

# 大数据时代数据挖掘技术分析

张凌燕

(常州工程职业技术学院 图文信息中心 江苏 常州 213164)

**摘要:** 数据挖掘技术是依托于大数据技术所诞生和发展的新技术,它隶属于人工智能技术,主要应用于商业领域。数据挖掘技术的诞生源自原始数据具有不规则性、复杂性、多样性等特点。运用数据挖掘技术可以将原始数据中具有价值的信息提取出来,使其发挥出更多的作用。通过聚类分析、分类和预测、关联分析等手段,发现数据挖掘技术在市场营销领域、制造业领域、电信领域、教育领域都得到了广泛的应用。在可预期的将来,数据挖掘技术还会在更多的领域中大放异彩,为社会经济发展提供技术支持和数据信息服务。

**关键词:** 大数据时代;数据挖掘技术;聚类分析;关联分析

**中图分类号:** TP311 **文献标志码:** A **文章编号:** 1674-8646(2020)18-0090-02

## Analysis of Data Mining in Big Data Era

Zhang Lingyan

(Graphic Information Center, Changzhou Vocational Institute of Engineering, Changzhou 213164, China)

**Abstract:** Data mining is the new technology, which relies on big data technology birth and development. It belongs to artificial intelligent technology, and is mainly applied in the business field. The birth of data mining originates from the irregular, complicated and diversified characteristics of original data. Data mining can extract the valuable information, and play more function. Through clustering analysis, classification and prediction and correlation analysis, it is found that data mining is widely applied in marketing, manufacture, telecom and education field. In the future, data mining will be applied in more fields, to provide technical support and data information service for social and economic development.

**Key words:** Big data era; Data mining; Clustering analysis; Correlation analysis

## 1 数据挖掘技术及其意义

数据挖掘技术是依托于大数据技术所诞生和发展的新技术,它隶属于人工智能技术,主要应用于商业领域。数据挖掘技术的诞生源自原始数据具有不规则性、复杂性、多样性等特点,运用数据挖掘技术可以将原始数据中具有价值的信息提取出来,使其发挥出更多的作用。从商业层面上看,原始数据虽然数量庞大,但通过数据挖掘技术对其抽取、转换、分析后,就可以为商业决策提供参考依据,从而促进相关行业及企业的健康发展。

## 2 数据挖掘技术的常见类型

### 2.1 聚类分析

聚类分析分为硬聚类和模糊聚类两种主要类型。硬聚类是在指向距离最短的聚类中划拨事物,且该项事物只属于某一类。模糊聚类则不归属于某一类,极

有可能存在多种类型。总体来说,聚类分析是对数据库或对象库进行抽选后,在每一组类当中放置一个相同或者相类似的数据,由此生成多个组类,这一过程就是聚类分析过程。聚类分析过程无须知识基础,也无须监督管控,只需要找到数据中有价值的部分并对其进行相应的分类即可。聚类分析与“分类”之间既存在着一定的相似之处又存在着差异性,比如聚类分析无法知道相应的属性,但可以分析出相间类聚的某些规律。因此,聚类分析技术在心理学领域、医学领域具有较为广泛的应用。

### 2.2 分类和预测

预测过程主要包括分类和预测两种类型。分类指的是预估分类标号,也可以将分类视为数据挖掘技术中的一个基础性步骤,在训练数据可以预计的情况下,分类可以很好地呈现出其相应的特性,只要完成一个类别描述,便可以进行下一步的分类构建。在分类的过程中,不仅存在着监督的必要,还需要来源于训练数据的确定性描述。目前的分类存在多种方法,决策树和神经网络是其中比较常用的两种方法。在神经网络系统当中存在着众多具有分布式存储信息、可大规模

收稿日期: 2020-04-29

作者简介: 张凌燕(1982-),女,本科,助理研究员。

并行处理、自组织自学习能力良好的神经单元。也可以将神经网络视为多个神经单元的集合,从而使轴突连接的生物神经元大群集问题得以有效解决。但事实上,每个神经单元还可以同时连接多个其它神经单元,特别是将相应的数值输入以后,神经单元之间便可以重新进行数值组合,也可以对其进行求和。预测是指通过连续值函数模型的构建再借助分类、回归、回归分析法、局势外推法等方法对数据规律进行预估。以回归分析法为例,它可以根据数据库当中某一数据的函数关系体现出数据之间的联系与区别,最终分析出数据信息特质的依赖程度。

### 2.3 关联分析

从理论的角度上分析,世界上所有事物都不是独立存在的,一旦其中某一方发生变化,就会导致另外一方或者多方受到不同程度的影响,从而衍生出更多的变化。数据挖掘技术中的关联分析就是依据这一项理论,以大数据技术作为支撑分析出事物之间的数据规律,再依据该规律开展预估工作。例如某电商企业针对购物篮进行分析时,需要先通过数据挖掘技术找到购物篮中各商品的排列规律及其内在联系,进而分析出消费者的消费层次、消费心理、消费行为特征,最终推出具有较强针对性的营销策略,推动销量的提升。在现实世界当中并不存在信息“非黑即白”的情况,所以灰色关联分析是所有关联分析当中最为常见的方法之一。该方法主要以各因素的发展趋势、相互之间的差异性或者相似度作为分析的依据,对各项数据进行归纳与评价,最终分析出各因素间的关联性。当通过灰色关联分析来描述数据信息之间的过渡阶段时,包含已知和未知两部分信息。由于未知信息的存在,导致分析结果存在较强的不确定性,而灰色分析则可以对解决方案进行优化,最终得出最佳解决方案,使数据分析中的各项问题得到妥善解决。

## 3 数据挖掘技术在各领域中的应用

### 3.1 市场营销领域

在目前所有的数据挖掘技术应用领域中,应用得最广泛、效果最显著的当属市场营销领域。企业在开展市场营销的过程中通过应用数据挖掘技术,可以更加全面细致地分析出消费者的消费行为习惯。这样一来,市场营销人员便可以制定出极具针对性的营销方案,促进销售量的显著提升。伴随着数据挖掘技术的不断发展,它在保险行业、电子商务行业中的应用价值也大幅提升。通过聚类分析方法可针对消费者无意识的、不规律的消费行为进行分析,从中识别出消费习惯和消费动机,赋予原本无效的数据信息新的应用价值,从而帮助企业制定出更加精准的营销策略,促进企业的健康发展。

### 3.2 制造业领域

数据挖掘技术应用于制造业领域与应用于其他社会领域相比有所不同。数据挖掘技术在制造业的应用价值主要体现在检查产品质量,如通过对产品数据进行分析找出其中的规律,或者通过对生产流程进行分析找出影响生产效率及产品质量的主要因素。分析后再有针对性地提出解决方案,使生产流程更加科学合理,使产品的生产成本得到有效控制。在制造业企业决策中,数据挖掘技术同样可以发挥非常重要的作用,利用相关技术能够获取具备使用价值的信息,再通过决策树算法确保决策的准确性与科学性。

### 3.3 电信领域

近年来,我国的电信行业实现了快速发展,用户规模也在不断扩大。在这种情况下,电信企业必须采取有效措施提高自身的技术水平及服务质量,以此来保证客户满意度与忠诚度。但是,对于整个电信行业来说,技术服务属于非常庞大的混合载体,唯有通过应用数据挖掘技术方可对各种复杂数据进行全面、深入的分析,找到其中的规律,进一步帮助电信企业对以往的技术服务体系作出优化调整,为用户提供更优质的服务。

### 3.4 教育领域

随着我国教育体制的全面改革,数据挖掘技术在教育领域内也得到了很好的应用。教育管理者可利用数据挖掘技术对学生心理、行为习惯进行系统化的分析,并依据分析结果对学校的教育管理方式进行优化改革,帮助教师找到更加科学有效的管理方法,从而促进教学成效的提升。数据挖掘技术还可以针对学生的成绩进行分析,找到大多数学生在学习中的薄弱之处,使教师为其提供有针对性的教学指导,促进教学质量的显著提升。

## 4 结语

伴随着我国数据挖掘技术的整体发展,该项技术在科研领域、生物制药领域、航空航天领域中也得到了大范围的应用,并且取得了较为理想的应用成效。数据挖掘技术使大数据信息变得更加实用,使大数据技术的优势在各行各业都得到了有效发挥。

### 参考文献:

- [1] 石静静,王书真,王艳佩.基于大数据的数据挖掘技术探讨[J].数字技术与应用,2016,(03):96-99.
- [2] 孙勤红,沈凤仙.大数据时代的数据挖掘及应用[J].电子技术与软件工程,2016,(06):86-88.
- [3] 杨华昆.大数据时代数据挖掘技术探讨[J].电脑编程技巧与维护,2015,(24):155-157.
- [4] 曹诚诚.大数据时代的数据挖掘技术分析[J].数字技术与应用,2017,(12):240-243.