Prediction of Internet User's Purchase Behavior Based on Mixed kernel SVM Model

Wen Hu¹, Yuxue Shi²,*

¹School of Computer and Information Engineering, Harbin University of Commerce, Harbin, China
²Heilongjiang Provincial Key Laboratory of Electronic Commerce and Information Processing, Harbin, China
*Corresponding author's e-mail: 18845778684@163.com

Abstract. In order to better predict the purchase behavior of online consumers and improve the purchase conversion rate of e-commerce companies, a model of consumer purchase intention prediction based on mixed kernel function support vector machine is proposed. In the research, the prediction model of consumer purchase intention is designed as a two-classification problem, and the feature selection algorithm mRMR is used to obtain the ranking of features, seeking to obtain better or similar classification results when choosing fewer features. On its basis, the particle swarm optimization algorithm (PSO) is used to optimize the parameters of the model. Experiments show that, to a certain extent, the constructed mixed kernel function can effectively improve the classification effect of SVM.

Keywords. SVM, Game theory, Kernel function, Behavior prediction

1. Introduction
In recent years, with the development of the Internet, my country's e-commerce industry has developed rapidly, and the number of online consumers has also been increasing. In fact, the conversion rate has not increased at the same rate. Therefore, how to use good analysis and forecasting methods to analyze the relevant data generated by consumers on the e-commerce platform, so as to predict and understand consumers' shopping habits and purchase intentions, as well as their behavioral rules is very important for companies. On this basis, companies can not only bring consumers a better shopping experience, but also change from passive display to active recommendation of products to consumers, so as to meet the specific needs of consumers and harvest more potential consumers to increase purchase conversion rate. Among them, Sakar et al. [1] proposed a real-time online shopper behavior analysis system based on multi-layer perceptron and recurrent neural networks. This system not only predicts visitors' shopping intentions, but also predicts the possibility of visitors abandoning the website. At the same time, Baati K et al. [2] made a supplement to this system and proposed a prediction analysis of online shoppers' buying intention based on random forest. The experimental results show that compared with other technologies, the random forest method has superiority. Liu et al. [3] used the support vector machine model to predict the future purchase status of online consumers through the feature extraction of consumer behavior data, and obtained the expected results.
Support vector machine is a method developed on the basis of statistical learning theory by Vapnik [4] and others. It not only has a strict theoretical basis, but also solves practical problems such as small samples, nonlinearity, high dimensions, and local minima. It exhibits many unique advantages, so it has become one of the fastest-growing research directions in the 1990s, and is widely used in pattern recognition and time series prediction and other fields [5]. SVM can handle non-linear problems, and the introduction of kernel function is the key. It solves the linear inseparable situation in the original space by mapping data to high-dimensional space. Although kernel functions have been applied and practiced in many fields, most of them are based on single-kernel methods. If a mixed kernel function can be constructed with the characteristics of two types of kernel functions at the same time, it will effectively improve the learning ability of the SVM classifier. Therefore, in recent years, many mixed kernel functions based on multiple kernels have also appeared one after another. Most of them use linear combination methods, multiplication combination methods, compound calculation methods and two-by-two combination methods to construct mixed kernel functions. It is verified through experiments that the mixed kernel has a better classification effect than single kernel in solving some problems [6-11].

Game theory is a science that studies the decision makers’ strategic choices and strategic equilibrium when there are competitors. It aims to help us understand the phenomena we observe when decision-making interactions. It has a wide range of applications in many disciplines and fields [12]. The game model mainly contains three elements: the player in the game, the strategy set of each player in the game, and the payoff function that calculates the “profit and loss” of each player in the game. The most concerned problem in the game is how to find the equilibrium situation in the problem through the game to ensure the interests of each participant in the game. Bargaining is the most common phenomenon in the market economy and the most classic dynamic game in game theory. The bargaining game model was proposed by Ariel Rubinstein [13] in 1982, also known as the Rubinstein model. Rubinstein regarded the bargaining process as a cooperative game process. He modeled this process by taking two participants splitting a piece of cake as an example. The two participants participating in the competition or confrontation each have different goals and interests, in order to achieve their respective goals, both parties must consider the opponent's various possible coping strategies and try to choose the most beneficial or reasonable coping strategy for themselves. Based on this, this article attempts to apply the bargaining model in game theory to the construction of mixed kernel functions, and proposes a prediction model of consumer purchase intention based on mixed kernel function SVM, and strives to balance the functions of the two types of kernel functions in the mixed kernel function. So that SVM classification has better learning ability and generalization ability.

2. Mixed kernel based on game theory

2.1. Choice of kernel function
As a learning mechanism, SVM algorithm can be divided into linear SVM and nonlinear SVM. Linear SVM determines the partition structure based on the Euclidean distance between samples. The nonlinear SVM replaces the inner product with an appropriate kernel function, and implicitly maps the nonlinear data to a high-dimensional space. Given a set of sample label pairs \((X_i, Y_i), (i=1, 2, ..., I)\) the optimization problem solved by SVM is:

\[
\begin{align*}
\text{minimize} & \quad \frac{1}{2} ||w||^2 + C \sum_{i=1}^{I} \varepsilon_i \\
\text{subject to} & \quad Y_i (w^\top \phi(X_i) + b) \geq 1 - \varepsilon_i, \varepsilon_i \geq 0
\end{align*}
\]  

(1)

In the above formula, the vector \(X_i\) is mapped to the high-dimensional space through the function \(\phi(X)\). \(c\) is the penalty parameter of the error term, which is used to adjust the ratio of the credible interval and the empirical risk in the feature space. Its size determines the degree of punishment for the wrong sample \(\varepsilon_i \geq 0\). \(v\) is called the slack variable, which represents the amount of function interval that the corresponding data point is allowed to deviate. \(k(X_i, Z_j) = \phi(X_i)^\top \phi(Z_j)\) is called the kernel
function, and represents the mapping from vector $X_i$ to the inner product feature space. In SVM, commonly used kernel functions are shown in Table 1. This article will also choose these two kernel functions to construct its mixed kernel.

| Kernel function                       | Formula                                                                 |
|---------------------------------------|-------------------------------------------------------------------------|
| Radial Basis Function Kernel          | $K_{RBF}(X_i, Z_j) = \exp\left(-\frac{||X_i - Z_j||^2}{2\sigma^2}\right)$, $\sigma > 0$ |
| Polynomial Kernel function            | $K_{Polynomial}(X_i, Z_j) = (X_i^T Z_j + c)^d$, $d \in N, c \geq 0$     |

Table 1. Kernel function names and formulas

For the RBF kernel function and Poly kernel function, they are two types of functions with opposite characteristics. The characteristics of the functions are shown in Figure 1 and Figure 2. It can be seen from Figure 1 that the RBF kernel only affects samples within a small range of the test point, but its interpolation ability is strong. In Figure 2, the interpolation ability of the Poly kernel is relatively weak compared with the RBF kernel, but it is good at extracting the global features of the sample points, and its function value is mainly affected by the sample points with a relatively scattered distribution and a relatively large range. It can not only affect the sample points within a small range of the test point, but also has a strong influence on the sample points relatively far from the test point, and has a strong generalization ability. Therefore, it can be said that the RBF kernel function and the Poly kernel function are two complementary kernel functions, and mixing them will produce stronger learning capabilities.

2.2. The construction of mixed kernel

2.2.1. Basic assumptions of the game model

(1) Because the global kernel function and the local kernel function are classified through different hyper-planes. Therefore, in the model, it is assumed that there are only two players in the game, namely the RBF kernel function and the Poly kernel function, denoted by $X$ and $Y$ respectively;

(2) Both players are rational, and the game information is complete;

(3) According to the different characteristics of the participants, suppose that in the game, let $X$ propose the plan first, and $Y$ propose the plan again;

(4) In this game, suppose $X$ and $Y$ are competitively divided into 1 unit in order to find the largest proportion of each;
(5) Although the game can be played indefinitely, there is cost consumption, so the discount factor \( \delta(0<\delta<1) \) is introduced;

(6) In the process of the game, let \( s_i(i=1,2,3...) \) be the percentage of \( x \), then the percentage of \( y \) is \( 1-s_i \), which is recorded as the partition plan \((s_i,1-s_i)\), \(0<s_i<1\);

2.2.2. The establishment of game model In order to construct a mixed kernel function, that is, to determine the proportion of the RBF kernel function and Poly kernel function in the mixed kernel function, assuming that the RBF kernel function and Poly kernel function take turns to propose a proportion plan, it can be roughly divided into three stages:

According to the hypothesis, the first stage \( x \) proposes the proportion plan \((s_i,1-s_i)\), \(0<s_i<1\). If \( y \) accepts this plan, the game ends at this time, the proportion of \( x \) is \( s_i \), and the proportion of \( y \) are \( 1-s_i \); at this time, the gains of both parties are \((s_i,1-s_i)\); if the plan is rejected, the game will continue and enter the second stage;

In the second stage, \( y \) first proposes the proportion plan \((s_i,1-s_i)\), \(0<s_i<1\). If \( x \) accepts this plan, the game ends at this time, the proportion of \( x \) is \( s_i \), and the proportion of \( y \) is \( 1-s_i \); at this time, the benefits of both parties are \((\delta s_{x2}, \delta(1-s_{y2}))\), \(0<\delta<1\). If \( x \) rejects this plan, the game will continue and enter the third stage;

In the third stage, \( x \) first proposed the proportion plan \((s_i,1-s_i)\), \(0<s_i<1\). If \( y \) accepts this plan, the game ends, the proportion of \( x \) is \( s_i \), the proportion of \( y \) is \( 1-s_i \); at this time, the benefits of both parties are \((\delta^2 s_{x3}, \delta^2(1-s_{y3}))\), \(0<\delta<1\); If \( y \) rejects this plan, the game will continue and enter the fourth stage. By analogy, \( x \) proposes plans at stages 1, 3, 5..., and \( y \) proposes plans at stages 2, 4, 6....

Assuming that the two sides proceed to the final stage of the game, it means that the game will end regardless of whether the outcome is agreed. When in the \( n \)th stage, \( n \) is an odd number, \( x \) proposes a-plan as \((s_{n1},1-s_{n1})\), \(0<s_{n1}<1\). If \( y \) accepts this plan, the game ends at this time, the proportion of \( x \) is \( s_{n1} \), and the proportion of \( y \) is \( 1-s_{n1} \). The two parties reach an agreement with their respective proportions \((s_{n1},1-s_{n1})\), and the benefits of both parties are \((\delta^{n+1} s_{x}, \delta^{n+1}(1-s_{y}))\), \(0<\delta<1\); If \( y \) refuses to accept this plan, it means that the division fails, the cooperation between the two parties is not reached, and the income from the bargaining of both parties is zero; when \( n \) is an even number, \( x \) proposes the plan at this time, and the same is described above.

2.2.3. Equilibrium analysis of game model In order to obtain the equilibrium solution of the above game model, this article will refer to Shaked and Sutton [14] to solve the indefinite bargaining game, and convert the indefinite game of the participants RBF kernel function and Poly kernel function into a three-stage game, and use the reverse Inductive method [10] to solve.

First, start the analysis from the third stage of the game. For rational \( x \), this is the last opportunity for it to cooperate with \( y \) to construct a mixed kernel. If it refuses, it means that the cooperation will fail, and the two parties have no chance to create more utility through cooperation. Therefore, the plan proposed by \( x \) is \((s_{n1},1-s_{n1})\), \(0<s_{n1}<1\), and \( y \) must accept it. At this time, the benefits of \( x \) and \( y \) are \((\delta^2 s_{x3}, \delta^2(1-s_{y3}))\), \(0<\delta<1\).

Then reverse to the second stage of the game, \( y \) will consider what kind of plan will be proposed to maximize its own benefits. If \( y \) proposes a plan that makes \( x \)'s return less than the third stage's return, he will definitely reject \( y \)'s proposal at this stage Scheme, the game will proceed to the third stage. Therefore, the \( s_{n1} \) in the plan \((s_{n1},1-s_{n1})\), \(0<s_{n1}<1\) proposed by \( y \) must not only make \( x \) accept it, but also make its own income greater than the income of the third stage. This is the optimal division plan. Therefore, \( s_{n1} \) should satisfy the equation \( \delta s_{x3} = \delta^2 s_{x3} \), that is, \( s_{n1} = \delta s_{x3} \), and the income of is \( \delta(1-s_{y3}) = \delta(1-\delta s_{y3}) \). And the income of \( y \) in the third stage is \( \delta^2(1-s_{y3}) \). Because of \( 0<\delta<1 \), and
\( \delta(1-\delta s_i) - \delta^2(1-s_i) = \delta(1-\delta) \) at this time, the income of \( Y \) in the second stage is larger than that in the third stage.

Finally, it is reversed to the first stage of the game. Since the information of the game is complete, \( X \) knows that its own profit in the third stage is \( \delta^2 s_i \), and also knows the strategy of the second stage \( Y \), so it proposes the plan \((s_i, 1-s_i)\) in the first stage, which is to make \( Y \) satisfied, you must also make your own gains greater than the gains in the second stage. Therefore, the optimality condition that \( s_i \) should satisfy is: \( 1-s_i = \delta(1-\delta s_i) \), that is, \( s_i = 1-\delta + \delta^2 s_i \).

In the end, the perfect equilibrium solution of the sub-game about the distribution ratio of the game is obtained as \((1-\delta + \delta^2 s_i, \delta - \delta^2 s_i)\), where \( 0 < \delta < 1, \ 0 < s_i < 1 \). Therefore, the mixed kernel function obtained based on the bargaining game model can be expressed as:

\[
K_m(X_i, X_j) = (1-\delta + \delta^2 s_i)K_x + (\delta - \delta^2 s_i)K_r
\]

3. SVM classification based on mixed kernel

3.1. Effectiveness of mixed kernel

When SVM deals with nonlinear problems, the kernel function is the key. For the constructed kernel function, to judge whether it is valid, it is necessary to check whether it satisfies the Mercer condition. The Mercer condition is a sufficient condition for testing whether the kernel function defines a feature space, and the kernel function that satisfies the Mercer condition is called an effective kernel function, that is, an allowable kernel [15].

According to Mercer's theorem [16], if the function \( \kappa \) is a mapping on \( R^n \times R^n \rightarrow R \) (that is, mapping from two \( n \) vectors to the real number field). Then if \( \kappa \) is an effective kernel function (also called Mercer kernel function), then if and only if for training example \( \{x^{(1)}, x^{(2)}, ..., x^{(m)}\} \), the corresponding kernel function matrix is symmetric positive semi-definite. From the basic properties of the kernel function, we can see that if \( k_i(x_i, z_i) \) and \( k_j(x_j, z_j) \) are both kernel functions on \( R^n \times R^n \), including \( \alpha \geq 0 \), then functions (3) and (4) are still kernel functions:

\[
k(x, z) = k_i(x_i, z_i) + k_j(x_j, z_j)
\]

(3)

\[
k(x, z) = \alpha \cdot k_i(x_i, z_i)
\]

(4)

Therefore, it can be known that the mixed kernel function obtained based on the bargaining game satisfies the Mercer condition, which makes the feasibility of the constructed mixed kernel function theoretically guaranteed.

3.2. Optimization of parameters

For SVM, the penalty parameter \( C \) and the relevant parameters of the kernel function determine its generalization ability. But at present, there is still no universal parameter selection method, most of which are based on prior knowledge, expert knowledge or continuous attempts in experiments to obtain the best parameters of SVM [17]. Based on this, in the mixed kernel function (2), let \( \delta = \sigma \), \( s_i = \frac{1}{d} \). When the value of \( \sigma \) of the RBF kernel function is 0.2, 0.4, 0.5, and the value of \( d \) of the Poly kernel function is 1, 2, and 4, the graph of the constructed mixed kernel function can be drawn, and the characteristics of the mixed kernel function can be obtained as shown in Figure 3. It can be seen from Figure 3 that when the test point is 0.5, the learning ability and generalization ability of the mixed kernel performs better than the RBF kernel and the Poly kernel, it can not only extract local information near the test point, but also obtain global information far away from the test point.
The classification performance of SVM is affected by many factors, and the more critical influencing factors are related parameters. Therefore, only by choosing the appropriate penalty factor and the parameters of the kernel function can the SVM classifier obtain good generalization performance.

From the explanation of the mixed kernel function in Section 1.2-2.2, the expression of the constructed mixed kernel function based on the game is obtained as:

$$K_M(X_i, Z_i) = (1-\sigma^2)\exp\left(-\frac{\|X_i-Z_i\|^2}{2\sigma^2}\right) + \sigma^2 \exp\left((X_i^T Z_i + c)^{\frac{1}{d}}\right)$$

(5)

It can be seen from the above formula (5) that the parameters to be optimized are penalty factors $C$, $\sigma = \sigma$, $d$ (where $c=0$ is taken as the homogeneous polynomial kernel function). In this paper, the particle swarm algorithm is used for parameter optimization, the number of iterations is set to 500, the population number is set to 50, and a 5-fold cross-validation method is used.

4. Experiment and analysis

4.1. Data set

In the research of this article, the prediction model of consumer purchase intention is designed as a two-category problem to measure the user’s willingness to complete the transaction. This data set is composed of feature vectors belonging to 12330 sessions. The formation of the data set is to make each session belong to a different user within a one-year time period, so as to avoid trends that tend to specific activities, special days, user profiles or time periods. In the 12330 sessions in the data set, there are a total of 18 features, including 10 numerical features and 8 categorical features. 84.5% (10422) of the negative samples that did not end with shopping are marked as "0", the rest (1908) are positive samples at the end of shopping, and are marked as "1". In order to improve the classification performance of the predictive model, feature selection technology will be applied to feature selection to study whether it is possible to obtain better or similar classification performance while selecting fewer features. For the ranking of features, the Max-Relevance and Min-Redundancy (mRMR) algorithm will be applied instead of using a wrapper algorithm that requires a learning algorithm, because it may lead to a reduction in the feature set specific to the classifier. Before feature selection, the data will be normalized. Finally, the feature ranking obtained using the mRMR algorithm is shown in Table 2.

In order to obtain the classification effect of different number of features, in the experiment process, the first 3 features in the feature ranking will be selected for experiment, and then 1 feature will be added one by one until all features are selected, in order to find the best classification effect. Feature number. Since the problem studied in this article is a two-class classification problem, in order to
better describe the classification effect of the model, a confusion matrix will be used to describe it to evaluate the classification effect of the model. All experiments in this article are performed on the platform of the libsvm software package designed and developed by Professor Lin from Taiwan in MATLAB.

Table 2. Feature ranking based on mRMR feature selection algorithm

| Rank | Feature name                  | Rank | Feature name                  |
|------|------------------------------|------|------------------------------|
| 1    | Page value                   | 10   | Product related duration     |
| 2    | Month                        | 11   | Special day                  |
| 3    | Exit rate                    | 12   | Informational                |
| 4    | Weekend                      | 13   | Traffic type                 |
| 5    | Informational duration       | 14   | Administrative               |
| 6    | Region                       | 15   | Bounce rate                  |
| 7    | Operating system             | 16   | Browser                      |
| 8    | Administrative duration      | 17   | Product related              |
| 9    | Visitor type                 |      |                              |

4.2. Experimental results and analysis

In order to verify the performance of the mixed-kernel SVM model in predicting consumer purchase behavior, this paper compares its model with the classification effect of the SVM model based on single-kernel RBF and Poly. The single-kernel SVM parameter optimization method will use a grid search algorithm based on cross-validation [18]. The relevant index results of each model training are shown in Figure 4-7.
As can be seen from Figures 4-7, as the number of features increases, the relevant indicators of each model are constantly changing. The classification accuracy of the single-kernel SVM model does not change much, and the model accuracy basically changes around 0.85. The classification accuracy of the mix-kernel SVM is increasing with the increase of the number of features. The classification accuracy is higher than that of the single-kernel SVM. Only when the number of features is 7, the classification accuracy of the SVM-Poly model is higher. Precision and recall are a pair of contradictory measures. Generally speaking, when the accuracy of the model is high, the recall is often low, and the performance of the model cannot be judged. Therefore, the F1 value will be used for comparison. It can be seen from Figure 7 that the F1 value of the SVM-Mix model is higher than that of the single-kernel SVM model, and it gradually increases with the increase of the number of features.

| Model       | Number of features | Accuracy | Precision | Recall | F1    |
|-------------|--------------------|----------|-----------|--------|-------|
| SVM-Poly    | 9                  | 0.87     | 0.62      | 0.60   | 0.61  |
| SVM-RBF     | 14                 | 0.86     | 0.57      | 0.60   | 0.58  |
| SVM-Mix     | 17                 | 0.89     | 0.74      | 0.57   | 0.65  |

For the single-kernel SVM, the SVM model based on the Poly kernel performs better. Table 3 shows the best training results obtained by different models and the number of input features in the corresponding models. The best result of SVM-Poly model training is when the number of features is 9, the F1 value reaches 0.61. The best classification effect obtained by the model SVM-RBF is that when the number of features is 14, the F1 value reaches 0.58. For the constructed game-based mixed kernel SVM model, the classification effect obtained is that when all the features are selected, that is, when the number of features is 17, the F1 value of the model reaches 0.65. Compared with the SVM by 0.04 and 0.07. It can be seen that the SVM-Mix model uses more features in the best model. Therefore, from the overall index of model training, compared with the single-kernel SVM model, the constructed mix-kernel model has a better classification effect, and performs better in solving the problem of consumer buying behavior prediction.

5. Conclusions
Aiming at the problem of predicting the purchase behavior of network users, a game-based mixed kernel function SVM model is created to predict the purchase behavior of users. The kernel function is an important part of the SVM. This paper solves the defects of the single-kernel kernel function by constructing a mixed kernel function, so that it can have the advantages of two types of kernels, give better play to its performance, and make the model have better learning ability and generalization ability, so as to more accurately predict the future purchase intention of online users, realize precision marketing, and improve the operational efficiency of e-commerce enterprises.

References
[1] Sakar C O, Polat S O, Katircioglu M, et al 2018 Real-time prediction of online shoppers’ purchasing intention using multilayer perceptron and LSTM recurrent neural networks (Neural Computing and Applications).
[2] Baati K, Mohsil M 2020 Real-time prediction of online shoppers’ purchasing intention using random forest vol 12 (Artificial Intelligence Applications and Innovations) p 518
[3] Liu Y, Pi D, Cheng Q 2016 Ensemble kernel method: SVM classification based on game theory vol 271 (Journal of Systems Engineering and Electronics) pp 251-259
[4] Vapnik V N 1995 The nature of statistical learning theory (New York: Springer-Verlag)
[5] Wang H Y, Li J H, Yang F L 2014 Summary of support vector machine theory and algorithm research vol 3105 (Computer Application Research) pp 1281-1286
[6] Chi E N, Li C X, Zheng X F 2017 Downwind non-Gaussian space wind pressure prediction based on wavelet and multiplicative mixed kernel function LSSVM vol 3609 (Vibration and Shock) pp 116-121

[7] Zhang L L, Shi Y K, Li J Y 2018 Study on water enrichment of roof sandstone based on mixed kernel function support vector machine vol 4502 (Mining Safety and Environmental Protection) pp 72-76

[8] Li Z S, Chen X X, Wen C J, et al 2020 Research on the classification model of amanita fungi morphological features based on machine learning vol 4101 (Chinese Journal of Agricultural Machinery Chemistry) pp 136-143

[9] Li S, Li L W, Zhuang D F, et al 2015 Research on mixed kernel function and its application in data modeling vol 3207 (Computer Simulation) pp 1-6

[10] Liu Z K 2016. An improved mixed kernel function support vector machine text classification method, vol 2906 (Industrial Control Computer) pp 113-114+117

[11] Zhou Z G, Guo Y, Li K Q 2016 Research on intelligent diagnosis method of support vector machine based on improved kernel function vol 3405 Light Industry Machinery pp 3405:23-26.

[12] Osborne M J, Rubinstein A 1994 A course in Game Theory (Economica) p 63249

[13] Rubinstein A 1982 Perfect equilibrium in a bargaining model vol 501 (Econometrica) pp 97-109

[14] Shaked A, Sutton J 1984 Involuntary unemployment as a perfect equilibrium in a bargaining model vol 526 (Econometrica) pp 1351-1364

[15] Smola A J, Bernhard Scholkopf 2004 A tutorial on support vector regression vol 143 (Statistics and Computing) pp 199-222

[16] Hofmann T, Scholkopf B, Smola A J 2008 Kernel methods in machine learning vol 363 (The Annals of Statistics) pp 1171-1220.

[17] Cherkassky V, Ma Y 2004 Practical selection of SVM parameters and noise estimation for SVM regression vol 171 (Neural Networks) pp 113-126

[18] Guo L N, Fan G S 2018 Surface soil bulk density prediction based on grid se-arch and cross-validation support vector machine vol 19 (Soil Bulletin) pp 512-518