

# Reflection on the Application of Mathematical Knowledge in the Era of Big Data

Ping Li

Zhoukou Technician College, Zhoukou, Henan, 466000, China

## Abstract

With the continuous upgrading of computer network technology, the new term “big data” has emerged. In the era of big data, it is not just about collecting and summarizing massive amounts of data, but also integrating data, scientific research, calculation, and development, the processing and use of data are gradually becoming new economic growth points, and human society has therefore entered the era of big data. From the perspective of the development of the times, big data is the result of the rapid development of contemporary information technology. Therefore, mathematical knowledge in the context of big data is crucial for data processing and mining. This paper intends to conduct in-depth research on the correlation between big data and mathematical knowledge, as well as the comprehensive application of mathematical knowledge in the big data environment.

## Keywords

big data era; mathematical knowledge; applied thinking

## 大数据时代数学知识的应用思考

李萍

周口技师学院, 中国·河南 周口 466000

## 摘要

随着计算机网络技术的不断升级,“大数据”这个新词汇应运而生。在大数据时代,不只是简单的海量数据搜集汇总,更是集数据、科研、计算和发展于一体,对数据的加工和使用正在逐步成为新的经济增长点,人类社会也因此步入了大数据时代。从时代发展的观点出发,大数据正是当代信息技术飞速发展的结果。因此,大数据背景下的数学知识,对于数据的处理和挖掘非常重要。论文拟通过对大数据与数学知识之间的关联关系和大数据环境下对数学知识的全面运用进行深入研究。

## 关键词

大数据时代; 数学知识; 应用思考

## 1 引言

从信息的视角来看,大数据是指现在的社会,人们利用某种现代的软件,高效地捕捉到网络中或实际中的信息,并对其进行深入的分析与处理,这个过程的目的从这些数据中获得更有价值的信息,从而推动社会的发展。在中国现阶段,大数据的分析与应用更适合中国目前海量数据的存储体系,以期推动当前大数据分析与应用水平的提高。把数学与大数据相结合,充分利用其效应,提高数据分析与应用的能力。所以,在当前的大数据发展进程中,研究数学知识应用,是非常有实际意义的。

## 2 数据与数学知识之间的联系

在探索大数据与数学知识的关系前,必须明确大数据

与数学知识的定义。从历史中可以了解到:前三次科技革命使人类的历史进程加快,而以电子信息为代表的信息时代也在迅速地改变着我们的生活。这类海量、复杂的数据,不能通过传统的数据处理手段来进行分析,因此被称为“大数据”。而数学则是“对数量、结构、变化、空间和信息的研究”。从这一点可以看到,大数据是一种拥有特有的要素的信息资产,因此必须有特殊的处理方式,而这一特殊的处理方式,就是用数学作为基础的学科和工具来构建的。换句话说,如果没有数学知识,大数据就失去了它应有的特征,也就没有对信息资产的利用效能。这说明,运用数学知识,才能让大数据成为现实。如今,各个行业都已经完全融入了大数据的时代,网络思维对我们的生活模式和行为习惯也在逐渐地发生着变化。开展对大数据的分析与研究,推动数学知识应用在当前大数据发展进程中的实用化进程<sup>[1]</sup>。

## 3 重要性分析

大数据可以存储海量的数据,拥有高速的数据采集和

【作者简介】李萍(1966-),女,中国河南周口人,本科,高级讲师,从事数学教学研究。

处理能力,数据覆盖的范围不再是单一的,而且数据的价值也很小,但是其质量和真实性均较高。据有关资料显示,全世界的数字增长率超过了50%,而且这个数字还在不断地增加。这就要求我们要利用数学的知识去分析、挖掘这些数据。

大数据时代的发展也促使数据挖掘方法产生和发展,在其最初的定义中,数据挖掘是指对资料中隐含或未开发的信息价值进行分析挖掘。但是,随着大数据的普及,数据挖掘已经从单纯的统计建模技术,变成了一种在各行各业都有广泛应用的方法合计数,包括教育、研究、市场、通信、制造、网络等行业,同时,也在商业人工智能的研究中扮演着重要的角色。但是,如果离开了数学知识的应用,数据挖掘就不能从数据中提取出隐含的信息和知识,从而无法为各个行业提供客观的数据基础<sup>[2]</sup>。

## 4 应用现状分析

从总体上讲,在当前阶段,数学知识应用已脱离了最初的纯粹数学研究的范畴。经过长期的发展,该方法已进入了从理论到实践相结合的阶段。随着教育事业的发展,学科体系的逐步健全,数学又衍生出了许多复合型学科,在新的的发展环境下,数学知识应用的发展也越来越被业界人士所关注。与传统的纯粹数学专业不同,数学知识应用不仅要深入地研究与发展数学理论,更要重视解决实际问题,推动它们在现实社会发展中的应用。目前,数学知识应用的研究方向各不相同,但其中最具活力的一种是以计算机为代表的。大数据科学与计算机科学有着密切的联系,而随着现代科学技术的飞速发展,迫切需要数学知识应用来支撑该方向的发展。因此,论文拟将数学知识应用与大数据时代相结合,研究大数据的分析与应用<sup>[3]</sup>。

## 5 具体的应用措施

### 5.1 在数据处理中的应用

在利用大数据时,要保证其时效性,必须在采集和处理数据的时候,对时间进行严格的控制,因此在数据处理过程中,需要更高的计算效率和结构。另外,原始数据本身所具有的特点,也会对其采集和处理产生一定的影响。例如,大数据在进行信息处理时,其原始数据含有噪音,完整性和一致性方面也有问题,因此大数据会对数据进行预处理,以确保其准确性。

在有效时间周期内,数据不仅要保持高效,而且要有较好的处理效果。当原始资料中含有噪音或不完全、不统一时,还要对资料进行预处理,以提高资料的准确度;当样本数量太多或者指数太多时,可以选择与研究成果有关的重要数据,或者是能够全面反映研究成果的重要指标。其中,回归分析方法一般与关联分析相结合,基于关联分析,通过观测两个或多个相关变量之间的定量变化关系,并构建相关的数学模型,从而可以根据已知信息对未知量进行推理。回归

分析的基本目标是使用抽样数据来估算参数,并建立一个用于检验,判断,或预测参数或模式的数学模型。

在资料的处理过程中,也要运用大量的数学原理。例如,在某个数据中,如果要对某个参数或者某个模型进行验证和判定,就可以采用回归分析的方法对其进行估算,并构建相关的数学模型。资料的处理还牵涉到许多数学原理。再如,在一个大数据中,如果要对某个数据进行降维,那么它就会利用度量理论,在大数据中,如果要对某个数据进行降维,就需要利用度量理论,把多个单一的指标集合起来,从而保持数据的完整。

### 5.2 在数据处理分析中的应用

在对大数据进行挖掘操作的时候,需要先进行需求数据的采集和整理工作。大数据具有极强的时效性,每一位工作人员都要充分意识到如何提升数据处理的质量与效率,对于采集到的原始资料,如有不全、不统一、有噪音等情况,也要对这些资料进行预处理,以提高资料的准确度。当数据数量太多或指标集过于庞大时,工作人员可根据相关指标选取若干有代表性的资料加以处理和分析,以保证所得到的研究成果得以体现<sup>[4]</sup>。

在进行资料的处理与分析时,应充分利用数学方面的知识,将不同的分析方法加以应用,主要采用了多元回归、描述性分析、关联分析等多种研究手段。在运用回归分析与相关性分析时,应将这两种数学分析方法有机地结合起来,并基于关联分析,有效地观察两个以上相关变量间的数量变异,建立相应的数学模型,再根据这些数据精确地推导待测量。在数据处理过程中,回归分析的主要目标是利用样本数据对各个参数进行科学估算,然后建立相应的数据模型,从而对参数数据进行判定和预测。同时,数据的处理与分析还要求有一定的数学理论基础知识。例如,数学测度论的知识,就是将两个或更多的单调度量组合起来,从而建立一个新的单调度量,使研究者在对其进行降维时,能够适当地使用度量理论。

### 5.3 数学在数据挖掘中的应用

数据挖掘是大数据研究的重要内容,表面上看起来杂乱无章,但要从海量数据中获得有用的、有用的信息,就需要进行数据挖掘。数据挖掘是一个应用、工程、集合、交叉的学科。数据挖掘的步骤主要分为确立挖掘目的、数据准备、数学建模、模型评估、模型应用,每一步都需要数学知识的参与。因此,数据挖掘是一种具有实际意义、集合性、工程性和交叉性特点的技术。在数据挖掘中,数学知识是不可替代的。在数据挖掘工作中,通常要用到神经网络、依存分析、聚类分析、决策树等数学方法。聚类分析作为一种常用的数学工具,已经在心理学、医学和市场营销等诸多领域得到了广泛的应用,基于“物理类取”的聚类分析方法,根据一定的准则,把关联程度高的事物归为一类,并尽可能地扩大不同类别间的差别,实现对数据集合或重要指标的科学划分。

换言之，聚类分析本质上就是把一个物体的私有集合划分为几个相似或相似的群体，这就是“物理类聚”。聚类分析是一种非常重要的数学工具。

其中，如图1所示为数据挖掘流程图。

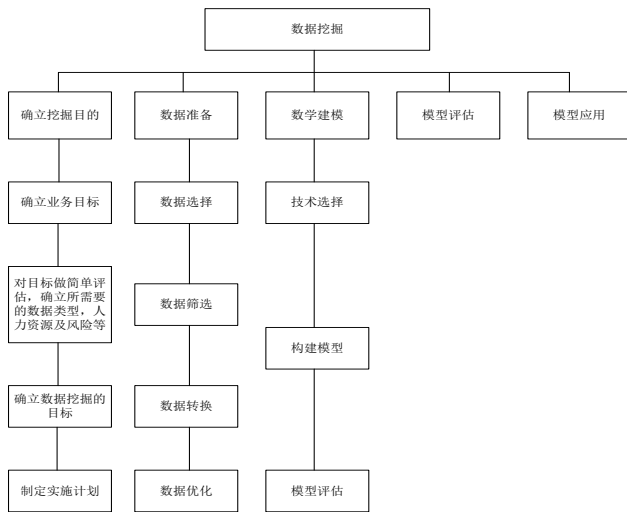


图1 数据挖掘流程图

第一，灰色关联分析法的运用。灰关联分析方法也是建立在“灰关联度”的基础上，用以度量各要素之间的相关性，是衡量各种因素间相互关系的一种数学方法，在数据挖掘过程中，在处理小样本、不完备或不完整情形时，常采用灰关联度分析方法。在灰关联分析中，灰关联系统的具体表示可以用  $S=(X, R)$  来表示。其中，字母  $X$  表示影响因素的集合，而字母  $R$  表示因素之间的趋向相关图。在数据挖掘工作中，学者们常用的方法是运用数理灰色关联分析方法，通过对不同几条曲线间的几个形态进行分析和对比，从而对数据进行分析处理。采用灰关联度分析法，对不同几何图形进行对比，判定各几何图形之间的相关程度，几何图形越是相似，说明相关程度就越高，反之就越小。在对样本的数量比较少，或者是有不完全的样本的情况下，使用灰色关联分析进行数据挖掘。例如，由于历史原因造成的数据样本不足，或者是因为样本的快速更新，造成了样本数据的不一致性。

第二，目标函数模糊聚类法的运用。该方法在对数据进行归一化和校正后，通过构造与之对应的模糊矩阵，采用直接聚类、模糊等价矩阵等手段，目标函数模糊聚类算法适合处理高维、高效、具有可扩展性的大数据集。此外，还可以采用最大法、网状法等多种方法对其进行聚类。由于其有效性、可扩展性、可处理维度等方面的优势，被广泛地用于

大多数的数据挖掘过程，然而，在实际应用中，模糊聚类技术在图像处理，数据分析，数据挖掘等领域得到了广泛的应用。例如，当处理图形时，就采用该方法来聚类。并且，由于其相对成熟的应用，使得其成为一种高效、可靠的解决方法。

第三，区间值算法的运用。在大数据的处理和挖掘中，区间算法可以转换某些“比较型”数据，也可以用来分析某一特定值的数据。该方法是一种较为普遍的方法，它能够挖掘和分析各种系统的习性。主要表现为对不完整的系统信息的挖掘与分析。大数据挖掘技术是一种基于区间值的数据挖掘技术，它可以对某些“比较型”的数据进行转换，也可以将其应用到特定的数值范围内。对于某些“比较型”的数据，或者具有一定的数值范围，可以通过区间值算法来分析。在数据挖掘中，区间值算法是一种常用的数据挖掘技术，其主要应用于不完备系统的挖掘和分析。在挖掘过程中，利用定量聚类、区间聚类等技术来解决不完整的系统信息。区间值的确定，可以通过有经验的专家或统计的方法来确定。在数据挖掘中，数量和区间聚类是应用最多的一种方法，它的合理应用，有助于科研人员对不完备的系统信息进行真实、快速和准确的科学分析。在确定区间的时候，研究者可以采用高级的统计学方法来科学地决定，也可以通过彼此之间的探讨和分析，并根据自己的工作经验来确定区间。

## 6 结语

综上所述，在大数据时代，数学知识和大数据之间存在着密切的关系，并且对大数据的实际应用起到了非常重要的作用，尤其是在数据处理和数据挖掘两个领域。数学是大数据挖掘的支持基础，科研工作者要善于将多种数学方面的理论知识与方法应用到大数据挖掘中，要想将大数据的应用真正做好，就必须将数学的应用理论牢固地把握住，从对数据的前期采集开始进行深度剖析，最后得出结论。因此，可以看到，在大数据中，运用数学知识是保证大数据有效性和实用性的重要保证。

## 参考文献

- [1] 王珂. 大数据时代数学知识的全面应用探索[J]. 佳木斯职业学院学报, 2020, 36(1): 199-201.
- [2] 刘英男. 浅析数学与应用数学在大数据中的应用[J]. 科学与财富, 2020(21): 147.
- [3] 姚磊, 袁维佳, 熊焕, 等. 大数据应用思考与探索[J]. 中国航天, 2023(增刊): 56-59.
- [4] 徐济科. 大数据时代应用型本科数学类课程教学改革研究[J]. 科教导刊, 2020(14): 17-18.