Universe Scientific Publishing
Computer System Networking and Telecommunications
计算机系统网络和电信，2019,1(2)
ISSN:2661-3719 (Print); 2661-3727 (Online)

Data Mining in Bank Credit Card Management

Qiang LIU Zhengde BAO Yawen TANG
School of Computer and Software, Jincheng College, Sichuan University, Chengdu, Sichuan 611731

Abstract

With the development of computer technology and the Internet, data flow is increasing rapidly every day, and the importance of data is beginning to be reflected in every aspect of our life, for example, shopping malls choose the right purchase quantity according to the sales information, and so on. How to find out valuable information to our daily life and enterprise development from many data, so data mining appears. This paper summarizes the functions and algorithms of data mining technology, and analyzes the application of data mining technology in bank credit card management.

Key Words

Data Mining, Credit Card Management, Cluster Analysis Classification Analysis

DOI:10.18686/jsjxt.v1i2.670

银行信用卡管理中的数据挖掘术

刘 强 鲍正德 唐娅雯
四川大学锦城学院计算机与软件学院，四川成都，611731

摘 要

伴随着计算机技术、互联网的发展，数据流量每天都在急速增加，数据的重要性也开始体现在我们生活的方方面面，例如商场根据销售信息选择合适的进货量等。如何从众多的数据中找出对我们日常生活和企业发展有价值的信息，就这样数据挖掘出现了。本文概述了数据挖掘技术的功能和算法，并结合数据挖掘技术浅析了在银行信用卡管理中的应用。

关键字

数据挖掘；信用卡管理；聚类分析；分类分析

1.引言

基于计算机网络的飞速发展，数据挖掘逐渐渗透于每个领域，了解数据挖掘的基本原理也是非常重要的，在生活方面，他可以为我们的生活提供便利，例如当我们在购物网站上购买一样商品同时还能收到与这件商品有很大联系的商品推荐，减去了我们更多的操作；在企业方面，它可以帮助我们对事物的提前预测，根据其制定良好的战略体系，例如在银行信用卡管理中通过合理的方式对客户信息进行分析，加强银行信息化，对银行决策和发展提供了很大的动力[1]。

2.数据挖掘的概述

数据挖掘又称为数据库知识发现，它通常是指在大量的、不充分、不规律的数据源（如数据库、文本、图片、万维网）中探寻隐含其中的、事先不知道的有用的模式或知识的过程[2]。

数据挖掘技术可以集成各种各样的数据，进行加工并从中提取出有价值的信息，可以帮助我们了解数据中的潜在规律，用历史来预测未来[3]。反映同一类事物的共同特征和不同类事物之间的差异。在各个领域都有其举足轻重的作用，尤其在商业化领域中，促进企业往更为有利的方向发展。同时这也是数据挖掘为什么能够在当今社会发展如此迅速的重要原因之一。

3.数据挖掘的特性

数据的价值化是几乎每个人都会关心的话题，正确又快速的完成这一系列工作是非常重要的，数据挖掘的

出现很大程度上解决了这一些列问题，总的来说，数据挖掘主要有以下几个主要特性在当今大数据时代中：

3.1 多功能性

数据挖掘不仅仅只由一种算法或功能组成，而是由多种.功能，且每个功能有不同算法支持分析和建模等的集合，例如以下几点：

3.1.1 分类分析

分类分析也称为分类预测，事先知道训练样本的特点标签，通过挖掘将属于不同类别标签的样本分离开，然后利用得到的分类模型，预测样本属于哪个类别，常见的方法有贝叶斯网络和决策树分析等。

3.1.2 关联分析

关联分析可用于发现大量数据中不容易被发现的有价值的联系，用置信度和支持度来衡量关联规则，我们可以通过实际情况制定最小置信度和最小支持度的值来筛选出我们所需要的关联规则，其主要步骤有在数据中找出满足最小支持度的项集，再由项集生产关联规则，最后找出满足最小置信度的管理规则。常用的算法有 Apriori 算法：在候选的基础上利用逐层迭代的方式产出频繁项集。FP-增长算法：对每个频繁项集构建它的条件模式基，构建其 FP-树，然后对所创建的数进行重复第一步步骤，该路径生产的所有子路径组合都是一个频繁项模式。

3.1.3 聚类分析

根据数据间的相似性分为多个组，不同组之间特点相异性大，常用的聚类方法有 K-Means 算法：采用划分法聚类技术采用距离来做评价指标，两个数据对象越近相似度就越大的，其具体实现有随机指定 n 个簇中心，然后根据簇中心生成 n 个簇，计算簇内所有数据对象的平均值生成新的簇中心覆盖原有的簇中心，然后重复计算簇中心，直到簇中心不变为止[4]。凝聚层次聚类算法：输入数据集，以个体点作为簇底层开始，依次合并相邻的簇构建为新的一层，直到剩下最后一个簇为顶层。其中定义方法有单链接（距离最近优先）、全链接方法（距离最远优先）、平均链接方法（距离平均值）。

3.2 应用性

数据挖掘的出现很大程度上是为了解决我们实际生活中的应用需求,数据大多来自现实生活中的人为数据，移动通信数据，企业交易数据，电商数据等，依赖于现实事务，数据挖掘就是对这样的数据进行处理，从中提取出不容易被人发现的有价值的信息，将适合的算法应用于实际中，然后将这些有价值的信息再应用到我们具体事务中，实现其价值，在实际中得到检验。例如提高销售预测和准确性和时效性，著名的沃尔玛啤酒和纸尿布案例就是一个应用于销售很好的例子，沃尔玛的工作人员发现以往期销售的数据来看啤酒的卖出的比例和卖出纸尿裤的比例总是成正比，所以根据这一数据分析商场将啤酒和纸尿裤放置在相邻的位置，顾客在购买啤酒和纸尿裤其中之一时，不会因为找不到另外一个商品而不买，为顾客提供了很大的便捷。这样就大大提高了商场在其两件物品上的销售情况，所以数据挖掘有很强的应用性。

3.3 完善性

数据挖掘不仅仅由分析和算法支持，其拥有完善的体系结构，囊括业务理解、对数据理解、数据的准备、建立模型、对模型进行评估、最终发布等过程，在实际应用中贯穿始终具有高度完善的体系结构。且数据挖掘综合大数据，、数据分析学，云计算等不同领域的学术思想和体系知识，虽然涉及众多领域知识，却又与其他领域有很大的差别，其高应用性就是其主要区别之一。

4.案例背景及目的

本案例获取通过某银行的客户信用卡记录，通过数据发掘，找出一些有利于银行发展的有价值的信息，为该银行的信用卡业务决策提供参考。该银行面临的信用卡欺诈和欠款拖欠现象比较严重。本案例希望通过对影响用户信用等级的主要因素进行分析，以及结合信用卡用户的人口特征属性对欺诈行为和拖欠行为的影响因素进行分析。

通过对银行的客户信用卡申请信息、是否存在欺诈、欺诈人口属性分析、拖欠行为记录等数据进行分析，对不同程度的客户进行归类，研究信用卡贷款拖欠、信用卡欺诈等问题与客户的个人信息、信用卡使用信息的关系，为银行提前识别、对信用卡业务风险进行防控提

供参考，从而减少银行在信用卡业欺诈和欠款等方面的损失。

5.数据挖掘技术在银行信用卡管理中的应用

5.1 数据准备及处理

每种事件所考虑的因数各不相如同，在银行信用卡管理中我们要得到较为准确的结果需要考虑的目标有，客户自然情况：年龄，性别，教育程度等；客户职业情况：职业类别，工作期限等；客户收入及财产：年收入，居住情况，车辆情况，保险缴纳情况等；客户银行记录；信贷情况等。然后我们可以根据其有目的性的对客户进行问卷调查，收集客户的每次交易记录和信贷信息，收集客户注册信用卡所使用的注册信息，收集客户在银行官网上进行的操作信息等，

数据处理的基本目的是从无规律的，大量的，难以被我们一眼发现的数据中提取并分析出对我们有价值的便于处理操作的数据。所以当我们在利用软件进行分类分析时，为保证结果准确性我们需要将我们所收集的数据处理为更适合软件操作的布尔类型数据。在软件中进行聚类分析更为适合连续类型的数据等。还有对我们结果影响不大的数据也可以选择性去掉。

5.2 聚类分析

5.2.1 K-Means 模型实现

可以通过 SPSS Modeler 软件通过 K-Means 模型进行聚类分析，将我们所处理的数据导入数据源节点，可以通过过滤节点将我们所不需要的数据过滤掉，通过类型节点根据我们的需求进行输入输出，建模选择 K-Means 算法，在模型设置中我们可以根据我们自己的需求选择聚类数，有些人希望分的类多就会将聚类数设置得较大一点，有些人希望分的类少就可以设置小一点。依赖于每个人的个人经验。运行节点我们就可以看到聚类结果，在聚类结果中我们可以看到我们根据设置的聚类数所生成的每一类所占总数的比例，并且还可以得到聚类质量评估，可以根据聚类质量看出我们所设置的聚类数合适与否，聚类质量高于平均数越高则说明我们所设置的属性越合适。在预测变量重要性中可以看到每个变量对结果所占用的重要性和影响效果。

5.2.2 结果分析

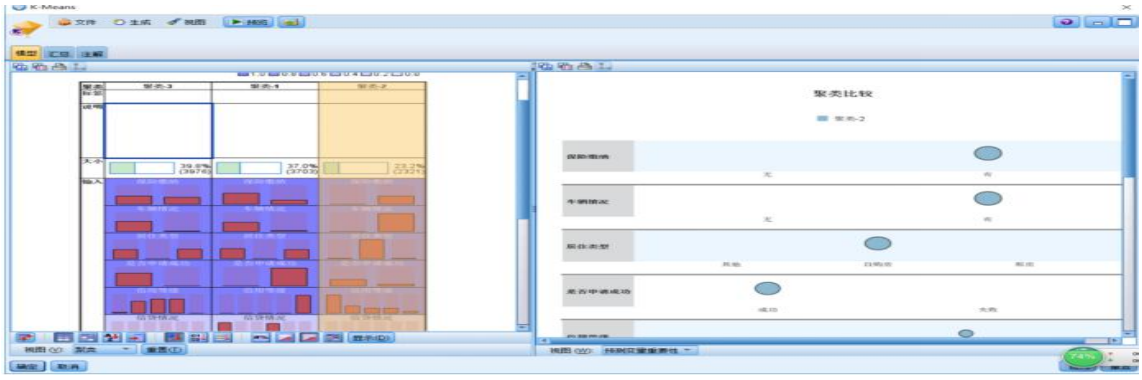

图 1 聚类结果

在银行信用卡管理中我们可以对所聚成得类进行特点分析，如图 1 所示，聚类二普遍为有车，有自购房，信贷情况为无拖欠行为，保险缴纳情况良好人群，很明显就可以将此类人定为低风险用户，银行可较为放心的借贷等，另一方面聚类而普遍为低收入，工作不稳定，信用贷款差，没有固定居所的人群，我们就可以将此类人定位为高风险用户，银行可以相应的采取一些措施，例如对该类用户的信用卡进行限额等避免银行损失。

5.3 关联分析

5.3.1 网络图模型实现

使用 SPSS Modeler 软件通过，将我们所处理的数据导入数据源节点，同样可以用过滤节点将对我们结影响不大的数据过滤点例如客户姓名等，在整个过程中我们可以随时使用输出节点对数据进行查看，例如输出表节点，可以在表中看到我们的数据变化，之后通过类型节点设置输入输出，首先分析各个特点属性之间的潜在联系，在图形中找到网络节点，构建一个网络图，将所有属性全选，且根据需求选择勾选仅显示真值标志单选

框，运行节点生成网络图，从生成的网络图中可以看到各个属性之间的关系强弱，并且汇总里面还可以看到各个属性在整个源数据中出现的次数等，例如居住情况属性和车量情况属性在网络图
中线段明显，则表明居住情况属性和车量情况属性关系度强。

5.3.2 Apriori 模型实现

接下来查看关联规则，在构建网络图模型的基础上，添加属性节点读取前面的值，将各个属性设置为两者，两者的意思就是即作为输入又可作为输出，紧随类型节点之后选择 Apriori 算法建模，可以根据自己的需求设置满足要求的最小支持度和最小置信度，运行该节点生成模型结果，可以从结果中明显的看出前项，后项，置信度，支持度。

5.3.3 结果分析

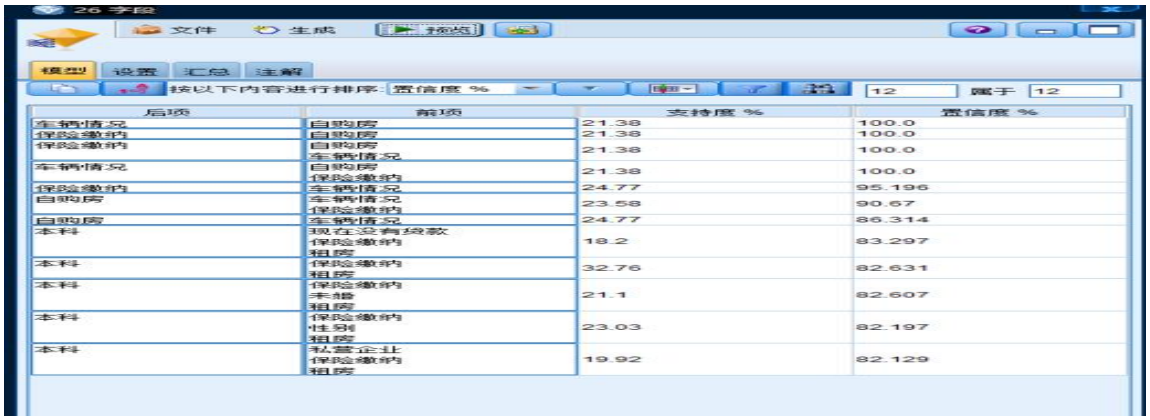

图 2 聚类结果

根据图 2 聚类结果所示，在银行行用卡管理中有房的客户一般是有车有保险的，并且这三项有一项一般便意味着另外两项也会有；交保险、租房的客户，如果没有贷款/未婚/男性/私企工作的话，往往是大学生。通过关联分析可以帮助银行进行提前识别，识别出存在信用贷款拖欠的客户中主要集中有什么特点，通过一个或多个特点能够较为准确的预测出其另外的特点，就可以对其进行风险防估等，减少银行损失。

6.结束语

总的来说，在当今大数据时代，数据就像一个钻石矿，当它的首要价值被发掘后仍能不断给予[5]。数据挖掘就是其挖掘的工具通过将模糊无规律的数据有效的转换成为规律的概念、信息和知识, 从而发挥着巨大的应用价值。数据挖掘在银行信用卡管理方面，根据客户的个人信息，是否有房，是否有车等属性预测客户偿还信贷的能力等为银行在完善银行体系上提供了有力的依据。

参考文献

[1] 杨小军。数据挖掘技术在银行信用卡业务中的应用研究[D].数据挖掘技术在银行信用卡业务中的作用，2015（10）：80-82.
[2] 李梅. 商务智能与数据挖掘[M]. 数据挖掘概述，2016（2）：46-47.
[3] 简祯富.大数据分析与数据挖掘[M].数据挖掘作用，2016（3）：15.
[4]张晓婷. 商务智能与数据挖掘[M]. 数据挖掘算法 2016（1）：118-119.
[5]维克托.迈尔 .大数据时代[M]. 数据的价值，2013（1）：127-128.

作者简介

第一作者：刘强（1999-），男，汉，四川省泸州市，本科，四川大学锦城学院，研究方向：信息管理，数据挖掘。
第二作者（通讯作者）：鲍正德（1989-），男，汉，黑龙江哈尔滨，研究生，四川大学锦城学院，研究方向：电子商务。
第三作者：唐娅雯（1999-），女，汉，四川省资阳市，本科，四川大学锦城学院，研究方向：信息管理、J2EE