PREDICTION OF INTERFERON EFFICACY IN CHRONIC HEPATITIS C TREATMENT USING SUPPORT VECTOR MACHINE

Anto Satriyo Nugroho[1], Jun Yang [2], Susumu Kuroyanagi[3], Akira Iwata[3], Kazunobu Yamauchi[2]

School of Life System Science & Technology, Chukyo University, Japan[1]
Dept. of Medical Information and Management Science, Nagoya University Graduate School of Medicine, Japan[2]
Department of Computer Science and Engineering, Nagoya Institute of Technology, Japan[3]

[Background] Interferon has been widely recognized as a new treatment for Chronic Hepatitis C and its efficacy are reported in many clinical studies. The use of interferon, however, is accompanied by many side effects such as flu-like symptoms, low blood count, and blood infection. If the efficacy of interferon treatment could be predicted earlier, the doctors would be able to select other alternative treatment for patients whose interferon-treatment is predicted not to be effective. The aim of this study is to develop a statistical pattern recognition system to predict the efficacy of the interferon treatment, based on the clinical data of the patients.

[Method and Experimental Results] The database used in this study is obtained from seven years observation to the Chronic Hepatitis C patients, from 1997 through 2004, conducted in Nagoya University Hospital, Japan. It comprises of 112 samples, 80 males and 32 females, age of interval 17-72 years. The observation includes HCV-RNA, HCV gene type, hepatobiopsy, total protein, albumin, ribavirin, and the other information of total 30 items. These observations were conducted before the patients were treated by interferon. Six months after the treatment, the HCV-RNA of the patients were measured. Based on this value, the data are grouped into two classes : positive (+1) and negative class (−1). Positive class is labeled to the patients of which the treatment success, and negative class is labeled to the data from not success patients. The number of examples of positive and negative class is 66 and 46, respectively. This database is then used to design the model, and evaluating its performance.

The proposed method consists of two parts : 1) feature subset selection (FSS) and 2) Support Vector Machine Classifier, as shown in Fig.1.

Fig.1 Scheme of the predictor system
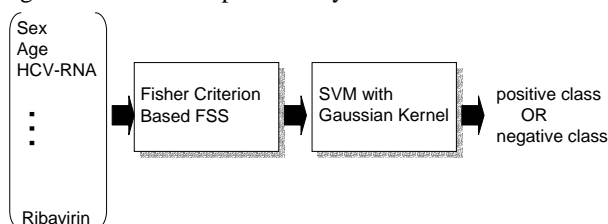
FSS-part works to select the feature subset that provides significant contribution to the classification task. The feature is selected based on its individual merit, and its significance is measured using Fisher Criterion. The best feature subset is determined and presented to the second part. The second part is a classifier that predicts the result of the treatment based on the selected features. It is implemented by Non-Linear Support Vector Machine (NL-SVM) with Gaussian Kernel. SVM in principle is a linier classifier which is trained to obtain an optimal classification hyperplane in the feature space. This optimal hyperplane is obtained by maximizing the "margin": a criterion which is defined by the distance between the hyperplane and the nearest training samples. To work with non-linear problem, the data are projected to a new feature space with higher dimensionality, increasing the computational power of SVM. The linear classification is then performed in this new space.

The performance of SVM is evaluated using the clinical database. Due to the limitation of available data, the performance was estimated by leave one out cross validation[1]. As a result, SVM showed a good score in predicting the efficacy of interferon treatment using the top 5 ranked features. The total score was 84%, by correctly classified 94 samples of total 112 samples. The classification rate of positive and negative class were 85%(56 samples) and 83%(38 samples). The evaluation of the top five ranked features selected in FSS part shows that HCV-RNA, hepatobiopsy, HCV gene-type, ALP and CHE appeared in the top five ranks. This result indicates that these are the most informative features, which also shows agreement with the opinion from clinical side.

[Conclusions] The performance evaluation shows the proposed method achieved a good result. Using the best five features, the system achieved 84% classification rate estimated by leave-one-out cross validation. The future work of this study is addressed to reducing the number of support vectors, and evaluating the performance on larger clinical dataset.

[References]

(1) Hastie T. et al., The Elements of Statistical Learning : Datamining, Inference, and Prediction, Springer Verlag, 2001