软件学报 ISSN 1000-9825, CODEN RUXUEW
Journal of Software, 2022,33(4):1412−1429 [doi: 10.13328/j.cnki.jos.006470]

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel: +86-10-62562563

类脑超大规模深度神经网络系统*

吕建成[1], 叶 庆[1], 田煜鑫[1], 韩军伟[2], 吴 枫[3]

[1](四川大学 计算机学院, 四川 成都 610065)

[2](西北工业大学 自动化学院, 陕西 西安 710129)

[3](中国科学技术大学 电子工程与信息科学系, 安徽 合肥 230022)

通信作者: 吕建成, E-mail: lvjiancheng@scu.edu.cn

摘 要: 大规模神经网络展现出强大的端到端表示能力和非线性函数的无限逼近能力, 在多个领域表现出优异的性能, 成为一个重要的发展方向. 如自然语言处理(NLP)模型 GPT, 经过几年的发展, 目前拥有 1 750 亿网络参数, 在多个 NLP 基准上到达最先进性能. 然而, 按照现有的神经网络组织方式, 目前的大规模神经网络难以到达人脑生物神经网络连接的规模. 同时, 现有的大规模神经网络在多通道协同处理、知识存储和迁移、持续学习方面表现不佳. 提出构建一种启发于人脑功能机制的大规模神经网络模型, 该模型以脑区划分和脑区功能机制为启发, 集成大量现有数据和预训练模型, 借鉴脑功能分区来模块化构建大规模神经网络模型, 并由脑功能机制提出相应的学习算法, 根据场景输入和目标, 自动构建神经网络通路, 设计神经网络模型来获得输出. 该神经网络模型关注输入到输出空间的关系构建, 通过不断学习, 提升模型的关系映射能力, 目标在于让该模型具备多通道协同处理能力, 实现知识存储和持续学习, 向通用人工智能迈进. 整个模型和所有数据、类脑功能区使用数据库系统进行管理, 该系统了还集成了分布式神经网络训练算法, 为实现超大规模神经网络的高效训练提供支撑. 提出了一种迈向通用人工智能的思路, 并在多个不同模态任务验证该模型的可行性.

关键词: 大规模深度神经网络; 脑科学; 多模态; 通用人工智能; 分布式计算

中图法分类号: TP18

中文引用格式: 吕建成, 叶庆, 田煜鑫, 韩军伟, 吴枫. 类脑超大规模深度神经网络系统. 软件学报, 2022, 33(4): 1412−1429. http://www.jos.org.cn/1000-9825/6470.htm

英文引用格式: Lü JC, Ye Q, Tian YX, Han JW, Wu F. Brain-inspired Large-scale Deep Neural Network System. Ruan Jian Xue Bao/ Journal of Software, 2022, 33(4): 1412−1429 (in Chinese). http://www.jos.org.cn/1000-9825/6470.htm

Brain-inspired Large-scale Deep Neural Network System

LÜ Jian-Cheng[1], YE Qing[1], TIAN Yu-Xin[1], HAN Jun-Wei[2], WU Feng[3]

[1](College of Computer Science, Sichuan University, Chengdu 610065, China)

[2](School of Automation, Northwestern Polytechnical University, Xi'an 710129, China)

[3](Department of Electronic Engineering and Information Science, University of Science and Technology of China, Hefei 230022, China)

Abstract: Large-scale deep neural networks (DNNs) exhibit powerful end-to-end representation and infinite approximation of nonlinear functions, showing excellent performance in several fields and becoming an important development direction. For example, the natural language processing model GPT, after years of development, now has 175 billion network parameters and achieves state-of-the-art performance on several NLP benchmarks. However, according to the existing deep neural network organization, the current large-scale network is difficult to reach the scale of human brain biological neural network connection. At the same time, the existing large-scale DNNs do not perform well in multi-channel collaborative processing, knowledge storage, and reasoning. This study proposes a

* 基金项目: 国家重点研发计划(2017YFB1002201); 国家杰出青年科学基金(61625204); 国家自然科学基金(61836006)
本文由"面向开放场景的鲁棒机器学习"专刊特约编辑陈恩红教授、李宇峰副教授、邹权教授推荐.
收稿时间: 2021-05-20; 修改时间: 2021-07-16; 采用时间: 2021-08-27; jos 在线出版时间: 2021-10-26

brain-inspired large-scale DNN model, which is inspired by the division and the functional mechanism of brain regions and built modularly by the functional of the brain, integrates a large amount of existing data and pre-trained models, and proposes the corresponding learning algorithm by the functional mechanism of the brain. The DNN model implements a pathway to automatically build a DNN as an output using the scene as an input. Simultaneously, it should not only learn the correlation between input and output but also needs to have the multi-channel collaborative processing capability to improve the correlation quality, thereby realizing knowledge storage and reasoning ability, which could be treated as a way toward general artificial intelligence. The whole model and all data sets and brain-inspired functional areas are managed by a database system which is equipped with the distributed training algorithms to support the efficient training of the large-scale DNN on computing clusters. This study also proposes a novel idea to implement general artificial intelligence, and the large-scale model is validated on several different modal tasks.

Key words: large-scale deep neural networks; brain science; multi-modal; general artificial intelligence; distributed computing

深度神经网络已经在多个领域广泛应用, 例如计算机视觉[1,2]、自然语音处理[3,4]、目标检测等, 并取得了领人瞩目的成绩. 为了提高神经网络在特定应用场景下的性能和泛化能力, 扩大神经网络参数规模和训练数据集是一种有效的方法. 例如: OpenAI 在 2019 年发布了通用语言模型 GPT2, 能够生成连贯的文本段落, 在多个语言建模基准上取得了 SOTA 性能. 该模型是对 GPT 模型的扩展, 参数规模达到了 15 亿. 随后, OpenAI 在 2020 年提出了拥有 1750 亿个参数的自然语言深度学习预训练模型 GPT3[5],该模型经过约 45 TB 文本数据的预训练, 可以在多个 NLP 任务上达到最先进的性能. 此外, BigGan[6], Bert[7]等最新成果表明: 神经网络模型越大, 任务处理表现就越好. 该结论在视觉识别任务上也得到了验证, 实验[8]表明, 神经网络规模大小与识别精度之间存在很强的关联性. 综上, 大规模神经网络展现出强大的端到端表示能力和非线性函数的无限逼近能力, 在多个领域表现出优异的性能, 成为一个重要的发展方向.

人工神经网络构建过程受人脑神经元工作原理启发, 一直以来, 研究者都在尝试构建复杂神经网络模型以模仿大脑功能. 但人脑大约 1 000 亿个神经元, 每个神经元通过数千甚至上万个神经突触和其他神经元相连接, 整个生物神经网络拥有约 100 万亿连接[9], 人脑生物神经网络的复杂程度远超目前的人工神经网络模型. 按照目前神经网络的组织方式, 难以到达人脑生物神经网络连接的规模. 同时, 现有的大规模网络在多通道协同处理、持续学习、知识存储和迁移方面表现不佳. 此外, 大规模神经网络的训练需要消耗巨大的计算资源和时间. 例如, GPT3[5]存储需求高达 350 GB, 单次训练成本超过 1 200 万美元, 这在很大程度上限制了大规模神经网络的推广和应用. 因此, 大规模神经网络模型需要一种新的组织方式, 以实现自身的高效训练和灵活使用.

现有大规模神经网络模型往往是针对特定任务精心设计. 由于特定任务确定了输入/输出空间的边界,任务相关的大规模输入数据为大规模神经网络学习提供了条件, 使得大规模神经网络展现出强大的端到端表示能力和非线性函数的无限逼近能力[10]. 但当输入/输出空间的边界不确定, 或者输入空间发生变化, 人工神经网络模型往往难以适应, 表现出较低的迁移性, 无法直接迁移到不同任务中并取得较好的结果. 同时, 与大脑学习机制不同, 神经网络不具有持续学习能力. 当任务发生变化, 输入/输出空间的相关关系随之改变, 神经网络难以持续学习更新已学习到的空间映射应对任务变化. 特定的任务、确定的输入/输出空间边界, 往往造成神经网络映射的鲁棒性不佳. 当前, 神经网络如何存储知识是存在争议的, 神经网络映射能否实现知识迁移, 怎么样去实现, 还在探索之中. 人脑生物神经网络的复杂性, 使得大规模神经网络模型设计存在瓶颈,需要一种新的类脑机制指导大规模神经网络设计和学习.

本文提出一种启发于脑功能机制的超大规模深度神经网络模型, 该模型以脑功能和神经网络研究为基础, 集成海量数据集和预训练模型, 借鉴类脑功能的协同机制, 利用模型设计组件, 完成针对不同任务场景的神经网络模型设计. 该神经网络模型关注如何构建输入/输出的映射关系, 而不是神经网络连接细节, 目标在于使神经网络模型具备多通道协同处理能力, 提升输入/输出关系的映射能力, 实现知识存储, 让模型具备可持续学习能力应对任务的变化. 具体研究内容包括: (1) 基于人脑功能研究成果, 研究如何基于大脑功能机制构建类脑神经网络功能区(以下简称类脑功能区); (2) 构建类脑功能区, 设计数据库系统, 实现类脑功能区灵活管理; (3) 基于脑功能协同工作机制研究模型设计组件平台, 完成针对特定场景的神经网络模型设计, 构建

分布式训练组件提供算法和分布式训练支持; (4) 完成超大规模神经网络模型构建, 研究从场景输入到构建神经网络模型作为输出的通路, 并在多个不同模态任务进行可行性验证. 整个类脑大规模神经网络系统集成了海量数据、知识、算法和算力, 启发于大脑功能和协同工作机制, 集成了有监督、半监督、无监督的学习算法, 直接用于构建单模态或者多模态神经网络模型, 旨在缩小数据特征空间学习和语义推理之间的差距, 向通用人工智能迈进.

1 相关工作

1.1 大规模神经网络模型设计

大量研究表明: 增加神经网络的参数规模或训练样本数量, 可以提高神经网络在特定任务的性能和泛化能力. 借鉴人脑中具有大量的神经元结构, 很多工作都在研究如何构建超大规模的神经网络[5,7,11], 目标是使用深度学习网络模拟大脑的学习过程, 利用少量标注数据的训练解决优化问题. 换句话说, 通过参数规模的量变, 引起神经网络学习能力的质变. 例如: Devlin 等人用基于 Transformer[12]的双向编码器结构, 提出了 Bert 模型[7], 整个网络大约有 3 亿的参数, 在传统 NLP 领域的 11 个子任务中刷新 SOTA 纪录; Brown 等人使用单向 Transformer 作为基础块, 堆叠设计出了 GPT3[5], 拥有超过 1 700 亿的参数量. GPT3 仅仅使用 few-shot 训练策略在数据集上进行迁移, 即可超越当前的 SOTA 模型. 这些例子说明了超大规模神经网络能够有效提升神经网络学习能力、降低对领域内标签数据的依赖、有效避免过拟合的能力, 启发更多的研究向此方向展开.

虽然超大规模神经网络从多个方向都证明了增加参数量对神经网络学习有帮助, 但大规模神经网络训练往往要涉及数据分割、模型分割、带宽限制、批处理问题等多种问题, 使得参数量与训练所需计算能力之间的关系是非线性的[13,14], 研究人员往往需要提升数倍的计算能力才能为模型提升一倍的参数量, 严重限制了超大规模神经网络的发展. 此外, 目前的超大规模神经网络往往也是人为预先设计好的, 主要解决的是某一个领域内的多个问题, 但跨领域泛化问题的能力较弱. 例如: Bert 模型虽然在 NLP 领域效果较好, 但若将其直接应用于其他领域, 则效果欠佳. 跨领域的超大规模神经网络的应用往往需要对原始模型做进行重构设计[15], 而反观我们的人脑, 不仅具有极强的持续学习能力, 更是拥有解决不同领域问题的能力. 本文提出的方案欲将人脑功能的协同机制和功能关系映射到大规模神经网络构建中, 结合已有的数据和预训练模型, 根据应用场景的不同设计大规模神经网络模型, 提高神经网络的迁移性和鲁棒性.

1.2 受脑启发的神经网络模型设计

随着人工智能研究的不断发展, 许多重要进展反映了一个趋势: 来自脑科学的启发,即使是局部的借鉴都能够有效地提升现有人工智能模型及系统的智能水平[16]. 不同脑区之间的协同, 使得高度智能的类人认知功能得以实现, 如哺乳动物脑的强化学习功能、长短时记忆功能等都是通过不同脑区功能协同实现的. 此外, 有些脑区负责融合来自不同脑区的信号, 从而使对客观对象的认识更为全面[17]. 而有些脑区在接收到若干脑区的输入后, 则负责在问题求解的过程中屏蔽来自问题无关脑区的信号. 传统的深度神经网络模型一般只有前馈连接, 尚缺乏对反馈的建模. 为了更好地模拟人脑, 最近有很多研究探索如何将反馈引入神经网络模型. 例如: Liang 等人[18,19]提出在 CNN 的卷积层加上层内连接的方法, 使每个单元可同时接收前馈和反馈的输入; Wang 等人[20]通过 top-down 的反馈连接和乘法机制引入注意力模型; Cao 等人[21]在 CNN 的卷积层加上层间的反馈连接, 将高级的语义和全局信息传到下层, 通过语义标签的反馈, 可以激活特定的与目标语义相关的神经元, 从而实现自顶向下的视觉注意, 定位复杂背景中的潜在目标.

此外, 通过脑结构启发并推动神经网络发展在自然语言处理领域也取得了重要突破. 近年来, 关于如何理解机器中的自然语言处理和大脑中的自然语言处理之间的关系问题受到广泛关注, 这可以看作是自然语言处理和类脑神经科学的交叉研究. 对于应用于自然语言处理任务中的各种人工智能模型, 人们总是希望它们能够在完成文本理解的任务上达到跟人类一样的水平. Devlin 等人在总结前人借鉴大脑注意力机制的成果上设计了 Bert[7], 改变网络学习到的权重, 使它们能够像大脑一样工作. 清华大学类脑团队在芯片上高效实现人

工神经网络 ANN 和脉冲神经网络 SNN[22], 目标就是为通用人工智能提供新架构.

众多研究表明: 类脑和神经网络研究的结合, 可能是实现通过人工智能的新途径. 目前, 人工神经网络模型研究借鉴了脑神经的部分概念和结构, 但是在信息处理机制上, 真正从脑科学借鉴的机制并不深刻. 本文的研究目标主要是通过借鉴人脑脑区和脑功能关系以及协同工作机制, 启发并指导构建大规模神经网络模型, 实现新一代神经网络模型的设计.

2 类脑大规模神经网络模型设计

本文提出的启发于脑功能机制的大规模深度神经网络模型不仅要学习神经网络输入/输出的关系, 同时需要具备有多通道协同处理能力. 脑功能机制的指导增强神经网络的可解释性. 大量的预训练模型实现知识存储和持续学习, 增强模型的迁移性和鲁棒性. 整个模型包含 4 层架构(研究基础、类脑功能区、算法平台、应用平台)的解决方案, 如图 1 所示, 所有数据和类脑功能区通过数据库系统进行管理(以下称为类脑大规模神经网络系统). 具体内容如下.

- 首先是对人脑脑区以及在特定场景下脑功能关系进行研究, 研究成果用于指导构建类脑功能区以及多个类脑功能区的协同工作, 是整个方案的构建基础;
- 其次, 结合大脑脑区特点构建了类脑功能区, 目前主要构建了视觉、听觉、语言、情感这 4 个大的类脑功能区, 类脑功能区的协同工作由脑功能机制指导. 随着人工深度神经网络功能的不断丰富, 类脑功能区可以继续扩展. 此外, 为了在不同应用场景中提供最优神经网络模型设计, 构建包含多个组件的算法平台中间件作为类脑功能区与应用平台的桥梁. 以场景作为输入, 依托类脑功能区构建面向任务的神经网络模型作为输出;
- 最后, 我们把所有依托类脑大规模神经网络系统解决的应用场景问题统一归纳至应用平台, 并在实际问题研究中验证系统的可行性.

在接下来的章节中, 我们将对各部分进行详细介绍.

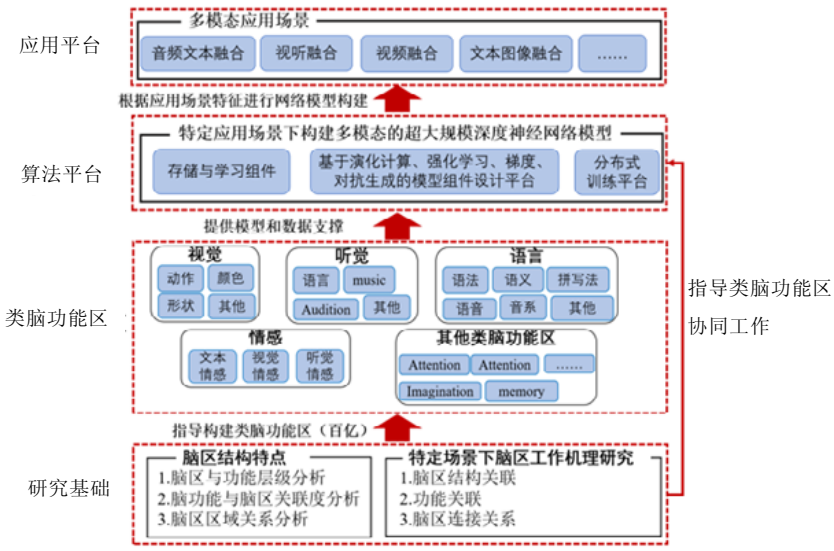


图 1 类脑大规模神经网络模型的总体架构

2.1 类脑研究基础

在现有的类脑研究中, 研究者将人脑根据区域、划分粒度不同, 大致可分为几百到上千个不同的脑区. 通过特定任务场景下脑区活动监测, 可以确定不同脑区的功能以及脑区间的协作机制. 研究[23]表明: 脑功能和

脑区存在多对多的关系, 即一个脑功能涉及多个脑区, 一个脑区可能服务多种功能. 我们对目前收集到的 100 多个有效脑区数据 Regions of Interests (ROI)进行了分析(该数据描述了脑区的坐标数据和功能描述), 结果显示: 大脑功能丰富, 包括感知、动作、情感、视觉、听觉、注意力、音乐、语言、颜色等近 60 种功能, 且每个功能涉及 1 个或者多个脑区. 从区域关系分析, 不同功能的脑区在大脑中分布式存在, 但是不同脑区的分布又存在一些特点, 例如: 视觉和语言的脑区分布相对较近, 在枕叶比较集中. 原因在于人脑对语言的理解很大一部分是和阅读相关的, 这需要视觉功能参与. 在实际功能性磁共振成像(fMRI)监测中, 也发现多模态场景下不同功能脑区需要一定的协同才能完成对任务的响应. 由于人生物脑神经网络的复杂性, 对于脑区划分和脑区协作机制还在探索中, 我们将借鉴现有人脑区域和脑功能的研究成果构建类脑功能区.

2.2 类脑功能区设计

在类脑功能区构建过程中, 我们无法对应所有大脑功能构建类脑功能区. 根据脑功能统计分析发现, 视觉、听觉、语言、情感这 4 个功能覆盖的脑区较多, 这也和目前神经网络研究的主要领域相符. 所以, 目前我们构建的超大规模的类脑网络模型主要包括了视觉功能区、听觉功能区和语言功能区以及情感功能区. 随着神经网络研究不断深入和扩展, 类脑功能区可以根据需要扩展或调整分类. 以视觉脑功能区为例, 介绍如何依据脑区的功能特点构建类脑功能区, 如图 2 所示. 根据目前掌握的数据, 视觉功能和 30 多个脑区相关, 与其相关的二级功能可以细分为动作、颜色、形状和其他, 如图 2 左侧所示. 在类脑功能区的构建过程中, 我们将现有视觉相关的神经网络网络和预训练模型进行收集、整理, 根据脑区功能层级关系, 对神经网络模型进行分类以构建子功能区, 形成和脑功能相似的层级关系如图 2 右侧所示, 视觉类脑功能区包含多个子功能区, 对应视觉功能相关的多个子功能. 类脑功能区的复杂程度取决于脑功能关联的脑区数量和神经网络规模. 如果一个脑功能丰富, 包含的脑区较多, 对应构建的类脑功能区就会有多个子功能区, 且相关的神经网络模型也会较多; 反之, 如果一个脑功能关联的脑区较少, 没有复杂的子功能, 对应构建的类脑功能区则相对简单. 类脑功能和区域的关系将指导我们构建大规模神经网络模型.

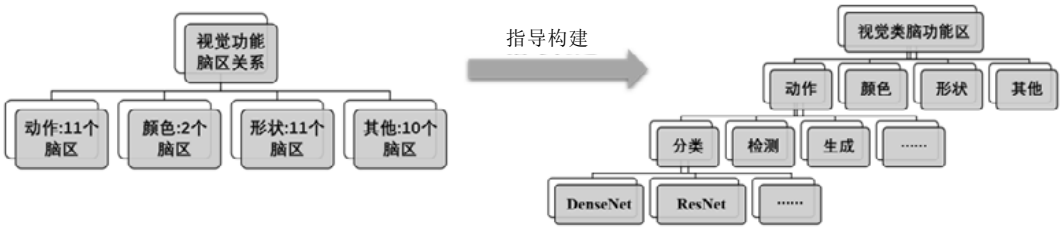


图 2 视觉脑功能为例构建类脑功能区

2.3 类脑功能区协同工作

为了使构建的类脑超大规模神经网络模型具有较好的多通道处理能力, 模型借鉴大脑脑区协作机制指导类脑功能区协同工作, 实现多通道数据处理. 在特定应用场景下, 对脑区活动进行监测, 分析脑区的在特定任务场景下的结构关联和功能关联, 并以此作为类脑功能区协作的基础. 以视听场景为例, 如图 3 所示, 假设具有视觉功能的脑区有 A, B, C, D, 具有听觉功能的脑区有 E, F, G, H. 与其对应的, 超大规模类脑功能区中, 构建的听觉子功能区应该有 a, b, c, d, 构建的听觉子功能区应该有 e, f, g, h. 如果在视听刺激下, 观测到大脑响应的脑区是 A, B, C, E, F, G, 则对应的类脑功能区中的 a, b, c, e, g, f 这 6 个子功能区应该被激活, 需要的神经网络架构和预训练模型应到从激活的类脑功能区中进行选择. 多个子功能区的协作关系,应当参照真实脑区的协作关系进行构建, 从激活的子功能区中选出多个神经网络进行融合, 最后应用到目标任务上,解决多模态问题.

综上, 脑区结构特点, 包括脑区功能的层级关系、脑功能与脑区关联度、脑区区域关系, 都将作为构建类脑功能区的重要依据. 在特定场景下, 脑区工作机制, 包括脑区结构的关联、功能关联和脑区连接关系, 将用于指导类脑功能区如何协同工作.

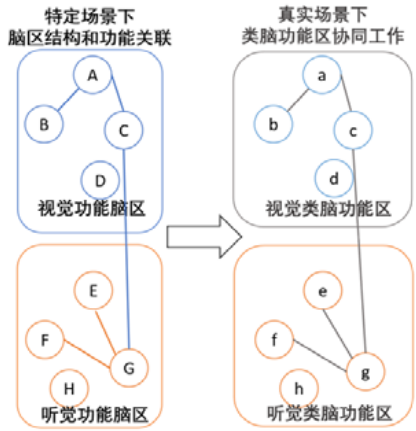


图 3 多个类脑功能区协同工作

2.4 类脑功能区数据存储和系统管理

在完成类脑功能区的建设后, 我们将目前开源或者可用的神经网络结构和预训练模型以及依赖的数据集进行整理和归纳, 按照类脑功能区的构建原则, 将所有神经网络模型映射到对应的脑功能区中. 为了实现类脑功能区的高效管理和使用. 我们设计了一套“人工编码”对神经网络模型重新进行描述, 见表 1. 为了实现类脑功能区的自动化运行, 所有神经网络都要求符合预先定义好的平台和格式才能归档至类脑功能区中.由于神经网络模型的数量和数据集规模比较大, 为了实现模型的高效访问, 神经网络以及相关的数据集以文件的方式用数据库进行管理.

表 1 人工编码-案例

模型名	功能区	功能	类别	数据集	模型规模	准确率	优化器	平台
ResNet18	图像、动作	检测、分类	ResNet	CIFAR10	2M	96%	SGD	PyTorch1.6

为此, 我们设计了一个完整的信息管理系统实现类脑功能区的管理: 首先, 将脑功能区的字典进行单独存储与维护, 作为基本模型的属性之一; 其次, 对于模型本身, 拆分成基本网络模型和预训练模型两个部分, 其中, 基本模型包含模型本身的各种属性, 包括所归属的子功能区、网络结构特征、评价指标、以及拥有的关键层等, 而针对预训练模型则记录其训练平台、数据集、优化器等信息; 最后, 对上传的神经网络模型和预训练模型分类存储, 一个简要的实体关系如图 4 所示.

图 4 类脑功能区系统的实体关系图

类脑功能区管理系统采取前后端分离的方式进行设计: 前台展示模型情况和统计信息, 同时提供模型下载功能; 后台进行类脑功能区维护. 使用模块耦合的方式对神经网络结构各个属性进行维护, 再构建预训练模型, 最后上传模型储存到对应的类脑功能区中. 对于模型的属性等其他信息, 采取分模块和字典维护, 分别存储, 最后再进行组装构建, 降低了系统各个模块的耦合性, 提高类脑功能区管理系统的可扩展性.

3 算法平台中间件

为了高效利用超大规模神经功能区, 根据不同应用场景进行神经网络模型设计, 我们构建了一个包含多个组件的算法平台中间件实现面向任务的神经网络模型设计, 完成超大规模类脑功能区和多模态应用场景衔接. 如图 1 所示, 该算法平台中间件主要功能是基于现有类脑功能区中的数据和模型, 根据应用场景需求完成新的神经网络模型设计. 目前主要包括 3 个组件.

(1) 神经网络模型设计组件, 构建面向任务的网络模型;

(2) 知识储存和学习组件, 实现预训练模型的存储和更新. 引入注意机制, 持续学习优化模型设计组件的性能;

(3) 分布式训练组件, 为神经网络模型的分布式训练和存储提供算法支撑.

3.1 模型设计组件

借鉴神经网络架构自动搜索相关研究, 整个类脑功能区包含的所有的子网络模型可以看作模型搜索空间. 模型设计组件集成了当前主流的自动化搜索方法包括: 演化计算、强化学习、梯度可微等, 更多的搜索算法正在重新设计实现、整合到模型设计组件中. 结合类脑功能区, 模型设计组件便可实现以场景为输入构建神经网络模型.

3.1.1 基于演化计算的模型设计方法

演化计算是一种基于种群的优化算法, 借鉴生物界自然选择和自然遗传机制而发展起来的通用问题求解方法. 它一般包括基因编码、种群初始化、交叉变异算子、选择保留等基本操作. 与传统的基于微积分的方法和穷举方法等优化算法相比, 进化计算是一种成熟的具有高鲁棒性和广泛适用性的全局优化方法, 具有自组织、自适应、自学习的特性, 能够不受问题性质的限制, 有效地处理传统优化算法难以解决的复杂问题, 目前被广泛用于神经网络设计、参数优化、神经网络架构设计等领域, 并取了不错的效果[24].

我们将用演化算法来实现多个网络模型的选择, 算法的示意图如图 5 所示(基因 1 和基因 2 分别对两个模态的类脑功能区进行编码, 每个基因有 3 个部分组成, 分别用来编码优化器、网络结构特征、网络规模大小. 基因 1 和基因 2 构成一个个体, 多个个体构成种群).

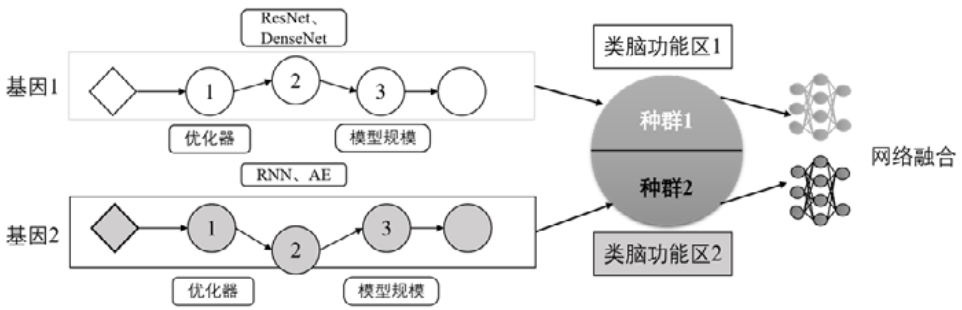


图 5 基于演化算法的模型设计组件示意图

首先是设计一种整数编码, 对神经网络模型特征进行编码, 例如优化器、网络结构特征、模型规模. 给定一个整数编码, 即可解码出一个神经网络模型. 当涉及到多个模态的网络, 根据多个功能区协同工作的机制(第 3.3 节), 分别对多个类脑功能区进行编码, 然后将多个编码信息组成一条基因, 即对应到一个个体. 个体

通过解码可以得到多个不同模态的神经网络模型, 多个网络融合后可以用于解决多模态的问题. 一个种群往往包含多个个体. 其次, 用个体在多模态目标任务上的性能指标作为个体的适应度值. 最后, 通过演化算法的演化操作: 选择、变异、交叉生成新的种群, 通过不断的演化, 最终找出适应度值最高的个体, 即是针对当前应用场景最佳的神经网络模型设计.

3.1.2 基于强化学习的模型设计方法

强化学习作为一种机器学习方法, 在很多领域获得了广泛的应用. 结合强化学习的强大能力, 我们也可以用来搜索模型的网络结构. 整个流程大概分为控制器(一般选用循环神经网络, recurrent neural network, RNN)以一定的概率生成结构 *A*, 训练模型结构 *A* 得到模型 *A* 的精度, 然后用强化学习的技术去计算梯度, 更新我们的控制器 RNN. 如此循环往复, 直到达到某个终止条件结束搜索. 利用控制器 RNN 去生成子网络方法正如文献[25]中提到的, 神经网络的结构和连接可以表示为可变长度的字符编码, 所以我们可以利用如 RNN 来生成这样的字符串, 用来表征不同的网络结构. 如图 6 所示, 以 CNN 网络为例, 控制器可以生成不同数量的卷积核、不同尺寸的卷积核(高度与宽度)、卷积滑动的高度与宽度等. 控制器本生的更新通过强化学习来实现. 强化学习的反馈(奖励)特别重要, 这里可以采用不同子网络的训练精度作为训练控制器的反馈.

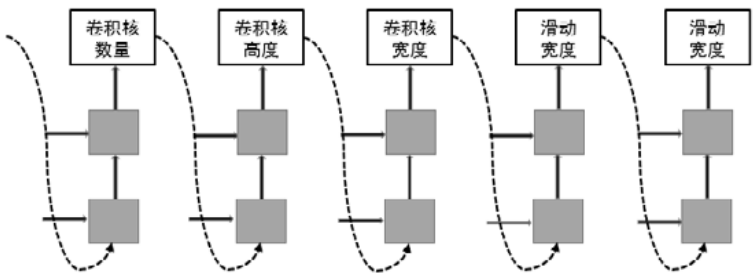


图 6 控制器生成神经网络结构序列表示

同时, 为了加速训练, 可以采用并行和异步参数更新. 我们可以用一些参数服务器(parameter server)用来存储控制器的副本. 每个控制器副本采样了几个子网络结构进行并行的训练. 然后, 控制器收集这个子网络模型的梯度发送给参数服务器, 用来更新所有控制器的参数. 由于该策略的通用性, 我们可以尝试用来联合多模态的任务进行多个神经网络模型的联合搜索. 例如在视觉和文本多模态任务中, 最简单的方案即用一些控制器来生成视觉网络, 一些控制器用来生成文本网络, 其融合网络模型在多模态任务上的精度, 作为反馈进行控制器的训练.

3.1.3 基于可微梯度的模型设计方法

基于梯度的网络结构搜索算法, 由于比强化学习和进化算法需要更少的计算资源和更少的时间, 而得到了快速的发展. 一般地, 基于梯度的网络结构搜索方法是把离散的网络结构搜索空间转化为连续的搜索空间表示. 这样, 在进行搜索和优化的时候, 能够加快搜索效率和节约搜索所消耗的资源. 其中, NAO 算法[26]就是其中的代表之一, 大致包括 3 个部分: 编码器、性能预测器以及解码器, 如图 7 所示. 其中, 通过编码器可以把离散的网络结构编码成连续的空间表示, 性能预测器则可以输出更好的网络结构嵌入表示. 最后就是解码器,是把搜索得到的更好的网络结构嵌入表示解码为最终的结构. 该算法在训练的时候, 为找到更好性能的网络结构, 一般会让这 3 个部分进行联合训练.

需要注意的是: 该方法通过编码器得到连续的空间表示后是直接通过梯度优化, 得到我们想要更好的网络结构表示. 除此之外, 在把离散的神经网络结构编码为连续的空间表示的时候, 可以采用一些数据增强的方式, 比如在文献[26]中, 作者通过增加额外的相似性能对, 提高了在 CIFAR-10 数据集[27]上的精度. 而且作者在解码器部分添加了注意力机制, 使得把连续的空间表示解码为离散的网络结构的时候变得更容易. 另外, 如果能把融合后的多模态网络进行编码, 就可以采用 NAO 算法来进行多模态的搜索.

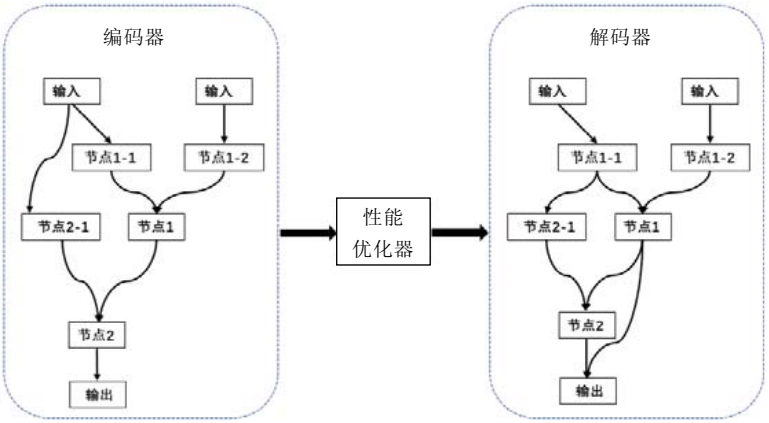


图 7 NAO 算法流程图

3.2 知识存储与学习组件

当前的神经网络是一种端到端的学习方式, 最终实现输入空间到输出空间的映射, 空间的映射质量直接表现为在特定任务上的工作性能. 为了实现神经网络的知识存储, 进而达到语义分析和推理, 我们认为: 除了脑启发的神经网络架构之外, 还需要大量训练数据和大规模网络结构将学习到的知识存储到神经网络中. 例如自然语言处理领域的 GPT3 模型[5], GPT3 通过足够多的数据训练和庞大的神经网络架构, 在自然语言处理领域多个子任务取得最佳效果. 所以在类脑超大规模神经网络模型中, 我们将构建知识存储和学习组件, 主要功能是对已有知识进行存储和学习、辅助类脑功能区在特定任务场景下的神经网模型设计. 该组件要存储的是经过预训练的、专门用于感知的(如视觉分类、听觉分类、自然语言处理)预训练模型, 并能通过新的数据对预训练模型进行更新. 此外, 在类脑神经网络模型应对新任务进行神经网络模型设计时, 如何选择预训练模块也非常关键, 决定了新神经网络模型对输入/输出空间的映射能力. 为了更好地利用这些预训练模型, 该组件引入人脑注意力机制, 在不同应用场景下, 选择一些关键的信息输入进行处理, 选出最佳的预训练模型, 辅助模型设计组件构建新的神经网络模型. 通过存储组件的更新和注意力机制的增量学习, 使得整个类脑神经网络模型拥有持续学习能力, 不断优化面向任务的神经网络模型设计过程.

3.3 分布式训练组件

神经网络训练需要大量的数据进行训练, 才能在目标任务达到很好的性能. 随着训练数据规模和神经网络规模的增大, 神经网络的训练需要耗费大量的时间. 在我们构建的大规模类脑功能区中, 预训练模型达到 100 多个, 并且还在不断扩充, 整个类脑的大规模神经网络模型的参数量已达百亿. 为了实现各个类脑功能区高效的存储和运用, 在算法平台中间件中, 我们构建了一个分布式训练模块用于支持整个大规模神经网络模型的运行, 主要包含两个部分: 基于多节点的分布式训练平台和大规模神经网络分布式训练算法.

3.3.1 基于多节点的分布式训练平台

在模型设计组件中, 每个多模态网络需要在目标数据集上进行验证, 得到的性能指标将作为反馈继续指导类脑功能区的协同工作. 为了提升多模态网络的验证过程, 我们构建了一个基于计算节点的分布式训练平台[28], 如图 8 所示. 整个集群由两类节点组成: 服务节点和计算节点, 所有节点通过 TCP/IP 网络连接. 在我们构建的分布式集群中, 有 1 个服务器节点, 该节点主要负责集群的管理以及类脑功能区的管理. 多个计算节点主要负责多模态网络在数据集上进行验证, 所有节点异步向服务器节点发起任务请求. 该平台的特点是对计算节点性能没有一致性要求, 任何性能的计算节点都可以加入该集群, 提升分布式计算能力.

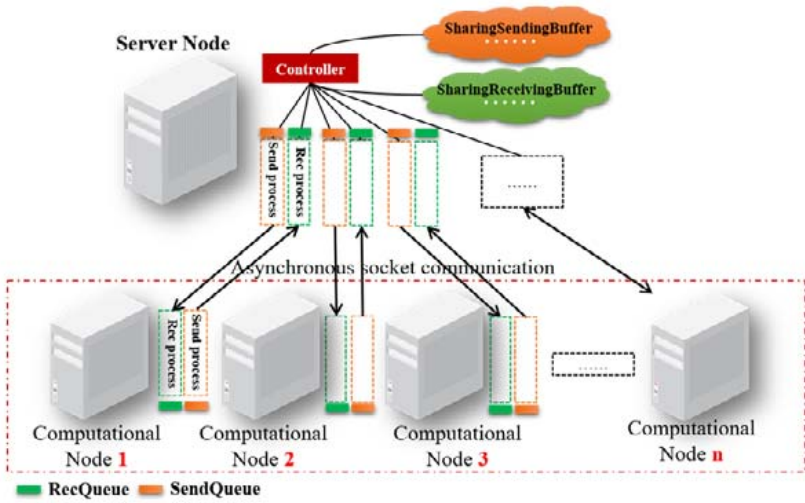
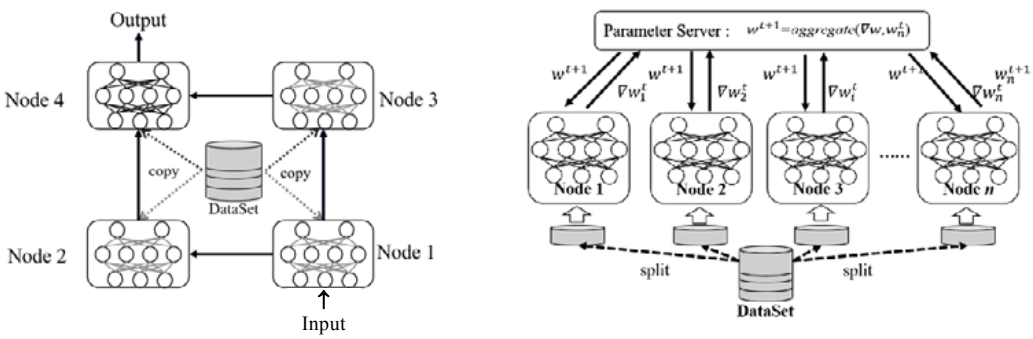


图 8 基于多计算节点的分布式平台

3.3.2 大规模神经网络分布式训练算法

针对单个神经网络模型的分布式训练算法, 主要分为两大类[29,30].

- 一是模型并行: 分布式系统中的不同节点负责单个网络不同部分的训练, 如图 9(a)所示, 神经网络的每个层可以被分配给不同计算节点进行训练, 整个训练过程如流水线. 该方法的优势在于神经网络模型可以分布式存储, 减小存储压力; 缺点在于神经网络在训练过程中, 各个节点间需要通信开销, 且实现困难, 模型一致性难以保证;
- 二是数据并行: 不同计算节点具有完整的模型副本和目标数据集的部分子集. 每个计算节点基于分配的子数据集对神经网络进行训练, 参数服务器收集各个计算节点的神经网络参数或者梯度, 并计算新的参数, 最后分发新参数给每个计算节点, 如图 9(b)所示, i 表示参数版本. 数据并行训练方法的关键在于每个计算节点训练结束后, 需要完成模型参数同步. 该方法的优势在于实现简单, 如果每次训练结束后的数据都同步, 则等同于单个计算节点的训练, 收敛效率有保障. 该方法的缺点在于: 同步频率过高, 会因为通信和同步开销过大导致集群训练效率降低; 同步频率过低, 会导致收敛效率降低甚至不收敛.



(a) 模型并行 (b) 数据并行

图 9 分布式训练架构

为了提高集群的训练效率, 降低同步的开销, 基于现有的数据并行模式的算法, 我们提出了一系列改进算法并集成到了分布式训练平台中. 例如: 基于粒子群的分布式训练优化方法[31], 该算法利用粒子群算法的

优化过程来建模分布式训练的参数更新过程, 提高分布式训练的准确率; 动态 Batch-size 的分布式训练方法[32], 该方法根据节点的计算能力动态调整集群的负载, 消除节点间等待的时间, 提高集群节点的利用率, 从而提高训练效率. 除此之外, 我们还提出一种并行训练过程和参数同步过程的方法[33], 在保证收敛效率不降低的前提下, 降低通信开销, 提升集群的训练效率.

4 应用场景

目前, GPT3, Switch Transformer[34]等大模型的训练和计算资源成本太高, 无法直接用到我们的对比实验中. 为了验证本文提出的类脑大规模神经网络模型的可行性, 我们分别在单模态和多模态选定了一个应用场景, 利用该模型构建神经网络进行实验, 并对结果进行简要分析.

4.1 单模态应用场景研究

类脑大规模神经网络系统搭建完成后, 算法平台需要在特定的应用场景下进行模型的搜索, 本节将介绍基于视觉功能区的单模态应用场景的构建. 我们构建的单模态应用场景为多目标跟踪, 并且这里的多目标追踪特指基于检测的多目标跟踪.

4.1.1 背景

多目标追踪是计算机视觉领域中的经典任务, 深度学习的发展, 极大地推动了这一领域的发展, 其中, 利用深度神经网络学习目标检测的表观特征, 是简单有效提升多目标跟踪的方法. 如今有多种多样的方法来实现多目标跟踪, 但本节介绍的是基于检测的多目标跟踪方法. 基于检测的多目标跟踪实现流程(以检测视频中的多目标为例)如下.

(1) 给定视频的原始帧;

(2) 运行目标检测模型获得给定视频原始帧中对象的边界框;

(3) 对于每个检测到的物体, 计算不同的特征;

(4) 对不同的特征进行相似度计算;

(5) 对不同帧的对象进行关联, 为每个对象分配数字 ID.

流程图如图 10 所示, 可以看出, 本节介绍的这种基于检测的多目标跟踪方法并不是一个端到端的过程. 我们可以将其看成是一个 two-stage 的过程: 第 1 步, 利用目标检测模型进行物体检测, 即上述实现流程中的步骤(1)、步骤(2); 第 2 步, 将检测到的物体进行关联, 即上述实现流程中的步骤(3)~步骤(5). 其中, 第 1 步中的基于深度学习的目标检测模型按照是否是端到端的过程可以分为 one-stage 和 two-stage 两大阵营, 比如常见的 one-stage 的目标检测模型有 YOLO, SSD 等, two-stage 的目标检测模型有 RCNN, Fast-RCNN 等. 基于深度神网络的目标检测模型有很多种, 而我们所构建的单模态应用场景即多目标跟踪又是一个 two-stage 的过程,因此, 我们可以使用某种搜索算法去搜索视觉功能区中的目标检测模型来替换多目标追踪框架的第一步中的目标检测模型.

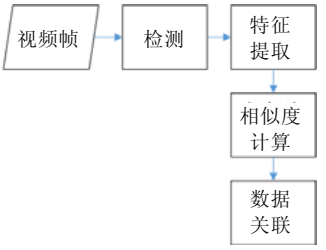


图 10 多目标追踪流程图

4.1.2 实验设计

视觉功能区的组织形式如图 11 所示, 其中, 检测这一层级中有 Faster R-CNN, SSD, YOLOv3 等目标检测模型. 这些模型的源文件和预训练好的参数都已经存储在我们的视觉功能区中. 实验选取的多目标跟踪框架

为 deep-sort[35], deep-sort 算法是由 sort[36]算法改进而来. sort 和 deep-sort 算法的实现流程与第 4.1.1 节所介绍的基于检测的多目标跟踪算法实现流程一致, 同样是分为两个步骤: 第 1 步, 先使用目标检测模型对物体进行检测; 第 2 步, 使用某种关联算法对检测到的物体进行关联. deep-sort 与 sort 不同的是, sort 算法在第 2 步使用的是简单的卡尔曼滤波处理逐帧数据的关联性以及使用匈牙利算法进行关联度量. 这种算法虽然在高帧速率下可以表现出比较好的性能, 但有一个比较大的缺陷就是它没有考虑到被检测物体的表面特征. 在 deep-sort 算法中, 采用目标运动与表面特征信息相结合的方式来代替 sort 算法中的关联度量; 并且使用卷积神经网络在大规模行人数据集上进行训练, 提取特征, 这样做的目的是可以提升网络对处理目标消失或遮挡的鲁棒性. 实验使用算法平台从视觉功能区中的检测子功能区中选择出一个目标检测模型, 用于构成 deep-sort 算法中的目标检测模块.

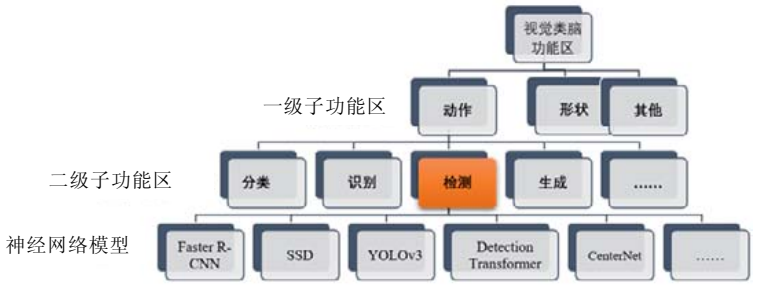


图 11 视觉功能区

对检测子功能区中的目标检测模型进行简单编码作为视觉功能区的搜索空间, 见表 2.

表 2 视觉功能区目标检测模型编码

目标检测模型	编码
DETR	001
SSD	002
YOLOv3	003
Faster R-CNN	004
...	...

4.1.3 实验结果

基于模型构建组件和视觉功能区中的检测子模型, 利用现有的 deep-sort 模型的主干网络构建新的跟踪模型, 并在 MOT16 数据集[37]上进行测试. MOT16 数据集是 2016 年提出的多目标跟踪 MOT Challenge 系列的一个衡量多目标检测跟踪方法标准的数据集, 其中, 搜索到的 DETR, SSD, YOLOv3 模型将用于 deep-sort 中, 并且在 MOT16 数据集上进行测试, 实验结果见表 3–表 5.

多目标追踪任务的评价指标有很多, 这里只选取了 4 个比较重要的评价指标, 其中, *idf*1 指的是正确识别的检测与真实数和计算检测的平均数之比, 取值范围为 0 到 1, 越接近 1 越好; switches 指的是 Ground Truth 所分配的 ID 发生变化的次数, 即 ID 被错误改变的次数, 值越小越好; mota 衡量的是检测物体和保持轨迹时的性能, 当跟踪器产生的错误超过了场景中的物体, 其值就为负数, mota 的取值范围理论上为负无穷到 1, 越接近 1 越好; motp 主要量化检测器的定位精度, 取值范围为 0 到 1, 越接近 1 越好. 实验结果表明: 使用模型设计组件选择视觉功能区中的目标检测模型用于替换 deep_sort 框架中的检测模块是可行的, 并且视觉功能区中不同的目标检测模型在多目标追踪任务中表现出了不同的性能. 根据实验结果可知, 使用模型设计组件在视觉功能区中搜索出的 YOLOv3 作为检测框架的 deep_sort 在 MOT16 上的表现要优于 DETR 和 SSD 作为检测框架的 deep_sort 在 MOT16 上的表现. 这说明在多目标追踪这一应用场景之下, 类脑功能区能够支撑我们的搜索算法, 并为我们的多目标追踪这一应用场景提供一个适合、高效的目标检测模型.

表 3 deep_sort_DETR

	idf1	switches	mota	motp
MOT16-02	0.28	81	0.21	0.31
MOT16-04	0.35	62	0.26	0.28
MOT16-05	0.37	66	0.42	0.30
MOT16-09	0.48	34	0.54	0.27
MOT16-10	0.40	83	0.36	0.30
MOT16-11	0.41	29	0.46	0.26
MOT16-13	0.24	95	0.10	0.36
OVERALL	0.35	**450**	0.29	0.29

表 4 deep_sort_YOLOv3

	idf1	switches	mota	motp
MOT16-02	0.28	72	0.20	0.32
MOT16-04	0.40	134	0.31	0.29
MOT16-05	0.40	66	0.44	0.30
MOT16-09	0.39	59	0.52	0.27
MOT16-10	0.31	120	0.29	0.30
MOT16-11	0.37	34	0.50	0.26
MOT16-13	0.22	131	0.10	0.35
OVERALL	0.35	616	**0.30**	**0.30**

表 5 deep_sort_SSD

	idf1	switches	mota	motp
MOT16-02	0.28	72	0.20	0.32
MOT16-04	0.40	134	0.31	0.29
MOT16-05	0.40	66	0.44	0.30
MOT16-09	0.39	59	0.52	0.27
MOT16-10	0.31	120	0.29	0.30
MOT16-11	0.37	34	0.50	0.26
MOT16-13	0.22	131	0.10	0.35
OVERALL	0.35	616	**0.30**	**0.30**

4.2 多模态应用场景研究

4.2.1 背 景

研究者们对于多模态有着各自不同的认识, 但都对其抽象定义有着普遍共识, 即: 对于同一个事物有不同角度的观察, 每一个角度成为一个模态; 多个角度构成多个模态. 当前, 研究者对于多模态领域内的问题的基本假设为: 单个模态对于事物的描述是有偏的, 通过多个模态来描述该事物能够减少这种偏移, 使描述或者表示更加饱满和准确. 多模态下, 事物的统一表示有联合表示(joint representation)和协同表示(coordinate representation)两种方法[38]. 通常, 构建联合表示的核心问题之一即为如何将多个模态的数据进行融合. 一般可采用的数据融合操作为向量的连接、加权和、多线性池化和基于注意力机制的融合等[39], 然而不同的融合操作有不同性质, 这些性质影响着不同任务的最终表现. 而协同表示则类似于坐标系下的坐标, 用单个模态的表示作为多模态表示的一个分量. 协同表示的各个模态分量一般需要通过结构或者相似性约束来建模各个模态之间的内在联系.

人脑每天都在接受大量多模态数据, 因此, 作为类脑大规模神经网络模型研究, 我们的平台同样需要面对多模态应用问题. 如果在这个平台上进行人为的多模态网络设计, 要求设计者在选取融合操作时具有专家知识. 同时, 平台在多模态应用上有着双重多样性, 即多模态任务的多样性和可选网络的多样性. 因此对于每个任务都人为地设计是不可行的和有偏的, 同时也是不符合人脑学习机制的. 当前最新的研究表明, 多模态数据融合可以通过网络结构搜索完成. 日前, 有研究者提出了用于多模态网络结构搜索的框架 MFAS (multi-modal fusion architecture search)[40], 该框架搜索了单模态数据与多模态融合操作之间的连接, 即多模态融合操作的数据来源.

4.2.2 基于类脑模神经网络模型的多模态应用

一般的, 多模态数据都选取高层特征来进行融合, 例如卷积神经网络(convolution neural network)的特征图(feature map)或者循环神经网络(recurrent neural network)的特征向量等. MFAS[40]利用不同特征提取器构建

不同模态的高层特征, 并从这些特征提取器中搜索了不同模态的特征组合来完成多次多模态特征融合. 我们将 MFAS[40]研究整合进入我们提出的类脑网络模型的算法平台, 并在多模态数据集 Audio-visual Mnist[41]上进行了分类任务验证. 在我们提出的类脑模型中, 特征提取器, 即模型组件由类脑网络模型的算法中间件提供. 同时, 每个模型组件都会进行单模态数据的预训练. 最后, 自动化地完成多模态融合网络的构建. 综上所述, 类脑神经网络模型的多模态问题的应用过程如下.

(1) 算法平台中间件提供单模态的模型组件设计;

(2) 对单模态模型进行预训练;

(3) 采用预训练模型组件进行多模态融合网络的自动化构建.

4.2.3 实 验

- 数据预处理

首先, 我们基于 CentralNet 提出的方法[41], 手工构建了多模态数据分类任务数据集 Adudio-visual Mnist. 它有音频和图像两个模态: 音频模态的数据来源于 FSDD[42]和 ESC-50 数据集[43]; 图像模态数据来源于 Mnist 数据集. 在构建多模态数据集时, 由于数据是离散的, 所以在构建时需要人为地进行数据对齐. 对于音频模态, 我们首先从 ESC-50 数据集中随机采样音频作为环境噪声; 接着, 将音频预处理成 112×112 大小的频谱图(spectrogram); 而图像模态会经过 PCA 预处理, 去除其中 75%的能量.

- 网络结构

实验过程中, 我们向类脑功能区中加入了多个单模态的模型, 其卷积层相关信息见表 6, 每一个卷积层后均有 kernel 大小为 2 的 MaxPool2d 层和 LeakyRelu 激活层, Input 表示输入示例大小.

表 6 模型组件网络结构

Name	Layer Ops	Output size
LeNet3	Input	28×28×1
	Conv2d, *kernel_size*=5	28×28×8
	Conv2d, *kernel_size*=3	14×14×16
	Conv2d, *kernel_size*=3	7×7×32
LeNet5	Input	112×112×1
	Conv2d, *kernel_size*=5	112×112×8
	Conv2d, *kernel_size*=3	56×56×16
	Conv2d, *kernel_size*=3	28×28×32
	Conv2d, *kernel_size*=3	14×14×64
	Conv2d, *kernel_size*=3	7×7×128
MediumNet	Input	112×112×1
	Conv2d, *kernel_size*=3	112×112×8
	Conv2d, *kernel_size*=3	56×56×16
	Conv2d, *kernel_size*=3	28×28×32
	Conv2d, *kernel_size*=3	14×14×64
SparseNet	Input	112×112×1
	Conv2d, *kernel_size*=3	112×112×12
	Conv2d, *kernel_size*=1	112×112×1
	DepthwiseConv2d[45], *kernel_size*=5	112×112×8
	DepthwiseConv2d, *kernel_size*=3	56×56×16
	DepthwiseConv2d, *kernel_size*=3	28×28×32
	DepthwiseConv2d, *kernel_size*=3	14×14×64
	DepthwiseConv2d, *kernel_size*=3	7×7×128

- 实验设计

由算法平台中间件从类脑功能区中选取两个单模态模型进行预训练, 随后利用它们进行多模态融合搜索, 尝试自动化的构建用于解决多模态问题的多模态融合网络. 为了验证自动化构建融合网络的可行性, 实验时, 我们对所有模型组件的组合情况均进行了实验. 表 7 展示了单模态模型的预训练分类结果. 在进行融合结构搜索时, 搜索算法采用由 Zhou 等人[44]提出的层次递进的算法, 同时设置最大搜索深度为 4, 即搜索 4 个特征融合层, 并采取特征向量拼接的方式进行融合.

表 7 模型组件预训练结果

Modal	Name	Acc
Image	LeNet3	52.52±0.24
	LeNet5	54.17±0.58
	MediumNet	53.63±0.16
	SparseNet	26.44±2.24
Audio	LeNet3	67.86±1.49
	LeNet5	95.50±0.23
	MediumNet	97.14±0.14
	SparseNet	48.36±6.52

- 实验结果

首先, 对于单模态的实验结果, 我们发现: 对于图像, LeNet3, LeNet5, MediumeNet 表现相差不大, 处于正常水平; 但是 SparseNet 的性能较差, 出现欠拟合现象. 欠拟合是因为 SparseNet 相较于 LeNet5 和 MediumNet 参数量更少而层数更深, 在优化时梯度容易消失, 较难训练至拟合; 即使其比 LeNet3 参数量有提升, 可是仍然难以遏制层数所引起的的梯度消失问题. 而至于音频, LeNet3 表现较差是因为其参数量较 MediumNet 和 LeNet5 更少, 使模型的编码能力下降. 其次, 在进行多模态特征融合搜索时, 实验结果表明: (1) 模型是否过拟合或者欠拟合会极大地影响融合特征的表达能力; (2) 合理收敛的单模态模型的特征在融合后表达能力将会得到提升, 即融合后的特征在具体任务上表现更好. 例如: SparseNe 所提取的图像特征在与其他模型提取的音频特征融合后, 得到的表示表现均为超过单模态模型的表现. 同时, 实验表明, 过拟合模型所提取的特征同样会影响融合表示的表达能力. 例如: MediumNet 提取的音频特征在与其他模型提取的图像特征融合时, 出现了较大的性能损失; LeNet5 提取的音频特征和图像特征融合时, 其实验结果同样也印证上述结论. 当图像和音频模态特征均被一个较好模型(既不是欠拟合模型也不是过拟合模型)提取时, 其融合表示的表达能力相较于单模态的表示有所提升, 如表 8 所示. 综上所述,当被用于搜索的模型是一个收敛到合理范围的模型时, 我们融合搜索算法能够搜索得到用于融合操作的最优特征组合, 并提升单模态模型的表现.

表 8 多模态融合结果

Image	Unimodal Acc	Audio	Unimodal Acc	Fusion Acc
LeNet3	52.52	LetNet5	95.50	95.45
		SparseNet	48.36	**68.60**
		MediumNet	97.14	52.18
LeNet5	54.17	LetNet3	67.86	**73.83**
		SparseNet	48.36	**69.98**
		MediumNet	97.14	54.06
SparseNet	26.44	LetNet3	67.86	63.52
		LetNet5	95.50	93.11
		MediumNet	97.14	47.35
MediumNet	53.63	LetNet3	67.86	**74.40**
		LetNet5	95.50	93.86
		SparseNet	48.36	**62.03**

5 总结和展望

神经网络面向任务端到端的学习方式确定了输入/输出的相关关系, 大规模神经网络可以为巨大的输入/输出空间之间建立良好的相关关系, 已经成为一个重要的发展方向. 但按照现有的神经网络组织方式, 目前的大规模网络是难以到达人脑生物神经网络连接的规模. 本文提出一种受脑功能机制启发的超大规模深度神经网络模型构建方法, 完成以场景作为输入自动构建模型的通路. 整个方案由 4 层架构组成, 包含理论基础、类脑功能区构建、算法平台中间件研究、应用场景可行性验证. 首先, 基于脑区功能研究成果, 模块化构建类脑功能区, 整个类脑大规模神经网络模型参数超过百亿, 并可以根据需要不断扩充; 其次, 基于脑功能区域协作关系, 建立超大规模神经网络模型的协同机制, 提出相应的学习算法; 依托构建的类脑功能区和算法平台中间件, 实现针对特定场景的神经网络模型设计; 设计分布式训练平台及其训练算法, 支持大规模神经网络训练; 最后, 在多个应用场景下, 对整个类脑大规模神经网络模型进行可行性验证, 所有数据和类脑功能

区使用数据库统一管理.

该类脑超大规模神经网络系统由脑功能协作机制指导, 集成了海量数据、蕴含知识的预训练模型、各类训练算法以及分布式训练算法, 可以直接用于构建单模态或者多模态神经网络模型.

新一代神经网络的目标是像人脑一样具备多通道协同处理、知识存储和迁移能力, 通过不断的学习, 构建有效、鲁棒的输入到输出空间映射关系, 解决复杂的多模态应用问题. 本文提出的大规模神经网络模型是受脑功能启发, 实现新一代神经网络模型设计的一次尝试, 旨在跨越数据特征空间学习到语义分析和推理间的障碍, 为实现通用人工智能提供一种研究思路. 鉴于目前有限的研究成果和计算资源, 该系统还需要在更多复杂应用场景下进行探索. 在未来的工作中, 我们将借鉴前沿的人脑工作机理研究成果, 不断对类脑大规模神经网络模型进行优化, 同时提高整个类脑神经网络规模、集成更多的预训练模型和数据集. 加强算法平台中间件的研究, 提升新一代神经网络模型对输入/输出空间的映射能力、持续学习能力, 提高整个神经网络模型的迁移性、鲁棒性.

References:

[1] He K, Zhang X, Ren S, *et al*. Deep residual learning for image recognition. In: Proc. of the 2016 IEEE Conf. on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society, 2016. 770−778.

[2] Krizhevsky A, Sutskever I, Hinton G. ImageNet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems. 2012, 25: 1106−1114.

[3] Kingsbury B. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. IEEE Signal Processing Magazine, 2012, 29(6): 82−97.

[4] Miao Y, Gowayyed M, Metze F. EESEN: End-to-end speech recognition using deep RNN models and wfst-based decoding. In: Proc. of the 2015 IEEE Workshop on Automatic Speech Recognition and Understanding. Los Alamitos: IEEE Computer Society, 2015. 167−174.

[5] Brown TB, Mann B, Ryder N, *et al*. Language models are few-shot learners. In: Advances in Neural Information Processing Systems. 2020, 33.

[6] Brock A, Donahue J, Simonyan K. Large scale gan training for high fidelity natural image synthesis. In: Proc. of the 7th Int'l Conf. on Learning Representations. 2019.

[7] Devlin J, Chang M-W, Lee K, *et al*. Bert: Pre-training of deep bidirectional transformers for language understanding. In: Proc. of the 2019 Conf. of the North American Chapter of the Association for Computational Linguistics. Stroudsburg: ASSOC Computational Linguistics, 2019. 4171−4186.

[8] Hu J, Shen L, Sun G. Squeeze-and-excitation networks. In: Proc. of the 2018 IEEE Conf. on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society, 2018. 7132−7141.

[9] Chen YJ, Huang TJ, Pan G, *et al*. Research progress and development trend of brain-like computing. In: China Computer Federation Proceedings CCFP 0027. Beijing: China Machine Press, 2016. 263−317 (in Chinese with English abstract).

[10] Yang GW, Wang SJ, Yan QX. Research of fractional linear neural network and its ability for nonlinear approach. Chinese Journal Of Computers, 2007, 30(2): 189−199 (in Chinese with English abstract).

[11] Huang Y, Cheng Y, Bapna A, *et al*. Gpipe: Efficient training of giant neural networks using pipeline parallelism. In: Advances in Neural Information Processing Systems. 2019, 32: 103−112.

[12] Vaswani A, Shazeer N, Parmar N, *et al*. Attention is all you need. In: Advances in Neural Information Processing Systems. 2017, 30: 5998−6008.

[13] Shazeer N, Mirhoseini A, Maziarz K, *et al*. Outrageously large neural networks: The sparsely-gated mixture-of-experts layer. In: Proc. of the 5th Int'l Conf. on Learning Representations. 2017.

[14] Ma L, Yang Z, Miao Y, *et al*. Towards efficient large-scale graph neural network computing. arXiv: 1810.08403, 2018.

[15] Qi D, Su L, Song J, *et al*. Imagebert: Cross-modal pre-training with large-scale weak-supervised image-text data. arXiv: 2001.07966, 2020.

[16] Xu B, Liu CL, Zeng Y. Research status and developments of brain-inspired intelligence. Bulletin of the Chinese Academy of Sciences, 2016, 31(7): 793−802 (in Chinese with English abstract).

[17] Deng H, Huang XQ, Zhang W, *et al*. A method and system for robot navigation based on multimode perception and reinforcement learning. CN202010157337.9, 2020 (in Chinese).

[18] Liang M, Hu X. Recurrent convolutional neural network for object recognition. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society, 2015. 3367−3375.

[19] Liang M, Hu X, Zhang B. Convolutional neural networks with intra-layer recurrent connections for scene labeling. In: Cortes C, Lawrence ND, Lee DD, *et al*., eds. Advances in Neural Information Processing Systems. 2015, 28: 937−945.

[20] Wang Q, Zhang J, Song S, *et al*. Attentional neural network: Feature selection using cognitive feedback. In: Advances in Neural Information Processing Systems. 2014, 27: 2033−2041.

[21] Cao C, Liu X, Yang Y, *et al*. Look and think twice: Capturing top-down visual attention with feedback convolutional neural networks. In: Proc. of the 2015 IEEE Int'l Conf. on Computer Vision. Los Alamitos: IEEE Computer Society, 2015. 2956−2964.

[22] Pei J, Deng L, Song S, *et al*. Towards artificial general intelligence with hybrid tianjic chip architecture. Nature, 2019, 572(106): 106−111.

[23] Liu T. A few thoughts on brain rois. Brain Imaging and Behavior, 2011, 5(3): 189−202.

[24] Al-Sahaf H, Bi Y, Chen Q, *et al*. A survey on evolutionary machine learning. Journal of the Royal Society of New Zealand, 2019, 49(2): 205−228.

[25] Zoph B, Le QV. Neural architecture search with reinforcement learning. In: Proc. of the 5th Int'l Conf. on Learning Representations. 2017.

[26] Luo R, Tian F, Qin T, *et al*. Neural architecture optimization. In: Advances in Neural Information Processing Systems. 2018, 31: 7827−7838.

[27] Krizhevsky A, Hinton G, *et al*. Learning multiple layers of features from tiny images. Technical Report, 2009.

[28] Ye Q, Sun Y, Zhang J, *et al*. A distributed framework for ea-based nas. IEEE Trans. on Parallel and Distributed Systems, 2021, 32(7): 1753−1764.

[29] Dean J, Corrado GS, Monga R, *et al*. Large scale distributed deep networks. In: Advances in Neural Information Processing Systems. 2012, 25: 1232−1240.

[30] Zhang S, Choromanska AE, LeCun Y. Deep learning with elastic averaging SGD. In: Advances in Neural Information Processing Systems. 2015, 28: 685−693.

[31] Ye Q, Han Y, Sun Y, *et al*. PSO-ps: Parameter synchronization with particle swarm optimization for distributed training of deep neural networks. In: Proc. of the 2020 Int'l Joint Conf. on Neural Networks. Los Alamitos: IEEE Computer Society, 2020. 1−8.

[32] Ye Q, Zhou Y, Shi M, *et al*. DBS: Dynamic batch size for distributed deep neural network training. arXiv: 2007.11831, 2020.

[33] Zhou Y, Ye Q, Zhang H, *et al*. HPSGD: Hierarchical local sgd with stale gradients featuring. In: Proc. of the 27th Int'l Conf. on Neural Information Processing. New York: Springer, 2020. 440−451.

[34] Fedus W, Zoph B, Shazeer N. Switch transformers: Scaling to trillion parameter models with simple and efficient sparsity. arXiv: 2101.03961, 2021.

[35] Wojke N, Bewley A, Paulus D. Simple online and realtime tracking with a deep association metric. In: Proc. of the 2017 IEEE Int'l Conf. on Image Processing. Los Alamitos: IEEE Computer Society, 2017. 3645−3649.

[36] Bewley A, Ge ZY, Ott L, *et al*. Simple online and realtime tracking. In: Proc. of the 2016 IEEE Int'l Conf. on Image Processing. Los Alamitos: IEEE Computer Society, 2016. 3464−3468.

[37] Milan A, Leal-Taixé L, Reid I, *et al*. MOT16: A benchmark for multi-object tracking. arXiv: 1603.00831, 2016.

[38] Baltrušaitis T, Ahuja C, Morency LP. Multimodal machine learning: A survey and taxonomy. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2019, 41(2): 423−443.

[39] Zhang C, Yang Z, He X, *et al*. Multimodal intelligence: Representation learning, information fusion, and applications. IEEE Journal of Selected Topics in Signal Processing, 2020, 14(3): 478−493.

[40] Rúa P, Manuel J, Vielzeuf V, *et al*. MFAS: Multimodal fusion architecture search. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society, 2019. 6966−6975.

[41] Vielzeuf V, Lechervy A, Pateux S, *et al*. CentralNet: A multilayer approach for multimodal fusion. In: Leal-Taixé L, Roth S, eds. Proc. of the Computer Vision—ECCV 2018 Workshops. New York: Springer, 2019. 575–589.

[42] Jackson Z, Souza C, Flaks J, *et al*. Jakobovski/Free-Spoken-Digit-Dataset: V1.0.8. Zenodo, 2018.

[43] Piczak KJ. ESC: Dataset for environmental sound classification. In: Proc. of the 23rd Annual ACM Conf. on Multimedia Conf. New York: Association for Computing Machinery, 2015. 1015–1018.

[44] Zhou Y, Wang P, Arik S, *et al*. EPNAS: Efficient progressive neural architecture search. In: Proc. of the 30th British Machine Vision Conf. 2019. Guildford: BMVA Press, 2019. 71.

[45] Chollet F. Xception: Deep learning with depthwise separable convolutions. In: Proc. of the 2017 IEEE Conf. on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society, 2017. 1800–1807.

附中文参考文献:

[9] 陈云霁, 黄铁军, 潘纲, 等. 类脑计算的研究进展与发展趋势. 见: 中国计算机学会文集 CCF 2015-2016 中国计算机科学技术发展报告. 北京: 机械工业出版社, 2016. 263–317.

[10] 杨国为, 王守觉, 闫庆旭. 分式线性神经网络及其非线性逼近能力研究. 计算机学报, 2007, 30(2): 189–199.

[16] 徐波, 刘成林, 曾毅. 类脑智能研究现状与发展思考. 中国科学院院刊, 2016, 31(7): 793–802.

[17] 邓寒, 黄学钦, 张伟, 等. 一种基于多模感知与强化学习的机器人导航方法及系统. CN202010157337.9, 2020.

吕建成(1973－), 男, 博士, 教授, 博士生导师, CCF 高级会员, 主要研究领域为神经网络基础理论, 自然语言处理, 智慧医疗, 智慧文旅, 工业智能化.

叶庆(1989－), 男, 博士生, 主要研究领域为神经网络分布式训练, 联邦学习.

田煜鑫(1998－), 男, 博士生, 主要研究领域为深度学习及其应用.

韩军伟(1977－), 男, 博士, 教授, 博士生导师, CCF 杰出会员, 主要研究领域为人工智能, 模式识别, 类脑计算, 遥感影像解译.

吴枫(1969－), 男, 博士, 教授, 博士生导师, CCF 杰出会员, 主要研究领域为视频编码与通信, 多媒体内容分析, 计算机视觉.