

# 基于神经网络算法的大数据分析方法研究

方 芳

(天津国土资源和房屋职业学院, 天津 300270)

**摘 要:** 伴随着我国科学技术的不断进步,我国已经步入了互联网信息时代。大数据作为互联网时代最为核心的组成部分,具有比较广泛的应用价值。基于此,针对神经网络算法的应用缺陷,本文主要是对大数据和神经网络算法之间的有机结合进行了设计,进而构建了一种安全高效的数据分析平台。神经网络算法具有广泛的理论基础,其中神经网络结构模拟、网络通信模型、记忆模型、学习算法等都是比较重要的概念,这也就表明基于神经网络算法的大数据分析具有良好的性能和应用前景。

**关键词:** 神经网络; 算法; 大数据; 分析方法

**中图分类号:** TP391.4 **文献标识码:** A

## Research on the Big Data Analysis Method Based on Neural Network Algorithm

FANG Fang

(Tianjin Land Resources and Housing, Career Academy, Tianjin 300270, China)

**Abstract:** With the continuous progress of science and technology, China has entered the era of Internet information. As the core component of the Internet era, big data has a wide application value. Based on this, aiming at the application defects of neural network algorithm, this paper mainly designs the organic combination of big data and neural network algorithm, and then constructs a secure and efficient data analysis platform. Neural network algorithm has a wide range of theoretical basis, in which neural network structure simulation, network communication model, memory model, learning algorithm are all important concepts, which shows that big data analysis based on neural network algorithm has good performance and application prospects.

**Keywords:** neural network; algorithm; big data; analysis method

## 1 引言(Introduction)

大数据时代是一个全新的时代,其表现出诸多的新特征,尤其是大数据分析方面,这为社会生产生活创造了巨大的价值。大数据分析贯穿于信息创造的整个过程,存在于信息创造的整个生命周期中。对于信息数据的应用,这主要可以涉及五个方面,“数据准备、数据存储、数据运算、数据分析、数据共享”,其中,数据分析是中坚环节,这也是数据信息创造价值的主要体现。从我的工作出发,神经网络算法在大数据分析中的作用比较突出。下面我就神经网络算法的大数据分析方法展开全面研究,以供业内人士的借鉴与参考。

## 2 大数据的相关概述(An overview of big data)

### 2.1 大数据的概念

大数据是基于互联网信息时代高度发展的产物,这也是历史发展的必然产物。大数据的概念于1997年诞生。一般来讲,大数据就是指具有庞大体系的数据集合,这是单体数据

无法比较的。我们完全可以从大数据的四个主要特点入手,对其基本概念进行全面的了解<sup>[1]</sup>。首先,海量性,这是大数据最基本的特征,也就是说大数据不是部分数据的集合,而是全样本所有数据的集合,具有整体代表性;其次,价值整合性强,大数据是依靠整体数据而产生价值,仅仅依靠单条数据并不能产生有效的利用价值,这也就表明大数据的个体数据之间存在着比较强的关联性;再次,广泛性,所谓大数据的广泛性主要是针对其来源而言,大数据具有多元化的来源渠道、差异化的产生途径,并且大数据多是处于非结构化的状态,这就使其无法轻易的进行分类、整合;最后,指数生长性,这主要是指大数据所包含的数据量会随着产出而发生指数的相关变化,其最大流量能够达到TB级<sup>[2]</sup>。

### 2.2 大数据分析核心技术支持

#### 2.2.1 数据平台

在大数据的技术支撑中,数据平台是最基础的部分,它能够对相关的数据信息进行搜集、整合、储存等,并且能够

为下一步的数据分析提供基础。我们在上文中已经提到,对于大数据信息的收集需要立足于全样本数据容量,而不是局部、部分的样本容量,这就在一定程度上减少数据误差,提高数据分析的精准度。同时,我们在完成数据收集环节后,还要对数据进行相关的分类、标记,其中涉及标记的数据都是具有比较特殊的特征,这也是其研究价值的体现<sup>[3,4]</sup>。大数据存储于传统数据存储存在着较大差异,这也是推进数据存储进程的重要途径。传统意义层面,我们主要是对原始数据进行存储,这就必然会消耗比较多的资源,增加了数据存储的成本,再就是不能够使复杂的数据清晰化。但是,大数据平台就完全解决了这一缺陷,“只存储知识”就是其最重要的特点,有效的提高了大数据平台的工作效率。

### 2.2.2 分析平台

分析平台是大数据分析中最为核心的环节,这也是数据价值发挥的主要环节。数据分析平台具有其自身的特定任务,主要是对相关大数据进行特定化的计算和分析。大数据分析平台需要建立在计算平台上,这就包括计算资源的建设和分析算法的设计<sup>[5]</sup>。所谓计算资源的建设,就是在CPU方面有所创新,实现从小型化的CPU逐向高性能CPU的转变,最终能够有效促进大数据分析的推进。一般来讲,分布式计算框架主要包括MapReduce、Parameter Server等技术。大数据分析主要可以划分为两种类别,第一种是基于人工经验而建立的分析模型,但是这种模式已经不符合时代发展的要求,存在着比较多的应用缺陷,尤其是在人力资源、物力资源的消耗方面,一直都是处于比较高的水平。同时,这种分析模型的知识无法迁移,对数据有所限制。第二种是基于人工智能的数据分析手段,其中神经网络算法就是最为核心的技术,这种分析模型已经在实际生产中得到了广泛的应用,为社会发展提供了强劲动力。

### 2.2.3 展示平台

在实际的工作过程中,我们在完成大数据分析环节之后,大数据主要是以知识形态而存在,这就需要特定的展示平台,为其推广提供基本条件。所以,展示平台的任务也就是在完成大数据分析后的产品推广。再就是,大数据在经过相对应的分析环节后,其主要是以直接知识和间接知识这两种形态而存在。所谓大数据直接知识就是指数据知识的具象化,这也是比较系统的分析结果,比如我们在对基因大数据进行相关分析后,能够对特定的基因规律进行挖掘。那么,所谓间接知识就是指可计算的分析模型,它完全可以应用到数据知识的获取过程中,最为典型的的就是“举一反三”,比如我们在掌握一种技巧后,就可以将这种技巧进行分散化操作,应用到更多的事物中<sup>[6]</sup>。总之,大数据展示平台就是要将知识进行展示和传递,如何明确区分间接知识和直接知识,并将其最为科学的展示给受众,这是大数据分析的重点。

## 3 神经网络算法的相关概述(An overview of neural network algorithm)

### 3.1 神经网络算法模型

顾名思义,神经网络算法就是与生物神经系统类似的模型,这是一种模拟大脑分析机制的算法。但是,神经网络算法仍然是以仿生计算的形态而存在,并没有达到生物神经的高度发展性。基于现有的科学技术,神经网络算法涉及诸多的研究方面,比如神经网络结构模拟、网络通信模型、记忆模型、学习算法等。我们现有的神经网络算法模型主要有前

馈神经网络、回复式神经网络、时序记忆神经网络这三个部分。这三种具体的算法模型都有其自身的特点,其中以前馈神经网络最为高效,它具有比较好的价值。同时,神经网络算法模型最为鲜明的特征就是其自身的学习能力,它能够在智能语音识别、生物医疗等行业中得到应用。前馈神经网络分为多层结构,每层由多组神经元构成,信息沿着前馈层输入,单向流动传输。

### 3.2 无限深度神经网络

深度神经网络就是指具有层级限制,表现出一定层级深度的前馈式神经网络,这种神经网络没有“时间参数”的限制,比较适用于静态数据的相关处理工作。无限深度神经网络就是着重凸显“无限深度”这一概念,就是指反馈结构的回复式互联,这也就能够在动态数据的处理方面具有重要意义。总之,神经网络算法在大数据分析中的应用需要充分发挥其固有的特点,利用好无限深度神经网络的作用,为其提供基本的技术支撑。

## 4 基于神经网络算法的大数据分析的应用实例(Application example of big data analysis based on neural network algorithm)

### 4.1 在计算机视觉中的实际应用

基于人工智能技术的快速发展,深度学习已经成为计算机视觉研究领域的标准配置,这也就成为当下人们的研究方向。深度学习自然就离不开神经网络算法,传统意义层面的计算机视觉研究主要是以图像特点为基本依据,并且将其与人工设计进行差异化区分,其中,最为关键的部分就是边缘化、颜色化、尺度不变化等。我们可以在这些不同的特征基础上,实现对计算机视觉任务的划分,对于视觉图像方面,我们可以尽可能的将其实现。同时,传统的计算机视觉需要借助人工元素,所涉及的任务也是出于比较低端、直观的状态,这也就不利于图像的全方位表达。而神经网络算法则是对其进行了改进,它可以充分集合大量的数据信息,将海量的图像进行集合,完全自动地进行图像特征的记忆。比如,在2012年以前,我们主要是利用“特征+支持向量机”的图片分类模式,在2012年后,卷积神经网络模型则成为计算机视觉领域的关键支撑。同时,卷积神经网络模型已经在人脸识别中得到了广泛的应用,这也就催生出了以深度学习为核心的人脸识别算法模型。人脸识别在实际生活中有诸多的应用场景,我们可以将其划分为三个层面,第一层面,1:1场景,这广泛应用于火车站、机场的安检工作中,利用旅客的身份证与其进行智能化比对;第二层面,1:N场景,公安部门在对犯罪嫌疑人锁定过程中,就会依据相关信息进行智能比对,这就需要从大量的人群中进行筛选;第三层面,大数据分析场景,这会在多个方面有所体现,比如表情分类等。总之,在神经网络算法的大数据分析中,各层数据形成了不同的层次划分,这就为数据分析提供了系统化的分析流程。

### 4.2 在智能语音识别方面中的应用

总结一下智能语音识别的发展现状,dnn、rnn/lstm、cnn是比较主流的发展方向,这也代表着智能语音识别的发展趋势。智能语音识别是神经网络算法的第一个技术突破,这也为神经网络算法在大数据分析中的应用提供条件。在神经网络算法与语音识别结合之前,我们主要是利用声学的低层特征,将高斯混合模型进行特征的相关提取,再就是利用隐马尔可夫模型对序列进行转移,最后在形成成熟的建模状态<sup>[7]</sup>。

经过多年的发展,神经网络算法与语音识别已经进行了有机结合,实现了两者之间的完美契合,全面提升了语音识别文字的正确率,为其实际应用奠定了基础。比如,灵云语音识别系统在公检法领域的智能语音转写中得到了广泛应用。基于灵云语音识别系统的识别准确率,灵云语音识别系统已经在上海、山东、浙江、广东、江西、福建等省市的30余家法院进行了应用,这便是神经网络算法在大数据分析中的良好应用案例。在实际的审讯过程中,灵云智能庭审语音转写系统能够将审判员、原被告等角色的发言进行实时转写,以文字的形式实时展现到书记员的电脑上。书记员可对转写内容进行实时校对、编辑,并具备丰富的功能。全面提升了法院案件办理的效率和质量,为推动智能法院建设提供了条件。

### 4.3 在医学医疗领域中的应用

医学医疗领域一直是人工智能的研究前沿,这也是大数据分析在其领域应用的重要基础。所以,为了进一步提升医学医疗行业的发展,我们必须加快推进神经网络算法在其中的应用,全面改善现有的医疗环境。如何利用好大数据充分解决医学医疗领域中的难题,提高现有的医疗水平,这是我们需要关注的焦点。比如,2018年1月,外科研究中心WEISS、UCL医疗影像计算中心、HIG等机构共同开源了NiftyNet,这也就是指卷积神经网络的医疗影像分析平台,这是对神经网络算法在医学医疗中的应用典例,并且其能够为研究社区提供开放机制,最终能够保障临床能够适应新型的医疗影像研究成果。NiftyNet具有其自身的优势,它不仅能够辅助研究人员进行医疗方案的快速开发,还能够在语义分割、图像回归等方面有所突破。

## 5 神经网络算法在大数据分析中的核心问题(Core problems of neural network algorithm in big data analysis)

### 5.1 数据展示

在大数据分析的过程中,我们必须为其提供一个良好的展示平台,也就是说,良好的数据展示平台将会对数据展示提供关键支持。所以,我们所提到的第一个核心问题就是数据的展示。我们都知道,数据原始空间是一个处于稠密状态的空间,在其内部包含诸多的原始细节。尤其是对于各种模态的数据而言,其存在状态更是复杂多变的,其中主要是数据噪声和数据缺失这两个层面。学习理论中的Cover定理旧引也是指这一方面的问题,数据在进行变换的过程中,一旦处于高维空间的状态,其可分性自然就会演变成更为重要的方面。但是,保证高维空间中的稀疏表达需要具有一定的技术性,这也是提升数据空间结构协调性的重要方面,这是我们在后期需要注重开发的方面<sup>[8]</sup>。

### 5.2 数据存储

在认知科学的研究领域,我们认为大脑主要是对知识进行客观的记忆处理,而不能对其进行存储。所以,神经网络算法也是与之相一致的,其主要是对数据知识进行存储,而不能对原始数据进行存储。在此基础上,这就会涉及大数据分析的第二个层面的问题,即数据的存储。但是,对于神经网络算法的相关记忆机制、记忆形式都是处于未知状态的,所以我们在这一方面也是需要重点研究的。一般来讲,基于神经网络算法的大数据分析方法是将突触进行连接的,最终会形成特定的网络记忆范式,这也就是大数据分析的大

致记忆过程。

### 5.3 数据预测

数据预测则是大数据分析的第三个层面的核心问题,这也就是对神经网络算法应用趋势的测定和展望。所以,在后期的方法开发中,加强对认知计算原理的相关应用,进一步开发基于神经网络算法的大数据预测,并且还要不断投入对该研究方面的成本投入,从而为大数据分析提供最为系统的软硬件支撑。

### 5.4 安全保障

互联网信息时代具有开放性,这自然会使得大数据分析的安全性受到威胁。所以,为了进一步提升神经网络在大数据分析中的实际应用,我们必须为其提供最为坚实的安全保障,最终促进数据挖掘价值发挥<sup>[9]</sup>。基于此,我们需要从两个方面着手,第一方面,我们要保障数据分析过程中的精准化,为社会生产生活提供最为精准的服务;第二方面,我们还要不断提升数据分析的安全性,确保神经网络数据挖掘能够安全稳定的状态。

## 6 结论(Conclusion)

综上所述,基于神经网络算法的大数据分析是一项系统工程,它会涉及诸多的研究方面,所以,我们必须采取全方位的发展策略,对其实际应用、算法模型、技术改进等进行研究,最终能够推动大数据分析技术的发展和革新。本文主要是对大数据和神经网络算法概述、实际应用案例、技术改进要点等进行研究,为后期大数据分析方法的开发提供理论依据。

## 参考文献(References)

- [1] A genetic-algorithm-based neural network approach for EDXRF analysis[J].核技术(英文版),2014(3):18-21.
- [2] Parameters Optimization of Plasma Hardening Process Using Genetic Algorithm and Neural Network[J].钢铁研究学报(英文版),2011(12):57-64.
- [3] Accelerating the Construction of Neural Network Potential Energy Surfaces:A Fast Hybrid Training Algorithm[J].化学物理学报(英文版),2017(6):727-734.
- [4] 栾玉飞,白雅楠,魏鹏.大数据环境下网络非法入侵检测系统设计[J].计算机测量与控制,2018(1):194-197.
- [5] 王春香,张勇,梁亮,等.BP和RBF神经网络对复杂型面零件点云漏洞的修补应用[J].组合机床与自动化加工技术,2018(3):118-120.
- [6] 赵广振,张翠肖,武辉林,等.基于主成分分析和概率神经网络的入侵检测方法[J].石家庄铁道大学学报(自然科学版),2018(1):91-95.
- [7] 项寅.基于改进神经网络的恐怖袭击风险预警系统[J].灾害学,2018(1):183-189.
- [8] 胡硕,赵银妹,孙翔.基于卷积神经网络的目标跟踪算法综述[J].高技术通讯,2018(3):207-213.
- [9] 郝怡然,盛益强,王劲林,等.基于递归神经网络的网络安全事件预测[J].网络新媒体技术,2017(5):54-58.

## 作者简介:

方芳(1986-),女,硕士,讲师.研究领域:软件开发,图形图像,网页设计.