

以因子實驗法發掘支援向量機的重要變數與建構最小成本之診斷模型

葉怡成, 杜榮原

資訊管理學系

資訊學院

icyeh@chu.edu.tw

摘要

The purpose of this study is to employ design of experiments (DOE) to discover important features to build simple but accurate model for support vector machine (SVM). Its basic principle is to regard selecting or do not selecting a feature as a two-level independent factor; the parameters of SVM as continuous noise factors; the accuracy of the SVM model as the dependent variable; to employ a fractional factorial experiment design to form the experiment which combines the two-level factors and the noise factors; after the experiment is finished, to analyze the effect of each factor to determine the effective independent factor combination, namely the best feature combination for the SVM model. To prove the performance of methodology, two artificial problems (one classification problem and one regression problem) as well as two real problems (one classification problem and one regression problem) were employed to verify the methodology. The results proved that the methodology can discover the important features to build simple but accurate model. Moreover, this study extended the methodology to cost optimization problem, which regarded the dependent variable as the total cost of the model (the diagnosis cost consisted of independent variables plus the risk cost consisted of misdiagnosis cost multiplied by misdiagnosis probability), and employed the heart disease diagnosis case study to verify it. The results showed that it can really discover the cost-effective independent variables, and build the minimum cost medical diagnosis model.

關鍵字 : support vector machine, fractional factorial experiments, response surface methodology, central composite design, feature selection.