Research on the Middle-Distance Running Achievement Prediction for College Students Based on LSSVMGAS

Xiaobing Yang, Suqiong Feng*

Sichuan Agricultural University
fengtingting0828@163.com

By the combination of the computer technology and the physical training, we can better predict the training achievement. Then we can make the targeted training for the college students. Therefore, the combination of the physical training for the college students and the computer technology has become a trend. Middle-distance running is one of the most important standards for the physical training of the college students. Improving the middle-distance running performance not only improves the physical quality of the college students, but also improves the perseverance of the college students. In this paper, we combine the LSSVM method with the GAS method and propose the improved LSSVMGAS method. In the experimental part, we use the LSSVMGAS method to predict achievement of the middle-distance running for the college students and achieve the better results.

1. Introduction

In June 30, 2015, health and family planning commission issues a report. The report shows that the average height of adult male and female is 167.1cm and 155.8cm in China. It is lower than the average level of Japan. In this report, we can see that the physical quality of Chinese people is not optimistic. In National Standard for Sports Training, the middle-distance running is easy to be neglected of the college students. To middle-distance running not only enhances the physical quality for the college students, but also hones the character and will. Qiuhe Huang, Lanyong Wei and Haifeng Huo (2014) researched the college teachers’ teaching ability.

Tang Chifei (2008) studied the running train for the college students. The author thought that the running train for the college students must be in accordance with their own growth and make the most effective use of their energy. Chi Hua (2013) et al. researched the college students’ running pain sensation scale. According to the item analysis, reliability and validity, the sensation scale could better reflect the degree and level of the feeling pain of running for the college students. Guan Qingli and Wang Haiyuan (2010) proposed that the long-distance running should take the physical education and the extra-curricular activities as the carrier. And it needed the participation of the teachers and the students. Zhao Fangye (2013) thought that the college physical education must adapt to the physical conditions of the contemporary college students. We should make the reform from the education philosophy, curriculum structure, teaching management link and the assessment indicator. In addition, Lin Zhenglan (2013), Pang Rong (2011) also studied the question.

Support vector machine (SVM) was a data mining method based on the statistical learning theory. Tao Lin (2015) et al. studies the SVM Least squares support vector machine (LSSVM) was a new support vector machine. The operation speed of the LSSVM was faster than other support vector machines. Therefore, it had the wide application. Now, the LSSVM had many applications in the industrial chemical by Mohammad Mesbah et al. (2015), Hossein, Safari et al. (2014) Hamidreza Yarveicy et al. (2014), energy by Xing Yan and Nurul A (2013) and weather by Yi Zhang et al.. Gravitational search algorithm (GSA) was an intelligent optimization algorithm which was proposed in 2009. The thought of GSA algorithm was derived from Newton’s law of gravity. It guided the optimization search according to the group intelligence which generated by the gravity among the particles. The gravitational search concept was simple, easy to achieve and needed to adjust a few parameters. Now, Oliveria, Zikang Su, Beatriz Gonzalez and other people (2015) studied the gravitational search algorithm.
In this paper, we combine the LSSVM with GAS and propose an improved LSSVMGAS method. Then, we use the method to predict the middle-distance running performance for the college students and achieve the good results. The structure of this paper is as follows. The first part is the introduction. In this part, we present the research status of the related content. The second part introduces the LSSVM. The third part is the basic principle of GSA. In the fourth part, we combine the LSSVM with GSA method and propose the improved LSSVMGAS method. The fifth part is the numerical analysis and the last part is the conclusion.

2. The LSSVM method

We suppose the training data set is \((x_i, y_i), i = 1, 2, \ldots, N\). \(x_i\) and \(y_i\) are the input vector and the corresponding output respectively. The input space \(\mathbb{R}^d\) is mapped into a feature space \(Z\) though the nonlinear function \(\phi(x_i)\).

The linear function of the feature space defines as follows.

\[
y = f(x, \omega) = \omega^T \phi(x) + b
\]

\(\omega \in Z, b \in R\)

According to the principle of the risk minimization, we can get the objective function

\[
\min_{\omega, b} \frac{1}{2} \omega^T \omega + \frac{1}{2} \sum_{i=1}^{N} \xi_i^2
\]

s.t. \(y_i = \omega^T \phi(x_i) + b + \xi_i\)

\(\gamma > 0, i = 1, 2, \ldots, N\)

\(\gamma\) is the penalty factor, \(\xi_i\) is an error variable to indicate an error between the true output value of the sample \(i\) and its estimated value.

We can construct the corresponding Lagrange function:

\[
L(\omega, b, \xi, \beta) = \frac{1}{2} \omega^T \omega + \frac{1}{2} \sum_{i=1}^{N} \xi_i^2 - \sum_{i=1}^{N} \beta_i [\omega^T \phi(x_i) + b + \xi_i - y_i]
\]

Where \(\beta_i\) is the Lagrange multiplier.

3. Basic principle of GSA

The gravitational search algorithm is a kind of group intelligence optimization algorithm which is proposed by Esmat professor. The magnitude of the gravitational force is proportional to the mass of the two particles. And it is inversely proportional to the distance of the two particles.

\[
F = G \frac{M_i M_j}{R^2}
\]

The gravitational force \(F_{ij}\) between the particle that the quality is \(i\) and the particle that the quality is \(j\) is expressed by the following function.

\[
F_{ij} = G \frac{M_a \times M_p}{R^2}
\]

\[a_i = \frac{F_i}{M_i}\]

\(M_a\) is the active gravitational mass of the particle \(i\). \(M_p\) is the passive gravitational mass of the particle \(i\). \(M_i\) is the inertia mass of the particle \(i\).

In the search space of \(D\) dimension, we assume that there are \(N\) particles. We define the location of the \(i\) particle as \(X_i = (x_{i1}, \ldots, x_{id})\) and \(i = 1, 2, \ldots, N\). Where, \(N\) is the population quantity. \(x_{id}\) is the location of the \(i\) particle in \(d\) dimension. Now, we assume that the gravity and the mass of the particle are determined by the target function value of the search space. We calculate the inertia mass of the particle according to the formula (7) and (8), and we get the normalized mass according to the formula (9).

\[
M_a = M_p = M_i = M, i = 1, 2, \ldots, N
\]

\[
m_i(t) = \frac{\text{fit}_i(t) - \text{worst}(t)}{\text{best}(t) - \text{worst}(t)}
\]
\[ M_{i(t)} = \frac{m_i(t)}{\sum_{j=1}^{N} m_j(t)} \tag{9} \]

Where, \( \text{fit}_i(t) \) is the fitness degree value of the \( i \) particle at \( t \) time.

Then, we can calculate the gravity of each particle in each dimension according to the following formula. The gravity of the \( i \) particle and the \( j \) particle in \( d \) dimension is,

\[ F_{d}^{i} = G(t) \cdot \frac{M_{i(t)} \times M_{j(t)}}{R_{ij(t)} + \epsilon} \cdot (x_{j}^{d}(t) - x_{i}^{d}(t)) \tag{10} \]

\( x_{j}^{d}(t) \) represents the location of the \( j \) particle in \( d \) dimension. \( x_{i}^{d}(t) \) represents the location of the \( i \) particle in \( d \) dimension. \( \epsilon \) is a very small constant which is to prevent the denominator is zero.

In the gravitational search algorithm, in order to increase the random properties of the algorithm, we think that the force of the particle \( i \) is equal to the sum of other particles. The value is defined as follows.

\[ F_{d}^{i}(t) = \sum_{j=1, j \neq i}^{N} \text{rand}_{j} F_{d}^{j}(t) \tag{11} \]

Where, \( \text{rand}_{j} \) is a random number which belongs to \([0,1]\).

According to Newton’s second theorem, the acceleration of the particle \( i \) in \( d \) dimension at \( t \) time is defined as follows.

\[ a_{d}^{i}(t) = \frac{F_{d}^{i}(t)}{M_{i(t)}} \tag{12} \]

Where, \( M_{i(t)} \) is the inertia mass of the particle \( i \) at \( t \) time.

We update the speed and the location according to the following formula.

\[ v_{d}^{i}(t+1) = \text{rand}_{k} \times v_{d}^{i}(t) + a_{d}^{i}(t) \tag{13} \]

\[ x_{d}^{i}(t+1) = x_{d}^{i}(t) + v_{d}^{i}(t+1) \tag{14} \]

We assume that the number of the particles is \( N_{\text{best}} \). And the \( N_{\text{best}} \) changes with the time. Therefore, \( N_{\text{best}} \) decreases from the initial \( N_0 \). In the initial time, the particles have forces. As the time goes on, some particles that the masses are bigger begin to force other particles that the masses are smaller. To the end, there is only particle forcing other particles. Therefore, the formula can be changed as follows.

\[ F_{d}^{i}(t) = \sum_{k \in K_{\text{best}}} \text{rand}_{k} F_{d}^{k}(t) \tag{15} \]

Where, \( K_{\text{best}} \) is the set of the particles that the inertia masses are bigger.

4. LSSVMGSA

This paper establishes the predicted model of LSSVM and GSA. Firstly, we improve the GSA algorithm. In addition, GSA will optimize the parameter of the LSSVM model. The linear combination of the intermediate nodes is the output and each intermediate node corresponding to a support vector.

In GSA algorithm, the expected particle is guided by the resultant. And it moves to the particle that the quality is heavy. At the same time, during the moving of the expected particle, it is also influenced by other particles. Therefore, when we calculate the resultant, there may be some deviations. Therefore, we introduce the correction factor to modify the resultant force in the algorithm.

\[ x = [M(i) - M(j)] \]

\[ y(x) = \begin{cases} x \geq C \\ \frac{\pi}{2} & x < C \\ k_1 \sin(C) & 0 \leq x < C \end{cases} \]

\[ \lambda = k_2 \cdot \cos(y(x)) \]

Where, \( \lambda \) is the correction factor that the force needs to add. \( k_1 \) and \( k_2 \) are the adjustable parameters. And they can be adjusted according to the different locations. \( C \) is a coefficient which is related to the \( k_2 \).

\[ C = \frac{\pi k_2}{2} \]

\[ y(x) \] is a function. And the new formula is as follows.
\[ F_j^S(t) = \sum_{i=1, i \neq j}^{N} \hat{\lambda} \cdot F_{ij}^S(t) = \sum_{i=1, i \neq j}^{N} k_i \cdot \cos(k_j \cdot (|M(i) - M(j)|)) \cdot F_{ij}^S(t) \]  

(19)

The fitness function is defined as:

\[ \text{fitness} = \frac{1}{N} \sum_{i=1}^{N} (y_i - y_p)^2 \]  

(20)

\( N \) is the number of the training samples. \( y_i \) is the true value, \( y_p \) is the fitted value.

The flow chart of LSSVMGAS is as follows.

![Flow chart of LSSVMGAS](image)

**Figure 1: Flow chart of LSSVMGAS**

### 5. Numerical analysis

In the numerical experiment, we predict the long-distance running performance for the college students. We choose 4 college students to run and predict their performances. There are two male college students and two female college students. The male college students run 1000 meters and the female college students run 800 meters. Firstly, we collect data of 30 groups' long-distance running. The first 20 groups are as the training sets and the last 10 groups are as the predicted sets. The data of the training sets is as follows.

**Table 1: The data of the training set**

<table>
<thead>
<tr>
<th>No</th>
<th>Male1</th>
<th>Male2</th>
<th>Female1</th>
<th>Female2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>224</td>
<td>210</td>
<td>214</td>
<td>225</td>
</tr>
<tr>
<td>2</td>
<td>226</td>
<td>209</td>
<td>218</td>
<td>226</td>
</tr>
<tr>
<td>3</td>
<td>218</td>
<td>203</td>
<td>211</td>
<td>217</td>
</tr>
<tr>
<td>4</td>
<td>226</td>
<td>208</td>
<td>208</td>
<td>220</td>
</tr>
<tr>
<td>5</td>
<td>230</td>
<td>205</td>
<td>205</td>
<td>221</td>
</tr>
<tr>
<td>6</td>
<td>217</td>
<td>214</td>
<td>210</td>
<td>214</td>
</tr>
<tr>
<td>7</td>
<td>214</td>
<td>212</td>
<td>213</td>
<td>218</td>
</tr>
<tr>
<td>8</td>
<td>210</td>
<td>210</td>
<td>211</td>
<td>219</td>
</tr>
<tr>
<td>9</td>
<td>215</td>
<td>208</td>
<td>216</td>
<td>225</td>
</tr>
<tr>
<td>10</td>
<td>223</td>
<td>213</td>
<td>213</td>
<td>227</td>
</tr>
<tr>
<td>11</td>
<td>225</td>
<td>211</td>
<td>218</td>
<td>225</td>
</tr>
<tr>
<td>12</td>
<td>228</td>
<td>214</td>
<td>223</td>
<td>224</td>
</tr>
<tr>
<td>13</td>
<td>219</td>
<td>207</td>
<td>220</td>
<td>221</td>
</tr>
<tr>
<td>14</td>
<td>220</td>
<td>194</td>
<td>213</td>
<td>218</td>
</tr>
<tr>
<td>15</td>
<td>223</td>
<td>205</td>
<td>211</td>
<td>217</td>
</tr>
<tr>
<td>16</td>
<td>216</td>
<td>203</td>
<td>206</td>
<td>219</td>
</tr>
<tr>
<td>17</td>
<td>224</td>
<td>206</td>
<td>214</td>
<td>223</td>
</tr>
<tr>
<td>18</td>
<td>225</td>
<td>218</td>
<td>215</td>
<td>224</td>
</tr>
<tr>
<td>19</td>
<td>233</td>
<td>220</td>
<td>218</td>
<td>221</td>
</tr>
<tr>
<td>20</td>
<td>228</td>
<td>216</td>
<td>220</td>
<td>227</td>
</tr>
</tbody>
</table>
Then, we predict the performance. The predicted performance and the actual performance are as follows.

Table 2: The actual values and the predicted values

<table>
<thead>
<tr>
<th>No</th>
<th>Male1 Actual values</th>
<th>Male2 Actual values</th>
<th>Male1 Predicted values</th>
<th>Male2 Predicted values</th>
<th>Female1 Actual values</th>
<th>Female1 Predicted values</th>
<th>Female2 Actual values</th>
<th>Female2 Predicted values</th>
</tr>
</thead>
<tbody>
<tr>
<td>21</td>
<td>225</td>
<td>227</td>
<td>215</td>
<td>213</td>
<td>217</td>
<td>219</td>
<td>224</td>
<td>223</td>
</tr>
<tr>
<td>22</td>
<td>223</td>
<td>224</td>
<td>217</td>
<td>215</td>
<td>213</td>
<td>215</td>
<td>224</td>
<td>224</td>
</tr>
<tr>
<td>23</td>
<td>220</td>
<td>220</td>
<td>208</td>
<td>210</td>
<td>214</td>
<td>214</td>
<td>221</td>
<td>223</td>
</tr>
<tr>
<td>24</td>
<td>218</td>
<td>220</td>
<td>210</td>
<td>211</td>
<td>216</td>
<td>215</td>
<td>218</td>
<td>220</td>
</tr>
<tr>
<td>25</td>
<td>221</td>
<td>218</td>
<td>213</td>
<td>211</td>
<td>218</td>
<td>217</td>
<td>220</td>
<td>221</td>
</tr>
<tr>
<td>26</td>
<td>224</td>
<td>222</td>
<td>214</td>
<td>214</td>
<td>217</td>
<td>215</td>
<td>218</td>
<td>219</td>
</tr>
<tr>
<td>27</td>
<td>223</td>
<td>224</td>
<td>212</td>
<td>213</td>
<td>213</td>
<td>214</td>
<td>223</td>
<td>222</td>
</tr>
<tr>
<td>28</td>
<td>225</td>
<td>224</td>
<td>208</td>
<td>209</td>
<td>211</td>
<td>212</td>
<td>226</td>
<td>225</td>
</tr>
<tr>
<td>29</td>
<td>224</td>
<td>223</td>
<td>207</td>
<td>209</td>
<td>208</td>
<td>210</td>
<td>227</td>
<td>226</td>
</tr>
<tr>
<td>30</td>
<td>224</td>
<td>224</td>
<td>210</td>
<td>211</td>
<td>207</td>
<td>208</td>
<td>225</td>
<td>224</td>
</tr>
</tbody>
</table>

We compare the actual performance with the predicted performance. The results are as follows.

Figure 2: The experiment results of male1

Figure 3: The experiment results of male2

Figure 4: The experiment results of female1

Figure 5: The experiment results of female2

From the above figure, we can see that the predicted value which is obtained by the LSSVMGAS method is very close to the actual values. The two curves of each training set are very similar. From the experiment, we can know that the LSSVMGAS method achieves the good results. The experiment shows that the LSSVMGAS method is feasible and effective.

7. Conclusions

The middle-distance running is popular by the teachers and the students. And the majorly of teachers pay more and more attention to the middle-distance running with the computer technology. In this paper, we propose the LSSVMGAS method to predict the middle-distance running performance and the targeted training. The main work of this paper is as follows. Firstly, we introduce the research status of the related content. Secondly, we introduce the LSSVM method and GAS method. Thirdly, we combine the LSSVM method with the GAS method and propose the improved LSSVMGAS method. Fourthly, we use the LSSVMGAS method to predict the middle-distance running performance for the college students. And it achieves the good experimental results. The content of this part has some reference value for the research of the LSSVM and the physical quality of the college students.
Acknowledgment

This research is supported by Undergraduate Theses Incubation Program of Sichuan Agricultural University.

References


