

Adaptive Credit Scoring with Kernel Learning Methods

Dr Yingxu Yang

SHS Informationssysteme

Fraunhoferstr. 12, 82152 Martinsried, Munich, Germany

yingxu.yang@shs.de

Abstract

Credit scoring is a method of modelling potential risk of credit applications. Traditionally, logistic regression, linear regression and discriminant analysis are the most popular approaches for building credit scoring models. Despite their popularity, quite a few limitations are known to be associated with these methods, such as being instable with high-dimensional data (also known as combinatorial explosion) and small sample size, the need for intensive data pre-processing through variable selection/reduction analysis and incapability of efficiently dealing with non-linear systems. Most importantly, based on these algorithms, it is difficult to automate the whole modelling process or design a continuous workflow, and when environment or population changes are introduced, the static models usually fail to adapt and may have to be rebuilt from scratch.

In this paper, a kernel method from machine learning is used to derive a novel and practical adaptive credit scoring system capable of adjusting the model on-line. The kernel method is based on a two-layered structure (data dependent form) and provides a powerful and elegant solution for non-linear systems by plugging in a kernel that maps patterns from non-linear input space to potentially high-dimensional feature space. A new kernel, the KGPF kernel, has been introduced and shown to be able to generate comparable results to the popular Gaussian kernel. The model is adjusted according to an on-line update rule and can always converge to the optimal solution. This approach is also robust for scoring data sets having a large number of attributes (high-dimensional patterns) and does not require variable selection or reduction effort. Experiment studies have demonstrated the effectiveness of the proposed approach. The present study is focused on application risk scoring, but without lack of generalization the method can also be extended to customer behaviour scoring and other related topics.