Accepted Manuscript

A hybrid system with filter approach and multiple population genetic algorithm for feature selection in credit scoring

Di Wang, Zuoquan Zhang, Rongquan Bai, Yanan Mao

PII: S0377-0427(17)30207-8
DOI: http://dx.doi.org/10.1016/j.cam.2017.04.036
Reference: CAM 11116

To appear in: *Journal of Computational and Applied Mathematics*

Received date : 2 November 2016
Revised date : 16 April 2017

Please cite this article as: D. Wang, et al., A hybrid system with filter approach and multiple population genetic algorithm for feature selection in credit scoring, *Journal of Computational and Applied Mathematics* (2017), http://dx.doi.org/10.1016/j.cam.2017.04.036

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Manuscript
Click here to view linked References

ACCEPTED MANUSCRIPT

A Hybrid System with Filter approach and Multiple Population Genetic Algorithm for Feature Selection in Credit Scoring

Di Wang[a,b], Zuoquan Zhang[a,*], Rongquan Bai[a], Yanan Mao[a]

[a]*School of Science, Beijing Jiaotong University, Beijing 100044, China*
[b]*Department of Basic Courses, Beijing Union University, Beijing 100101, China*

Abstract

With the financial crisis happened in 2007, massive credit risks are exposed to the banking sectors. So credit scoring has attracted more and more attention. Bank owns a lot of customer data. By Using those data, credit scoring model can judge applicants credit risk accurately. But those data are often high-dimensional, and have some irrelevant features. Those irrelevant features will affect classifiers accuracy. Therefore, feature selection is an important topic. This paper proposes a two-phase hybrid approach based on filter approach and multiple population genetic algorithm-HMPGA. In phase 1, it introduces the idea of wrapper approach into three filter approaches to acquire some important prior information for initial populations setting of MPGA. In phase 2, it takes advantage of MPGA's characteristics of global optimization and quick convergence to find optimal feature subset. This paper uses two real credit scoring data sets of UCI databases to compare HMPGA, MPGA and GA. It verifies that the accuracies of feature subsets acquired from HMPGA, MPGA and GA are superior to three filter approaches. Meanwhile, nonparametric Wilcoxon signed rank test is held to confirm that HMPGA is better than MPGA and GA. HMPGA not only can be applied to feature selection of credit scoring, but also can be applied to more fields of data mining.

Keywords: `Credit scoring`, Feature selection, Hybrid approach, HMPGA

1. Introduction

Financial crisis happened in 2007 has shaken financial market deeply. It undermines the confidence of customers and investors, raises the attention of the stability of the financial institutions for financial markets, and influences global economic environment. The financial crisis causes a lot of difficulties in operation of the banks, or even bankruptcy. Investors begin to make decision very cautious in order not to get into any trouble. Regulators require banks to strengthen their capital reserves so that it can have the ability to resist risk. So as not to appear worse, governments and international organizations have to rescue some banks[1].

*Corresponding author
Email addresses: `ldtwangdi@buu.edu.cn` (Di Wang), `zuoquanzhang@163.com` (Zuoquan Zhang)

Download English Version:

https://daneshyari.com/en/article/8902311

Download Persian Version:

https://daneshyari.com/article/8902311

Daneshyari.com