

数据挖掘的智慧财务决策支持方法设计及应用

林森

黑龙江省鹤岗市宝泉岭农垦山林粮食加工有限责任公司

DOI:10.12238/ej.v7i3.1439

[摘要] 随着数据挖掘技术的不断发展,智慧财务决策支持方法在金融领域的应用越来越广泛。本文主要介绍了基于数据挖掘的智慧财务决策支持方法设计及应用,包括数据挖掘技术的基本原理、智慧财务决策支持系统的构建流程,以及实际应用案例。通过对内容的分析,帮助金融机构更好地利用数据挖掘技术,提高财务决策的效率和准确性。

[关键词] 数据挖掘; 智慧财务; 决策支持; 金融

中图分类号: F234.4 **文献标识码:** A

Design and application of intelligent financial decision support method based on data mining

Sen Lin

Hegang City, Heilongjiang Province Baoquanling Farming Forest Grain Processing Co., LTD

[Abstract] With the continuous development of data mining technology, the intelligent financial decision support method is more and more widely used in the financial field. This paper mainly introduces the design and application of intelligent financial decision support method based on data mining, including the basic principle of data mining technology, the construction process of intelligent financial decision support system, and practical application cases. Through the analysis of the content, it helps financial institutions to make better use of data mining technology and improve the efficiency and accuracy of financial decision-making.

[Key words] data mining; Smart finance; Decision support; finance

引言

在当今信息化时代,金融行业面临着巨大的挑战和机遇。随着金融市场的不断发展和金融产品的不断创新,金融机构需要更加灵活和高效地进行财务决策,以应对市场的变化和风险。而数据挖掘技术的不断成熟和普及应用,为金融行业提供了新的机会,帮助金融机构更好地利用数据资源,提高财务决策的效率和准确性。

1 数据挖掘技术的基本原理

1.1 分类

数据挖掘分为两种类型,一种是基于数据挖掘技术的决策支持系统,另一种是基于数据仓库技术的决策支持系统。

1.1.1 基于数据挖掘技术的决策支持系统

基于数据挖掘技术的决策支持系统,是指利用数据库管理系统、信息检索系统等软件来完成各种数据分析和挖掘工作。这种决策支持系统通常是面向特定领域、特定需求设计的,适用于某个企业内部的财务数据分析和挖掘工作。在实际应用中,主要有以下几种方法:一是基于统计分析的方法,利用统计分析软件对财务数据进行分析 and 挖掘;二是基于数据挖掘技术的方

法,如决策树、神经网络等;三是基于数据仓库技术的方法,如联机分析处理、数据挖掘等。

1.1.2 数据仓库

由众多不同类型和不同规模的数据库组成的集合。由于财务数据具有较强的稳定性、连续性和不确定性,在财务数据仓库中进行挖掘工作时,通常采用数据挖掘技术。具体来说,数据仓库主要有以下几种应用:一是利用数据挖掘技术对财务数据进行分析 and 挖掘,根据特定需求选择合适的工具和算法对财务数据进行处理;二是利用数据挖掘技术对财务数据进行预处理,使之符合挖掘标准;三是利用数据挖掘技术对财务数据进行挖掘和分析,以发现其中隐含的知识。

1.2 聚类

聚类分析是一种无监督的学习方法,通过对数据进行划分,将数据分类,进而为数据挖掘提供有效的数据集,对同一簇中的数据进行分析,实现对不同类别中数据的差异化处理。聚类分析的应用领域非常广泛,如在金融领域中,通过聚类分析来对客户进行分类,对客户进行分层管理。在电信领域中,通过聚类分析来进行客户分类管理。聚类结果的好坏直接影响到数据挖掘结

果的准确性和可靠性。目前在聚类分析过程中广泛应用的方法主要有五种: 层次聚类、层次网格聚类、密度聚类和基于密度的聚类方法(DBSCAN)。层次聚类是将数据对象分成多个层次, 每个层次包含一个数据对象及所属类别或属性。层次聚类主要是利用数学公式将数据对象映射到低维子空间, 形成一个有序的簇。适用于小样本聚类问题; 层次网格聚类是利用网格来进行聚类分析; 密度聚类是通过调整密度函数来形成一个簇; 基于密度的聚类是根据距离最小化原则和相似性原则来形成簇。DBSCAN算法是一种基于密度和空间距离相结合的无监督聚类算法。DBSCAN算法中主要采用了K-means算法思想和密度函数来对数据对象进行处理, 将原始数据划分为K个集合, 随机选择K个点作为初始聚类中心, 然后通过迭代计算各个点到每个聚类中心的距离, 将数据对象分为不同类别。

1.3 关联规则

关联规则挖掘就是将数据库中的事务数据和属性数据用一种关联关系来表示, 利用关联规则找出事物之间的某种内在联系。是数据挖掘的重要内容, 主要包括两个方面: (1) 从大量数据中发现相关性, 即找出事务数据集中的项与项目之间的相互关系; (2) 通过发现与已知项目之间的联系, 发现新的知识。对于给定的数据集D, 找到一个最大的支持度为FP(D)=FP-D_i, 其中FP表示项目集合中的第i个项目, D_i是属性集中对应项。在关联规则挖掘中, 最小支持度定义为1-FP, 最大支持度定义为FP-Max(D)=MAX(D)/MAX(D)。

关联规则挖掘算法一般指Apriori算法和FP-growth算法

1.3.1 Apriori算法

由美国西北大学教授Don Apriori于1984年提出的。Apriori算法的基本思想是通过产生频繁项目集, 频繁项目集产生更多的频繁项目集来完成整个挖掘过程。Apriori算法适用于处理非整数维数据和不确定维数据。

1.3.2 FP-growth算法

由Donald Ruder于1992年提出的, FP-growth算法更好地发现任意维数据中潜在的、有价值的关联关系, 较好地处理非整数维数据和不确定维数据。FP-growth算法最大优点是只需要进行一次扫描即可获得所有可能的关联规则, 是关联规则挖掘中效率最高的方法之一。

2 智慧财务决策支持系统的构建流程

2.1 数据挖掘与知识发现的结合

在财务管理中, 决策的主要依据是财务报表, 而财务报表的数据量非常大, 在处理大量的财务数据时, 很难将所有的数据直接导入到决策模型中去。为了有效地利用大数据量, 提高决策效率, 采用数据挖掘与知识发现相结合的方法。在数据挖掘中, 主要使用关联规则挖掘算法和决策树等算法。在知识发现中, 主要使用聚类、关联规则、神经网络等方法。

在数据挖掘中, 先进行数据预处理, 然后根据预处理后的数据进行知识发现。比如: 利用关联规则挖掘方法对原始数据进行预处理后, 再使用决策树算法对这些数据进行挖掘。由于知识

发现需要大量的有价值的知识, 在数据挖掘中采用基于规则库的方法对海量数据进行挖掘。比如: 在智慧财务决策支持系统中建立一个基于规则库的知识库, 利用数据挖掘技术对财务信息进行分析后得出财务报表所需要的规则和信息。由于财务报表具有非常强的时效性, 所以在利用规则库对海量数据进行挖掘时要注意对数据进行实时更新。

2.2 数据预处理

数据预处理是数据挖掘的第一步, 主要目的是消除或减少数据中存在的异常值、离群值等, 使数据集成为符合要求的数据集。常见的数据预处理方法包括: 删除和替换法、归一化、标准化和数值化等。中, 数据归一化是将多个数值归为一个数值; 标准化是将一个变量转换成另一个变量; 数值化则是将连续值转化成离散值。

对于数据集中的异常值, 通过直接删除或替换来解决。直接删除法就是直接将缺失值用0或1来替代, 然后进行运算, 减少处理时间, 提高运算效率。通过设定阈值来达到同样的效果。

2.3 特征选择

特征选择是选择那些对分类结果影响最小的特征, 以提高分类器性能的一种过程。根据所选择特征在分类问题中的重要程度, 可将特征选择方法分为基于数据、基于模型和基于属性三类。

数据是最常见的特征选择对象, 但本身不能作为有效的特征。在数据集上建立模型需要大量的处理过程和计算资源, 如何对数据进行预处理并提取特征是数据挖掘领域研究的重点。以特征选择法为例, 是指利用各种方法对原始数据进行某种处理, 使得有用信息尽可能多地被提取出来, 避免对原始数据进行冗余处理, 降低数据处理对样本学习能力和存储能力的影响。

2.4 模型构建

2.4.1 回归分析法

根据输入变量与输出变量之间的定量关系建立模型, 利用样本数据来描述和解释因变量变化而导致的因变量变化的现象。根据回归分析的原理, 建立了预测模型。在实际应用中, 回归分析法通常被用于处理大规模数据集, 能够有效地发现隐藏在数据集中的数据规律, 实现对数据的挖掘。常见的回归模型有: 线性回归、逻辑回归和神经网络等。

2.4.2 决策树法

一种树形结构的分类方法, 通过建立决策树实现对不确定分类问题进行有效的分类。决策树根据不同分类要求分为有监督学习、半监督学习和有监督半监督学习等不同类型。

2.4.3 神经网络法

一种模仿人脑神经系统结构和功能而建立起来的数学模型, 在学习过程中, 根据输入和输出变量之间的关系, 建立一个对应关系。在实际应用中, 神经网络通常被用于解决复杂问题和预测未知数据的分类问题。神经网络主要由输入层、隐藏层和输出层组成。例如: 在对某一公司进行预测时, 需要先将该公司的财务指标输入到神经网络中进行训练。

2.4.4 随机森林

具有自适应特征选择能力的分类算法，是一种新型集成算法。随机森林不需要大量的数据就能实现对未知数据的分类，基本思想是将训练集中在所有样本点上，按照某种规则进行随机分组，然后根据不同组中样本点的分布情况来调整每个样本点的权重，最后得到最优解。

2.5 模型评估

2.5.1 预测准确性

模型评估主要是评估在实际场景中的预测能力。对此，从两个方面进行：一是构建更精确的特征提取方法。采用更丰富的特征提取方法来提升模型的预测能力，例如利用聚类、关联规则、多分类等方法来提取更加复杂的特征，提升模型的预测能力；二是提高对数据进行建模时所需的特征数量，例如在随机森林中，将数据分为训练集、测试集以及验证集三部分进行建模。

2.5.2 预测能力

评估指标从两个方面进行：一是预测精度。通过构建评价指标体系来评价模型的预测精度；二是拟合曲线。拟合曲线在一定程度上反映出模型在训练集和测试集上的拟合效果。该指标主要用于衡量模型对实际场景中数据所表现出来的拟合能力，为模型选择提供参考依据。

2.5.3 预测性能

评估指标主要包括两个方面：一是拟合优度，即通过拟合曲线来评价模型对实际数据的拟合效果；二是预测精度，即通过拟合结果来衡量实际数据与模型预测之间的误差。在实际场景中，采用一系列指标来度量该指标，例如召回率、F1分数等。

2.5.4 综合评估

主要包括三个方面：一是评价指标体系构建是否科学合理；二是模型选择是否准确；三是模型在实际场景中是否具有较好的适用性。在构建评价指标体系时，将多个评价指标结合起来进行综合评价；在选择具体模型时，通过对比分析不同模型对实际数据的拟合效果来选择最优模型。

2.5.5 适应性

主要包括两个方面：一是在实际场景中是否具有较好的适用性；二是在实际场景中是否能表现出较好的性能。在构建评

价指标体系时，根据实际需求进行调整。例如在随机森林中，根据数据集规模大小、数据集本身复杂程度来进行指标权重设置；在交叉验证过程中，将多个实验结果进行对比分析来选择最佳交叉验证结果。

2.5.6 应用效果

主要包括三个方面：一是通过对实际场景中的数据进行建模分析来评估模型对实际数据表现出来的拟合效果；二是通过对预测精度、预测误差等指标进行对比分析来评估模型在实际场景中的性能；三是通过对实际场景进行应用分析来评估该方法的适用性。

3 结束语

本文主要介绍了基于数据挖掘的智慧财务决策支持方法设计及应用。通过对数据挖掘技术的基本原理、智慧财务决策支持系统的构建流程和实际应用案例的分析，可以看到数据挖掘技术在金融领域的重要性和应用前景。未来，随着人工智能和大数据技术的不断发展，智慧财务决策支持方法将会更加智能化和个性化，为金融机构提供更多的决策支持和帮助。希望本文对金融从业者和研究人员有所启发，促进智慧财务决策支持方法的进一步发展和应用。

[参考文献]

- [1]那晓红.基于数据挖掘的智慧财务决策支持方法设计及应用[J].中国总会计师,2022(4):36-40.
- [2]林宝容.基于数据挖掘的智能财务决策支持系统研究[D].杭州电子科技大学,2012.
- [3]林宝容.基于数据挖掘的智能财务决策支持系统研究[D].杭州电子科技大学,2024.
- [4]汪诗怀.基于数据挖掘的智能财务决策支持系统设计研究[D].郑州大学,2006.
- [5]郭永强.决策树方法在农业智能决策中数据挖掘的应用与实现[D].中国农业科学院,2012.

作者简介:

林森(1991—),男,汉族,辽宁省黑山县人,硕士,悉尼科技大学,会计,研究方向:智慧财务。