1. Introduction

Internet micro-blog is a new internet media with real-time and interactivity, which make the information spread more timely and faster on the internet, and have important significance to perform analysis and modeling to the micro-blog hot topics and predict its development trend. People have paid close attentions to the micro-blog hot topic prediction, however, being influenced by various factors, it is a complicated non-liner dynamic system[1].

In the recent years, many scholars have researched the micro-blog hot topic prediction and proposed various prediction methods. The traditional methods mainly adopt the time series method to analyze and predict the micro-blog hot topics[2]. The time series method correctly reflects the variation laws of the micro-blog hot topics, but the prediction accuracy is relatively lower. With the development of artificial intelligence technology, there have emerged the micro-blog hot topic prediction methods such as the artificial neural network and support vector machine, etc. However, the neural network has the defects of local minimum, overfitting and weak generalization capacity, etc, in order to improve the prediction accuracy of the micro-blog hot topics, a micro-blog hot topic prediction model (MSO-LSSVM) improving the quantum-behaved Particle Swarm optimization (QPSO) algorithm LSSVM has been proposed. Firstly, it adopts the MQPSO algorithm to optimize the parameters of LSSVM, and then use the optimized LSSVM to perform modeling to the trend of the micro-blog hot topics, and finally select the micro-blog hot topic data for simulation experiment. The experiment results show that, MQPSO-LSSVM has improved the prediction accuracy of the micro-blog hot topics, and the prediction results have certain practical values.

Keywords: MICRO-BLOG HOT TOPIC; QPSO; PARAMETER OPTIMIZATION; LEAST SQUARES SUPPORT VECTOR MACHINE

Abstract

The micro-blog hot topic prediction is a complex prediction issue of small sample and instability, the traditional linear method can’t characterize the variation laws of the micro-blog hot topics, and the neural network has the defects of overfitting and weak generalization capacity, etc, in order to improve the prediction accuracy of the micro-blog hot topics, a micro-blog hot topic prediction model (MSO-LSSVM) improving the quantum-behaved Particle Swarm optimization (QPSO) algorithm LSSVM has been proposed. Firstly, it adopts the MQPSO algorithm to optimize the parameters of LSSVM, and then use the optimized LSSVM to perform modeling to the trend of the micro-blog hot topics, and finally select the micro-blog hot topic data for simulation experiment. The experiment results show that, MQPSO-LSSVM has improved the prediction accuracy of the micro-blog hot topics, and the prediction results have certain practical values.

1. Faculty of Computer Engineering, Huaiyin Institute of Technology, Huaian, Jiangsu, 223003, China
2. College of Computer and Information, HOHAI University, NanJing, Jiangshu, 211100, China
cress of optimizing the research, particle swarm optimization (OSP) algorithm is a swarm intelligence search algorithm with relatively simple structure and faster convergence speed\(^5\). However, it is easy for the PSO algorithm to produce the premature convergence, and fall into the local optimum value, and the quantum-behaved Particle Swarm optimization (QPSO) algorithm is a new optimization algorithm\(^6\) proposed when researching the PSO algorithm convergence. However, as the same with the standard PSO algorithm, QPSO algorithm still can produce the premature convergences. When the QPSO algorithm converges certain accuracy, it fails to continue to optimize the search, hence falls into the local optimum solution. MQPSO algorithm has strong capacity of global search which has solved the defects of premature convergence and local optimum value in PSO algorithm and QPSO algorithm, it can discover the LSSVM optimum parameter by adopting the MQPSO algorithm to further improve the prediction accuracy of the micro-blog hot topics\(^7\).


2.1. Least Squares Support Vector Machine

LSSVM seeks a optimum function relationship reflecting the sample data, and the un-liner regression function of LSSVM is:

\[
f(x) = w^T \cdot \varphi(x) + b
\]

(1)

In this function, \(\omega\) is the weight vector, and \(b\) is the offset value\(^8\).

In order to seek the coefficients \(\omega\) and \(b\), introduce the slack variable, the regression model of LSSVM solved in function (1) is:

\[
\min \|w\|^2 + \frac{1}{2} \gamma \sum_{i=1}^{n} (\xi_i + \xi_i^*)
\]

s.t. \(y_i - w^T \varphi(x) + b = e_i\)

(2)

In this function, \(\gamma\) is the positive associated parameter; \(e\) is the error.

Basing on the principle of structure risk minimization, introduce the Lagrangian multiplier to perform the dual optimization problem to the variation in function (2), namely:

\[
L(w, b, \xi, \alpha) = \frac{1}{2} w^T w + \frac{1}{2} \gamma \sum_{i=1}^{n} \xi_i^2 + \sum_{i=1}^{n} \alpha_i (w^T \varphi(x_i) - b + \xi_i - y_i)
\]

(3)

In this function, \(\alpha_i\) is the Lagrangian multiplier according to the optimization condition

\[
\frac{\partial L}{\partial w} = 0, \frac{\partial L}{\partial b} = 0, \frac{\partial L}{\partial \xi_i} = 0, \frac{\partial L}{\partial \alpha_i} = 0
\]

we can obtain that:

\[
\begin{align*}
\left\{ \begin{array}{l}
\omega = \sum_{i=1}^{n} \alpha_i \varphi(x_i) \\
\sum_{i=1}^{n} \alpha_i = 0 \\
w \varphi(x_i) + b + \xi_i - y_i = 0
\end{array} \right.
\]

(4)

According to Mercer condition, kernel function definition \(K(x_i, x_j) = \varphi(x_i)^T \varphi(x_j)\), t last, the regression model of LSSVM ca be obtained as:

\[
f(x) = \sum_{i=1}^{n} \alpha_i K(x_i, x_j) + b
\]

(5)

In this article, it selects RBF kernel function as the LSSVM kernel function, the RBF kernel function is:

\[
k(x_i, x_j) = \exp \left(-\frac{||x_i - x_j||^2}{2\sigma^2} \right)
\]

(6)

And at last, the LSSVM regression model is:

\[
f(x) = \sum_{i=1}^{N} \alpha_i \exp \left(-\frac{||x_i - x||^2}{2\sigma^2} \right) + b
\]

(7)

In this function, \(\sigma\) is the width of RBF kernel function.

2.2. Parameter Analysis of LSSVM

The parameters of LSSVM have huge influences on the generalization ability and prediction accuracy of model, in this article, LSSVM adopts the RBF kernel function (RBF), the major factors influencing the model performances including the positive factor \(\gamma\) and the kernel parameter \(\sigma\); \(\gamma\) is mainly used to adjust the proportion of the confidence range and empirical risk of the support vector machine in the determinate feature space to make the generalization ability of the support vector machine the best. The selection of \(\gamma\) has huge influences on the stability and complicity of the model. The kernel parameter \(\sigma\) mainly reflects the degree of correlation between the support vectors, and has huge influences on the generalization ability of the model. If the value of \(\sigma\) is too small, the model will be relatively complex, and the generalization ability can’t be guaranteed; if the value of \(\sigma\) is too large, it will be difficult for the model to reach the enough prediction accuracy\(^9\).

The optimization selection of parameter is a very important issue in the field of LSSVM model research, whose essence is a process of optimization
research. Hence, it can adopt the swarm intelligence
optimization algorithm to perform the optimization
selection to the parameter of LSSVM model. Deter-
mine the optimum parameter of LSSVM model and
improve the prediction accuracy of LSSVM model.

2.3. QPSO Algorithm

In the classic PSO algorithm research process, the
particle research can’t cover the whole feasible space,
hence the classic PSO algorithm can’t guarantee to
discover the global optimum solution. In the quantum
space, for the quantum satisfies the qualities of the
clustering status, and can search in the whole feasible
space, hence, after researching the convergence be-
haviors of particles, Sun et al. Proposed the quantum
particle swarm optimization (QPSO) algorithm from
the mechanical aspect.

Assume the research space as D-dimension space,
and the swarm has n particles. QPSO algorithm up-
dates the location of particle \( i \) through the following
formula:

\[
\begin{align*}
    x_i(t + 1) &= p - \alpha \times |m_{best} - x_i(t)| \times \ln \frac{1}{u} \quad \text{if} \,(u > 0.5) \\
    x_i(t + 1) &= p + \alpha \times |m_{best} - x_i(t)| \times \ln \frac{1}{u} \quad \text{if} \,(u \leq 0.5)
\end{align*}
\]

(8)

\[m_{best} = \frac{1}{n} \sum_{i=1}^{n} p_{best_i} = \left(\frac{1}{n} \sum_{i=1}^{n} p_{best_{1i}}\right) + \frac{1}{n} \sum_{i=1}^{n} p_{best_{2i}} + \frac{1}{n} \sum_{i=1}^{n} p_{best_{3i}}\]  

(11)

2.4. MQPSO algorithm optimizing LSSVM

Although QPSO algorithm has stronger global re-
search ability than the standard PSO, as the standard
PSO algorithm, there also have premature convergen-
ces in QPSO algorithm, and fall into the local opti-
mum solution, the crossover operations in the genetic
algorithms can make the progeny particles inherit the
strong points of the parental particles, and enhance
the research ability to the particle areas, and increase
the particle diversity, and make the particles better
discover the global optimum solution. For this, in this
article, it firstly improves the QPSO algorithm, and
introduces the crossover operations of the genetic al-
gorithm into the QPSO algorithm, and then adopts the
improved QPSO (MPSO) algorithm to perform the
optimization selection to the parameters of LSSVM,
and establish the MQPSO-LSSVM prediction model,
and finally use the MQPSO-LSSVM prediction mod-
el for the prediction of the micro-blog hot topics. The
steps of MQPSO algorithm optimizing the parameter
\( \gamma \) and \( \sigma \) of LSSVM are as follows:

(1) Initialization. Randomly produces M particles,
each particle includes parameter \( \gamma \) and \( \sigma \). Assume the
global sample set needs to be solved as U, namely the
swarm. Initialize various parameters required in the
algorithm. The initializing position \( x_i \) of the particle;
the maximum iteration number \( N \) of the algorithm;
convergence curacy.

(2) Decode each particle as the training param-
eter of LSSVM, study the training samples. Use the
adaptability function to evaluate each particle and
calculate the adaptability value of each particle. The
adaptability function is defined as the liner functions
of all sample error squares and reciprocals, and the
details are as shown in formula(12), from the adapt-
ability formula, it can be obtained that the less the
error, the bigger the value of the corresponding adapt-
ability, and the better the adaptability.

\[Fitness = \frac{A}{\sum_{i=1}^{N} (t_i - y_i)^2} + B\]  

(12)

In this formula, \( Fitness \) is the adaptability func-
tion, both \( A \) and \( B \) are constants, \( H \) is the sample
number, \( y_i \) means the prediction value of the micro-blog
hot topics, \( t_i \) means the actual value of the micro-blog
hot topics.

(1) After each updating of the swarm, sort the
adaptability of all particles according to the descend-
ing order. Evenly divide the sorted swarm into 2 parts.
The part with better adaptability, namely the M/2 particles with bigger adaptability value is sub-swarm U1, and the M/2 particles with smaller adaptability value is sub-swarm U2.

(2) Introduce the crossover operations of the genetic algorithm. The sub-swarm U1 with bigger adaptability uses the formula (8) to update the particle positions. Sub-swarm U2 will perform the crossover operations of the genetic algorithm and randomly cross match between the two, and perform the crossover operations by using the crossover probability pc. The realization formula performing the crossover operations to the positions of particle i and particle j is as follows:

\[
\begin{align*}
    x_i(t+1) &= \theta_1 \times x_i(t) + (1 - \theta_1) \times x_j(t) \\
    x_j(t+1) &= (1 - \theta_2) \times x_i(t) + \theta_2 \times x_j(t)
\end{align*}
\] 

(13)

In this formula, \( \theta_1 \) and \( \theta_2 \) are the random values between [0, 1].

(3) Compare the adaptability values of the crossover operated progeny particles and the adaptability values of the parental particles, and retain the particles with bigger adaptability into the next iterative evolution.

(4) Update the individual extremum. After introducing the crossover operations of the genetic algorithm through step (4), compare the new adaptability values of the particles with the adaptability values of the individual extremumpbest, if the new adaptability values of the particles are superior to the adaptability values of the individual extremumpbest, then update the pbest, and endow the current new positions of the particles to pbest.

(5) Update the swarm global extremum. After introducing the crossover operations of the genetic algorithm through step (4), compare the new adaptability values of the particles with the adaptability values of the global extremumpbest, if the new adaptability values of the particles are superior to the adaptability values of the global extremumpbest, then update the pbest, and endow the current new positions of the particles to pbest.

(6) Perform the repeated iteration. Repeat the above mentioned steps (2)–(6), till the objective function reaches the convergence accuracy or the iteration number reaches the set maximum number, the optimum parameters of LSSVM will be obtained when the training is over.

3. Simulation Examples

3.1. Data Sources

In P4 2.8GHz CPU, 4G RAM, the operation system is P4 2.8GHz CPU, 4G RAM environment and realize the algorithm through Matlab 2012 program.

Sina Micro-blog is the most widely used micro-blog in China at present, due to its huge user involvement amount, it makes the emergency events rapidly spread on the internet with very sensitive response to the emergency events, the experimental data is from the Star in Danger: the High Dive micro-blog data in April, 2013, and the details are as shown in Figure 1. In order to better compare the prediction results, select the BP neural network (BP-NN) prediction model, PSO-LSSVM prediction model, QPSO-LSSVM prediction model and MQPSO-LSSVM prediction model for the prediction of the micro-blog hot topics.

![Figure 1. Time series of Internet public opinions](image)

3.2. Data Pre-treatment

Due to the various vectors have different amount grades in the original micro-blog hot topic data samples, and the amount grades have huge differences, hence, it is necessary to perform the generalization treatment to the original micro-blog hot topic data sample, and in this article, it uses the mapminmax function to perform the generalization treatment to the original micro-blog hot topic data sample to make the treated data evenly distribute within the range of [0, 1], and the generalization formula is:

\[
x' = \frac{(y_{\text{max}} - y_{\text{min}}) \times (x - x_{\text{min}})}{x_{\text{max}} - x_{\text{min}}} + y_{\text{min}}
\] 

(14)

In this formula, \( x \) is the actual value of the micro-blog hot topics, \( x_{\text{max}} \) and \( x_{\text{min}} \) are respectively the maximum value and minimum value of the micro-blog hot topics. \( y_{\text{min}} \) and \( y_{\text{max}} \) are the mapping range parameters.

The prediction result of micro-blog hot topics obtained at last needs to be performed with the uni-generalization treatment before obtaining the final prediction results, and the uni-generalization formula is:
3.3. Parameter Setting

(1) The value range of $\gamma$ in LSSVM is $[0.1, 100]$, and value range of $\sigma$ is $[0.01, 10]$.

(2) BP neural network selects network structure of three node numbers of the input layer, seven node numbers of hidden layer and one node number of output layer, the transfer function of the hidden layer is Sigmoid-type tangent function $\text{tanh}$, the transfer function of the output layer is purelin function, and the momentum coefficient $\eta = 0.8$, learning rate $\mu = 0.01$.

(3) The swarm amount $M = 40$ in PSO algorithm, inertial weight $w = 0.8$, learning factor $c_1 = 1.4$, $c_2 = 1.6$, maximum iteration number $N = 200$.

(4) The value of the contraction-expansion coefficient $\alpha$ in QPSO algorithm will decrease from 1 liner to 0.5 with the increase of the iteration numbers.

(5) The crossover probability $c_p$ used in the crossover operations of MQPSO algorithm is $c_p = 0.4$.

3.4. Evaluation Index of the Prediction Results

Select the average relative error $\text{MAPE}$ and the determination coefficient $R^2$ as the evaluation indexes of the prediction results, $\text{MAPE}$ and $R^2$ are defined as:

$$MAPE = \frac{1}{l} \sum_{i=1}^{l} \left| \frac{y_i - t_i}{y_i} \right|$$  \hspace{1cm} (16)

$$R^2 = \frac{(l \sum_{i=1}^{l} t_i y_i - \sum_{i=1}^{l} t_i \sum_{i=1}^{l} y_i)^2}{(l \sum_{i=1}^{l} t_i^2 - \sum_{i=1}^{l} t_i^2)(l \sum_{i=1}^{l} y_i^2 - \sum_{i=1}^{l} y_i)^2}$$  \hspace{1cm} (17)

4. Result and Analysis

Firstly put the micro-blog hot topic training sample set into LSSVM for learning, and select the optimum LSSVM parameters with PSO algorithm, QPSO algorithm AND MQPSO algorithm, and the obtained LSSVM optimum parameters are seen in Table 1.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>$\gamma$</th>
<th>$\sigma$</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSO Algorithm</td>
<td>102.89</td>
<td>0.31</td>
</tr>
<tr>
<td>QPSO Algorithm</td>
<td>95.96</td>
<td>0.30</td>
</tr>
<tr>
<td>MQPSO Algorithm</td>
<td>208.44</td>
<td>0.69</td>
</tr>
</tbody>
</table>

From the prediction results of micro-blog hot topics in Figure 2 and Figure 3 as well as the prediction result evaluations in Table 2 it can be seen that, LSSVM model has better prediction performances and effects than BP neural network model. Comparing with the BP neural network prediction model, PSO-LSSVM prediction model and QPSO-LSSVM prediction model, the average relative error $\text{MAPE}$ of MQPSO-LSSVM has been respectively reduced by 0.48%, 1.17% and 1.47%, while the determination coefficient has been respectively increased by 4.23%, 3.08% and 0.95%, the smaller the value of the average relative error $\text{MAPE}$, means the smaller the prediction error of the model, and the higher the prediction accuracy. The size of the determination coefficient $R^2$ determines the relevant close relationship among the variables. The lager the determination coefficient...
Table 2. Prediction effect evaluation of each model

<table>
<thead>
<tr>
<th>Prediction model</th>
<th>MAPE(%)</th>
<th>$R^2$(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BP-NN</td>
<td>4.18</td>
<td>94.92</td>
</tr>
<tr>
<td>PSO-LSSVM</td>
<td>3.70</td>
<td>96.07</td>
</tr>
<tr>
<td>QPSO-LSSVM</td>
<td>3.01</td>
<td>98.20</td>
</tr>
<tr>
<td>MQPSO-LSSVM</td>
<td>2.71</td>
<td>99.15</td>
</tr>
</tbody>
</table>

Conclusions

Because it is difficult for the traditional liner prediction method to correctly reflect the variation laws of the micro-blog hot topics, and the neural network has the defects of local minimum, overfitting and weak generalization, a micro-blog hot topic prediction model optimizing LSSVM based on MQPSO is proposed, and the experimental results show that, MQPSO-LSSVM micro-blog hot topic prediction method can perform the predictions to the micro-blog hot topics more accurately, and has wide application prospect in the micro-blog hot topics.

References


---

Research on Enterprise Comprehensive Evaluation Model of Subjective Trust Based on Fuzzy Theory

LIU Ta¹, HANG Dong-Ping¹, ZHOU Hang²

¹.Department of Management, Harbin Institute of Technology , Harbin,150000, China
².Department of Accounting, Harbin University of Commerce, Harbin,150000, China

Abstract
In view of the features of information trust such as its subjectiveness and uncertainty in the enterprise comprehensive evaluation environment, the author presents a subjective trust evaluation model based on fuzzy theory. The model uses fuzzy theory to get the calculation formula of comprehensive trust evaluation between nodes, introduces time factor and constraint mechanism for bad faith node in trust calculation, calculates comprehensive trust value by using similarity degree reverse weight, and achieves cluster analysis of trust value using fuzzy equivalence relations. The simulation results analysis proves the effectiveness and feasibility of the model mentioned above, and the simulation comparison verifies that the model can objectively reflect the situation close to the real.