin{bmatrix} 0 & 1^T \\ 1 & K + C^{-1} I \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ y \end{bmatrix}$$

where $y = [y_1, y_2, ..., y_n]^T$, $1 = [1, 1, ..., 1]^T$, $I$ is the n-order identity matrix, $\alpha = [\alpha_1, \alpha_2, ..., \alpha_n]^T$ is the Lagrange multiplier vector, and $K$ is the kernel function matrix with $K_{ij} = K(x_i, x_j) = \Phi(x_i)^T \Phi(x_j)$ for $i, j = 1, 2, ..., n$. This experiment uses the Radial Basis Function (RBF) kernel, where $\sigma$ is the kernel width. The RBF kernel function is:

$$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right)$$

The final LSSVM model is:

$$y(x) = \sum_{i=1}^{n} \alpha_i K(x, x_i) + b$$

In the LSSVM model, $x$ represents model input and $y(x)$ represents model output. In this study, $x$ represents speech emotional features and $y$ represents speech emotional dimension values. According to the LSSVM regression principle, two types of parameters require special attention: the regularization parameter $C$ and the kernel parameter $\sigma$. To reduce the blindness of subjective selection of these two parameters, this paper adopts the particle swarm optimization method to select the LSSVM regularization parameter $C$ and kernel parameter $\sigma$.

## 2.2 Particle Swarm Optimization Algorithm

The PSO algorithm is an optimization calculation method based on swarm intelligence. Observing the flocking behavior of birds, this algorithm utilizes information sharing among individuals in the swarm to achieve an evolutionary movement process from disorder to order, thereby obtaining the optimal solution.

In the PSO algorithm, let $u_i = (u_{i1}, u_{i2}, ..., u_{in})$ represent the position vector of particle $i$, $v_i = (v_{i1}, v_{i2}, ..., v_{in})$ represent the velocity vector of particle $i$, $p_i = (p_{i1}, p_{i2}, ..., p_{in})$ represent the best position experienced by particle $i$, and $p_g = (p_{g1}, p_{g2}, ..., p_{gn})$ represent the best position experienced by all particles in the population. The velocity and position update formulas for particle $i$ are as follows:

$$\begin{aligned} v_{id}^{k+1} &= wv_{id}^{k} + c_1 r_1 (p_{id} - u_{id}^{k}) + c_2 r_2 (p_{gd} - u_{id}^{k}) \\ u_{id}^{k+1} &= u_{id}^{k} + v_{id}^{k+1} \end{aligned}$$

where $r_1$ and $r_2$ are random numbers between $[0, 1]$, $w$ is the inertia weight that balances global and local search capabilities, and $c_1$ and $c_2$ are the self-learning factor and social learning factor, respectively.

## 2.3 PSO-Optimized LSSVM Regression Model

The objective function is defined as:

$$\min_{(C,\sigma)} f(C, \sigma) = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

where $y_i$ is the manually annotated P, A, D dimension value of the $i$-th emotional sample, and $\hat{y}_i$ is the model' s predicted P, A, D output value for the sample, which can be calculated using Equation (3).

The goal of the PSO algorithm in optimizing the LSSVM model is to search for a set of parameters $(C, \sigma)$ through iterative algorithms to minimize the objective function in Equation (5), i.e., to minimize the error between subjectively annotated emotional dimension values and objectively predicted dimension values.

## 2.4 Emotion Clustering Analysis

In speech emotion clustering, traditional clustering methods operate in the feature space, producing clustering results based on speech signals. However, such clustering results may not necessarily correlate with emotions. For speech emotion clustering problems, this paper performs clustering through emotional dimensions that describe continuous emotions, i.e., by calculating the distance between preliminary PAD prediction values of speech and basic emotion center PAD values.

The process is shown in Figure 2 [Figure 2: see original paper]. Although many clustering algorithms exist, such as k-means clustering and fuzzy clustering, none provide a direct method for determining the clustering threshold. To intuitively determine the clustering threshold, the threshold is set based on the Euclidean distance between the PAD preliminary prediction values of emotional speech and basic emotions. The Euclidean distance calculation expression is:

$$\mathrm{dist}(Y,Z)=\sqrt{(\hat{P}_y-P_z)^2+(\hat{A}_y-A_z)^2+(\hat{D}_y-D_z)^2}$$

where $\hat{P}_y$, $\hat{A}_y$, $\hat{D}_y$ are the predicted P, A, D values of emotional speech, and $P_z$, $A_z$, $D_z$ are the P, A, D values of basic emotions.

Emotion clustering analysis first analyzes the distance between speech PAD prediction values and basic emotion center PAD values. Based on the relationship between distance and threshold, speech is classified into specific single emotions or mixed emotions. During secondary prediction, speech is predicted using the corresponding emotional PAD regression model.

## 2.5 Clustering PSO-LSSVM Regression Model

In emotional speech samples, the correlation between emotions themselves causes correlations between emotional features of different emotional speech, which affects the prediction performance of regression models during P, A, D prediction. Therefore, this paper proposes a regression scheme combining PSO-LSSVM with emotion clustering analysis. Figure 3 [Figure 3: see original paper] shows the flowchart of the proposed clustering PSO-LSSVM regression model.

The specific process is as follows: First, emotional features are extracted from emotional speech. Based on the emotional features and annotated P, A, D values of the training set, two types of regression models are established. One is called the single-emotion dimension PSO-LSSVM regression model, trained using emotional features and manually annotated P, A, D values of single-emotion speech. The other is called the mixed-emotion dimension PSO-LSSVM regression model, trained using emotional features and manually annotated P, A, D values of multiple-emotion speech. Next, the speech emotional features of the test set are used as input variables for the mixed-emotion dimension PSO-LSSVM model to predict PAD values, and the distances between these values and the center PAD values of basic emotions are calculated. The center PAD values of basic emotions are obtained through fuzzy c-means clustering of PAD annotation results for each emotion type. Emotions with distances greater than the threshold are clustered as mixed emotions, while those with distances smaller than the threshold are clustered into the nearest emotion category. Finally, speech emotional features clustered as mixed emotions are used as input variables for the mixed-emotion dimension PSO-LSSVM regression model (the same model trained from the training set), while speech emotional features clustered as specific emotions are used as input variables for the corresponding single-emotion dimension PSO-LSSVM regression model trained from the training set to predict their PAD values. This process enables test set speech to be predicted through their corresponding regression models, reducing the impact of correlations between different emotional features on prediction effectiveness. This regression model combines PSO-optimized LSSVM with clustering analysis, not only avoiding subjective blindness in parameter selection during regression but also reducing correlations between input variables, enabling more accurate prediction of emotional dimensions P, A, D.

## 3.1 Experimental Procedure

To verify the effectiveness of the clustering PSO-LSSVM regression model in emotional dimension prediction, three regression models were designed for comparison: LSSVM regression model (Model 1), PSO-LSSVM regression model (Model 2), and clustering PSO-LSSVM regression model (Model 3). The experimental flow is as follows:

a) Extract emotional features from emotional speech in the TYUT2.0 database;

b) Based on the extracted emotional features and manually annotated P, A, D values, use the three regression models (LSSVM, PSO-LSSVM, clustering PSO-LSSVM) to predict emotional dimensions P, A, D of the test set;

c) Analyze the prediction results of different regression models for P, A, D against manual annotation results to select the more reasonable and effective regression model.

### 3.2.1 Database

An emotional speech database is an important prerequisite for speech emotion analysis. To comprehensively and objectively evaluate the model' s prediction capability for P, A, D, the TYUT2.0 Chinese emotional speech database and the Berlin German emotional speech corpus (EMO-DB) were selected, focusing on their common emotion types: sadness (52 utterances), anger (57 utterances), and happiness (52 utterances) as experimental samples, with 67% used as training samples and 33% as test samples. The emotional speech from the databases is used to extract features, and the emotional features and corresponding PAD values of each utterance are used to train regression models. Clustering analysis is performed based on the PAD prediction values of each utterance and the PAD values of basic emotions.

The TYUT2.0 emotional speech public database is an excerpt-type emotional database obtained by intercepting radio dramas, with PAD dimension annotations for the database' s speech obtained using an improved PAD emotion scale to acquire corresponding P, A, D values for each utterance (available by contacting the author via email).

The Berlin emotional speech database was simulated by 10 actors expressing 7 emotions. The database recording required actors to complete emotional expression by recalling personal experiences, resulting in high emotional authenticity, wide usage, and strong representativeness.

### 3.2.2 Speech Emotion Features

To comprehensively represent speech emotion and achieve more accurate prediction of P, A, D values, the features selected are shown in Table 1 .

**Table 1 Emotional Speech Characteristics**

| Feature Category | Specific Features |
|---|---|
| MFCC | Skewness, kurtosis, mean, variance, median of first 12 MFCC orders |
| Prosodic | Maximum, minimum, mean of energy and its first-order difference; Maximum, minimum, mean of pitch and its first-order difference; Maximum, minimum, mean, variance of first three formants and their first-order differences |
| Nonlinear | Maximum, minimum, mean, median, variance of Hurst exponent; Maximum, minimum, mean, median, variance of minimum delay time; Maximum, minimum, mean, median, variance of correlation dimension; Maximum, minimum, mean, median, variance of Kolmogorov entropy; Mean, median, variance of maximum Lyapunov exponent |

As shown in Table 1, this experiment selects features from two perspectives: acoustic features extracted based on the short-term stationary characteristics of speech signals, i.e., MFCC features (60 dimensions) and prosodic features (38 dimensions), and nonlinear features (23 dimensions) extracted based on the chaotic characteristics of speech. After fusion, a 121-dimensional feature set is obtained.

### 3.3 Clustering Analysis Results

The purpose of emotion clustering analysis is to improve PAD prediction accuracy for speech. To ensure that the selection of the clustering distance threshold is not affected by a single database, clustering analysis statistics were performed on speech PAD prediction values from both the TYUT2.0 database and the Berlin database against basic emotion PAD values to select a clustering distance that yields better clustering effects. Table 2 shows the number of emotional speech utterances contained in different clustering threshold intervals based on different clustering distances.

**Table 2 Number of Speech Utterances in Different Cluster Thresholds**

| Distance Interval | TYUT2.0 Database | Berlin Voice Library |
|---|---|---|
| [0.0, 0.3] | 20 | 18 |
| [0.3, 0.6] | 35 | 32 |
| [0.6, 0.9] | 28 | 30 |
| [0.9, 1.2] | 22 | 25 |
| [1.2, 1.5] | 15 | 18 |
| [1.5, 1.8] | 12 | 10 |
| [1.8, 2.1] | 8 | 5 |
| [2.1, 2.4] | 5 | 0 |
| [2.4, 2.7] | 3 | 0 |
| [2.7, 3.0] | 2 | 0 |

From the distribution of speech utterances in different intervals for the two databases shown in Table 2, when distances are distributed within [0, 1.8], the number of speech utterances in both databases remains basically consistent. However, when greater than 1.8, the distribution numbers fluctuate, and in the Berlin voice library, there are no corresponding speech distributions. Therefore, the clustering distance threshold in this paper is set to 1.35.

### 3.4 Experimental Results and Analysis

The prediction results of the clustering PSO-LSSVM regression model are compared with those of the LSSVM regression model and the PSO-LSSVM regression model. The prediction performance for P, A, D is evaluated based on the data distribution of predicted values versus annotated values and experimental performance metrics.

### 3.4.1 Data Distribution

To more intuitively compare the data distribution of prediction results versus annotated values for the three regression models, taking the prediction results for the TYUT2.0 database as an example, Figures 4 [Figure 4: see original paper] through 6 show the comparisons between predicted and annotated values for the P, A, and D dimensions, respectively.

As can be seen from Figure 4, for sad and angry speech, the predicted values of Model 2 and Model 3 are relatively similar and closer to the annotated P values than Model 1, while for happy emotional speech, Model 3' s prediction results are significantly closer to the annotated P value distribution than the other two models. From Figure 5, the variation trends of Model 2 and Model 3 prediction results are closer to the variation trend of annotated A values than Model 1, with Model 3' s predicted values being closer to annotated values and having smaller relative errors than Model 2, indicating that Model 3 provides more accurate and stable A value predictions for each utterance compared to the other two regression models. From Figure 6, Model 3's prediction results for the D dimension of the three emotions are closer to annotated values, especially for angry and happy emotions, where the dispersion of Model 3' s predictions is closer to the dispersion of actual D values. These results demonstrate that Model 3, the clustering PSO-LSSVM regression model, produces predictions closer to the annotated P, A, D values.

### 3.4.2 Performance Metrics

Experimental performance metrics use Mean Absolute Error (MAE) and model coefficient of determination ($R^2$) as evaluation criteria, with Pearson correlation coefficient ($r$) introduced to measure the variation trend between model predictions and annotated values. Smaller MAE is better, while $R^2$ and $r$ closer to 1 are better. Their expressions are:

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

$$r = \frac{\sum_{i=1}^{n}(y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2 \sum_{i=1}^{n}(\hat{y}_i - \bar{\hat{y}})^2}}$$

where $n$ is the number of samples, $y_i$ is the actual annotated value, and $\hat{y}_i$ is the model predicted value.

Tables 3 and 4 show the experimental performance comparisons of the three regression models for P, A, D prediction in the TYUT2.0 database and Berlin voice library, respectively.

**Table 3 Comparison of Experimental Performance Indexes in TYUT2.0 Database**

| Dimension | Model | MAE | $R^2$ | Pearson $r$ |
|---|---|---|---|---|
| P | Model 1 | 0.42 | 0.68 | 0.83 |
| P | Model 2 | 0.38 | 0.75 | 0.87 |
| P | Model 3 | 0.31 | 0.84 | 0.92 |
| A | Model 1 | 0.35 | 0.72 | 0.85 |
| A | Model 2 | 0.32 | 0.78 | 0.89 |
| A | Model 3 | 0.26 | 0.86 | 0.93 |
| D | Model 1 | 0.45 | 0.65 | 0.81 |
| D | Model 2 | 0.41 | 0.71 | 0.85 |
| D | Model 3 | 0.34 | 0.80 | 0.90 |

**Table 4 Comparison of Experimental Performance Indexes in EMO-DB**

| Dimension | Model | MAE | $R^2$ | Pearson $r$ |
|---|---|---|---|---|
| P | Model 1 | 0.38 | 0.71 | 0.85 |
| P | Model 2 | 0.40 | 0.73 | 0.86 |
| P | Model 3 | 0.33 | 0.81 | 0.91 |
| A | Model 1 | 0.41 | 0.69 | 0.83 |
| A | Model 2 | 0.37 | 0.74 | 0.87 |
| A | Model 3 | 0.30 | 0.82 | 0.92 |
| D | Model 1 | 0.36 | 0.74 | 0.86 |
| D | Model 2 | 0.33 | 0.79 | 0.89 |
| D | Model 3 | 0.27 | 0.87 | 0.94 |

Based on Tables 3 and 4, the following conclusions can be drawn:

a) Comparing Model 2 with Model 1 for P, A, D dimension prediction: The Pearson correlation coefficient shows that Model 2' s correlation coefficient in the TYUT2.0 database is improved to some extent compared to Model 1. The coefficient of determination shows that Model 2' s $R^2$ is improved in both databases compared to Model 1, indicating better data fitting. The MAE shows that Model 2' s experimental errors are reduced in the TYUT2.0 database, while in the Berlin database, prediction errors for A and D are reduced, but the error for P increases. This indicates that Model 2 provides more stable prediction effects for A and D dimensions, demonstrating that the PSO algorithm' s optimization of LSSVM regression parameters improves prediction performance to a certain degree.

b) Comparing Model 3 with Model 2 for P, A, D dimension prediction: The Pearson correlation coefficient shows that Model 3' s prediction correlation coefficients for both databases are improved compared to Model 2, indicating that Model 3' s prediction variation trends are more similar to annotated value trends. The coefficient of determination shows that Model 3 provides better data fitting than Model 2. The MAE shows that Model 3' s experimental errors are smaller than Model 2' s errors in both databases. This demonstrates that Model 3' s prediction performance for PAD dimensions is superior to Model 2' s, verifying that emotion clustering analysis can improve prediction accuracy by reducing correlations between features.

c) Comparing the three models' prediction results for P, A, D dimensions across both databases: In the TYUT2.0 database, the three regression models show better prediction performance for the A dimension than for P and D dimensions. In the Berlin voice library, the three regression models show better prediction performance for the D dimension than for P and A dimensions. This difference is related to the different database construction methods: the Berlin voice library is a performance-based recorded database, while TYUT2.0 is obtained by intercepting radio dramas. Although the two databases have different emphases in emotional expression, resulting in varying prediction effects, Model 3 consistently provides varying degrees of improvement for P, A, D dimension predictions. Moreover, compared with literature [15] results for P, A prediction, Model 3' s coefficient of determination for P, A prediction in both databases is higher than that of the KNN-based prediction model in the literature by 0.24 and 0.35, respectively. This confirms that Model 3' s predictions are superior.

In summary, Model 3 not only significantly improves the prediction performance for emotional dimensions P, A, D but also applies to different types of emotional databases, demonstrating that the clustering PSO-LSSVM model has strong PAD prediction capability and good data universality. The reasons are: PSO algorithm optimization of LSSVM model regression parameters avoids subjective blindness in parameter selection, improving prediction results to a certain extent; emotion clustering analysis reduces the impact of correlations between emotional features on prediction accuracy by preliminary emotion clustering; therefore, the proposed clustering PSO-LSSVM regression model not only makes the regression process more objective but also enables more accurate P, A, D dimension predictions by reducing feature correlations.

## 4 Conclusion

To achieve objective quantification of emotions from dimensional perspectives, this paper proposes a prediction method combining PSO-optimized LSSVM with clustering analysis—the clustering PSO-LSSVM regression model—for predicting emotional dimensions P, A, D, and compares its results with those of LSSVM and PSO-LSSVM regression models. Experimental results show that the clustering PSO-LSSVM regression model provides better prediction results for emotional dimensions P, A, D because it integrates the advantages of PSO algorithm and emotion clustering analysis. The PSO algorithm' s optimization of regression parameters effectively avoids subjective blindness in parameter selection, while emotion clustering analysis processing of PSO-LSSVM' s initial prediction results reduces correlations between emotional features, thereby achieving more accurate and intelligent P, A, D prediction. These results can more precisely reveal the emotional dimensions of speech. Future work can predict the coordinate points of speech samples in PAD space and analyze their relationship with basic emotions in the dimensional space based on the distribution of emotional speech in PAD three-dimensional space, providing a quantitative method for analyzing the composition elements and proportions of complex emotions.

## References

[1] Bu Zhan, Wu Zhiang, Cao Jie, et al. Affective computing and game theory based prediction for online reviews [J]. Acta Electronica Sinica, 2015, 43 (12): 2530-2535.

[2] Li Chunxiang, Ding Xiaoda, Ye Jihong. Fluctuating wind velocity forecasting based on LSSVM with hybrid ACO & PSO [J]. Journal of Vibration and Shock, 2016, 35 (21): 131-136.

[3] Geng Liyan. Forecast of logistics demand using LSSVM combining GRA with KPCA [J]. Journal of Transportation Systems Engineering and Information Technology, 2015, 15 (1): 137−142.

[4] Zhao P C, Liu B, Gao W, et al. Multiple kernel least square support vector machine model for prediction of cement clinker lime content [J]. CIESC Journal, 2016, 67 (6): 2480−2487.

[5] Zhao Y, Niu R, Ling P, et al. Prediction of landslide deformation based on rough sets and particle swarm optimization-support vector machine [J]. Journal of Central South University, 2015, 46 (6): 2324−2332.

[6] Yu H, Chen Y, Hassan S G, et al. Prediction of the temperature in a Chinese solar greenhouse based on LSSVM optimized by improved PSO [J]. Computers & Electronics in Agriculture, 2016, 122: 94−102.

[7] Gorjaei R G, Songolzadeh R, Torkaman M, et al. A novel PSO-LSSVM model for predicting liquid rate of two phase flow through wellhead chokes [J]. Journal of Natural Gas Science & Engineering, 2015, 24: 336−343.

[8] Song Jing, Zhang Xueying, Sun Ying, et al. Emotional speech recognition based on PAD emotion model [J]. Microelectronics & Computer, 2016, 33 (9): 128-131.

[9] Wang Li. Research on emotional prediction method of V-A spatial continuous dimension [D]. Zhenjiang: Jiangsu University, 2015.

[10] Chu Yuwei, Luo Xiaobo, Qu Ke, et al. DBSCAN and K-means hybrid clustering based automatic dental feature detection [J]. Journal of Computer-Aided Design & Computer Graphics, 2018 (7): 1276-1283.

[11] Tan Feigang, Liu Weiming, Huang Ling, et al. Object re-identification algorithm based on weighted euclidean distance metric [J]. Journal of South China University of Technology: Natural Science Edition, 2015 (9): 88-94.

[12] Zhang Xueying, Zhang Ting, Sun Ying, et al. Emotional speech database optimization and quantitative annotation based on PAD emotion model [J]. Journal of Taiyuan University of Technology, 2017, 48 (3): 469-474.

[13] Burkhardt F, Paeschke A, Rolfes M, et al. A database of German emotional speech [C]// Proc of INTERSPEECH 2005-Eurospeech, European Conference on Speech Communication and Technology. Lisbon, Portugal: DBLP, 2005: 1517-1520.

[14] Yao Hui, Sun Ying, Zhang Xueying. Research on nonlinear dynamics features of emotional speech [J]. Journal of Xidian University, 2016, 43 (5): 167−172.

[15] Giannakopoulos T, Pikrakis A, Theodoridis S. A dimensional approach to emotion recognition of speech from movies [C]// Proc of IEEE International Conference on Acoustics, Speech and Signal Processing. 2009: 65-68.

*Note: Figure translations are in progress. See original paper for figures.*