

Applying Sequential Pattern Mining Technique to Construct a Pairwise Sequence Similarity Kernel for Support Vector Machine Classifiers

Student : Yu-Yu Yao

Advisor : Dr. Chi-Yang Tsai

Institute of Industrial Engineering and Management
Yuan-Ze University

ABSTRACT

Sequence classification problem can be found and discussed in many real world applications such as protein function prediction, text classification, and so on. SVMs (Support Vector Machines) have been used to deal with sequence classification problem, since SVMs can deal with the nonlinear data and possess high efficiency in classification. However, the most difficult part in SVMs is to design an appropriate kernel function. Therefore, a pairwise sequence similarity kernel is proposed which takes sequential patterns instead of taking k-mers as reference sequences and evaluates the similarity scores between reference sequences and sequence data by a map function. To obtain sequential patterns, three different sequential pattern mining methods are used to extract frequent sequential patterns, frequent closed sequential patterns, and frequent maximal sequential patterns from sequence databases. The three sequential patterns are then evaluated to know which one could achieve higher accuracy. A map function, which is edit distance algorithm, is used in the proposed kernel to calculate the similarity score. Next, the sequence SVM classifier is built according to the proposed pairwise sequence similarity kernel. Through the proposed sequence SVM classifier with pairwise sequence similarity kernel, the class label of a new sequence will be predicted precisely. The artificial dataset and the real protein sequence dataset are employed to test the proposed SVM classification model using pairwise sequence similarity kernel with three different sequential patterns. The experiment results indicate the proposed SVM classification model using pairwise sequence similarity kernel is efficient and feasible.

Keyword: Sequential pattern mining 、 Sequence classification 、 Kernel method 、 SVM classifier