

智能大数据 与深度学习

朱定局◎著

Intelligent Big Data and Deep Learning

非外借



中国工信出版集团



电子工业出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY
<http://www.phei.com.cn>

智能大数据与深度学习

朱定局 著



電子工業出版社

Publishing House of Electronics Industry

内 容 简 介

本书是一本原创性的学术专著，分为两部分，第一部分介绍智能大数据与深度学习的理论方法；第二部分介绍智能大数据与深度学习的应用实践。本书的原创性在于提出、研究并给出了 12 种智能大数据与深度学习的新理论、新方法和新应用。

本书可作为高等院校、科研院所和企事业单位的科研及教学用书。

未经许可，不得以任何方式复制或抄袭本书之部分或全部内容。

版权所有，侵权必究。

图书在版编目 (CIP) 数据

智能大数据与深度学习 / 朱定局著. —北京：电子工业出版社，2018.10

ISBN 978-7-121-34286-8

I. ①智… II. ①朱… III. ①数据处理—高等学校—教材 IV. ①TP274

中国版本图书馆 CIP 数据核字 (2018) 第 111033 号

策划编辑：刘 瑀

责任编辑：章海涛

印 刷：北京盛通印刷股份有限公司

装 订：北京盛通印刷股份有限公司

出版发行：电子工业出版社

北京市海淀区万寿路 173 信箱 邮编：100036

开 本：720×1000 1/16 印张：12.5 字数：160 千字

版 次：2018 年 10 月第 1 版

印 次：2018 年 10 月第 1 次印刷

定 价：69.00 元

凡所购买电子工业出版社图书有缺损问题，请向购买书店调换。若书店售缺，请与本社发行部联系，联系及邮购电话：(010)88254888，88258888。

质量投诉请发邮件至 zlts@phei.com.cn，盗版侵权举报请发邮件至 dbqq@phei.com.cn。

本书咨询联系方式：liuy01@phei.com.cn。

前 言

因为 AlphaGo 战胜了人类围棋冠军，深度学习一举成名，再加上 BAT、谷歌等国内外知名公司的推崇，深度学习的热浪一波高过一波。深度学习之所以能使神经网络重新受到学者的青睐，主要是因为它用非监督式或半监督式的特征学习和分层特征提取高效算法，替代了手工获取特征，从而使过去被废弃的大量非标签数据得到利用。更重要的是，这也使深度学习的深度得到了极大增加，从而提高了深度学习的处理能力。深度学习虽然是一个初生儿，但地位极高，因为其前身是具有悠久历史的神经网络。20 年前，在我上大学期间，就有神经网络的课程。由于深度学习发展时间较短，所以不论在理论上还是在应用上，它都存在很多盲点和缺陷，都需要开拓、创新、完善和升级。

深度学习的春雨过后，无人驾驶、图像识别等人工智能相关应用，如雨后春笋般层出不穷。随着应用的深度发展，其理论研究也在如火如荼地进行。很多企业和高校已经成立了深度学习的研究中心或小组，专门研究深度学习的相关理论和应用，同时，深度学习的培训课程也在国内外掀起了热潮，但其困境是，关于深度学习理论和应用的相关著作仍然比较缺乏。

在深度学习之前，大数据已经火热了几年，深度学习的出现为大数据智能处理提供了一个新的出路。深度学习只有基于大数据才能发挥威力，才能达到很高的准确度。而大数据在应用了深度学习之后，所能达到的处理效果也是传统算法无法企及的。可以说，两者相得益彰、相互促进、互利共赢。

本书可以作为学习和研究深度学习的参考用书，其中，理论部分可以用于完善现有深度学习方法的不足，为理论创新奠定基础；应用部分可以为企业提供更多深度学习应用的新思路和新方案，进而使深度学习在实践中产生更大的价值。

本书的所有章节都是作者原创性的研究成果。本书内容的原创性在于：首次提出并研究给出了两种增强型深度学习神经网络方法及应用，包括基于顶层生成深度学习的数据转换方法和系统，基于双向深度学习的数据对应关系判断、生成方法和系统；首次提出并研究给出了两种高效智能型深度学习神经网络方法及应用，包括精简输入的深度学习方法和系统，基于数据分割的深度学习方法和系统；首次提出并研究给出了两种动态智能型深度学习神经网络方法及应用，包括基于有向图的深度学习构建方法和系统，深度学习神经网络训练方法及层数调整方法和系统；首次提出并研究给出了两种深度学习的大数据智能计算应用，包括基于大数据与深度学习的信息隐藏、提取方法和系统，基于大数据与深度学习的婚恋对象匹配数据处理方法和系统；首次提出并研究给出了两种深度学习的大数据智能诊断应用，包括基于深度学习的面向自动诊断的医疗数据处理方法和系统，基于双深度学习的数据处理方法和疾病诊断装置；首次提出并研究给出了两种深度学习的大数据智能预测应用，包括基于大数据与深度学习的气象预报方法和系统，基于大数据与深度学习的用户数据处理方法和系统。

本书选取的是人工智能的前沿领域和热点领域——深度学习，并将其与各个应用领域进行了交叉创新。在交叉创新的过程中，发现和改进了现有深度学习理论的不足，同时拓展了现有深度学习应用的范围，进而实现了深度学习技术从理论到应用的创新。

本书的特色在于，以深度学习应用牵引深度学习理论的创新，同时以深度学习理论的创新促进深度学习应用的创新。本书在深度学习与各行各业之

间进行了大胆的交叉创新，从而使深度学习技术更智能、更普适、更精准、更强大，加速深度学习智能应用的发展。当然，深入后必能浅出，读者也能通过本书，学习到深度学习的新理论和新应用，加深对现有深度学习技术的理解。

书中所述的研究得到了国家级新工科研究与实践项目(粤教高函〔2018〕17号)、国家社会科学基金重大项目(14ZDB101)、国家自然科学基金重点项目(41630635)、教育部-腾讯公司产学合作协同育人项目(201602001001)、广东高校重大科研项目(粤教科函〔2018〕64)、广东省新工科研究与实践项目(粤教高函〔2017〕118号)、广东省高等教育教学研究和改革重点项目(粤教高函〔2016〕236号)、广东省学位与研究生教育改革研究重点项目(粤教研函〔2016〕39号)、广东省联合培养研究生示范基地(粤教研函〔2016〕39号)的支持。

华南师范大学

朱定局

目 录

第一部分 理论方法

第 1 章	增强型深度学习神经网络	2
1.1	基于顶层生成深度学习的数据转换方法和系统	2
1.1.1	研究现状	3
1.1.2	基于顶层生成深度学习的数据转换方法	4
1.1.3	基于顶层生成深度学习的数据转换系统	7
1.2	基于双向深度学习的数据对应关系判断、生成方法和系统	9
1.2.1	研究现状	10
1.2.2	基于双向深度学习的数据对应关系判断、生成方法	11
1.2.3	基于双向深度学习的数据对应关系判断、生成系统	28
第 2 章	高效智能型深度学习神经网络	37
2.1	精简输入的深度学习方法和系统	37
2.1.1	研究现状	37
2.1.2	精简输入的深度学习方法	38
2.1.3	精简输入的深度学习系统	48
2.2	基于数据分割的深度学习方法和系统	50
2.2.1	研究现状	50

2.2.2	基于数据分割的深度学习方法	51
2.2.3	基于数据分割的深度学习系统	57
第 3 章	动态智能型深度学习神经网络	60
3.1	基于有向图的深度学习构建方法和系统	60
3.1.1	研究现状	60
3.1.2	基于有向图的深度学习构建方法	61
3.1.3	基于有向图的深度学习构建系统	70
3.2	深度学习神经网络训练及层数调整方法和系统	72
3.2.1	研究现状	72
3.2.2	深度学习神经网络训练及层数调整方法	74
3.2.3	深度学习神经网络训练及层数调整系统	80

第二部分 应用实践

第 4 章	深度学习的大数据智能计算应用	86
4.1	基于大数据与深度学习的信息隐藏、提取方法和系统	86
4.1.1	研究现状	86
4.1.2	基于大数据与深度学习的信息隐藏、提取方法	89
4.1.3	基于大数据与深度学习的信息隐藏、提取系统	102
4.2	基于大数据与深度学习的婚恋对象匹配数据处理方法和系统	108
4.2.1	研究现状	108
4.2.2	基于大数据与深度学习的婚恋对象匹配数据处理方法	109
4.2.3	基于大数据与深度学习的婚恋对象匹配数据处理系统	120
第 5 章	深度学习的大数据智能诊断应用	123
5.1	基于深度学习的面向自动诊断的医疗数据处理方法和系统	123

5.1.1	研究现状	124
5.1.2	基于深度学习的面向自动诊断的医疗数据处理方法	124
5.1.3	基于深度学习的面向自动诊断的医疗数据处理系统	126
5.2	基于双深度学习的数据处理方法和疾病诊断装置	132
5.2.1	研究现状	132
5.2.2	基于双深度学习的数据处理方法	133
5.2.3	基于双深度学习的疾病诊断装置	145
第 6 章	深度学习的大数据智能预测应用	148
6.1	基于大数据与深度学习的气象预报方法和系统	148
6.1.1	研究现状	148
6.1.2	基于大数据与深度学习的气象预报方法	149
6.1.3	基于大数据与深度学习的气象预报系统	155
6.2	基于大数据与深度学习的用户数据处理方法和系统	157
6.2.1	研究现状	157
6.2.2	基于大数据与深度学习的用户数据处理方法	158
6.2.3	基于大数据与深度学习的用户数据处理系统	175
结束语		181
参考文献		182

第一部分 理论方法



第1章 增强型深度学习神经网络

本章提出的增强型深度学习神经网络将现有技术中深度学习的功能进行了增强，能够实现更为强大的功能。本章给出了两种增强型深度学习神经网络及其应用，包括基于顶层生成深度学习的数据转换方法和系统，基于双向深度学习的数据对应关系判断、生成方法和系统。

1.1 基于顶层生成深度学习的数据转换方法和系统

本节研究了一种基于顶层生成深度学习的数据转换方法和系统，所述方法包括：① 建立深度学习神经网络，学习目标分辨率的样本数据；② 将源分辨率数据转换为第一目标分辨率数据；③ 将第一目标分辨率数据作为深度学习神经网络的输入，通过深度学习的认知过程得到深度学习神经网络的顶层概念；④ 通过深度学习将第一目标分辨率数据获取到顶层概念中，并生成第二目标分辨率数据，将第二目标分辨率数据作为源分辨率数据的转换结果。

本节所述的内容能够满足数据转换时对数据细节一致性的需求，便于人们查看数据的细节或全貌，有利于计算机更精细化、更节省资源地处理数据。

1.1.1 研究现状

在研究中,人们采集的数据分辨率往往受环境和采集设备的影响而不能满足需求,例如,照片由于雾霾、光线、摄像机分辨率等因素而分辨率不高,扫描出来的图片由于扫描仪的分辨率限制而分辨率不够。这些情况下,提高数据的分辨率至关重要。现有的提高数据分辨率的方法是插值或拟合,插值或拟合一般是指对相邻的已知数据值进行加权平均,得到需要插入的未知数据值。但实际上数据的分布并不是固定的,而是变化的,插值或拟合的方法只能采用固定的插值或拟合公式,因此,以固定的方式对分布变化的数据进行插值或拟合显然会在提高分辨率的同时造成数据的失真^[1-10]。

另外,从高分辨率到低分辨率,现有技术往往通过删除一些数据来实现,可能会导致得到的低分辨率数据不连贯。一种方式是通过高分辨率数据加权平均统计或拟合得到低分辨率数据,但实际上数据的分布并不是固定的,而是变化的,统计或拟合的方式只能采用固定的统计或拟合公式,因此,以固定的方式对分布变化的数据进行统计或拟合显然会在降低分辨率的同时造成数据的失真。

我们知道,人类往往能根据低分辨率的数据想象出高分辨率的数据,也能根据高分辨率的数据想象出低分辨率的数据。深度学习类似人的大脑,是从神经网络技术发展而来的。

现有的深度学习技术能通过输入数据得到输出标签(例如,输入头像输出身份证号,输入语音输出身份证号),但在自顶向下的监督训练阶段,需要通过带标签的数据(例如,带有身份证号标签的头像或带有身份证号标签的语音)进行监督训练^[11-17]。但是,现有深度学习技术无法通过输入一类数据得到另一类过于复杂的标签数据输出,因为只有比较简单的标签数据(如

类别标签)才能用于现有的深度学习技术,过于复杂的标签数据输出会使深度学习的计算复杂度剧增,无法在有效的时间内完成计算。因此,过于复杂的标签数据无法作为现有深度学习神经网络的输出。也就是说,通过现有深度学习技术不可能实现输入低分辨率数据得到高分辨率数据,或输入高分辨率数据得到低分辨率数据。同时,深度学习的认知过程只能认知出输入数据的特征,而无法从输入数据中认知出比输入数据更高或更低分辨率的数据,所以,通过深度学习的认知过程也无法实现从低分辨率数据得到高分辨率数据,或从高分辨率数据得到低分辨率数据。

为了解决上述技术的缺陷,本书提供了一种基于顶层生成深度学习的数据转换方法,该方法相对于现有技术具有如下优势。

在学习阶段,通过高(低)分辨率样本数据对应的低(高)分辨率样本数据,学习和训练出能够认知出低(高)分辨率样本数据顶层概念的深度学习神经网络。在应用阶段,将待转换的高(低)分辨率数据转换为第一低(高)分辨率数据,然后将第一低(高)分辨率数据作为深度学习神经网络的输入,得到深度学习的顶层概念,再通过深度学习的顶层概念,由深度学习神经网络生成第二低(高)分辨率数据,并将第二低(高)分辨率数据作为给定的待转换的高(低)分辨率数据的转换结果。满足了高(低)分辨率数据转换为低(高)分辨率数据时对数据细节一致性的需求。

1.1.2 基于顶层生成深度学习的数据转换方法

基于顶层生成深度学习的数据转换方法可将照片、遥感图像、语音、点云数据等数据的分辨率提高或降低,便于人们查看数据细节或全貌,有利于计算机更精细化、更节省资源地进行处理。

如图 1-1 所示,本技术方案提供了一种基于顶层生成深度学习的数据转换方法。

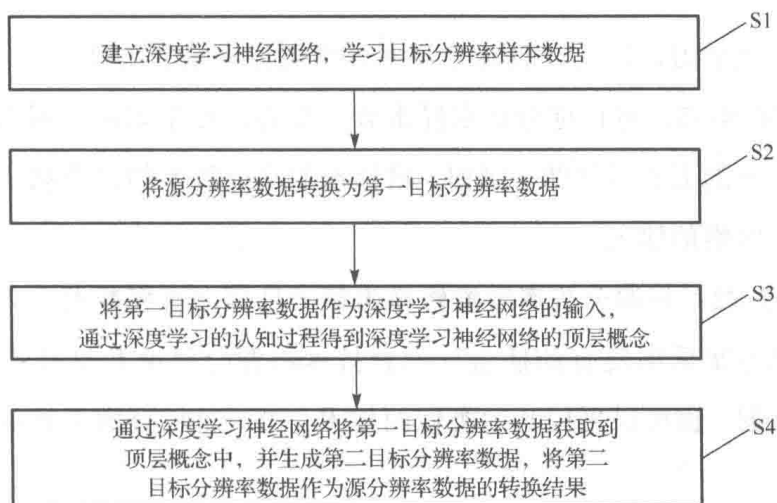


图 1-1 基于顶层生成深度学习的数据转换方法流程图

该方法包括以下步骤。

(1) 步骤 S1：建立深度学习神经网络，学习目标分辨率样本数据。具体实现步骤如图 1-2 所示。

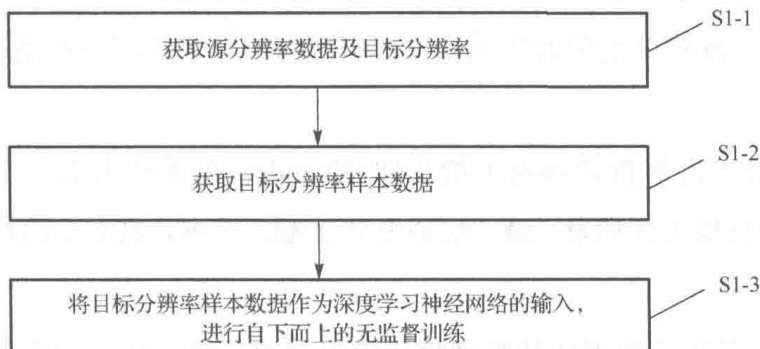


图 1-2 建立深度学习神经网络流程图

① 步骤 S1-1：获取源分辨率数据及目标分辨率。例如，源分辨率是预设的低分辨率，输入数据是猪头像，目标分辨率是预设的高分辨率。

② 步骤 S1-2：获取目标分辨率样本数据。目标分辨率样本数据需与源分辨率数据的类别相同。例如，源分辨率是预设的低分辨率，输入数据是猪头像，目标分辨率是预设的高分辨率，也需要用猪头像样本学习，而不能用

狗头像样本来学习。因此，目标分辨率样本数据应是猪头像。

③ 步骤 S1-3：将目标分辨率样本数据作为深度学习神经网络的输入，进行自下而上的无监督训练。例如，将预设的高分辨率的每个猪头像作为深度学习神经网络的输入。

(2) 步骤 S2：将源分辨率数据转换为第一目标分辨率数据。

本技术方案采用现有的插值、拟合技术将源分辨率数据转换为第一目标分辨率数据，也可以采用其他类似的技术，第一目标分辨率数据记为第一数据。

例如，将预设的低分辨率的猪头像转换为第一预设高分辨率数据。

(3) 步骤 S3：将第一目标分辨率数据作为深度学习神经网络的输入，通过深度学习的认知过程得到深度学习神经网络的顶层概念。

例如，将第一预设高分辨率数据作为深度学习神经网络的输入，通过深度学习的认知过程得到深度学习神经网络的顶层概念。

具体地，通过深度学习的认知过程得到深度学习神经网络的顶层概念，包括：

① 通过下层的概念和向上的认知 (Encoder) 权重产生上层的概念，上层概念比下层概念更抽象，最下层的概念是输入数据，最上层的概念是顶层概念。

② 首先从输入数据认知得到隐层第一层的概念，然后由隐层第一层的概念认知得到隐层第二层的概念，以此类推，直到得到隐层最后一层的概念，将隐层最后一层的概念作为深度学习神经网络的顶层概念。

(4) 步骤 S4：通过深度学习神经网络将第一目标分辨率数据获取到顶层概念中，并生成第二目标分辨率数据，将第二目标分辨率数据作为源分辨率数据的转换结果。

例如，通过深度学习神经网络将第一预设高分辨率数据获取到顶层概念

中,并生成第二预设高分辨率数据,将第二预设高分辨率数据作为源分辨率数据的转换结果。

具体地,通过深度学习神经网络生成第二目标分辨率数据,包括:

① 通过上层的概念和向下的生成(Decoder)权重产生下层的概念,下层概念比上层概念更具体,最上层的概念是顶层概念,最下层的概念是输入数据。

② 首先从顶层概念(隐层最后一层)生成得到隐层倒数第二层的概念,然后由隐层倒数第二层的概念认知得到隐层倒数第三层的概念,以此类推,直到得到输入层的概念,将输入层的概念作为第二目标分辨率数据。

可以理解的是,由于这个深度学习神经网络已经学习过很多未失真、无缺陷的目标分辨率样本数据,因此已具备了根据顶层概念生成逼真目标分辨率数据的能力。第一目标分辨率数据是采用现有技术转换成的,虽然有失真,但其中主要特征未变,可以被深度学习神经网络获取到顶层概念中,再由深度学习神经网络生成(如同人的想象)逼真的第二目标分辨率数据,这个逼真的第二目标分辨率数据即作为待转换的源分辨率数据的转换结果。这样就可以将低分辨率图像生成逼真的高分辨率图像了。

同样地,源分辨率也可以是预设的高分辨率,此时目标分辨率是预设的低分辨率。

1.1.3 基于顶层生成深度学习的数据转换系统

如图 1-3 所示,本技术方案提供了一种基于顶层生成深度学习的数据转换系统,所述系统包括深度学习神经网络建立模块 1、数据转换模块 2、顶层概念获取模块 3 和数据生成模块 4,各个模块的具体功能如下。

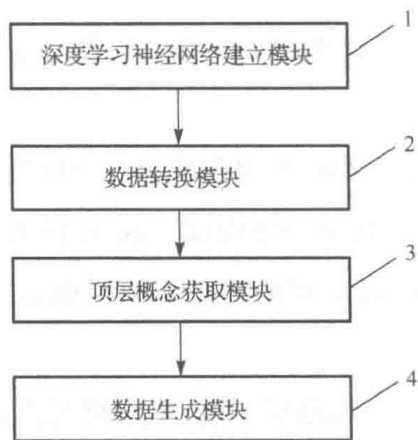


图 1-3 数据转换系统的结构图

1. 深度学习神经网络建立模块

深度学习神经网络建立模块用于建立深度学习神经网络，学习目标分辨率样本数据，如图 1-4 所示。具体包括：

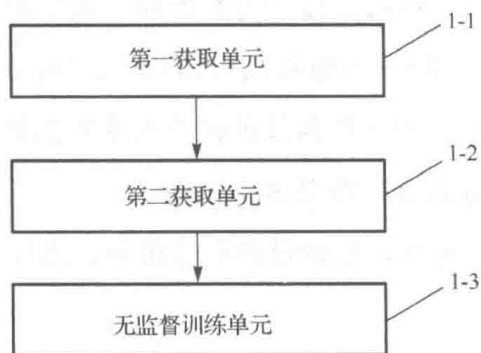


图 1-4 深度学习神经网络建立模块的结构图

- ① 第一获取单元 1-1，用于获取源分辨率数据及目标分辨率。
- ② 第二获取单元 1-2，用于获取目标分辨率样本数据；其中，目标分辨率样本数据与源分辨率数据的类别相同。
- ③ 无监督训练单元 1-3，用于将目标分辨率样本数据作为深度学习神经网络的输入，进行自下而上的无监督训练。