

大数据挖掘分析

程陈

(中国地震局第二监测中心, 西安 710054)

摘要: 同“互联网”、“云计算技术”相比, 大数据的发展更为迅速, 它被用于社会的各个方面, 例如交通、企业管理、安全监测以及我们的生活。相关人员声明, 全球的数据量将在 2020 年达到 35ZB, 因而, 大数据的涵义以及数据挖掘带给社会经济的挑战成为了人们关注的要点。

关键词: 大数据; 数据挖掘; 挑战

中图分类号: TP311

文献标识码: A

DOI: 10.3969/j.issn.1003-6970.2014.04.038

本文著录格式: [1] 程陈. 大数据挖掘分析 [J]. 软件, 2014,35(4): 130-131

Analysis of data mining

CHENG Chen

(China Earthquake Administration second monitoring center, Xi'an 710054, China)

【Abstract】 With the “Internet”, “cloud computer technology”, development of large data more quickly, it is used in all aspects of society, such as transportation, enterprise management, security monitoring and our life. Relevant staff statement, global data volume will reach 35ZB in 2020, therefore, the meaning of data and data mining to bring social and economic challenges become points of attention.

【Keywords】 large data; data mining; challenge

0 引言

人类自 2010 年便跨入了大数据时代, 许多同我们生活息息相关的仪器都步入了智能化。而在数据中生活的我们在制造数据的同时也在通过数据加快经济发展速度, 提高社会文明。因此, 数据的战略意义已经同人力资源、自然资源相同。所以, 正确理解大数据的涵义及其发展前景将对社会发展具有极大的裨益。

1 大数据的概念

大数据并非产品也不是技术, 只是数字化时代的一种现象。而且, 以战略眼光分析大数据可得出, 大数据不单指其庞大的数据量, 还包括如何用专业化方式处理这些数据。关于大数据的概念有不同的定义, 基维百科将大数据定义为规模巨大到无法用目前的软件工具处理的资料量, 在相应时间内完成摄取、管理、处理后利于企业完成经营目标的资讯; 研究机构将大数据定义为必须用新型的处理模式才能更好地洞察、决策并优化流程的巨量、多样化且高速增长的信息资产; 而麦肯锡的定义为大数据是一种数据集合, 且在一定时间内无法利用传统的数据软件对其进行数据采集、存数、管理及分析^[1]。无论是那种定义都有一定的狭义性, 依据广义的定义, 大数据不但包括大数据技术、应用还包括大数据科学以及大数据工程。在大数据时代, 如何深层次开发大数据并提供相关服务能力将成为竞争的关键。

2 大数据的特征

数据无处不在, 且大数据能够超越“物联网”、“云计算”开创自己的时代, 这与其自身的特征密不可分。第一, 种类多。随着社会进步, 传感器的种类与日增多且社交网络、智能设备被更多人认可, 数据类型也相对增多。目前, 数据除去传统的关系数据还包括视频、网页、文档、音频以及邮件等尚未处理、不具备结构模式或者半结构模式的数据^[2]。第二, 高速流动。传统的数据流动速度是指对数据摄取、存数及分析具有价值信息的速度。然而, 大数据因为其数据量的巨大, 快速变动的数据形成数据流的特点, 传统的处理方式已经无法处理这样高速流动的数据, 进而数据处理已经由 TB 级上升到 PB 级。第三, 数据量巨大。一般, 大数据指的是超过 10TB 规模的数据量。而导致这种结果的原因有三, 一是我们为能够了解更多事物而不断使用各类仪器, 并存储这些事物部分或者全部的数据; 二是集成电路的成本降低使得很多仪器智能化发展, 自行存储数据; 三是我们为能够随时传递信息而使用各类的通信工具, 尤其是机器到机器传递方式的诞生更是导致了交流数据激增。第四, 低价值密度。虽然数据量不断增长, 但是这些数据中具有意义的信息却没有以相应比例进行增长, 这会加大我们获得需要信息的难度。例如, “4V”不仅代表数据量巨大, 同时也代表数据分析将会更加复杂, 更难达到要求的效率。

作者简介: 程陈 (1987-), 男, 助理工程师, 主要研究大数据挖掘分析方向

3 数据挖掘的功能及应用

数据挖掘是大数据时代的关键技术，是指从非完整的、海量的、有噪音的、模糊且随机的数据中挖掘隐含在内且人们未提前得知的有用信息的过程。一般，数据挖掘的功能有两类，即描述和预测^[3]。描述性挖掘用于展现集体数据的一般特性，而预测性挖掘用于推算处理数据，完成预测目的。数据玩具功能同目标数据的类型有关，有些功能适用于不同类型的数据，有些功能则只适用于某种特定数据。数据挖掘功能能够让人得知未知信息，提升数据价值，从而应用到了不同领域。

3.1 数据挖掘在金融业的应用

金融业因其业务特性需要搜集大量数据。一般，这些数据比较真实、完整，分析之后便能够发觉其中隐含的模式和特点，掌握目标客户或者组织的信息，进而洞察到金融市场的动态。数据挖掘在金融业中主要被用于市场分析、预测、数据清理、分类账号、评估信誉等。

3.2 数据挖掘在市场业的应用

市场业主要通过数据挖掘技术分析市场定位以及消费者的行为，通过集中分析消费者的信息从而得出某些消费群体的规律、需求，并根据这些信息确定相应的市场定位和营销计划。同传统营销模式相比，大数据挖掘功能能够降低企业的成本，使其获得更高的利益。

3.3 大数据挖掘在医学中的应用

人类的某些疾病同人类的基因相关，有些是单个基因造成的，有些则是多种基因共同影响的。为找到治疗这些疾病的方法，必须进行基因研究，而基因研究是建立在区分编码序列和非编码序列上，区分这两种序列必须经过大量实验以及演算。其实，区分这两种序列就相当于分类，而分类问题在数据挖掘中已经有了一定的发展。因此，可将数据挖掘分类问题的某些研究运用到基因序列分类上，有利于人类医疗事业的发展。

4 大数据挖掘的发展前景

数据库因为大数据时代的来临而突破了传统模式，转变成可进行数据变换、连接及共享的数据库^[4]。因此，我们可借助大数据挖掘技术利用曾无论利用的多类数据，抓住机会让数据为自己服务。可见，大数据挖掘基于其数据的特征和数据挖掘的作用而拥有良好的发展趋势。

4.1 大数据挖掘将成为信息安全发展的契机

如今，互联网、云计算等新兴技术提升了数据的价值，逐

步成为了社会发展的重要力量。然而，随处可见的数据降低了自身信息的安全性。一般，大量数据会存储于云端，无法集中管理，从而无法单独管理用户信息并分别其是否合法，这就导致非法入侵或者窃取、篡改数据信息的危险性提高。所以，对于信息安全领域而言如何保证信息安全成为了首要问题。对此，大数据领域研发出了各种为信息安全服务的技术和产品，保证各大数据产业链的数据安全。所以，大数据挖掘对信息安全发展而言具有极大的意义。

4.2 大数据挖掘将成为企业及教育机构转折点

随着大数据挖掘技术在企业管理中的应用以及其带来的经济效益，企业若想在新的时代中继续保持自身利益，就必须重新制定管理模式，将大数据挖掘运用到企业管理当中，这样才能紧随时代发展的脚步。同时，大数据时代的到来也给人才资源带来一定的挑战。因为，企业必须拥有三类人才才能够从大数据中获得关键信息，即分析人才、管理人才以及技术型人才。相关人才的匮乏和需求对教育机构而言也是一项冲击，如何调整自身教育内容，为社会提供相应的人才才是教育机构面临的问题。

4.3 大数据挖掘将成为创造价值的核心

同传统数据相比，大数据将数据挖掘以及应用作为数据的中心，这对企业商业模式有着极大的影响。麦肯锡曾预测，欧洲政府以及美国医疗业可通过运用大数据而分别节省 1000 亿欧元和 3000 亿美元，此外，大数据中潜在个人信息价值高达 6000 多亿美元，可见大数据挖掘及应用能够从不同方面为社会创造价值。

5 结束语

大数据发展的时间虽然短却打开了一个新的时代，引领全球进入新的竞争。我国也应当洞悉大数据时代发展的核心，及时调整发展政策，开展人才培养、研究大数据挖掘技术、信息安全系统研究等工作，通过对大数据以及数据挖掘的研究及合理利用让我国经济快速发展。

参考文献

- [1] 葛春燕. 数据挖掘技术在保险公司客户评估中的应用研究[J]. 软件, 2013,34(1): 116-118
- [2] 杨泽民. 数据挖掘中关联规则算法的研究[J]. 软件, 2013,34(11): 71-72
- [3] 韩少锋, 陈立潮. 数据挖掘技术及应用综述[J]. 机械管理开发, 2006(02)
- [4] 钟瑛, 张恒山. 大数据的缘起、冲击及其应对[J]. 现代传播(中国传媒大学学报), 2013(07)