

SỬ DỤNG GA TỐI ƯU THAM SỐ CỦA SVM CHO ĐA PHÂN LỚP DỮ LIỆU

Using Genetic Algorithms to Optimize Parameters of Support Vector Machines for Data Multi Classification

Lữ Đăng Nhạc^{a*}, Nguyễn Hà Nam^b, Hà Văn Sang^c

^a Bộ môn Tin học, Học viện Báo chí và Tuyên truyền

^b Khoa Công nghệ Thông tin, Đại học Công nghệ Hà Nội

^c Khoa Hệ thống Thông tin Kinh tế, Học viện Tài chính

* e-mail: nhaclld@ajc.edu.vn

TÓM TẮT Bài toán phân lớp được ứng dụng phổ biến trong khai phá dữ liệu. Trong các thuật toán phân lớp, phương pháp Suport Vector Machine (SVM) là một phương pháp phân lớp nhị phân rất hiệu quả. SVM sử dụng kernel fuction chuyển dữ liệu phi tuyến tính vào không gian đặc trưng để phân lớp. Do vậy, kết quả phân lớp bị ảnh hưởng bởi việc lựa chọn các kernel nhưng hiện nay, chưa có cách chọn các hàm nhân tốt nhất cho các ứng dụng. Trong bài báo này, giải thuật di truyền (GA) được sử dụng để tối ưu các tham số của kernel trong đa phân lớp với SVM. Các thực nghiệm được tiến hành với dữ liệu thực tế và cho kết quả tốt hơn so với giải thuật cơ sở.

ABSTRACT The problem of classification is used commonly in data mining. Among classification algorithms, Suport Vector Machine (SVM) is very effective ones for binary classification. SVM is known as the algorithm that use kernel function to transform non-linear input data into high dimensional feature space. The results are influenced by kernel selection. However, there is no explicit way to choose the optimistic kernel for specific application. In this paper, Genetic algorithm (GA) was applied to optimize parameters of Kernel in SVM multiclassification. The experiments were done with published data and yielding better results than basic algorithms.