Poster Abstract

Adding Comprehensibility to Support Vector Machine Models Using Rule Extraction Techniques: a credit scoring case By David Martens, Bart Baesens, Johan Huysmans, Tony Van Gestel, Jan Vanthienen

In recent years, Support Vector Machines (SVMs) were successfully applied to a wide range of applications. Their good performance is achieved by an implicit non-linear transformation of the original problem to a high-dimensional (possibly infinite) feature space in which a linear decision hyperplane is constructed that yields a nonlinear classifier in the input space. However, since the classifier is described as a complex mathematical function, it is rather incomprehensible for humans. This opacity property prevents them from being used in many real-life applications where both accuracy and comprehensibility are required, such as credit risk evaluation and medical diagnosis. To overcome this limitation, rules can be extracted from the trained SVM that are interpretable by humans and keep as much of the accuracy of the SVM as possible. In this paper, we will provide an overview of the recently proposed rule extraction techniques for SVMs and introduce two others taken from the artificial neural networks domain, being Trepan and G-REX. The described techniques are compared using publicly available datasets, such as Ripley's synthetic dataset and the multi-class iris dataset. We also look at credit scoring and medical diagnosis where comprehensibility is a key requirement and even a regulatory recommendation. Our experiments show that the SVM rule extraction techniques lose only a small percentage in performance compared to SVMs and therefore rank at the top of comprehensible classification techniques.