APPLYING SUPPORT VECTOR MACHINE AND FEATURE SELECTION TO CHINESE PCB R&D DOCUMENT CLASSIFICATION AND RETRIEVAL

蔡禎騰,林水順,張志宇,莊英慎,彭泉
Business Administration
Management
ysjuang@chu.edu.tw

Abstract

The operation of product research and development (R&D) is relatively essential for a technology company. R&D knowledge is typically presented as documents, either in paper or electronic forms. Handling knowledge generated from the R&D processes is difficult for its complexity and abundance. If the R&D documents can be managed effectively, it will be helpful to knowledge accumulation and utilization. Along with the knowledge advance rapidly, it is beneficial for enterprise to manage R&D knowledge with information technology.
The objective of this study is threefold: (1) Identifying problems encountered in the classification and retrieval of Chinese R&D documents in the printed circuit board (PCB) industry. (2) Developing a document classification process compliant with the properties of the management of Chinese PCB R&D documents. (3) Proposing an R&D document classification methodology and analyzing the accuracy of different models.
The results indicated that applying support vector machine (SVM) to document classification and retrieval in PCB R&D document is feasible and SVM with future selection outperformed other algorithms, such as kNN and neural network.

Keyword：Support Vector Machine (SVM), Future Selection, Printed Circuit Board (PCB), Document Classification