

COMPUTER SCIENCE SEMINAR SERIES

Department of Computer Science
City University of Hong Kong
(Departmental Seminar Seminar 2007/2008 - No 1)

Dynamic Distance-based Active Learning with SVM

Mr JIANG Jun
PhD Student
Department of Computer Science
City University of Hong Kong

Date :

7 September 2007 (Friday)

Time :

2:30pm - 3:30pm

Venue :

CS Seminar Room, Rm Y6405, 6th Floor Yellow Zone, Academic Building, City University of Hong Kong, Tat Chee Avenue, Kowloon Tong.

Abstract

In supervised learning tasks, often the most time-consuming and costly process in designing classifiers is data labeling when we face a large number of unlabeled data. Instead of traditional randomly picking data to be manually labeled for our training set, active learning is a new mechanism for selecting unlabeled data based on the result of past labeled data. By this mechanism, active learning methods can provide fewer “optimal” data to the learner to construct a classifier as quickly as possible.

In this seminar, we will describe a novel active learning strategy, named dynamic active learning with SVM to improve the effectiveness of learning sample selection in active learning. The algorithm is divided into two steps. The first step is similar to the standard distance-based active learning with SVM in which the sample nearest to the decision boundary is chosen to induce a hyperplane that can halve the current version space. In order to improve upon the learning efficiency and convergent rates, we propose in the second step, a dynamic sample selection strategy that operates within the neighborhood of the “standard” sample. Theoretical analysis is given to show that our algorithm will converge faster than the standard distance-based technique and using less number of samples while maintaining the same classification precision rate. We also demonstrate the feasibility of the dynamic selection strategy approach through conducting experiments on several benchmark datasets.

The paper is presented in the 5th International Conference on Machine Learning and Data Mining in Pattern Recognition (MLDM 2007), July 18-20, 2007

Supervisor: Prof Horace Ip (CS)

Research Interests: Machine learning, pattern recognition, content based image retrieval and human face recognition

All are welcome!

* * * * *

In case of questions, please contact Prof Horace Ip at Tel: 2788 8641, E-mail: cship@cityu.edu.hk, or visit the CS Departmental Seminar Web at <http://www.cs.cityu.edu.hk/>.