A SVR Model for Differentiation of Syndromes in Traditional Mongolian Medicine

Biao WANG¹,*, Aoqi², Nashunwuritu², Li-jun YANG¹, and Narisu¹

¹School of Computer Information Management, Inner Mongolia University of Finance and Economics, Hohhot, China
²Inner Mongolia International Mongolian Medicine Hospital, Hohhot, China

*Corresponding author

Keywords: Support vector regression, Traditional Mongolian medicine, Differentiation of syndromes, Cause of disease.

Abstract. Traditional Mongolian medicine (TMM) is the unique medicine of Mongolia nationality. It is meaningful to research, protect and develop TMM. In TMM, there are different causes for a disease. A doctor of TMM must not only diagnose what disease that a patient has but also what syndrome of the disease. Differentiation of syndromes (DS) in TMM is to distinguish diseases and their syndromes. DS is the key and difficult point of medical diagnosis and treatment. Doctors mainly do it according their experiences. So, it is necessary and useful to establish aided diagnosis models of DS for TMM. This study aims to establish a Support Vector Regression (SVR) model of DS. Firstly, we deduced the SVR. Then, we used medical cases to train the SVR model. Finally, we did experiments of validation. The experiments shown its correct rate was 76.6%.

Introduction

Traditional Mongolian medicine (TMM) is the unique medicine of Mongolia nationality and is widely used in Mongolian regions. It is the essence of Mongolia nationality’s brilliant culture and an important part of the world’s traditional medicines [1]. So, it is meaningful to protect and develop TMM.

In TMM, the key thing is to identify syndromes of diseases. Only if patient’s syndrome is correctly and accurately diagnosed, the corresponding treatment plan can be determined. Therefore, one of the key points of TMM is differentiation of syndromes. It also is the difficulty of TMM.

The syndrome type is directly related to the cause of the disease. TMM believes that although diseases are varied and complex, their reasons can summarized as HaoYi, Xila, Badagan, Chusi, Xieriwusi, Nian, Baori, Halun and Huiteng etc., and each can be further refined into several kinds. After careful analysis and accurate identification, any disease can be attributed to some of those causes. Then a treatment program can be found, and the disease can be cured [2, 3].

For example, in TMM, the palpitation disease can be classified as Haoyi syndrome, Xila syndrome, Huiteng syndrome, and Halun syndrome. Each syndrome needs different treatment methods. Therefore, it is the key to determine the treatment plan and cure the disease [4].

At present, researches on TMM mainly aim at the syndrome differentiation for a certain disease, analysis its statistical rules and the symptom indexes. Some typical studies are as follows.

Morigentu did the statistic analysis according to 382 cases of acute diarrhea patients in TMM [5] and found that the Nian syndrome had 131 cases, accounting for 34.3%. Xila syndrome 113 cases, accounted for 29.6%, Haoyi 71 cases, accounting for 18.6%, and Badagan 67 cases, accounted for 17.5%.

Jimusi did statistic analysis on angina pectoris of coronary heart disease of Mongolian medicine [6], screened meaningful indicators for syndrome diagnosis, and suggested a diagnostic criteria of angina pectoris of coronary heart disease in TMM.
Suhe studied the distribution of syndromes and the etiologic relationships according to 410 cases of coronary heart disease using statistical analysis [7].

In order to study the syndromes characteristics of the diabetic nephropathy of Mongolian medicine, Saren Liu collected and collated 25 Mongolian modern literatures and data and 320 cases of medical record [8]. The conclusion is that Mongolian medicine classification of diabetic nephropathy is lack of uniform standards.

In fact, most doctors of TMM mainly diagnose syndromes based on their experiences. It is necessary to research models of syndrome differentiation of Mongolian medicine. However, the models can rarely be found in literatures. So, in this paper, we established SVR model of syndrome differentiation in TMM.

**SVR Model of DS**

**TMM and SVR Model**

Like other medicines, TMM also have surgery, internal medicine, gynecology, pediatrics and so on. Comparatively, for DS, internal medicine is complex and representative. So, we chose internal medicine syndromes as experimental data.

Qigeqitu, in his book Modern Mongolian Medicine [4] which is a representative literature for TMM, systematically discussed and summarized the theory of Mongolian medicine. In the book, there are at least 25 kinds of diseases of internal medicine introduced, each containing some syndromes, and there are about 86 kinds of syndromes. For example, asthma disease includes Huiteng syndrome, Halun Syndrome, Nian, and Haoyi etc. In this paper, the 86 syndromes of internal medicine were used to carry out the experiments of DS.

TMM mainly uses diagnoses of inspection, inquiry, pulse, palpation, smelling and olfaction to obtain clinical symptoms of patients. A total of 45 categories of symptoms and 288 symptoms about internal medicine are concerned in the book. For example, the symptoms of "tongue" include red tongue, tongue dry and rough, dark tongue, purple tongue, pale tongue, thin and white coating on the tongue, thick yellow coating on the tongue, thin yellow coating on the tongue, yellow coating on the tongue, white greasy tongue coating, thin coating on the tongue, white slippery, yellow and greasy tongue coating, and yellow white coating on the tongue etc.

A syndrome $d$ is related to a group of clinical symptoms. For example, the Huiteng syndrome of asthma generally has clinical symptoms of (livid face, white and slippery tongue, sunken pulse, breath hold fullness in the chest, shortness of breath, colorless and odorless urine, greasy and sticky sputum, throat singing phlegm, thin sputum, rich white foam sputum, cold extremities and fear of cold).

Support vector machine (SVM) mainly includes support vector classification (SVC) and support vector regression(SVR) [9].

In TMM, there are more than 86 kinds of syndromes introduced only in the book Modern Mongolian Medicine [4]. The syndromes are too many to use SVC, so we use SVR model.

**SVR Model of Syndrome**

SVR model [9-11] is a logical and reasonable method, and its main idea is shown as following. Assume \( x = (x_1, x_2, x_3, \ldots, x_n) \) represents symptoms and is an input variable, \( y \) represents syndrome and is an output variable, \( w = (w_1, w_2, w_3, \ldots, w_n) \) is a coefficient, \( b \) is a constant. Then, we want to find a function:

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

The main idea about SVR is shown as the figure 1.
For given sample set \( D = \{(x_1, y_1), (x_2, y_2), \ldots, (x_m, y_m)\} \) of medical cases, the objective function of SVR is

\[
\min_{w, b} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{m} \xi_i \quad \text{(2)}
\]

Where, \( C \) is a regular constant, \( \xi = \begin{cases} 0, & \text{if } |f(x_i) - y_i| \leq \varepsilon \\ |f(x_i) - y_i| - \varepsilon, & \text{otherwise} \end{cases} \) is an \( \varepsilon \)-insensitive loss function, \( \varepsilon > 0 \) is a deviation. Introduce slack variables \( \xi_i \) and \( \xi_i' \), the objective function (2) can be

\[
\min_{w, b, \xi, \xi'} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{m} (\xi_i + \xi_i') \quad \text{(3)}
\]

And constraint conditions are

\[
f(x_i) - y_i \leq \varepsilon + \xi_i,
\]

\[
y_i - f(x_i) \leq \varepsilon + \xi_i',
\]

\[
\xi_i, \xi_i' \geq 0, i = 1, 2, \ldots, m.
\]

Introduce Lagrangian multipliers \( \mu_i \geq 0, \mu_i' \geq 0, \alpha_i \geq 0, \alpha_i' \geq 0 \), we can get a lagrangian function

\[
L(w, b, \alpha, \xi, \xi', \mu, \mu') = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{m} (\xi_i + \xi_i') - \sum_{i=1}^{m} \mu_i \xi_i - \sum_{i=1}^{m} \mu_i' \xi_i' + \sum_{i=1}^{m} \alpha_i (f(x_i) - y_i - \varepsilon - \xi_i) + \sum_{i=1}^{m} \alpha_i' (y_i - f(x_i) - \varepsilon - \xi_i') \quad \text{(4)}
\]

Substitute formula (1) into (4), and let partial derivatives of \( w, b, \xi, \xi' \) be 0, we can get the following formulas,

\[
w = \sum_{i=1}^{m} (\alpha_i' - \alpha_i) x_i, \quad \text{(5)}
\]
\begin{align}
\sum_{i=1}^{m} (\alpha_i^j - \alpha_i) = 0, & \quad (6) \\
\alpha_i + \mu_i = C, & \quad (7) \\
\alpha_i^j + \mu_i^j = C. & \quad (8)
\end{align}

Using Lagrangian multipliers and dual methods, its dual problem is,

$$
\max_{\alpha} \sum_{i=1}^{m} y_i (\alpha_i^j - \alpha_i) - \varepsilon (\alpha_i^j + \alpha_i) - \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} (\alpha_i^j - \alpha_i) (\alpha_j^j - \alpha_j) x_i^T x_j 
$$

And constraint conditions are

$$
\sum_{i=1}^{m} (\alpha_i^j - \alpha_i) = 0, \\
0 \leq \alpha_i, \alpha_i^j \leq C, i = 1, 2, \ldots, m.
$$

And it also needs to satisfy KKT (Karush-Kuhn-Tucker) conditions,

$$
\alpha_i (f(x_i) - y_i - \varepsilon - \xi_i^i) = 0, \\
\alpha_i^j (y_j - f(x_j) - \varepsilon - \xi_j^i) = 0, \\
\alpha_i \alpha_i^j = 0, \\
\xi_i^i \xi_i^j = 0, \\
(C - \alpha_i) \xi_i^i = 0, \\
(C - \alpha_i^j) \xi_i^j = 0, \quad i = 1, 2, \ldots, m.
$$

For question (9), use SMO (Sequential Minimal Optimization) algorithm [12] to get all \( \alpha_i, \alpha_i^j, i = 1, 2, \ldots, m. \)

According the KKT conditions, for any support vector \((x_j, y_j) (0 < \alpha_j < C),\) we can get \( b, \)

$$
b = y_j + \varepsilon - \sum_{i=1}^{m} (\alpha_i^j - \alpha_i) x_i^T x_j. \quad (10)
$$

In order to \( b \) being more robust, let \( S = \{ j | 0 < \alpha_j < C, j = 1, 2, \ldots, m \}, \) then \( b \) can be

$$
b = \frac{1}{|S|} \sum_{j \in S} (y_j + \varepsilon - \sum_{i \in S} (\alpha_i^j - \alpha_i) x_i^T x_j). \quad (11)
$$

In order to SVR question being linear, eigenvector \( x \) can be mapped into a higher dimension space \( \phi(x) \). So the formula (1) becomes,

$$
f(x) = w^T \phi(x) + b. \quad (12)
$$
Similarly, \( w = \sum_{i=1}^{m} (\alpha_i - \alpha_i^*) \phi(x_i) \). \hspace{1cm} (13)

Further, \( f(x) = \sum_{i=1}^{m} (\alpha_i - \alpha_i^*) \kappa(x, x_i) + b \). \hspace{1cm} (14)

Where, \( \kappa(x_i, x_j) = \phi(x_i)^T \phi(x_j) \) is a kernel function.

Experiments

In order to verify the effectiveness of the SVR model in TMM, the following experiments were carried out.

We use 45 categories of symptoms as input variable, 86 syndromes as output value. \( \varepsilon = 0.2 \), \( \kappa(x_i, x_j) = (x_i^T x_j)^d \) is a polynomial kernel function, \( d = 5 \).

We used 300 medical cases as samples from some books [13-16]. We did 5 experiments. In every experiment, we selected 200 cases randomly as training samples, 100 cases as test samples. The average correct rate was 76.6%.

Conclusions

It is very important and necessary to research TMM with modern information technology. In our research, we consider the characters present situation of TMM. In consideration of less medical data of TMM, we introduced a SVR model of DS in TMM. Further, because doctors of TMM are accustomed to their experiences in clinical practice, the model can use those experiences.

In the experiments, we chose data from books of experts of TMM. The experiment results supported the model and showed the model is efficient. It is meaningful for differentiate syndromes in TMM.

At present, research results could rarely be found about the syndrome differentiation and typing model in Mongolian medicine. So, in order to improve the correct rate of syndrome differentiation, further study of this work will continue to establish more different machine learning models. At the same time, we hope to collect more data of TMM.

Acknowledgements

Teachers Pengxuan Ma, Juping Zhang, Haiying Yu and other colleagues from Inner Mongolia University of Finance and Economics provided many excellent comments on discussions and experiments. Inner Mongolia International Mongolian Medicine Hospital gave many suggestions about the work. So, thank them for their help.

Funding

This study was supported by project of the natural science foundation of the Inner Mongolia Autonomous Region (no.2015MS0607) and the application research development project of Inner Mongolia Autonomous Region (no.20120412).

References


