

# 数据挖掘技术在现代经济统计中的应用思路分析

廖友国 熊建益 沈波

**摘 要:** 数据挖掘技术的含义是对于目前的数据管理活动中的数据缺失情况严重、数量组成烦琐以及数量基数大等问题而深入研发出来的一类高效率的技术。近年来,伴随国内经济的进一步发展,经济、社会活动中的各类经济因素逐渐增多,整体经济关系也更加烦琐,为了让数据挖掘技术可以有效地作用于经济统计中,提升经济统计工作的效率,本文重点分析了在经济统计中如何运用数据挖掘技术,以供参考。

**关键词:** 数据挖掘技术; 现代经济统计; 应用思路

近年来,伴随国内经济的进一步发展,经济统计也在持续的进步,截至现在,国内经济统计所需要进一步统计的数据量很庞大,这严格要求了数据统计人员,针对不会充分挖掘数据的人来讲,其数据统计结果就不太正确。数据量在不断扩增的时候,数据的种类也逐渐增多,这造成以前的数据统计方式不能顺应如今的数据统计需要。为了提升数据统计的综合性以及正确性,提升数据统计的质量以及效率,本文重点分析了在经济统计中如何运用数据挖掘技术。

## 一、数据挖掘技术含义

从表面上理解数据挖掘属于一种深入开发数据信息的技术,在详细的运用实践中,能够从烦琐的数据库中进一步找出有意义的数据信息,并且对其进行探究与整理,使散乱数据信息发挥其价值。如今国内经济社会活动的统计数据信息具有的核心特征就是构成烦琐、随机性很强、信息量大以及一些数据信息不完善等。这些统计数据信息通过原生态的方式有效保存于数据库中,其价值就会被进一步浪费、闲置,无法施展其应有的功能,数据统计工作也就变得毫无意义。数据挖掘技术可以探究以及归纳这些烦琐的原始数据,最后产生可用的数据形态,便于数据使用人员的获取以及运用。这样一个深入加工统计数据信息的进程被叫作数据挖掘。从技术上说数据挖掘技术可以对原始数据进行深加工以及精加工,是一种探究原始数据的技术,要达到这么高水平的体系设计,需要借助于大量先进的技术,其中涵盖模糊概念技术、模式识别系统以及数据库知识等。

## 二、数据挖掘技术运用于经济统计的优势

### (一) 很强的综合性运用能力

现阶段,数据挖掘技术在大部分领域都已获得了普遍的运用,在探究以及总结数据的时候,施展着非常关键的作用,其中涵盖数学统计、物理分析以及经济统计等方面。数据挖掘技术不但需要具有很强的数据挖掘能力以及统计能力,而且需要很强的后续数据自动探究以及统计的作用,可以配合数据的使用人员进一步处理数据。在经济统计的具体运用中,数据挖掘技术发挥了超强的作用,其有

利于分类、整理以及开发经济数据,会为数据的使用者带来越来越快捷、便利的服务。

### (二) 很强的有效性

现阶段,国内运用数据挖掘技术的时间十分有限,技术上并未达到较高的水平,不过就如今的数据挖掘技术来说,对国内的有关经济统计工作已经发挥了良好的运用效果,充分强化了统计工作的性能,并且提升了笔者整理、统计和探究经济数据领域的的能力,并且会借助于数据挖掘技术进一步挖掘越来越多有意义、价值的信息,并最后体现为数据的正确性以及高效性。

### (三) 运用宏观型的数据库

数据挖掘技术也可以进一步运用于目前的宏观型数据库中,并获取优良的成效。现阶段国内的经济数据统计分析所采取的策略一般还被以前的经济统计方式所限制,大部分收集和总结数据信息的方式还很过时,不能与相关信息产生一个完善的整体。大部分数据的使用人员在综合归纳和管理数据的过程中,会产生很多的问题,不能进一步实现数据统计和探究的目的。于是,就应该采取越来越高效与新式的策略来提高对经济数据的统计效率。在宏观型的数据库中数据挖掘技术的适应性非常强,宏观型的经济统计数据库能够进一步为数据挖掘技术提供一个良好的平台,便于其施展出本身的具体特性。针对经济统计行业来说,一定要确保其数据探究的正确性,其速度也应该开展完善处理,于是在数据宏观经济统计中导入数据挖掘技术,需要借助于很多真实的数据资源,最后统计探究出高效的宏观数据。

## 三、在经济统计中如何运用数据挖掘技术

### (一) 集成化处理策略

数据集成的含义是整合各种数据,从杂乱进一步变成整体。伴随经济的飞速发展,搜集到更加巨大的数据信息量,获取信息的渠道更加广泛,不利于信息的集成。因为社会经济活动里面数据的相关来源并不只是官方统计局,多来源的数据体现为不一样的模式,从而产生了经济数据信息实体的相关识别问题。

比如,挖掘数据的时候明确数据库里面的“sfid—id”与

其他数据库里面的“std—no”是否象征一样的实体,经过探究比较两个数据库之间的含元数据可以确保实体数据识别的品质。有效处理原始数据中体现正相关关系的数据,进而确保数据库里面的量保持在很低的水平上,才可以方便有关单位的应用以及管理。在现实挖掘数据的时候应该精简数据,降低数据量。

#### (二) 预处理策略

原始数据一般都是不完整的,某些数据不同,需要开展数据预处理策略。数据预处理是最基本的处理形式,因为数据挖掘唯有在给予基本信息之上开展智能探究,其自身拥有基本信息局限的特征,不能取代数据搜集的功能,唯有经过预处理深入清理不可靠以及不正确的数据,才可以缩短统计数据信息之间的区别。数据清理的含义是去除原始数据信息里面空值、残缺以及具有噪声的数据,通常使用平滑法、预测法以及均值法进行处理,对于不同数据信息实施针对性地选择。如具有空值以及噪声的数据借助于平滑法或者均值法实施清理,两类策略之间的区别就是平滑法中一般采取加权平均数、而均值法中通常采取平均数,平滑法越来越重视各个数据所形成的权重影响,于是获得越来越准确的数据结果。

#### (三) 转换策略

数据转换的含义是借助于某类方式把数据转变为满足数据相关挖掘要求的方式,转变的核心策略包含数据的泛化以及规范化,后者又涵盖了零均值规范化、最小规范化与最大规范化。而数据的泛化主要是指借助于高层次的数据信息进一步取代低层次的数据,其中涵盖对数据实施连续性处理,然而因为如今的手段不利于达到数据的连续性,于是一般选取离散化,即对数据进行区间划分,直接以标号表示某些数据。进一步降低数据的搜集量,也被叫作概念分层。

#### (四) 决策树策略

在运用数据挖掘技术的时候,通过有效的探究以及归纳之后,探究数据的相关输出是一个重要的步骤,其输出的数据方式会直接影响使用人员的经济管理决策。决策树属于一种很直观的以及常见的快速分类策略。其运用的重点是建立决策树,详细来说重点分成两个部分:第一部分是借助于训练集创建且精简一棵决策树,创建输出探究的模型;第二部分是借助于建立好的决策树对输入数据实施分类,这个分类属于一个递进的过程,从决策树的根部逐渐进入到枝丫与树干,直至输入数据的具体分类进一步满足了某类条件而停止。在详细的运用中终止分割的条件包含以下两个方面。首先,当一个节点上的全部数据都属于同种类别时;其次,没有分类属性能够对输入数据实施再分割。在建立好决策树后,还要按照使用人员的详细要求对决策树实施“剪枝”,剪枝重点是为了减少由于采用训练集而对决策树自身数据输出形成的起伏影响。

#### (五) 回归分析策略

回归分析是明确两种及以上变量间互相依靠的定量关系的一类策略。借助于回归分析可以创建变量之间的相关回归方程,经过因素探究对次要变量与主要变量之间的具体关系进行判断,并且唯有经过各类检验,并且预测

误差很小,才可以把回归方程当作预测模型进一步实施预测。在挖掘数据的时候,回归分析一般被运用于解释销售额、品牌偏好与市场占有率等,经过函数有效表达其中两个及以上的数量关系,进而方便深入处理问题。挖掘数据的时候能够进一步解释因变量之间的相关变化关系与强度,有效预测自变量值。

(1) 简单线性回归分析,如因变量 $y$ 与自变量 $x$ 之间拥有正相关关系,能够获得直线方程,让全部的数据点接近于这条直线,能够以 $y = a + bx$ 表示,其中 $b$ 则是相关系数 $a$ 为截距。(2) Logistic 回归分析,此回归分析法的运用条件:第一,需要每个观测对象之间是互相独立的;第二,Logit  $P$ 和自变量之间具有线性关系;第三,当研究设计进一步为队列研究,横断面对照分析,在采取 logistic 回归分析的时候,应拥有一致的观察时间;第四,伴随统计学技术以及有关软件的不断发展,样本量不多的形势下也能够采取精确 logistic 回归分析,然而要求分析变量有效控制于一定范围内,且变量的分类宜少。判别分析群体中与正态分布不相符的形势下,借助于 Logistic 分析可以获取更佳的结果。

#### 四、结语

一言以蔽之,数据挖掘技术可以自动发现且收集有价值的信息,接着分析与统计这些数据,进一步达到数据统计的实用性、准确性以及有效性。在经济统计中运用数据挖掘技术,不但可以进一步节省经济数据统计的人力、精力以及时间,而且有效提升了经济数据统计的准确性以及有效性。

#### 参考文献:

- [1] 张爱国. 基于大数据的数据挖掘技术与应用[J]. 现代工业经济和信息化, 2017(03): 88-89.
- [2] 赵青. 大数据挖掘面临的挑战及思考[J]. 现代经济信息, 2017(09): 379.
- [3] 张丽丽. 试析数据挖掘技术在经济统计中的应用[J]. 中国商论, 2018.
- [4] 马万智. 企业数据挖掘技术在企业信息化中的应用[J]. 现代经济信息, 2017(16): 355.
- [5] 刘芬. 计算机数据挖掘技术的开发及其应用探究[J]. 现代工业经济和信息化, 2018, v. 8; No. 162(06): 49-50.
- [6] 李婧. 大数据背景下企业数据挖掘技术的应用研究[J]. 现代经济信息, 2017(23): 320.
- [7] 黄庆生. 计算机数据挖掘技术的开发及其应用探究[J]. 科学技术创新, 2018(11): 71-72.
- [8] 吕纪荣, 王士虎, 张贵彬, et al. 数据挖掘技术在高校教学评价中的应用分析[J]. 现代经济信息, 2017(14): 380-381.

#### 作者简介:

廖友国, 达州职业技术学院, 四川达州; 熊建益, 成都信息工程大学, 四川成都; 沈波, 达州职业技术学院, 四川达州。