

Accepted Manuscript

Title: A Principle Component Analysis-based Random Forest with the Potential Nearest Neighbor Method for Automobile Insurance Fraud Identification

Authors: Yaqi Li, Chun Yan, Wei Liu, Maozhen Li



PII: S1568-4946(17)30438-6
DOI: <http://dx.doi.org/doi:10.1016/j.asoc.2017.07.027>
Reference: ASOC 4355

To appear in: *Applied Soft Computing*

Received date: 21-9-2016
Revised date: 6-7-2017
Accepted date: 14-7-2017

Please cite this article as: Yaqi Li, Chun Yan, Wei Liu, Maozhen Li, A Principle Component Analysis-based Random Forest with the Potential Nearest Neighbor Method for Automobile Insurance Fraud Identification, *Applied Soft Computing Journal* <http://dx.doi.org/10.1016/j.asoc.2017.07.027>

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.

A Principle Component Analysis-based Random Forest with the Potential Nearest Neighbor Method for Automobile Insurance Fraud Identification

Yaqi Li^a, Chun Yan^a, Wei Liu^{b,*}, Maozhen Li^c

^aCollege of Mathematics and Systems Science, Shandong University of Science and Technology, Shandong, Qingdao, China

^bCollege of Computer Science and Engineering, Shandong University of Science and Technology, Qingdao, China

^cDepartment of Electronic and Computer Engineering, Brunel University London Uxbridge, UB8 3PH, UK

Graphical abstract

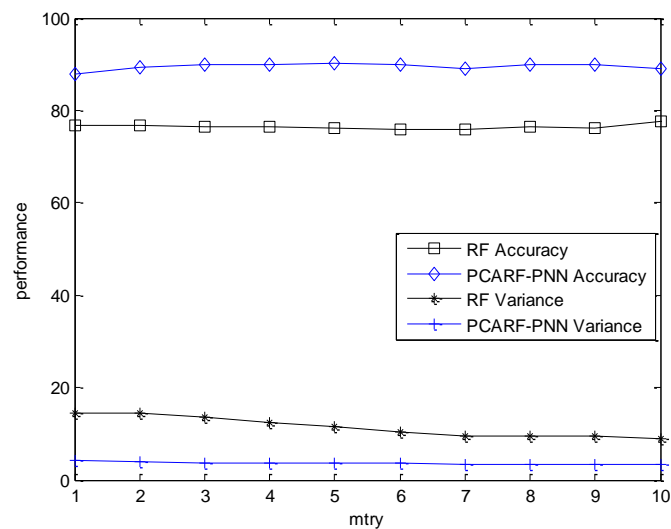


Fig. 1. Performance of the Two Models with Different “*mtry*” Parameters on Automobile Insurance Claims Data Set

Highlights

- A multiple classifier system based on Random Forest, Principle Component Analysis and Potential Nearest Neighbor is proposed. It improves the classification accuracy of the ensemble classifier by improving the difference of the base classifiers.
- The PCA transforms features in each node to increase diversity thereby resulting in decision trees with low correlation. It selects a tree with higher accuracy from larger number of decision trees in the ensemble feature spaces of PCARF - PNN.
- Random Forest is learned from the perspective of Adaptive Nearest Neighbor. It can be seen as an adaptive learning mechanism of *k* Potential Nearest Neighbors.
- A new voting mechanism, based on the Potential Nearest Neighbor, is presented to replace the traditional majority vote. It makes full use of the out-of-bag samples information on the decision tree.
- The proposed method is also applied to detect the automobile insurance fraud and the fraud rules are obtained.

Download English Version:

<https://daneshyari.com/en/article/11002715>

Download Persian Version:

<https://daneshyari.com/article/11002715>

[Daneshyari.com](https://daneshyari.com)