Available online at www.sciencedirect.com

ScienceDirect

Procedia Engineering 7 (2010) 366–370

Procedia Engineering

www.elsevier.com/locate/procedia

2010 Symposium on Security Detection and Information Processing

Prediction of Customer Classification Based on Rough Set Theory

Ju Li [a*], Xing Wang [b], Shan Xu [a]

[a] *School of Computer Science and Engineering, Chang Shu Institute of Technology, Changshu, Jiangsu, 215500, China*
[b] *School of Computer and Software, Nanjing University of Information Science and Technology, Nanjing, Jiangsu, 210044, China*

Abstract

This paper proposed customer classification prediction model based on rough set, we can do the customer classification based on a few properties, the analysis reduces the complexity of decision-makers. This model could help companies predict in advance the new customer or potential customer value level. Empirical analysis results indicate complexity of data calculation is decreased effectively and veracity of customer prediction is improved.

Keywords: rough set; attribute reduction; value reduction; prediction of customer classification mode

1. Introduction

Enterprise customers include historical clients, current customers and potential customers. Accordingly, customer relationship management[1-6], customer should also be divided into the history management, management of current customers and potential customers in the management.

Current customers and potential customers is the focus of corporate research. Access to rough set classification rules, the classification rules applied to a large number of the original customer data，we can determine a high value customers, medium value and low value customers what type of customer in. To carry out targeted marketing campaigns，Enterprises particularly concerned about how to identify and retain high value customers. Accurately assess and predict customer value、correct choice of target market is the key of enterprise that can effectively manage customer relationship, is an important part of enterprise core competence too.

There are a lot of factors that affect customer value，many factors impact on customer value is difficult to determine，Thus, customer value analysis methods difficult to give accurate values. This article discusses the customer classification based on rough set method solves this problem.

* Corresponding author. Tel.:+86-0512-52166136.
E-mail address: liju284532@163.com.

1877-7058
doi:10.1016/j.proeng.2010.11.059

2. Customer Classification Based on Rough Set

2.1 Data acquisition and representatio

The data stored in the database due to human or physical, and other factors, there are interference factors such as noise data, vacancy data, and inconsistent data, while a large number of raw data in the database is in the concept of a lower level, not suitable for knowledge discovery .Therefore, the preprocessing of customer data is necessary. The experimental data in this paper come from membership details of consumption of a store. Through data preprocessing，the cluster analysis of 30 selected customers, finding the most valuable customers accounted for 19.2% of total, sales profits accounted for 63.7%, the general value customer accounted for 34.6% of total sales profits accounted for 22.9%, low-value customer accounted for 46.04% of total, sales profits accounted for 13.25%. 12 customer detail property of cluster as condition property, the customer group from cluster analysis as a decision attribute. NS discrete decision table ,using this method discreted data, obtained customer discretization results, shown in Table 1.

Table 1 NS discretization results

Symbol	Meaning	Discrete values			
		0	1	2	3
A	Sex	Female	Male		
D	Revenue	Under 20 000	20 000 -50 000	50 000 -10 000	100,000 or more
….	….	….	….	….	….
M	Housing environment	Poor	General	Better	Very good
N	Decision attribute	Low-value	General value	High-value	

2.2 Condition attribute reduction

In order to mine classification rules from the large amount of collected data, reduce workload, under the condition required to maintain the condition property relative to the classification ability of decision attribute, minimize the redundant rules, remove the repeated condition attributes in the decision table. Further reduction to table 1 ,we get the initial decision table, as shown in Table 2. in Table 2, there is exactly the same value of the property records of the different customers, as redundant objects merge. Then calculated the condition attribute reduction relative to decision-making attributes, based on reduction of rough set theory and method, I used a program written in MATLAB redu.m, to achieve reduction of rough functions. Specific procedures are as follows:

```
function y = redu(c,d,x) % Defined function
clear all;
close all;
y = core(c,d,x);
q = ind(d,x);
p = ind(c,x);
pos_cd = pospq(p,q);
re = y;
red = ind(y,x);
pos_red = pospq(red,q);
while pos_cd~ = pos_red
    cc = setdiff(c,re);
    [c1,c2] = size(cc);
    for i = 1:c2,yy(i) = sgf(cc(i),re,d,x);
    end
    cd = setdiff(c,y);
    [d1,d2] = size(cd);
    for i = d2:-1:c2+1,yy(i) = [];
```