Applying Neural Networks for Diagnosis of Cancer Relapse

Kenneth J. Mackin
Department of Information Systems
Tokyo University of Information Sciences
1200-2 Yatoh-cho, Wakaba-ku, Chiba 265-8501, Japan
mackin@rsch.tuis.ac.jp

Abstract

Early detection of cancer relapse is very important for cancer patients, but prevention or early diagnosis of cancer relapse can be very difficult, depending on the type of cancer involved. In this research, we use real data for cancer operation patients, and investigate the possibility of applying artificial neural networks to predict early relapse. Artificial neural networks (ANN) can be characterized by its "black box" approach to learn and classify complex data patterns.

For this research, we propose applying 3 layer network structure (1 input layer, 1 hidden layer, 1 output layer) for the training of early cancer relapse classification, using the neural network to learn the complex relationship between risky and safe patient data. The input will include all the provided patient data including blood test results and the diagnosis of the cancer, and the single output of the neural network will be the risk factor or possibility of cancer relapse.

We analyze the data by using artificial neural networks using back propagation training, and compare the validity of predicting early relapse from the provided cancer patient data. We conclude with a discussion of the results and the validity of applying neural networks to assist physicians in predicting early cancer relapse.

1 Introduction

Recently, there is increased interest in applying data analysis methods to medical data, in order to find useful information from patient data which can be applied to medical diagnosis. On the other hand, medical data is complex and often contain error or noise in each data, and is therefore difficult to analyze using simple statistical methods. For example, early detection of cancer relapse is very important for cancer patients, but prevention or early diagnosis of cancer relapse can be very difficult, depending on the type of cancer involved. In this research, we use real data for cancer operation patients, and investigate the possibility of applying artificial neural networks to predict early relapse.

Artificial neural networks (ANN) can be characterized by its "black box" approach to learn and classify complex data patterns. For this research, we propose applying 3 layer neural network structure (1 input layer, 1 hidden layer, 1 output layer) for the training of early cancer relapse classification, using the neural network to learn the complex relationship between risky and safe patient data. We use actual medical data of kidney cancer (renal cell carcinoma) patients provided by Kitasato University Hospital, and aim at predicting the occurrence of cancer relapse in kidney cancer patients within 5 years.

2 Kidney (renal cell) cancer data

For the experiment data, we used actual kidney cancer (renal cell carcinoma) cancer patient data provided by Kitasato University Hospital.

Surgery to remove part or all of the kidney (nephrectomy) is the most popular treatment for kidney cancer, and in general 80% of the patients do not have cancer relapse, or recurrence of kidney cancer. The provided patient medical records also show the same ratio, and approximately 20% of the data record relapse within 5 years. The patient medical records show no strong 'markers' for patients with high relapse risks, and there is no simple correlation between medical test results and occurrence of relapse within 5 years. Therefore, we propose applying artificial neural networks which can learn non-linear patterns in data, for categorizing high risk and low risk patients from the medical test data.

3 Artificial Neural Network

Artificial neural networks (ANN) can be characterized by its "black box" approach to learn and classify complex data patterns. For this research, we propose applying 3 layer network structure (1 input layer, 1 hidden layer, 1 output layer) for the training of kidney cancer relapse risk classification, using the neural network to learn the complex relationship between medical test data of kidney cancer patients.

First we describe the basic artificial neural network for land cover classification. We considered the 3 layer artificial neural network (1 input layer, 1 hidden layer, 1 output layer) as the basic training classifier. We use a sigmoid function for the synapse function of the neuron, with back propagation (BP) training of the medical test data. The number of hidden neurons was decided by results of preliminary experiments of the neural network.

For the network training we used the database of collected MODIS sensor data, and applied BP training based on the difference between classified land cover and land-truth data provided by the Japanese Ministry of the Environment.

The constructed ANN structure took 12 medical test and diagnosis data for input, and output 1 value for the categorizing the risk of cancer relapse. The medical database contained 162 patient data, of which 2/3 were used for training, and the remaining 1/3 were used for untrained testing data.

4 Conclusion

Figure 1 shows the training curve of trained and untrained data using the artificial neural network. The results showed that for the untrained data the error rate was 0.2 . The error rate was still quite high, showing that the medical data was indeed difficult to learn from.

For future works, we will consider processing the input data prior to applying to neural network training, in order to improve the categorization accuracy.

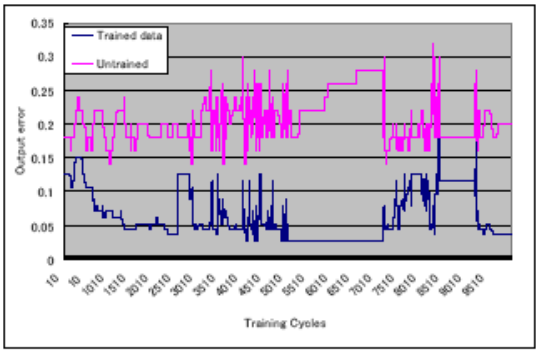

Figure 1. Training curve results