

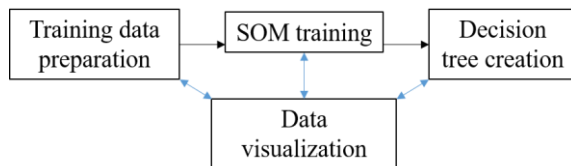
Fraud Detection for Credit Card Transactions

Shyh-Kuang Ueng and Chun-Yi Peng

Department of Computer Science and Engineering
National Taiwan Ocean University
skueng@mail.ntou.edu.tw

This article proposes a fraud detection procedure for uncovering fraud credit card transactions. It is composed of the following steps: (1) A training data set is produced from a collection of credit card transactions. (2) The key features of transactions are identified and used for fraud detection. (3) The proposed procedure train a Self-Organizing Map (SOM) to classify normal and fraud purchases via credit cards. (4) Following the SOM training process, a decision tree is constructed to further classify normal and anomaly transactions.

Key words: fraud detection, credit card transaction, self-organizing map, decision tree, visualization.



Flowchart of the proposed fraud detection procedure.

A. Training Data and Feature Extraction

The training data contains about 900 normal transactions and 900 fraud transactions. Each transaction of the data set contains 34 features. After going through a feature reduction process, the number of relevant features is reduced to 14.

B. SOM Training

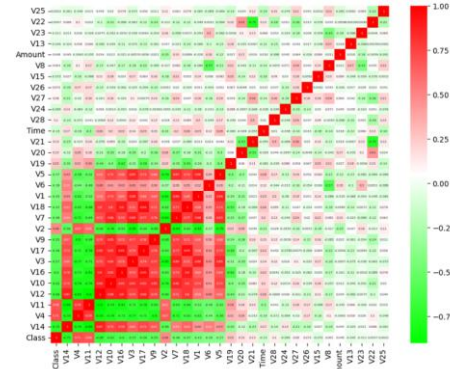
The SOM is a 2D lattice, which is composed of 32x32 neurons (nodes). Then, the SOM is undertaken an unsupervised training.

C. Decision Tree

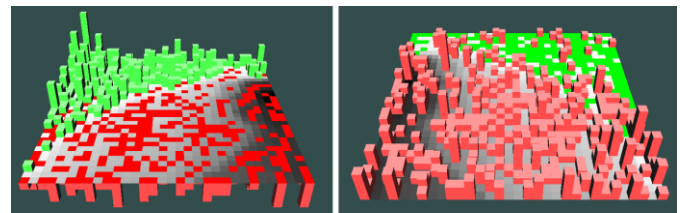
At each level of the decision tree, a feature is selected for split the data set. The best threshold of the feature is searched to minimize the Gini values.

D. Visualization Method

First, we illustrate the correlations among the features by using the heatmap. The numbers of transactions favored by the SOM nodes are represented by vertical bars. The lengths of the bars reveal the number of transactions. A bar is rendered in green color, if the transactions favored by its master node are normal. Otherwise, the bar is shaded in red color.



Heatmap of the 35 features of a test data.

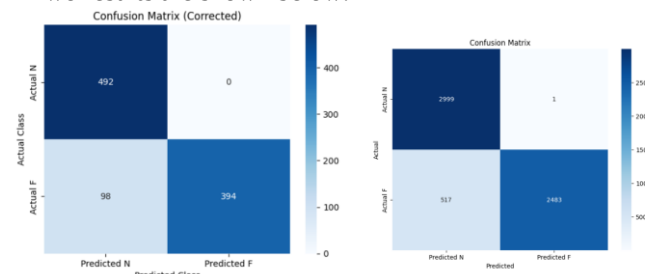


(a) Distribution of normal transactions, shown by green bars.

(b) Distribution of fraud transactions, shown by red bars.

Results and Analysis

Two results are shown below:



- [1] Visa Int. Managing Risk in the 21st Century. Strategies for Issuing & Acceptance, 156, October,2000.
- [2] Zaslavsky, V., & Strizhak, A. Credit card fraud detection using self-organizing maps. Information and Security, 18, 2006.
- [3] T. Kohonen. The self-organizing map. Proceedings of the IEEE, 78(9), pp. 1464–1480, 1990. doi: 10.1109/5.58325
- [4] Quinlan, J. R. Induction of decision trees. Machine learning, 1(1), pp. 81-106, 1986.
- [5] Liang Jiye, Feng Chenjiao, Song Peng. Review of Big Data Correlation Analysis. Chinese Journal of Computers, 2016.
- [6] Breiman, L. Classification and regression trees. Routledge, 2017.
- [7] Ren Lei, Du Yi, Ma Shuai, Zhang Xiaolong, Dai Guozhong. Overview of Big Data Visual Analysis. Journal of Software, 25 (9), pp. 1909-1936, 2014.