李社宏．气象领域深度学习知识体系框架及前沿应用[J]．陕西气象，2018(1)：21-25.
文章编号：1006-4354(2018)01-0021-05

气象领域深度学习知识体系框架及前沿应用

李社宏
（陕西省气象局，西安　710014）

摘　要：介绍了数据挖掘、机器学习和深度学习的概念和相互关系，按照整体性学习理论建立了气象领域深度学习知识体系框架，简要介绍了当前主流的深度学习框架工具 Caffe 和 TensorFlow，以及深度学习在气象领域的几个前沿应用，最后提出了推进深度学习技术在气象领域研究应用应当重视的三个关键环节。

关键词：数据挖掘；深度学习；卷积神经网络；知识体系框架
中图分类号：TP301：P413　　**文献标识码**：B

近年来，深度学习理论与技术取得了重大突破，日臻完善，其大规模研究应用的浪潮奔涌而至，在语音识别、人脸识别、自然语言处理、自动驾驶、搜索广告 CTR 预估等众多领域的应用日益广泛，并迅速向经济发展、社会生活等各个领域渗透，深度学习已经成为各行各业战略转型、创新发展的重要源头和强大动力。在气象领域早日引入深度学习技术，推进气象科技创新发展，重要而迫切，也面临严峻挑战。面临挑战的原因主要在于，深度学习和气象科技分别都是非常庞大且复杂的知识体系，两个知识体系叠加后的复杂程度远远超出了个人（甚至超级专家）可以掌控的范围。要同时驾驭这两个复杂知识体系，需要转变传统思维方式和学习方法，运用整体性学习策略，首先构建气象领域深度学习的知识体系框架，并依此组建团队，依靠团队的力量快速进入新领域。本文按照整体性学习的理论和方法，给出了一个初步的气象领域深度学习的知识体系框架，并简单介绍了当前主流的深度学习框架工具 Caffe 和 TensorFlow，以及深度学习在气象领域的几个前沿应用，最后提出了推进深度学习技术在气象领域研究应用应当重视的几个关键环节。

1　数据挖掘、机器学习与深度学习的联系与区别

1.1　数据挖掘

数据挖掘（data mining，DM）是指从海量数据中通过算法搜索隐藏于其中的信息和知识的过程。数据挖掘需要运用大量机器学习领域提供的数据分析技术和数据库领域提供的数据管理技术。

1.2　机器学习

机器学习（machine learning，ML）是一门专门研究计算机如何模拟或实现人类的学习行为，以获取新的知识或技能，重新组织已有的知识结构，使之不断改善自身性能的学科。机器学习是人工智能的核心，是使计算机具有智能的根本途径。

1.3　深度学习

深度学习（deep learning，DL）是机器学习的一个新领域，通过建立模拟人脑进行分析学习的多层神经网络，模仿人脑的机制来解释数据，例如视频、图像、声音和文本等数据。深度学习的优势是能够通过学习算法自动获取数据特征，同时由于模型的层次、参数很多，容量足够，因此模型擅长于表示大规模数据。深度学习方法主要包括监督学习与无监督学习两种。例如，卷积神经网络

收稿日期：2017-11-08
作者简介：李社宏（1969—），男，汉族，陕西周至人，高级工程师，主要从事气象业务管理。

(convolutional neural networks, CNN)就是一种监督学习模型,而深度置信网(deep belief nets, DBN)就是一种无监督学习模型。

数据挖掘、机器学习和深度学习三者之间既有联系,区别也十分明显。首先,数据挖掘是一个很宽泛的概念,不仅要研究、拓展、应用机器学习技术,还要使用数据库技术、数据清洗等非机器学习技术。其次,机器学习不仅涉及对数据的分析处理,还涉及对人的认知学习过程的探索,是人工智能的核心研究领域之一,也是数据挖掘的重要工具。再次,深度学习是机器学习的一个新的热门领域,本质上来源于多隐层的神经网络。总之,数据挖掘的技术成分更重一些,机器学习的科学成分更重一些。数据挖掘是从目的而言的,侧重于把数据转化为有用的信息和知识,而机器学习是从理论和方法而言的,侧重于如何科学高效的实现这种转化。

2 气象领域深度学习知识体系框架

从技术角度而言,深度学习是一系列复杂知识和技术的组合,真正全面掌握深度学习技术是一件很难的事情。何况,把深度学习应用到气象领域,还需要掌握气象领域的复杂的知识和技术,难度大幅度增加。如何克服困难,跨越这个难度,需要同时从两个方面着手。一是通过搭建气象领域深度学习知识体系框架,高效学习、快速进入深度学习领域。二是建立团队,发挥团队成员在各自领域的技术优势,分工合作。这两方面措施需要同时采取,同时发力,才能取得效果。这里重点探讨第一个方面。

整体性学习是一种高效学习理论,它在深入研究和科学借鉴人类大脑工作机理的基础上,充分发挥大脑结构和大脑中丰富的神经元的作用,通过创建知识网络,将一个个孤立的知识关联起来,以达到对知识的完全理解和轻松驾驭[1]。运用整体性学习理论,可帮助我们快速高效学习有关深度学习知识,并与气象知识建立关联,快速进入深度学习庞大而复杂的知识领域。整体性学习首要的是建立知识体系框架,作者通过阅读大量相关资料并进行梳理,建立了一套气象领域深度学习知识体系框架。气象领域深度学习知识体系框架主要包括:大数据平台、数据获取、实时数据流处理、历史数据交互式处理、批量数据处理、机器学习、深度学习和应用云化,共八个部分,每部分还可继续细分,如图1。图1中阴影部分

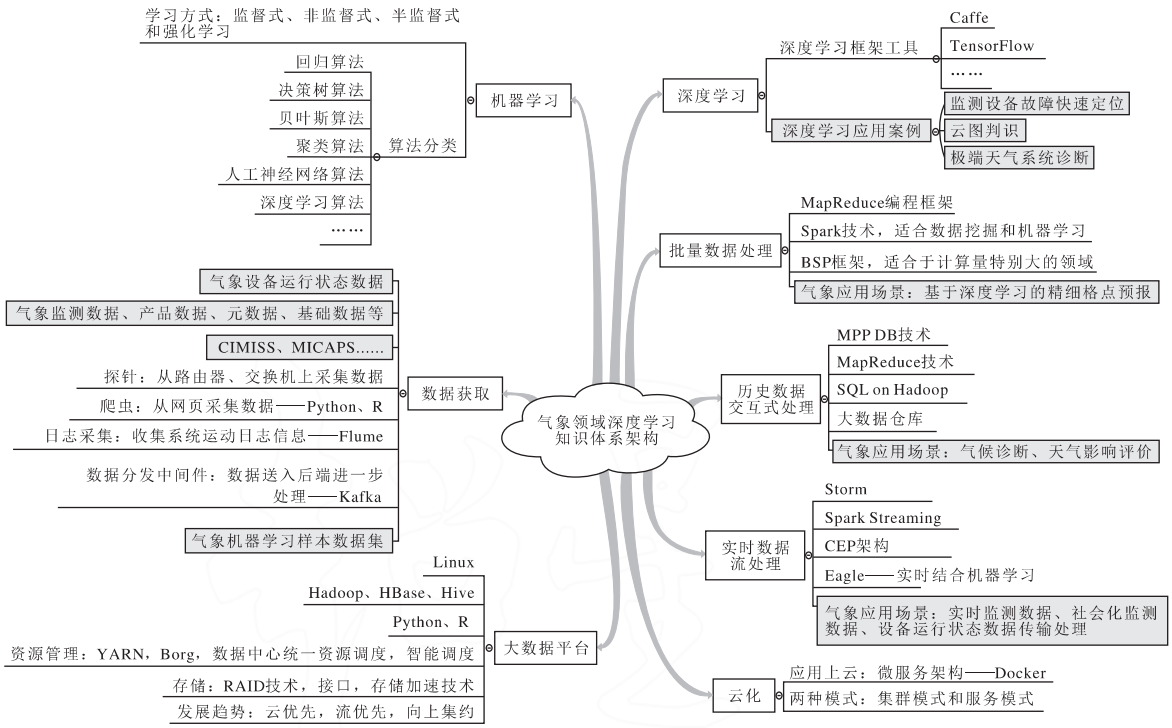

图1 气象领域深度学习知识体系框架

的内容是与气象领域有关的知识,其余为通用的深度学习知识。

3 主流深度学习框架工具

目前,深度学习正处于蓬勃发展阶段,各种深度学习框架工具层出不穷,发展迅速,迭代周期短。常见的深度学习框架工具有 Caffe、TensorFlow、Theano、Torch、DeepLearning4j、Marvin 等。这里介绍两个成熟、易于掌握、使用普遍的通用深度学习框架,即 Caffe 和 TensorFlow,以便对深度学习框架有一个基本认识。

3.1 Caffe

Caffe(全称 convolutional architecture for fast feature embedding,快速特征嵌入卷积框架)是一个清晰、高效、高可读的深度学习框架,作者是贾扬清。Caffe 支持命令行、Python 和 Matlab 接口,可以在 CPU 和 GPU 之间无缝切换运行。Caffe 的优势是:网络结构定义简单、容易上手、运行速度快、模块化、开源和活跃的社区等,是目前图像识别领域的主流框架。

3.2 TensorFlow

TensorFlow 是谷歌研发的第二代人工智能学习系统,其命名来源于本身的运行原理。Tensor(张量)意味着 N 维数组,Flow(流)意味着基于数据流图的计算,TensorFlow 为张量从流图的一端流动到另一端的计算过程。TensorFlow 是将复杂的数据结构传输至人工智能神经网中进行分析和处理的系统,其优势是:灵活性和可延展性,同时还支持异构设备分布式计算,可在小到一部智能手机、大到数千台数据中心服务器的各种设备上运行。TensorFlow 被广泛用于语音识别或图像识别等多项深度学习领域。TensorFlow 也是开源的,且社区繁荣。

需要说明的是,Caffe 和 TensorFlow 都可以在官网上下载到源码、测试数据集(如 CIFAR-10、MNIST 数据集等)和详细的安装配置指南,也可在网络社区找到丰富的学习资料和操作案例,二者都可在单机上调试运行,非常方便初学者入门与提高。

4 气象领域深度学习的几个前沿研究与应用

这里,简要介绍国内外把深度学习技术应用于气象领域的几个前沿研究与应用,以概要了解相关领域的发展思路、发展方向、发展态势及进展。

4.1 天气系统自动识别

在视频流中实时发现目标及其运动情况,是目前深度学习技术的一个热门应用领域。在气候数据集中发现极端天气系统,与在视频流中发现目标及其运动情况非常相似。不同之处在于,在气候数据集中,“视频”有 16 个或更多的“通道”信息(如气压、温度、湿度、风向风速等),而传统的视频中只有 3 个通道(RGB)。基于这样的思路,Liu Yunjie 等开发了深层卷积神经网络天气系统识别系统,通过 Caffe 建立了卷积神经网络并调优,用历史气候数据集进行了网络训练[2]。该 CNN 共有 4 个学习层,其中包括 2 个卷积层和 2 个全连接层,每个卷积层之后紧跟一个最大池化层。研究中使用了两种资料:气候模拟资料和再分析资料,时间范围从 1908 年至 2009 年,时间跨度超过了 100 年,样本数 5 000 到 10 000 不等,各种天气系统的真实标注数据来自 TECA 分析输出,由专家手动标注完成。该 CNN 网络识别热带气旋、水汽输送带、锋面等天气系统的准确率达到了 89%~99%。Evan Racah 等则更进一步考虑用一个统一的网络对多种类型天气系统进行识别,这是对这个问题的一个更高级的、类似于 3D 卷积神经网络的、半监督学习的思路,已取得了良好进展[3]。

4.2 卫星云图识别和云量计算

王舰锋等运用卷积神经网络开展了卫星云图判识研究,并在此基础上进行了卫星云量计算[4]。首先,在中国资源卫星网下载了 HJ-1A/1B 卫星资料,经过预处理,分别建立了训练样本集和测试样本集。其中训练样本集包括厚云、薄云和晴空三类各 3 000 个样本,测试样本集包括厚云、薄云和晴空三类各 1 000 个样本。对所有云图样本图像都进行了归一化处理,处理后的图像像素大小为 32×32,以此作为卷积神经网络的输入。其次,建立卷积神经网络 CNN,其架构包含输入层、卷积层、池化层、全连接层、Softmax 分类层和输出层,利用该 CNN 对 9 000 个训练样本进行训练,对 3 000 个测试样本进行测试。第三,对卷积神经网络从网络层数、滤波器个数、滤波器大小等

几个方面进行了优化。结论是:当卷积神经网络的层数为6层时,网络分类的准确率最高;当第一层滤波器个数为12个,第二层滤波器个数为16个时,网络分类的准确率最高;当第一层滤波器大小为5×5,第二层滤波器大小为7×7时,网络分类的准确率最高。优化后,网络对云分类的准确率达到了91.4%。第四,在运用训练后的CNN对多通道卫星云图进行检测基础上,采用基于反射率检测的算法开展总云量计算,最终得到云量分布图。第五,经比较传统阈值法、动态阈值法、极限学习机模型和卷积神经网络,结果表明基于卷积神经网络的云量计算准确率最高,接近90%。

4.3 地基全天空云图分类

张振等研究了基于深度学习的数字地基全天空云图分类方法[5]。研究使用了中国气象科学研究院和北京交通大学共同发布的全天空云图数据集,该数据集由布设在西藏的全天空云图成像仪采集获得,数据采集时间为2012年8月至2014年7月,共筛选出5 000张全天空云图像,定义了5种云图像类别,分别是:卷状云、积状云、层状云、晴空和混合云,每种类别1 000张云图。研究使用的深度学习框架为Caffe,卷积神经网络结构包括:三个卷积层、两个全连接层以及分类器层,每个卷积层后边都有池化层,池化层采用了最大值策略。针对小样本数据集标注样本较少的局限性,研究采用了迁移学习的思想和样本图像扩容技术。迁移学习即采用CaffeModel参数(ImageNet网络参数)初始化CNN前两层卷积层参数,然后再微调整个网络,此方法取得了最好的分类准确率。数据扩容即对原始全天空图采用拉伸、旋转、镜像、切割、改变纵横比等方法,产生新的样本。采用数据扩容使得原始数据集扩展了10倍。研究结果表明,基于深度学习的分类方法的准确率达到了98.4%,比基于统计学的传统分类方法提高了20%以上,这是目前已知的最好的分类结果。研究还建立了全天空云图分类平台,用户可将单张或多张云图图像提交平台,经过后端运算,返回分类结果。

5 跨入气象领域深度学习时代的关键环节

毋庸置疑,深度学习技术在气象领域有着广阔的应用前景,气象领域也呼唤着深度学习技术能够早日加入。加快推进深度学习技术在气象领域的研究应用,积极主动迎接气象领域深度学习时代的到来,需要优先重视以下三个关键环节。

5.1 建立气象标准数据集和开放网络社区

2012年以来,深度学习技术得以快速发展,主要取决于三个方面的有利因素:深度学习技术自身的突破、更大的数据集以及GPU等计算机硬件能力的提升。其中,深度学习技术自身的突破、计算机硬件能力的提升不是我们能左右的,只有更大的数据集与气象有关。由此可以相信,在气象领域引入深度学习技术的最大瓶颈将是适合机器学习的标准数据集建设,同时需要建立开放网络社区,吸引更多的爱好者参与社区交流。

5.2 改进创新管理方式

一个值得思考的现象是,目前气象领域深度学习的研究与应用多数是由高校、研究机构、企业等非气象机构的团队主导,这些团队先掌握了深度学习技术,然后把这项技术扩展应用到了气象领域,本文所列举的三个前沿应用都属于这种情况。相反,来自气象机构的团队以气象技术为基础,主动向深度学习技术领域扩展的则较少,或者说难度较大。克雷顿·克里斯滕森(美)在《创新者的窘境》一书中提出了“传统业务天生对创新具有绞杀功能”的观点,通过大量案例分析,作者还归纳了这种现象产生的深层次原因,并给出了应对法则,以改进突破性技术创新管理,很值得借鉴。

5.3 克服畏难心理

具有气象技术背景的人员要引入深度学习技术,大多会认为难度太大、难以跨越,对深度学习产生畏惧感。事实上,近年来深度学习技术发展非常迅速,通用学习工具简单易用,样本数据集丰富,操作说明齐全,对硬件的要求降低到可单机运行,深度学习的门槛已大大降低。团队和个人只要采取针对性措施,克服畏难心理,完全可以早日跨越深度学习的技术门槛,在深度学习的海洋中遨游。

6 小结

建立深度学习知识体系框架有利于从技术层面快速进入深度学习时代。本文给出了一个气象

张侠,胡琳,王琦,等. 2017年陕西气象条件对大气环境质量影响分析[J]. 陕西气象,2018(1):25-29.
文章编号:1006-4354(2018)01-0025-05

2017年陕西气象条件对大气环境质量影响分析

张 侠,胡 琳,王 琦,杜怡心
(陕西省气候中心,西安 710014)

摘 要:利用2017年1月1日—7月31日陕西省十地市空气质量资料和气象站地面观测资料,分析了2017年1—7月陕西省空气质量时间变化特征及影响大气环境质量的气象条件。结果表明:全省城市空气质量与2016年同期相比较差,1—3月全省首要污染物为颗粒物($PM_{2.5}$和PM_{10}),5—7月为臭氧。1—3月各市平均风速均在3.0 m/s以下且小风频率较高;全省冷空气活动较上年同期减少3次且强度偏弱;全省平均混合层高度与上年同期相比降低22 m。与上年同期相比,平均风速小,小风日数增多,冷空气活动次数减少且强度偏弱,混合层高度偏低,是颗粒物污染过程增多的主要因素。5—7月臭氧质量浓度与高温显著正相关,当日平均气温≥30 ℃或日最高气温≥35 ℃时,臭氧显著超标;臭氧质量浓度随日照时数增加而升高,日照时数≥6 h时,各市臭氧平均质量浓度均较高,日照时数≥10 h时臭氧超标率最高;臭氧质量浓度随日平均相对湿度的升高而降低,当相对湿度<60.0%时,臭氧平均质量浓度超过140 $\mu g/m^3$,当相对湿度≥70.0%时,臭氧超标率明显降低。与上年同期相比,气温偏高,日照充足,湿度减小是造成臭氧超标日增多的主要因素。

关键词:大气环境质量;气象条件;颗粒物;臭氧
中图分类号:X16 **文献标识码**:A

近年来,城市大气环境污染问题日趋严重,霾天气现象增多[1],严重威胁着社会可持续发展和人们身体健康。对大气环境污染的特征及气象条件影响因素的研究,一直是众多学者关心的焦点问题。在大气污染排放源稳定的前提下,气象条件是影响大气环境质量的一个主要因素。国内外

收稿日期:2017-08-23
作者简介:张侠(1984—),女,汉族,陕西渭南人,硕士,工程师,从事大气环境方面的研究。
基金项目:陕西省自然科学基础研究计划项目(2014JM2-4038)

领域深度学习的知识体系框架,在此基础上介绍了两种通用深度学习工具,以及天气系统识别、卫星云图识别和地基全天空云图分类等三个基于深度学习的气象应用,最后指出了建立气象标准数据集、改进创新管理方式和克服畏难心理的重要性。

参考文献:

[1] 斯科特·扬. 如何高效学习[M]. 程冕,译. 北京:机械工业出版社,2016:223-224.
[2] LIU Yunjie, RACAH Evan, PRABHAT, et al. Application of deep convolutional neural networks for detecting extreme weather in climate datasets(R/OL). (2016-05-04)[2017-11-03]. http://world-comp-proceedings.com/proc/p2016/ABD6152.pdf.
[3] RACAH Evan, BECKHAM Christopher, MAHARAJ Tegan, et al. Semi-Supervised detection of extreme weather events in large climate datasets(R/OL). (2016-12-07)[2017-11-03]. http://pdfs.semanticscholar.org/3fae/be9d5c47fc90998811c4ac768706283d605c.pdf.
[4] 王舰锋. 基于卷积神经网络的卫星云量计算[D]. 南京:南京信息工程大学, 2016.
[5] 张振. 基于深度学习的全天空云图分类方法研究[D]. 北京:北京交通大学,2016.