

数据挖掘技术在房地产行业的应用综述

岑希

(厦门大学经济学院计划统计系, 福建 厦门 363105)

摘要 房地产行业是复杂的非线性系统, 数据挖掘技术的应用对政府的决策、企业的运营和个人的投资选择都有重要的影响。本文从房地产行业的价值链出发, 旨在从房地产行业贷款风险分析、公司经营等级划分和客户关系管理三个方面着手, 综述数据挖掘技术这一新的研究领域在房地产行业中的应用。

关键词 数据挖掘; 房地产行业; 多元统计分析

房地产行业是一个数据量大、关联性强、影响因素多的复杂非线性系统。数据挖掘技术在房地产行业的应用是一个年轻且充满希望的研究领域, 人们对它的研究正日益广泛和深入。解决好这些问题, 对于政府部门合理分析产业发展, 制定产业政策及开发企业和个人正确判断房地产市场形势、做出投资或购房决策具有重要意义。

总体来说, 数据挖掘(Data Mining, 本文下面均简称 DM 技术)在房地产行业的应用主要使用了三种技术:

- (1) 分类技术
- (2) 聚类技术
- (3) 关联规则发现技术

本文主要从商业贷款风险评估技术、房地产公司经营等级划分和客户关系管理这三点综述数据挖掘技术在房地产行业的应用。

1 房地产商业贷款风险分析——基于神经网络技术

随着房地产市场的逐步走热, 投身于房地产项目开发的企业日益增多, 在贷款企业情况的多数不确定因素面前, 银行的商业贷款风险也随之提高。信息不对称情况下, 银行在贷款的过程中不得不承担由于缔约方违约或经营的失败而使得银行不能回收贷款利息与本金的风险。因此, 银行机构为了对贷款风险的有效管理与控制, 必须寻求有效的贷款风险预测量化的方法和技术。

针对此问题, 可以构建一种基于神经网络的客户模糊分类数据挖掘方法, 旨在对客户的贷款风险进行分类, 为银行贷款风险分析部门准确分析目标客户提供一种有效的方法。该方法对客户进行了合理区分, 试图为充分做好客户关系管理, 增强企业竞争优势, 提高企业的核心竞争力提供有效的解决方案。该方法的思想是运用模糊神经网络分析探索影响因素之间共同作用产生的交互效应, 用过去的事实进行综合评价, 训练网络, 来确定权重的大小。这样将会使结果更加准确可信。

1.1 贷款风险评估中指标体系的建立。

数据挖掘针对不同行业的具体运作具有不同的分类方法, 输入量也具有各自的特点, 对于建设企业的贷款风险系数等级的确定, 建立需要影响风险系数的指标体系, 依据以下 8 个方面的分析方法。即: (1) 行业发展趋势 (2) 国家政治与监管环境; (3) 管理层素质及承担风险态度 (4) 公司经营及竞争地位 (5) 财务状况及流动资金来源 (6) 公司结构框架; (7) 母公司担保及支持协议 (8) 突发事件风险。指标性质分为定性指标和定量指标。

1.2 客户贷款风险模糊神经网络分类器体系的建立。

模糊逻辑与神经网络从概念到方法都不相同, 但是它们都是为了表示预处理不确定性知识而引入的方法, 而且都是一种模型的估计器。因此, 在不确定知识处理中, 模糊逻辑与神经网络可以结合。模糊逻辑要求把使用的知识以模糊命题、模糊规则的形式表示在知识库中, 模糊命题和模糊规则是模糊概念的一种逻辑表示形式。模糊知识只能通过领域专家提供或其他途径获取, 模糊逻辑系统本身不具备学习能力。模糊神经网络以学习实例的形式隐含所需要的知识, 学习实例是以确定的输入输出模式表示的数值集合。神经网络具备学习能力, 经过学习后获得的知识的表示形式是网络的互连结构及其相应的邻接权矩阵。

模糊逻辑的推理是以模糊匹配和模糊集的运算方法为基础, 得出的推理结论是模糊概念之间的模糊逻辑关系。神经网络的推理是以网络的非线性数值运算为基础, 得出的推理结论也是数值的, 需要通过对输入输出模式的解释才能得出相应的逻辑结论。

由于本文只做综述性汇总和概括, 原理及应用的具体问题请参阅相关文献。

具体过程可以用 Matlab 软件程序编辑实现 BP 神经网络, 网络的输入层的神经元对应问题产生影响的测试变量, 输出层分别对应各种风险等级, 其中还包括隐层设置、学习率、最小误差的设定等。

对建筑行业企业贷款风险等级的确定, 可以引入一个基于

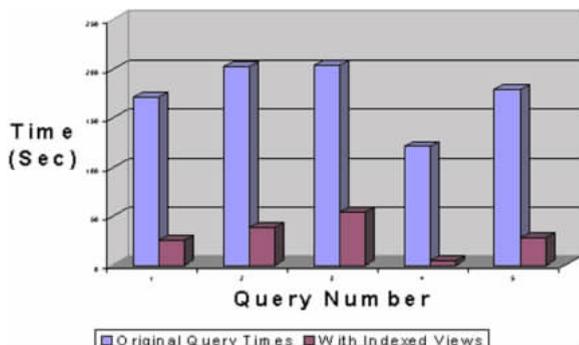
引视图和分区策略, 从而简化了物理数据库的设计过程。

3 结束语

下图说明了使用以上方法通常所能取得的性能改进程度: 实践证明, 灵活使用上述方法, 能提高数据库整体性能。

参考文献:

- [1] Microsoft SQL Server 2005 Enterprise Edition English version helper document.
- [2] ROBIN DEWSON.《SQL SERVER 2005 基础教程》[M].北京:人民邮电出版社, 2006.
- [3] SOLID QUALITY LEARNING[美].SQL SERVER 2005 从入门到精通(应用技术基础)[M].北京:清华大学出版社, 2006.



模糊数学与神经网络的解决体系构架。通过利用模糊逻辑模仿人脑模糊思维、模糊推力的能力,与神经网络模仿人脑结构映射输入特征与输出结论的非线性关系的优点相结合,取长补短,最终准确地预测客户的贷款风险等级,将能够为企业带来两方面的好处:

(1)有效地识别有价值客户,提高获利机会,并且有针对性的向顾客提供满意服务,维持良好的合作关系,从而延长客户生命周期。(2)减少银行不良贷款比率,规避贷款风险,以备扩大利润空间。

2 房地产业上市公司的统计数据挖掘分析——聚类分析和因子分析

对上市公司而言,对外披露的财务指标中往往隐含很多能具体反映出公司在某年度财务状况的信息,但如果仅仅是笼统的、盲目的去查看这些繁杂的财务指标,一般很难从中发现更多对投资者做出投资决策有帮助的信息。故这种分析将以多元统计的方法对这些财务指标进行数据挖掘工作。首先选取若干个能够反映公司财务状况主要参考指标进行因子分析,缩减变数并提取出对这一系列指标做出主要贡献的因素(即主成分),然后再对各上市公司进行聚类分析,划分营运等级,找出各类公司存在的财务状况的相似性。

这种利用多元统计的方法分析房地产业上市公司的好处是:投资者可以根据划分出来的上市公司的等级情况,透视企业的财务状况,做出合理的投资决策;而各上市公司也可以根据自己的营运等级,找出问题所在,弥补自身不足,修正自己公司的经营模式。

2.1 因子分析

可以借助若干家房地产板块上市公司的年度财务报表,从中选取若干个具有重要参考价值的指标,首先利用 SPSS 统计软件分析包进行因子分析,寻找这些财务指标背后是否有公共因子的存在,并籍以缩减变数进行分析。可以提出如下一些指标做参考:流动比率、总资产周转率、应收帐款增长率、速动比率、主营业务收入增长率、应收帐款周转率等。

在因子分析过程中,将上市公司的若干个财务指标通过提取其公因子,得到几个能较好反映原指标所包含信息的因子,根据财务管理的专业知识,可以对这几个公因子加以归纳分析。如:代表房地产企业盈利能力的“盈利能力因子”,还有“偿债能力因子”、“周转运营能力因子”以及“未来发展能力因子”等等,相应的指标可以分别为每股收益、流动比率、总资产周转率、每股收益增长率等。

由于这种利用多元统计的方法对各种指标进行抽象地提炼和划分,能够较客观地反应各房地产企业的财务能力。还可以通过因子分析得出因子得分矩阵,将各指标量化汇总,得出各公司的总得分,是能够较充分反映公司营运能力的指标。

2.2 聚类分析

根据选取的所有财务指标,对若干家公司进行聚类分析,得到聚类分析的龙骨图。将具有相似财务状况的房地产公司聚成一类,通过与其他类公司的比较和分析,得出自己竞争的优势和劣势,以及该企业在整个同行业中所出的地位,这对企业未来发展有较大的指导作用。

基于多元统计分析的视角,可以从大量繁杂的数据信息中,挖掘到很多实际的、有价值的东西,从而更有利于管理者和投资者决策活动,这也就是数据挖掘技术的含义所在。总之,若企业能够及早发现自己的经营绩效上已出现危机,就必须找出原因加以改善,提高自己的市场竞争力,投资者也可以对公司历来的

业绩情况进行分类,判断该公司是否值得投资,以降低投资风险。

3 DM技术在房地产客户关系管理系统中的应用

3.1 DM技术在房地产客户关系管理系统中的应用

随着市场竞争的进一步加剧,房地产企业对客户关系管理的关注比以往更高,能否对客户数据进行合理、有效地利用,决定着一个企业的成败。面对海量数据,DM技术越来越成熟地应用于房地产 CRM 系统中,为企业制定营销策略、开展营销活动提供决策支持。

房地产行业的客户信息有许多特点,如客户群广泛、客户特征描述结构复杂,又如,客户需求层次不一,易受外界因素影响等,因此,有必要针对这种复杂、多变和多层次性的客户信息进行分析挖掘,以识别客户购买行为,发现客户购买模式和趋势,从而进一步提高企业的服务质量,降低企业营销成本。下面是 DM 技术在房地产行业不同营销阶段上的应用:

| 营销阶段 | 所使用的数据挖掘技术 | 要达到的目标 |
|----------|-------------|--|
| 客户细分 | 分类、聚类技术 | 划分目标客户群 |
| 客户保留 | 分类回归、关联分析 | 识别无效客户及流失客户的消费决策模式,帮助企业识别现有客户中相似的潜在流失客户,以便采取预防措施。 |
| 客户拓展 | 分类、关联分析 | 有针对性地发现客户的反应行为模式,将潜在的客户列表排序,找出对所提供的服务最感兴趣的客户。 |
| 现有客户升级服务 | 分类、预测 | 通过对现有客户行为特点的分析,对客户将来的行为进行预测,用预测得到的评分为每个客户选出最合适的几种服务方案。 |
| 活动分析 | 多维分析 | 通过比较热销期间的销售量和交易数量等有关情况,分析营销手段的有效性。 |
| 销售预测 | 分类、特征化、关联分析 | 分析在项目开发的不同阶段,如开盘、热销、封顶、现房,成交客户和意向客户的属性特点,以便进行销售预测。 |

3.2 将 DM 技术应用于房地产 CRM 系统的具体步骤

3.2.1 明确商业目标。

描述出需要解决的问题,并且确定所要解决的问题属于哪种应用类型,是属于关联分析、时序模式、分类、聚类、偏差分析和预测,还是综合应用。

3.2.2 数据准备。

数据挖掘是由可以获取的数据驱动的,其成功很大程度上取决于所收集数据的数量和质量。建立可靠数据资源的第一步就是确定系统将要使用数据的数据字典。它们应包括如下工作:确定数据库每一列预期存放的内容;数据来源的描述;对缺失数据的处理等。

3.2.3 利用数据挖掘技术建立模型。

利用数据挖掘的一些算法对数据进行分析,建立模型,并选择好相应的实现工具。假如明确了主要任务是分类,可以采用的算法有遗传算法、决策树和人工神经网络等。选择了模型的类型及实现的工具后,有时还需对数据进行预处理。

3.2.4 反复验证。

从较小的系统(如较小的客户群体)开始,但要完成全过程(即从数据收集和处理,到数据挖掘,最后到产生促销数据结构,如邮件发送列表)。通过验证,可以纠正系统中的错误,并有利于用户对系统功能的进一步理解。

3.2.5 实施与维护。

在应用模型之后,还要不断监控模型的效果。因此随着模型使用时间的增加,要不断的对模型做重新测试,有时甚至需要重新建立新的模型。

4 结束语

房地产商之间激烈的竞争导致了面向市场营销和顾客服务的数据仓库的建设以及数据挖掘技术的应用,本文对数据挖掘技术在房地产 CRM 系统中应用的具体方式与方法进行了探讨。随着数据挖掘技术的进一步发展,必将使得房地产 CRM 系统更加成熟完善。