ABSTRACT

Although Support Vector Machines have been used to develop highly accurate classification and regression models in various real-world problem domains, the most significant barrier is that SVM generates black box model that is difficult to understand. The procedure to convert these opaque models into transparent models is called rule extraction. This thesis investigates the task of extracting comprehensible models from trained SVMs, thereby alleviating this limitation. The primary contribution of the thesis is the proposal of various algorithms to overcome the significant limitations of SVM by taking a novel approach to the task of extracting comprehensible models. The basic contribution of the thesis are systematic review of literature on rule extraction from SVM, identifying gaps in the literature and proposing novel approaches for addressing the gaps. The contributions are grouped under three classes, decompositional, pedagogical and eclectic/hybrid approaches. Decompositional approach is closely intertwined with the internal workings of the SVM. Pedagogical approach uses SVM as an oracle to re-label training examples as well as artificially generated examples. In the eclectic/hybrid approach, a combination of these two methods is adopted.

The thesis addresses various problems from the finance domain such as bankruptcy prediction in banks/firms, churn prediction in analytical CRM and Insurance fraud detection. Apart from this various benchmark datasets such as iris, wine and WBC for classification problems and auto MPG, body fat, Boston housing, forest fires and pollution for regression problems are also tested using the proposed approach. In addition, rule extraction from unbalanced datasets as well as from active learning based approaches has been explored. For classification problems, various rule extraction methods such as FRBS, DT, ANFIS, CART and NBTree have been utilized. Additionally for regression problems, rule extraction methods such as ANFIS, DENFIS and CART have also been employed. Results are analyzed using accuracy, sensitivity, specificity, fidelity, AUC and t-test measures. Proposed approaches demonstrate their viability in extracting accurate, effective and comprehensible rule sets in various benchmark and real world problem domains across classification and regression problems. Future directions have been indicated to extend the approaches to newer variations of SVM as well as to other problem domains.