DOI:10.3969/j.issn.1009-6868.2017.04.008
网络出版地址:http://kns.cnki.net/kcms/detail/34.1228.TN.20170705.1457.008.html

本期专家论坛栏目策划人

钟义信

北京邮电大学智能科学技术中心教授,发展中世界工程技术科学院院士,国际信息研究学会名誉主席及中国分会主席,曾任国家“863”计划通信主题首届首席专家,国务院信息化专家委员会常委,中国人工智能学会理事长,中国神经网络委员会主席;长期从事信息科学和人工智能基础理论的研究和教学;先后创立“全信息理论”“语义信息论”和“机制主义人工智能理论”,获多项国家级和部级科技奖励;已发表信息科学及人工智能领域学术著作16部,学术论文510余篇。

深度学习的能与不能
What Deep Learning Can or Cannot

于剑/YU Jian
(北京交通大学,北京 100044)
(Beijing Jiaotong University, Beijing 100044, China)

中图分类号:TN929.5 文献标志码:A 文章编号:1009-6868 (2017) 04-0041-003

摘要: 深度学习技术的应用日渐广泛,在语音、图像、文本处理、搜索引擎、广告推荐等领域都取得了巨大的成功。认为深度学习自身具有盲点,无法解决全部的机器学习问题,并指出了深度学习的优缺点,为深度学习的使用者提供了一定的理论指导。最后,还展望了深度学习的未来发展趋势。

关键词: 深度学习;傻瓜型学习算法;专家型学习算法;白箱算法;黑箱算法;相关性;因果性

Abstract: Deep learning is now widely used and has achieved huge successes in many fields, such as speech, image, natural language processing, search engine and advertising recommendation. In this paper, considering the achilles' heel of deep learning, we propose that deep learning cannot solve all problems in machine learning discipline. Moreover, the advantages and the disadvantages of deep learning are illustrated, and some suggestions for users are proposed. Finally, the future trend of deep learning is pointed out.

Key words: deep learning; fool-type learning algorithm; expert-type learning algorithm; white box algorithm; black box algorithm; relevance; causality

机器学习作为单独的研究方向,应该说是在20世纪80年代第1届国际机器学习大会(ICML)召开之后才有的事情。机器学习存在很多不同的定义,常用的有3个:第1个常用的机器学习定义是“计算机系统能够利用经验提高自身的性能”[1],第2个常见定义是“学习就是一个基于经验数据的函数估计问题”[2],第3个常见的机器学习定义是“提取重要模式、趋势,并理解数据,即从数据中学习”[3]。这3个常见定义各有侧重:第1个聚焦学习效果,第2个亮点是给出了可操作的学习定义,第3个突出了学习任务的分类。

虽然机器学习的定义晚至20世纪才出现,但是广义上来说,机器学习任务,或者说学习任务,在人类出现伊始就已有之。在日常生活中,人们每天都面临如何从自己采集的数据中提取知识进行使用的问题。比如:大的方面,需要观察环境的变化来学习如何制定政策使得地球可持

收稿时间:2017-05-27
网络出版时间:2017-07-05

续发展；小的方面，需要根据生活的经验买到一个可口的柚子或者西瓜，选择一个靠谱的理发师等。在计算机出现以前，数据采集都是人直接感知或者操作的，采集到的数据量较小，人可以直接从数据中提取知识，并不需要机器学习。如：对于回归问题，高斯在19世纪早期（1809）就发表了最小二乘法；对于数据降维问题，卡尔皮尔逊在1901年就发明了主成分分析（PCA）；对于聚类问题，K-means算法最早也可追踪到1953年[4]。但是这些算法和问题被归入机器学习，也只有在机器收集数据能力越来越成熟，导致人类直接从数据中提取知识成为不可能之后才变得没有异议。

在过去的30年间，机器学习从处理仅包含上百个样本数据的玩具问题起步，一直发展到今天，已经成为了从科学研究到商业应用的标准数据分析工具，机器学习的研究热点也几经变迁。

1 机器学习发展简史

机器学习最早的目标是从数据中发现可以解释的知识，在追求算法性能的同时，强调算法的解释性。早期的线性感知器、决策树和最近邻等算法可以说是这方面的典型代表作。但是，1969年Minsky指出线性感知器算法不能解决异或问题[5]。由于现实世界的问题大多是非线性问题，而异或问题又可以说是非线性问题中最简单的问题，由此可以推断线性感知器算法的实际用处不大。这对于以线性感知器算法为代表的神经网络研究而言可以说是致命一击，直接导致了神经网络甚至人工智能的第1个冬天。感知器算法的发明人、神经网络先驱Rosenblatt于1971年因故去世，则更加增添了这个冬天的寒意。

需要指出的是，很多实际应用并不要求算法具有可解释性，比如：机器翻译、天气预报、打卦算命等。在这种需求下，如果1个算法的泛化性能能够超过其他同类算法，即使该算法缺少解释性，则该算法依然是优秀的学习算法。20世纪80年代神经网络的复苏，其基本思路即为放弃解释性，一心提高算法的泛化性能。神经网络放弃解释性的最重要标志是其激活函数不再使用线性函数，而是典型的非线性函数，如Sigmoid函数和双曲函数等，其优点是表示能力大幅提高，但相应的复杂性也极度增长。众所周知，解释性能好的学习算法，其泛化性能也要满足实际需求，如果其泛化性能不佳，即使解释性好，人们也不会选用。在20世纪80年代，3层神经网络的性能超过了当时的分类算法，如：决策树、最近邻等，虽然其解释性不佳，神经网络依然成为当时最流行的机器学习模型。在神经网络放弃解释性之后，其对于算法设计者的知识储备要求也放到了最低，因此神经网络在20世纪80年代吸引了大批的研究者。

当然，也有很多实际应用要求算法具有可解释性，如因果关系发现、控制等。应该说，同时追求解释性和泛化性能一直是非神经网络机器学习研究者设计学习算法的基本约束。一旦某算法既具有很好的解释性，其性能又超过神经网络，神经网络研究就将面临极大的困境，这样的事情在历史上也曾真实地发生过。1995年Vapnik提出了支持向量机分类算法，该算法解释性好，其分类性能也超过了当时常见的3层神经网络，尤其需要指出的是，其理论的分类错误率可以通过Valiant的概率近似正确（PAC）理论来估计。这导致了神经网络研究的10年沉寂，有人也将其称为人工智能的第2次冬天。在这期间，大批原先的神经网络研究者纷纷转向离开，只有少数人坚持研究神经网络。这个时间段对于机器学习来说，显然不是冬季。在这10年间，人们提出了概率图理论、核方法、流形学习、稀疏学习、排序学习等多种机器学习新方向。特别是在20世纪末和21世纪初，由于在搜索引擎、字符识别等应用领域取得的巨大进展，机器学习的影响力日益兴旺。其标志事件有：1997年Tom Mitchell机器学习经典教科书的出现，2010年和2011年连续两年图灵奖颁发给了机器学习的研究者Valiant和Pearl。

“三十年河东，三十年河西”。2006年以后，神经网络突破了3层网络结构限制，即所谓的深度学习，大幅提高了模型的表示能力，又适逢大数据时代相伴而生的高计算能力，神经网络化身为深度学习，再次将分类能力提高到同时代其他模型无法匹敌的程度，有人将其称为人工智能的第3次春天。在机器学习的许多应用领域，深度学习甚至成为机器学习的代名词。虽然如此，时至今日，深度学习仍然只是机器学习的分支，无论其沉寂或者过热，都不可能逆转，而只能加速全部机器学习本身应用越来越普及，理论越来越深入的发展趋势。

2 深度学习的适应范围

理论上，神经网络深度越大，其表示能力越高，但是深度学习对于计算能力和训练数据的规模提出了极高的要求。2008年以前，计算机的计算能力和训练数据规模不具备大规模进行深度学习研究的条件。随着云计算、大数据的普及，具备了研究深度学习的外在技术条件。在2010年以后，人们通过采用新的激励函数，如ReLU，以及Dropout[6]，Batch Normalization[7]等新训练方式，还有特别设计的新网络结构Deep Residual Networks[8]等，逐渐克服了梯度消失或者发散问题，研究深度学习的内在技术条件也日渐成熟。这使得化名为深度学习的神经网络研究进入了另1个春天。

虽然如此，深度学习在理论上并没有突破以往神经网络的理论架

构。所有对于经典神经网络的理论分析对于深度学习也依然成立。1986年，Rumelhart等人提出了自编码器，该模型可以用来对高维数据进行降维[9]。2006年，Hinton等人在Science上发表了1篇文章，该文章通过改进的自编码器学习算法构建了1种深层自编码器[10]，自此深度学习的影响力日渐增大。常见的几种典型的深度学习网络包括：自编码器、卷积神经网络、循环神经网络、长短时记忆网络等。

感知器算法可能是最早的神经网络算法，该算法显然属于典型的白箱算法，但是其表示能力有限，连异或问题也解决不了。为了解决异或问题，主流的神经网络技术放弃了解释性，在黑箱算法的道路上越走越远。实际上，机器学习算法对于普通人来说，可粗分为两类：一类是傻瓜型学习算法，即只要输入一定，任何人都可得到同样的结果，如主成分分析等算法；另一类是专家型学习算法，即使输入相同，不同人由于参数设置不同，也会得到大不相同的结果。显然，神经网络学习算法是典型的专家型学习算法。

总而言之，机器学习有两个基本任务。一是试图发现输入和输出之间的因果关系，其主要功用是解释，最终目的是控制，即一旦发生问题，必须找出问题发生的原因，这样就可以通过控制学习算法输入使得输出满足需要。解决此类任务的学习算法是白箱算法，要求解释能力强。二是力图发现输入输出的相关关系，其主要功用是预测，最终目的是验证，即一旦做出判断，就可以根据外界反应判断预测是否准确，但是出现错误之后，并不要求根据输入来追踪错误发生的原因。解决此类任务的典型学习算法是黑箱算法，并不需要解释能力。

真实现实生活中这两类任务都是存在的。第1类任务，如各种高风险任务，包括无人驾驶（火车、飞机、汽车等）、医疗手术等，一旦发生错误，由于成本巨大，必须能够分析出发生错误的原因，以避免类似错误再次发生。完成这类任务，不但需要提高完成任务的性能，更重要的是能够发现输入与输出之间的因果关系，一旦发生错误，能追踪学习算法发生错误的原因，显然适宜解决此类问题的学习算法是白箱算法。第2类任务，如各种低风险甚至无风险性任务，包括搜索引擎、各种棋牌游戏等，显然这类任务即使发生错误，后果也不严重，成本可以承担，因此更重要的是提高其性能，特别是预测能力，而并不要求算法去解释这些错误为什么会发生。

显然，对于一个具体的学习任务，一旦白箱算法的性能超过黑箱算法，黑箱算法就再也不会是完成此类任务的优先考虑对象。但是，许多学习任务，由于具有极高的复杂性，难以设计1个性能满足需要的白箱算法，黑箱算法由于放弃了解释能力的约束而可能在性能上有较大优势。如今深度学习的表示能力已经十分强大，2015年卷积神经网络已经达到152层[8]，2016年卷积神经网络达到了1 207层，迄今为止没有任何一个白箱算法的表示能力可以与现今的深度学习相媲美。故可以预测，深度学习在不需要发现因果关系的学习任务上在可见的未来不再有被替代的可能。

另外需要指出的是，相关性的挖掘是目前大数据面临的典型任务。甚至有人认为，在大数据时代，数据相关性的重要程度远超数据因果性。由此可知，相关性任务在大数据时代应用广泛。当前深度学习的快速发展和应用领域的日渐扩大，从侧面证实了这一点。当然，这并不意味着不需要研究数据因果性，更不意味着数据因果性的消失。

3 结束语

深度学习不仅是目前热度最高的人工智能研究方向，也是工业应用最广泛的学习范式。在未来，随着深度学习与特定相关性学习任务的耦合程度越来越高，可以想像深度学习会有更多的变型出现。但是解释性的学习算法无论在工业界还是学术界同样也不会被放弃。

参考文献

[1] MITCHELL T. Machine Learning[M]. New York: MaGraw Hill, 1997
[2] VAPNIK V N. The Nature of Statistical Learning Theory[M]. New York: Springer, 1995
[3] HASTIE T, TIBSHIRINI R, FRIEDMAN J H. The Elements of Statistical Learning[M]. New York: Springer, 2003
[4] THORNDIKE R L. Who Belongs in the Family [J]. Psychometrika, 1953,18(4):267–276
[5] MINSKY M, PAPERT S. Perceptons[M]. MA: The MIT Press, 1969
[6] HINTON G E, SRIVASTAVA N, KRIZHEVSKY A, et al. Improving Neural Networks by Preventing Co–Adaptation of Feature Detectors[J]. Computer Science, 2012, 3(4): 212–223
[7] IOFFE S, SZEGEDY C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift[C]// Proceedings of the 32 nd International Conference on Machine Learning, 2015
[8] HE K, ZHANG X, REN S, et al. Deep Residual Learning for Image Recognition[C]// IEEE Conference on Computer Vision and Pattern Recognition. USA:IEEE, 2016:770–778. DOI: 10.1109/CVPR.2016.90
[9] RUMELHART D E, HINTON G E, WILLIAMS R J. Learning Internal Representations by Error Propagation[M]. Neurocomputing: Foundations of Research. MA:MIT Press, 1988:318–362
[10] HINTON G E, SALAHUTDINOV R R. Reducing the Dimensionality of the Data with Neural Networks[J]. Science, 2006, 313 (9): 504–507. DOI:10.1126/science.1127647

作者简介

于剑，北京交通大学计算机学院教授，交通数据分析与挖掘北京市重点实验室主任；主要研究方向为机器学习、自然语言处理等；曾著有《机器学习：从公理到算法》一书，已发表论文100余篇。