SVM Method Used To Study Gender Differences Based On Microelement*

Yang Chun¹, Liu Yuan², Du Jun², Tang Bin² *

¹The Hospital Attached to Sinohydro Bureau 10 Dujiangyan, Sichuan Province, China
²Department of Chemistry Luzhou Medical College Luzhou, Sichuan Province, China
76858236@qq.com, tangbin8888@163.com

Abstract

[objective] Intelligent Algorithm of SVM is used for studying gender differences based on microelement data, which provide reference for the application of microelement in healthy people, such as providing technical support for the investigation of cases. [Method] Our Long-term test results on hair microelement of health people were consolidated. Support vector machine (SVM) is used to classified model of male and female based on microelement data. The radical basis function (RBF) is adopted as a kernel function of SVM, and the model adjusts C and σ to build the optimization classifier. [Result] Healthy population of men and women of manganese, cadmium and nickel are quite different, The classified model of Microelement based on SVM can classify the male and female, the correct classification ratio set to be 81.71% and 66.47% by SVM based on 7 test date and 3 test data selection. [conclusion] The classified model of microelement data based on SVM can classify male and female.

© 2012 Published by Elsevier B.V. Selection and/or peer review under responsibility of ICMPBE International Committee. Open access under CC BY-NC-ND license.

Keywords: Female; Male; Microelement; Support vector machine (SVM)

1. Introduction

The human body consists of many an element, which can be divided into macroelements and microelements. The former are the ones which take up more than 0.01% of total weight in the body like carbon, hydrogen, oxygen, nitrogen, etc. while the latter less than 0.01% like iron, zinc, copper, manganese, etc. Although the amount of microelements is much little, they still play an important part in human life. The little amount of microelements, even less than match point, can do giant physiological action. On the other hand, an excess or lack of them can result in physiological anomaly or sick. Microelements have close relationship with human health so that there have been many researches on it.

* Thanks for the sponsoring of Key Project by Office of Science and Technology in LuZhou city [the planning number of project: 2009-S-15(77)] and Scientific Research Project by Youth Fund in LuZhou Medical College (the application of SVM in disease forecasting based on the medical testing data).
The paper manages the testing methods and results for microelements in hair of healthy population and discusses the analysis result. It provides the references for the application of microelements in healthy population, for example, the technical support for crime solving.

2. Samples

The sample in the study contains 202 cases, which are 202 lots of natural hair with 3cm length approaching scalp on the place of occipital from both male and female with healthy body in the south of Si Chuan province from December, 1989 to April, 1998. The seven clinical testing items for microelements contain six elements and the age.

3. Data processing and the principle of method

3.1 The preprocessing for criterion data of microelements of nasopharyngeal

Criterion data of microelements is programmed and calculated by the matlab7.0 and Microsoft Office Excel 2007. The statistical results for the mean value of microelements (except gender) is represented in table 1.

From the form one, it contrasts the mean value of microelements in healthy male with that of healthy female and the relationship between them is the followings:
1. The number of the sample of male and female is same, that is, 101 persons;
2. The mean value of age of male and female has small differences;
3. The content of iron in both male and female is close;
4. The content of zinc and popper in both male and female has a few differences;
5. The content of manganese, cadmium and nickel in both male and female has big gap. The male’s is 1.79 times, 1.5 times and 1.3 times respectively than female’s.

From those numbers, we can see that there is a big gap of microelements in healthy male and female so that we can distinguish male and female according to the testing results of the content of manganese, cadmium and nickel in their body.

3.2 Support vector machine

Support vector machine is a machine learning algorithm based on statistical learning theory. Support vector machine develops from optimal separating hyperplane in the situation of linear separability. The so-called optimal separating hyperplane requires separating hyperplane not only to separate two classes correctly (training error rate is zero) but also make the class interval biggest. The former assures the minimization of empirical risk (zero) while the latter makes the minimization of confidence interval, which results in the minimization of real risk. The vector closest to optimal separating hyperplane is called support vector. The kernel function in the paper is Gauss radial basis function: radial basis function: $K(x, x') = \exp(-\frac{||x - x'||^2}{2\sigma^2})$

3.3 The analysis of data

The testing items for collected microelements in patients contain seven parts, that is, zinc, copper, iron, manganese, cadmium, nickel and age. At first, we treat the testing results of female as negative sample while the male as positive one. Then we process the data with Excel, program it in Matlab,
construct model by Libsvm and classify the data model above respectively with leave-one-out cross validation.

4. Results and discussion

4.1 The recognition rate of parameter collecting of SVM

We train nasopharyngeal and other miscellaneous diseases. At first, it should be the optimization parameters with 10-fold cross validation. The SVM parameters mainly contain penalty factor C and kernel bandwidth \( \sigma \). When to define C and \( \sigma \), it judges the merits and shortages of model according to the correct discrimination rate of model to training set and forecast set respectively.

4.2 The results

From the form above, we can see that the recognition accuracy is 81.77% if it constructs models with seven testing data of age, zinc, copper, iron, manganese, cadmium, nickel; and if it does this with three testing data of manganese, cadmium, nickel, on which there is a big gap between male and female, the recognition accuracy is 78.22% which is little lower than the former one. However, in this way, the cost of testing is saved and the analysis of computer speeds up because it has less testing items. Besides, if the time span of testing for patients can be cut down, the recognition accuracy can still be raised.

This model can distinguish male and female perfectly so that it can be the assistant reference for gender definition. The paper takes use of cheap microelements to construct forecast model for sex and obtains relatively ideal result. However, the data couldn’t include all situations because the microelements have differences in terms of region, race, etc. The research is worth being deepened as more data collected.

References


[4] Zhou Keyuan, Jie Xinming, etc · The Relationship Between the Content of Zinc, Copper and Iron in hair and gender and age in Zhan Jiang[J] · Studies of Trace Elements and Health,1993,04:52 ·


[10] Liu Ke, etc. Tentative Analysis for Microelements in Hair of 131 Nasopharyngeal Patients [J]. Modern Preventive Medicine , 2010(9).


Table 1 the testing results of microelements in both healthy male and female

<table>
<thead>
<tr>
<th>item number</th>
<th>gender (person)</th>
<th>Age(year)</th>
<th>zinc (ppm)</th>
<th>copper (ppm)</th>
<th>iron (ppm)</th>
<th>manganese (ppm)</th>
<th>cadmium (ppm)</th>
<th>nickel (ppm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>male</td>
<td>Age</td>
<td>Zn</td>
<td>Cu</td>
<td>Fe</td>
<td>Mn</td>
<td>Cd</td>
<td>Ni</td>
</tr>
<tr>
<td>101 person</td>
<td>44.40</td>
<td>177.96</td>
<td>8.85</td>
<td>21.09</td>
<td>1.61</td>
<td>1.31</td>
<td>2.30</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>female</td>
<td>Age</td>
<td>Zn</td>
<td>Cu</td>
<td>Fe</td>
<td>Mn</td>
<td>Cd</td>
<td>Ni</td>
</tr>
<tr>
<td>101 person</td>
<td>40.65</td>
<td>195.62</td>
<td>9.43</td>
<td>21.38</td>
<td>0.90</td>
<td>0.86</td>
<td>1.73</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>healthy male</td>
<td>Age</td>
<td>Zn</td>
<td>Cu</td>
<td>Fe</td>
<td>Mn</td>
<td>Cd</td>
<td>Ni</td>
</tr>
<tr>
<td>101 person</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>healthy female</td>
<td>Age</td>
<td>Zn</td>
<td>Cu</td>
<td>Fe</td>
<td>Mn</td>
<td>Cd</td>
<td>Ni</td>
</tr>
<tr>
<td>101 person</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>average of healthy male/ the average of healthy female</td>
<td>1.00</td>
<td>1.09</td>
<td>0.91</td>
<td>0.94</td>
<td>0.99</td>
<td>1.79</td>
<td>1.5</td>
</tr>
</tbody>
</table>

Testing items

<table>
<thead>
<tr>
<th>testing items</th>
<th>(Positive / negative sample size)</th>
<th>(best C)</th>
<th>(best δ)</th>
<th>accuracy rate (ACC)</th>
</tr>
</thead>
<tbody>
<tr>
<td>the seven data with six kinds of microelement and age</td>
<td>101/101</td>
<td>8.0</td>
<td>2.0</td>
<td>~81.77</td>
</tr>
<tr>
<td>the four data with three microelements with big differences and age</td>
<td>101/101</td>
<td>32.0</td>
<td>2.0</td>
<td>78.22</td>
</tr>
</tbody>
</table>