

地质领域机器学习、深度学习及实现语言

侯 玮 何黄金 李易念

(重庆邮电大学, 重庆 400065)

摘要: 地质大数据一般分为核心地质数据和公共地质数据。前者包含大量机密信息, 因此无法进行公开, 主要存储在局域网中; 而后者主要用于科学普及, 对其潜在价值进行挖掘已经成为地质大数据应用的重要方面。当前, 机器学习已经成为地质大数据研究的热门方向, 其让地质大数据得到了飞速发展。作为机器学习的重要分支, 深度学习是通过建立具有许多隐藏层和大量训练数据的学习模型, 从而完成更多有用功能的学习, 提高预测或分类的准确性。本文深入分析了地质领域的机器学习、深度学习, 并在此基础上对其实现语言进行了探讨。

关键词: 地质领域; 机器学习; 深度学习; 实现语言

近年来, 地质大数据的增长十分明显。无论是水文地质、环境地质、工程地质、矿产地质、灾害地质以及相应的地质科学研究, 还是各种地质的基础地质调查和勘探、环境和地球灾害的监测预防、空间遥感观测活动等, 均涉及广泛的地质大数据。地质大数据可以是结构化的, 同时也包含了众多的半结构化以及非结构化数据, 如矿床、标本、古生物、遥感光谱、矿物、岩石、地震视频、海啸音频、地质图表等等^[1]。在这种情况下, 想要以传统方式处理地质大数据是不可能的, 先进技术的应用不可逆转。只有开发更加智能的数据处理方法, 才能满足地质大数据飞速增长的需要。因此, 智能化的地质学已经成为重要研究领域, 更加符合当前地质大数据的处理和应用需求。

一、机器学习

机器学习是让计算机通过学习而不是使用直接编程来解决问题的一种模式。例如, 网络空间的运作机制主要是将网络实体的行为数据提供给计算机。根据这些行为数据, 计算机可以形成行为模式, 该阶段正是机器学习的训练过程。然后根据形成的行为模式对后续行为进行判断, 准确识别出异常行为, 该阶段是机器学习的应用过程。在机器学习的过程中, 如果计算机已经有了样本, 那么就可以利用这些样本来指导学习过程。如果计算机不存在样本, 可以尝试让计算机自己进行学习, 并总结可能的行为类

别, 形成一种无监督学习的模式。当前, 不管是成熟的监督式机器学习算法, 还是非监督式学习算法, 均已广泛应用于包括网络空间安全在内的许多领域^[2]。例如, 通过现有的僵尸网络样本, 可以使用监督式机器学习来识别服务器的控制命令, 该行为与通过非监督式机器学习来收集僵尸网络控制机器的行为类似。

二、深度学习

在传统机器学习的基础上, 其中的重点是立足于原始问题, 对特征向量进行探究, 提取关键数据, 确定学习模型。以僵尸网络为例, 为了对它进行准确识别, 我们需要使用到 NetFlow 的相关特征, 帮助完成模型的构建。因此, 由上述分析得出, 对于机器学习而言, 特征质量是主要的影响因素, 直接限制最终的学习效果。通常, 有必要选择具有交叉语义的向量, 用来改善计算范围并确保实际性能不会增加。这种问题可以通过主成分分析法来实现其提出和解决, 从而达到维度降低的目的。此外, 它也可以通过深度学习来完成。从实际应用情况来看, 深度学习已应用于图像以及语音的识别、语言处理和许多其他领域^[3]。目前, 其在基础概念方面并未达到完全统一, 这主要是因为研究重点不同。有的研究关注复杂非线性模型的形成, 有的研究则强调机制发现的特征, 还有一些研究认为深度学习应按照从低到高的顺序完成概念形成的过程。

作者简介: 侯玮, 男, 汉族, 四川省人, 重庆邮电大学, 研究方向: 汽车电子;
何黄金, 男, 汉族, 广东省人, 重庆邮电大学, 研究方向: 汽车电子;
李易念, 男, 土家族, 重庆人, 重庆邮电大学, 研究方向: 汽车电子。

对于深度学习自身而言,它可以被视为机器学习中的一类。从实际应用情况来看,它具有无限的发展潜力。特别是从计算机的角度出发,应该更加重视算法的实际应用价值,这与人工神经网络在本质上具有关联性。在人工神经网络的概念中,其中间层突破了单层的局限,呈现出多层次的外在特征。基于深度学习领域,可以得知人工神经网络的首层功能在于原始数据的接受,而不是仅依靠特征层来完成。换言之,深度学习领域的中间层包含多个神经元,每个神经元均具备叠加值变换以及输入信息加权叠加两个基本功能。同时,信息值继续进行传送,输出层往往会根据多级操作的结果进行最终决策。从本质上讲,它所表现出来的强大功能主要是基于相关数据所进行的多级操作。同时,它还可以以非线性方式,使其特征得到最大化发挥,凸显出适应变化复杂环境的独特优势。

三、地质领域机器学习、深度学习实现语言分析

(一) 通过深度学习处理异常数据

在传统的地质勘探过程中,确定成矿目标时地球物理、地球化学、遥感异常、寻找地质均非常重要。在此过程中,重磁异常是非常重要的地球物理异常状况。针对重磁异常的处理通常可以分为两个阶段,一是数据的反演,二是过滤和增强。在反演技术过程中,可以将常用方法划分为全局随机优化的非线性反演算法以及线性反演的迭代优化方法两大类。在这两种方法运用的过程中,线性反演的迭代优化方法在收敛速度上表现较好,但容易存在局部误差的问题。因此,将全局随机优化的非线性反演算法引入其中,取得了更为明显的效果,这也正是通用深度学习算法的真正本质。与浅层算法相比,如BP神经网络,深度学习算法在处理异常数据时具有精准度高、收敛性好的特点。因此,深度学习算法可用于重磁异常的地球物理反演。

(二) 应用Python实现机器学习算法

人工神经网络是一个强大的模型,但是其原理非常简单,在语言处理、图像识别以及其他领域均有着重要作用。作为Python的一个库,Theano是由深度学习领域的专家Yoshua Bengio所带领的实验室进行开发的,主要是用于定义并有效解决与多维数组数据相对应的数学表达式的仿真估计问题。Theano的应用具有优化速度快、符号分解效率高、稳定性好等特点,更为重要的是,它还完成了GPU的加速,这使得密集数据的处理速度比CPU快数十倍。Theano还可以应用在构建有效的神经网络模型上,但对于普通研究者而言,该领域涉及专业度较高。因此,可以通过Keras库的应用来构建神经网络。Keras库的应用简化了构建各种神经网络模型的步骤,可以使普通用户更容易地构建和求解复杂的深度神经网络。但它并不是简单的神经网络库,而是基于Theano的深度学习库。它不仅应用于构建普通的神经网络,还可以构建更多的深度学习模型,如循环神经网络、卷积神经网络、自编码等。因为它有着Theano的基础,所以具备速

度快的优势。就像积木搭建一样,使用Keras库构建神经网络模型的过程非常简单且直观。通过相关代码,就可以构建非常强大的神经网络模型,甚至是构建相对完整的深度学习模型。

(三) 深度学习在地质工作中的实现

深度学习可以应用于地质工作领域,特别是遥感图像处理。对此本文提出了三种方式:一是借助于深度学习方法,帮助确定遥感异常区域,由此描述成矿目标区域位置。卷积神经网络算法的功能使其可以应用于遥感图像的数据处理过程,目前其实际应用已经取得了一些积极成果。在此过程中,可以通过两种方式进行,也就是监督学习以及非监督学习。前者需要通过已成图像手动划定异常区域,然后通过训练数据对未知图像进行判断。后者只需将遥感数据输入到深度学习模型中,对图像信息进行聚类,然后勾画出异常区域;二是使用深度学习算法来描述岩石的薄片数据。当前主要是通过岩石薄片镜下图像中显示的特定信息来进行识别,并提取特定信息,之后使用监督学习的方法将信息输入到深度学习模型中。该信息包含在岩石薄片,主要是由深度学习模型来确定,并汇总相关信息进行识别;三是使用深度学习算法对样本数据进行必要分类。借助卷积神经网络算法,我们可以从三种主要岩石开始,也就是常说的岩浆岩、变质岩以及沉积岩。首先,应该从最为基本的数据收集开始,然后基于卷积神经网络算法构建相应的学习模型,使用图像数据进行模型训练,并利用生成的模型文件完成后续检测。如果显示结果良好,那么基本可以确定模型文件是可用的。如果所得效果一般,则需要重新设计模型。

四、结语

在信息时代引领全球发展的现实环境下,机器学习以其优势引领着人工智能发展的新方向。在对相关学习模型进行修正和不断改进的过程中,应用到了神经网络,由此形成了深度学习概念,可以有效应对当前复杂多变的环境。信息时代所带来的影响是多方面的,我们关注的不仅仅是其应用优势,更要解决该过程中可能出现的问题。在此基础上,要注意对机器学习和深度学习等相关理念和模型的综合分析,把握其中的内涵,分别明确其应用优势,正确应用各种算法,在提高算法利用率、合理性和针对性的同时,更好地发挥技术特点,并在地质大数据处理方面具有更多优势,建立更加稳定高效的分析系统。

参考文献:

- [1] 周永章,王俊,左仁广,等.地质领域机器学习、深度学习及实现语言[J].岩石学报,2018,034(011):3173-3178.
- [2] 周子扬.机器学习与深度学习的发展及应用[J].电子世界,2017(23):74-75.
- [3] 蹇宇澄,刘昭策.深度学习的实现与发展——从神经网络到机器学习[J].电子技术与软件工程,2017(11):30-31.