DOI:10.11992/tis.201701006
网络出版地址:http://kns.cnki.net/kcms/detail/23.1538.TP.20170419.1448.002.html

智能交互的物体识别增量学习技术综述

李雪[1,2],蒋树强[2]
(1.山东科技大学 计算机科学与工程学院,山东 青岛 266590; 2. 中国科学院计算技术研究所 智能信息处理重点实验室,北京 100190)

摘　要:智能交互系统是研究人与计算机之间进行交流与通信,使计算机能够在最大程度上完成交互者的某个指令的一个领域。其发展的目标是实现人机交互的自主性、安全性和友好性。增量学习是实现这个发展目标的一个途径。本文对智能交互系统的任务、背景和获取信息来源进行简要介绍,主要对增量学习领域的已有工作进行综述。增量学习是指一个学习系统能不断地从新样本中学习新的知识,非常类似于人类自身的学习模式。它使智能交互系统拥有自我学习,提高交互体验的能力。文中对主要的增量学习算法的基本原理和特点进行了阐述,分析各自的优点和不足,并对进一步的研究方向进行展望。

关键词:人工智能;人机交互;计算机视觉;物体识别;机器学习;多模态;机器人;交互学习
中图分类号: TP391　　**文献标志码**:A　　**文章编号**:1673-4785(2017)02-0140-10

中文引用格式:李雪,蒋树强. 智能交互的物体识别增量学习技术综述[J]. 智能系统学报, 2017, 12(2): 140-149.
英文引用格式:LI Xue, JIANG Shuqiang. Incremental learning and object recognition system based on intelligent HCI: a survey [J]. CAAI transactions on intelligent systems, 2017, 12(2): 140-149.

Incremental learning and object recognition system based on intelligent HCI: a survey

LI Xue [1], JIANG Shuqiang [2]
(1. College of Information Science and Engineering, Shandong University of Science and Technology, Qingdao 266590, China; 2. Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China)

Abstract: Intelligent HCI systems focus on the interaction between computers and humans and study whether computers are able to apprehend human instructions. Moreover, this study aims to make the interaction more independent and interactive. To some extent, incremental learning is a way to realize this goal. This study briefly introduces the tasks, background, and information source of intelligent HCI systems; in addition, it focuses on the summary of incremental learning. Similar to the learning mechanism of humans, incremental learning involves acquiring new knowledge on a continuous basis. This allows for the intelligent HCI systems to have the ability of self-growth. This study surveys the works that focus on incremental learning, including the mechanisms and their respective advantages and disadvantages, and highlights the future research directions.

Keywords: artificial intelligence; human-computer interaction; computer vision; object recognition; machine learning; multimodality; robotics; interactive learning

智能交互系统最为重要的一项任务就是捕获和理解外界环境信息,从而完成交互方任务。近年来,由于人工智能和机器人学等相关领域技术的进步,智能交互系统得到了广泛的关注,高性能智能交互系统的实现也更加现实。智能交互系统感知外界环

收稿日期:2017-01-09.　**网络出版日期**:2017-04-19.
基金项目:国家"973"计划项目(2012CB316400).
通信作者:蒋树强.E-mail: :sqjiang@ict.ac.cn.

境比人类困难得多,而准确感知外界环境可以提高智能交互系统的交互性能,因此许多智能交互系统相关的工作探索了提高对外界环境感知性能的问题,主要的思想策略包括多模态信息融合和增量学习两个方面。多模态的信息融合可以使智能系统增加对外界环境的确定性,同时,不断变化的外界环境要求智能系统拥有不断自我学习的能力。通过交互不断学习外界信息也使智能系统的性能得以不断提升。在计算机视觉、智能交互系统等领域,增量学习都已引起了广泛的关注。本文基于智能交互系统的物体识别,对增量学习的进展进行综述。首先,对智能交互系统的研究背景和现状进行简要介绍,在此基础上,对增量学习主要算法进行综合对比与分析。最后讨论了增量学习可扩展和待解决的问题,以及进一步的研究方向。

1　智能交互系统对环境的感知

对于人类来说,我们可以精确地感知周围环境变化并作出相应的反应,但对于计算机来说,获取并分析周围环境信息,同时通过模仿人类行为来实现与人的交互,这是一个极具挑战性的任务。它包括场景理解、活动分类、运动分析、物体识别、自然语言理解、语音合成等方面。每个方面都可作为一个独立研究的任务。

准确感知外部环境可以使智能交互系统提高任务的完成度、完成的准确度和交互者对交互体验的满意度。多模态的外部信息,信息中较多的干扰和噪声,外界环境的复杂多变,都对智能系统建立对外部环境的准确感知提出了挑战。

为了增强交互系统对外部环境的感知性能,两个方面的相关工作被广泛研究:1)多模态信息融合;2)通过交互增量学习,自我改进。

2　多模态输入与信息融合

人类为了精确感知周围环境,往往会结合多种感知信息,如视觉、听觉、触觉等。认知科学的研究表明通过结合感官信息,人类可以增强对环境的感知。因此在多模态信息输入的智能交互系统中,互补的输入模式给系统提供了冗余的信息,而冗余输入模式增加了系统融合信息的准确性,降低系统对外界环境的不确定性,增加对环境感知的可靠性,从嘈杂的信息中产生一个单一的整体状态[1-3]。

2.1　自然语言理解

智能交互系统常常需要通过理解自然语言来对交互者的语言进行分析,从而获取到对方的指令。自然语言处理是计算机科学领域与人工智能领域中的一个重要方向。它研究能实现人与计算机之间用自然语言进行有效通信的各种理论和方法。自然语言处理是一门融语言学、计算机科学、数学于一体的科学。其常用的方法有:1)关键词匹配;2)使用有标注的语料库;3)语义分析。在文献[4]中,该系统使用关键词匹配技术实现自然语言理解,并假设相应的单词有某种特定的序列。文献[5]和文献[6]使用语义分析技术实现对自然语言的理解和分析。文献[5]的语言模型从现有的语料库[7]中训练得到,而文献[6]通过系统与交互者对话的过程不断获得语料,并逐步学习,不断改进其语言模型。

2.2　计算机视觉

由于获取外界信息的另一个主要渠道是视觉,所以计算机视觉是当前人机交互中一个非常活跃的领域。这一学科的基本假设是:可以通过计算的方式来模拟人类的视觉机制。如图 1 所示,智能系统模拟人类视觉机制的过程主要包括两个方面:1)智能系统要有能力将外部视觉信息转化为智能系统的内部表示;2)从外部环境获取到的视觉信息到语义方面的文字需要一个可用的映射。

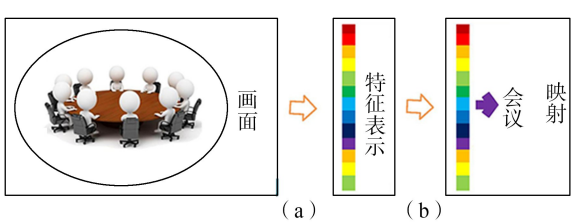


图 1　计算机模拟人类视觉机制

Fig.1　Computer simulation of human visual mechanism

第 1 个方面主要要求智能系统可以从图像中提取出有判别能力的特征。图像特征基本包括两种:手工设计的浅层特征和使用深度模型提取的深度特征。SIFT[8]、FPFH (fast point features histogram)[9]和 ensembles of shape features[10]等都属于手工设计的浅层特征。这种特征对图像变化如图像旋转、尺度变化等具有不变性。但是浅层特征只能捕捉到一部分图像信息[11]。与此相反的是,由于近年来深度学习模型(如卷积神经网络[12])方面的进步,由深度学习模型提取的深度特征可以捕获图像语义等更高层面的信息,具有更强的区分能力。因此,在计算机视觉方面,深度特征被广泛使用。

智能系统模拟人类视觉机制的另一个要求是可以对图像特征进行分类识别。在图像识别方面存在一系列的分类、聚类算法,如决策树、SVM、混合高斯模型等。

2.3 多模态信息融合

自然语言理解和计算机视觉是智能交互系统获取外界信息的两个主要途径。单一模态信息使智能系统难以对外界环境产成一个准确的认识,多模态信息融合可以增加系统对环境信息的确认度,通过多模态信息融合,智能系统摆脱了单一模态的限制,使人机交互更加智能。当前已经有很多工作关注于多模态融合这一方面的研究[13-20]。

2.4 多模态信息融合与增量学习

多模态信息融合帮助智能交互系统最大程度上地利用了可获取的外部信息,消除了单一模态中噪声带来的不一致性,从而可以准确地感知和理解外部环境。

对外部环境信息的准确感知使得智能交互系统在交互的过程中产生合情合理的语言或行为,这有助于提升系统的交互性能,得到更加良好的用户体验,如表 1 所示。

表 1 智能交互系统主要交互方式

Table 1 Major interaction of intelligent HCI systems

交互方式	面向任务	主要算法
自然语言	通过交互者从自然语言中获取到相应的指令;将任务结果转化为自然语言回馈给交互者	自然语言理解、语音合成
计算机视觉	通过对图像或视频进行分析“看到”周围环境	场景理解、活动分类、运动分析、物体识别
多模态融合	通过结合视觉、听觉等多方面信息,获得一个对周围环境更加准确的判断	特征层面的信息融合、语义层面的信息融合

优秀的交互性能和良好的用户体验使得智能系统可以从交互者处得到正确并且及时的反馈,这为智能系统在交互中进行增量学习打下了坚实的基础。

3 通过交互学习

由于外界环境复杂多变,智能交互系统无法在训练前获取到所有可能情形的全部有效信息作为训练数据(如图 2 所示,应用环境中的“书籍”在训练环境中出现过,属于旧类别的新实例,而“香蕉”则未曾在训练环境中出现,属于新类别。智能系统无法识别这两种未经学习的物体)。这就要求智能系统拥有自我学习的能力,可以在交互的过程中获得新的信息,学习到新的知识。

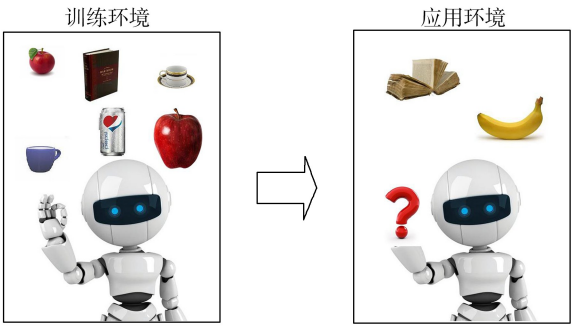


图 2 实际环境的复杂多变和非增量方法的局限性

Fig.2 The complex of environment and the limitation of constant model

多模态的交互引导多模态的学习,反过来多模态的学习又会改善多模态的交互。这是一个相互促进,共同提高的过程。

3.1 从交互中学习新知识的机器人

当前已经有许多相关工作展开了关于智能系统通过交互进行增量学习的研究[21-25]。

多方社交智能机器人在酒吧中使用自然语言与客人对话,根据客人的需要为他们提供相应的饮品[26]。它的学习任务在于引导一个多方互动对话,其目标为:当机器人的视野中同时出现多位客人时,以社会可接受的行为来尽可能为客人提供正确的饮品。

室内路线说明机器人[27]基于预定义的室内地图通过语音和手势向交互者提供方向引导他们到达相应的位置。它的学习任务是通过交互不断学习进入,维持和解除与它面前的人进行交互的恰当时机。

移动机器人[28]被用来获取物体和相关属性的新知识。它的任务包括发现未知的物品,询问物品的外形并获取相关的新知识。其学习任务为通过交互者获得新物品的物理外形描述,以此来扩充其知识库。

3.2 智能交互系统自我学习的策略

智能交互系统自我学习的能力需要通过某种探索和学习新知识的策略来实现。

增量学习是近年来备受关注的一种学习新知识的策略,旨在利用新数据来不断更新原有模型,使学习具有延续性,从而实现增量式的学习。

增量学习使智能交互系统可以进行持续性的学习,外部环境和交互者充当“老师”的角色,而系统则通过多模态的交互不断获得并学习新信息。

4 增量学习

4.1 增量学习的背景

由于真实的交互环境是开放并且复杂多变

的[29],在训练模型之前无法获取到所有可能情形的有效信息作为训练数据。除此之外,数据标签的获取也需要耗费大量人力、物力、财力和时间。最为重要的一点是,新的物体类别不断产生,已有物体类别的新实例不断出现,甚至有的物体类别的意义不断迁移变化,这都在数据方面要求智能系统需要具有不断学习的能力。另一方面,自我学习的能力可以使智能系统在获得新数据时随时学习,不需要重新训练全部数据[30]。这又在模型方面要求智能系统需要具有不断学习的能力。

4.2　增量学习的现状

学习新数据基本可以分为两种策略:一种是抛弃原有模型,在现有数据上学习新知识;另一种是基于原有模型,在此基础上继续学习新知识。这两种策略可以引出著名的稳定性-可塑性定理(stability-plasticity dilemma)[31]。

这个定理指出,一个完全稳定的模型可以保存已经学到的知识不忘记,但无法学习到新的知识;而一个完全可塑的模型可以学习新知识,但无法保存以前学到的知识(如图 3 所示)。而优秀的增量学习方法就是在可塑性和稳定性之间寻找一个合理的权衡。

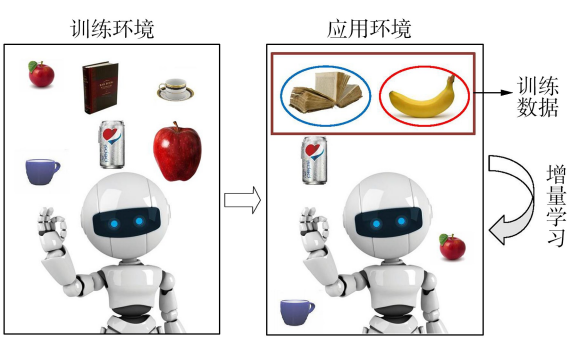


图 3　稳定性-可塑性定理

Fig.3　Stability-plasticity dilemma

文献[32]提出真正的增量学习应该满足 4 个条件,如图 4。

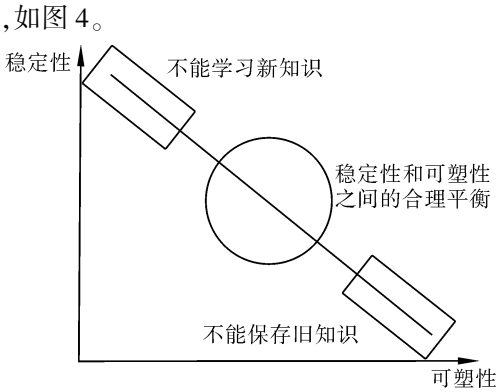


图 4　增量学习的条件

Fig.4　The conditions of incremental learning

1)可以学习旧类别的新数据。“书籍”概念在训练环境已经出现过,应用环境中的“书籍”是旧类别的新实例。

2)可以学习新类别。“香蕉”概念在训练环境未出现过,应用环境中的“香蕉”属于新类别。

3)在学习新知识时,旧的训练数据不是必须的。增量学习时只使用应用环境中的新数据(“书籍”)和新类别(“香蕉”)作为训练数据,而不需要已经学过的“罐”、“苹果”和“杯子”数据。

4)学习新知识后,不会忘记已经学到的旧知识。在应用环境中仍能识别以前在训练环境中学到的旧概念:“罐”、“苹果”、“杯子”和“书籍”概念的旧实例。

当前有许多增量学习方面的工作并不严格满足以上 4 个条件。

4.3　抛弃原有模型

对于学习新数据的第 1 种策略:抛弃原有模型,在现有数据上学习新知识。这种完全可塑的策略面临的最大问题是灾难性的遗忘(catastrophic forgetting)。它在现有新数据上学习知识,可以学到新的数据和类别,并且可以不需要原来的训练数据,满足增量学习的前 3 个条件。但它抛弃原有模型,则会导致旧知识的遗忘,不能满足第 4 个条件。神经网络常常使用这种策略的模型,例如多层感知机、径向基函数网络,小波网络和 Kohonen 网络。

4.4　基于原有模型继续学习

对于学习新数据的第 2 种策略:基于原有模型,在此基础上继续学习新知识。这种策略也常因关注于不同的方面而不能完全满足增量学习的 4 个条件。

根据增量学习算法学习的内容来看,新数据主要来源于两个方面:1)数据来源于已经学习过的类别,是旧类别的新实例;2)数据来源于没有学习过的类别,是新类别的数据。

4.4.1　学习旧类别的新实例

学习旧类别的新实例这一任务在某种程度上与迁移学习有些相似之处但又有不同,如表 2。

表 2　增量学习与迁移学习的比较

Table 2　Comparison between transfer learning and incremental learning

类别	相同点	不同点
迁移学习	将已学习的知识转移到新的任务	训练集领域与测试集领域不同;新领域的数据未经过学习
增量学习		训练集领域与测试集领域相同;新数据经过学习

迁移学习的任务是将某一领域学到的特征或信息应用到另一个不同但相似的领域上,如文献[33]。增量学习旧类别新实例的目标是利用现有的特征在相同任务(需要识别的类别不变)但规模扩大的数据集上学习新的知识。

文献[34]修改了原 SVM 目标函数中的损失项,使修改后的 SVM 可以在原模型的基础上修改分类面,实现增量学习旧类别新实例;文献[35]提出了一个基于 SVM 框架增量学习的精确解,即每增加一个训练样本或减少一个样本都会对 Lagrange 系数和支持向量产生影响,以此来调整分界面;文献[36]介绍了 HME(hierarchical mixture of experts)框架,这种框架在特征空间的不同区域训练了多个分类器,将各个分类器的输出通过一个网络进行加权得到最终结果,它利用线性最小二乘法(linear least squares)和加权线性最小二乘法(weighted linear least squares)通过递归来增量的更新每个数据点的参数,从而实现增量式的在线学习;文献[37]每次从候选训练数据集中选取一部分新的信息,并把选取出的新数据添加到当前数据集中;文献[38]扩展了文献[37]的增量学习方法,通过对候选训练数据集进行无监督的聚类,每次选出最有信息量的一部分数据加入当前训练数据中;文献[39]提出了一种结构学习算法,它使用数据集中的一小部分作为训练数据来建立一个具有最优隐藏层节点数目的前馈网络,该方法以训练数据集中较少的一部分数据作为初始的训练数据,通过有效的选择训练数据,最终产生一个最少但对所有数据有效的训练集。

这些增量学习方法更加关注于学习旧类别的新实例,它们都无法完全满足增量学习的 4 个条件。首先,这些方法无法学习新类别的数据。其次,有些方法在增量学习的同时必须使用部分或全部原始数据。

4.4.2 学习新类别的数据

与学习旧类别的新实例相比,学习新类别明显更加具有挑战性。

这个任务的目标是利用现有的特征在更加复杂的任务(需要识别的类别增加)并且规模扩大的数据集上学习新的知识。

对迁移学习的关注使得更多的研究工作注重于使用更少的数据来学得泛化性能更好的模型。由此转化到学习新类别方面的两个较为典型的研究领域为:one-shot learning 和 zero-shot learning。文献[40]提出了一种贝叶斯迁移学习方法,这种增量学习方法可以使用少量新数据学习到新类别。文献[41]提出了一种基于多模型的知识迁移算法,这种增量学习方法可以依靠已经学习的类别使用少量新数据来有效的学习新类别。通过求解一个凸优化问题,该方法自动选择利用哪一部分旧知识传递多少信息最为有效并确保在可用训练集上达到最小误差。文献[42]通过使用属性分类器来实现 zero-shot learning 的目标。

文献[43]指出,在其之前的大多数增量学习的工作都专注于二分类问题,这篇文章提出了一个多类分类的方法,在保存已学到的知识的基础上把当前的 N 类分类器转化为一个 $N+1$ 类分类器;文献[44]提出了一种具有层级关系的增量学习模型 NCMF(nearest class mean forest classifier)。这种方法以层级关系来组织概念,使得学习新类别时可以更新局部节点来达到增量的目的。文献[45]结合 SVM 算法最大分类间隔的策略和半监督学习算法低密度分隔符技术,来增加新的分界面以此识别新类别。

这些增量学习方法更加关注于学习新类别,它们对旧类别的新实例的学习效果尚未得到验证,同时有些方法在学习新数据的同时必须使用部分或全部原始数据,无法完全满足增量学习的 4 个条件。

表 3 增量学习算法对比分析

Table 3 Comparative analysis of incremental learning algorithms

算法	新类别	旧类别新实例	不需要原始数据	实现技术
文献[32]	√	√	√	多模型组合
文献[34]		√	√	调整模型参数
文献[35]		√	√	调整模型参数
文献[36]		√	√	多模型组合
文献[38]		√		选取有效数据
文献[39]		√		选取有效数据
文献[40]	√		√	调整模型参数
文献[41]	√		√	多模型组合
文献[42]	√		√	多模型组合
文献[43]	√			调整模型参数
文献[44]	√		√	调整模型参数
文献[45]	√			调整模型参数
文献[46]	√	√	√	多模型组合
文献[47]	√	√	√	调整模型参数
文献[48]		√	√	调整模型参数
文献[49]		√	√	调整模型参数
文献[50]		√	√	调整模型参数
文献[51]	√			调整模型参数

4.4.3　实现增量学习的 3 种技术

总体来说增量学习算法使用的技术可以总结为 3 类[32](图 5):

1)选择最有信息量的数据;

2)使用多模型集合实现模型的加强;

3)改变模型的参数或结构。

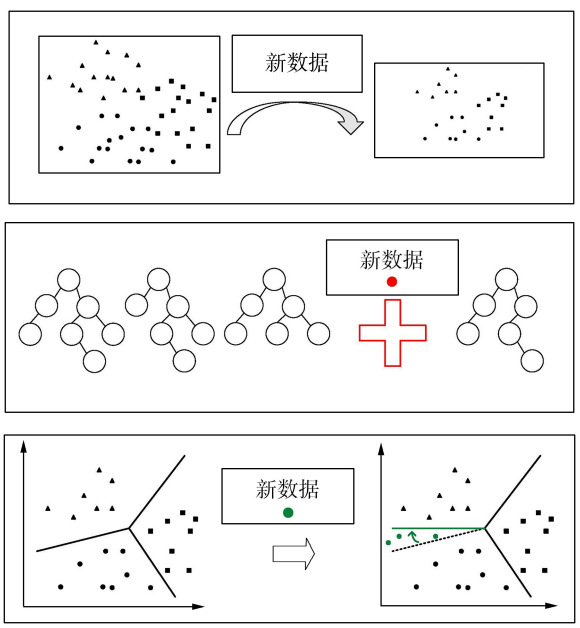


图 5　增量学习的 3 种技术

Fig.5　Three techniques of incremental learning

其中第 1 类方法往往用于实现旧类别新实例的增量,并且需要使用部分或全部原始数据。其目的是在一段信息流中选取最有效的数据,使用最少的数据完成学习任务。这种方法无法实现真正的增量学习。而第 2 类方法可以实现完全的增量学习。文献[46]提出了一种基于分类器集合的算法,该算法为与学习过的实例差别较大的新数据建立新的决策集群,每个集群以无监督的方式在特征空间中学习一个不同的超矩形部分,这个部分与要学习的目标类别相对应。但是这个方法对阈值的选取,训练数据中的噪声和训练数据学习的顺序都十分敏感;文献[47]提出了一种基于再生希尔伯特空间的增量学习算法。但是它需要数据分布的一个先验知识,这对于增量学习任务本身来说并不容易获得;文献[32]受 Adaboost 的启发,提出了一个由分类器集合构成的增量模型。这个算法的核心在于维护一个训练数据的分布,使得分类错误的数据更容易被采样,以此学习一个新的分类器加入集合中,而在增量学习的过程中,错误率较高的数据则恰恰是尚未见过或学习过的数据。但第 3 类方法需要训练多个模型进行组合,计算代价大大增加,而且随着增量学习的进行,不断增加的基模型也是一个未解决的问题。

4.4.4　通过改变模型参数实现增量学习

因此我们更为关注第 3 种方法:通过调整模型参数实现增量学习的单一模型。

文献[34]修改了原 SVM 目标函数中的损失项,使修改后的 SVM 可以修改原模型的分类面,并且在不需要原始数据的前提下,近似实现全局数据(新数据和已经学习过的旧数据)上的损失最小化。SVM 使用支撑向量来描述分界面,并将支撑向量作为参数存储在模型中。该方法利用支撑向量来代替原始数据,同时通过权重使支撑向量可以更好的模拟原始数据。文献[35]提出的 C&P 算法实现了 SVM 框架下增量学习的一个精确解。训练 SVM 相当于求解一个二次规划,二次规划的系数个数与训练数据个数相同。增量学习时,每增加一个训练数据,可以迭代求解一个新的系数。C&P 算法的关键在于,每增加一个实例,都要求学习过的所有数据全部满足 KKT 条件,来求解一个确定的增量模型。此后,许多研究基于 C&P 算法,逐渐展开了两方面的工作:一方面的工作专注于算法本身,文献[48]提出了该算法的扩展版本,每次迭代更新参数时可以同时处理多个数据;另一方面的工作使用 C&P 算法解决其他问题。文献[49]和文献[50]使用该算法实现了单类 SVM 的增量学习问题。

与文献[51]中修改损失项的方法相似的是,文献[43]修改了 SVM 目标函数的正则项,在增加新的分界面的同时,控制已有分界面的变化。该方法通过建立新的分界面学习到新类别,同时通过控制已学到的分界面的变化,确保学到的知识不会受新类别的影响而丢失。文献[45]借鉴 SVM 中最大分类间隔和半监督学习中低密度分隔符的思想,在所有低密度分隔符中选取一个分界面使得模型的经验损失,结构损失和增广损失(新类别的损失)整体最小。文献[51]将卷积神经网络组织成层级树形结构,每个节点由一些相似类别的聚类构成,该方法通过树形结构使得模型更新时只需要调整模型局部,并可以严格控制模型调整范围,增添新节点时此方法通过克隆原有节点进行调整,使得已学到的知识不会被遗忘。

这些通过修改原模型参数而实现增量学习的算法也没有完全满足增量学习的 4 个条件,它们都解决了灾难性遗忘的问题,但都更加侧重于学习旧类别新实例或者新类别中的某一方面,有些方法也没有解决需要原始数据的问题。

5 增量学习未来研究方向展望

目前,增量学习在智能交互、物体识别等许多方面都得到了广泛的研究,但由于应用环境远比训练环境更加复杂多变,离智能交互系统真正走出实验室,进入真实应用场景还有一段距离。本文将对增量学习未来的研究方向进行展望。

5.1 面向大规模数据集的增量学习

近年来,随着信息技术的发展,数据呈现爆炸式增长的趋势,这使得模型的训练和更新都变得更加困难并且耗时。

在面向大规模数据集时,增量学习的优点尤为突出。一方面,在训练数据规模扩大的同时,训练需要的时间和计算能力都随之增加。当新数据或新类别出现时,非增量的离线方法需要重新训练已经学习过的数据,这会导致资源的浪费。而增量学习方法则可以在原始模型的基础上继续学习,不需要重新训练所有数据。另一方面,非增量方法重新训练全部数据,这也就意味着全部的或绝大部分的数据或都必须保留,当数据量非常庞大时,数据的存储也是一个问题。而增量学习不需要原始数据,所以不需要考虑数据存储的问题。

5.2 面向深度学习的增量学习

深度学习技术被大量应用到图像、视频、文本等多媒体相关的任务上。一方面,深度网络可以直接完成图像分类,物体识别等任务。另一方面,这些任务所产生的标签又可以应用到图像检索相关的任务中。深度网络又可以间接地扩展到其他任务中去。所有这些任务的真实场景中,数据及其标记的总是以增量的方式进行收集的。因此在数据方面来说,面向深度学习的增量学习是合理的。

深度学习技术在图像分类任务中的应用取得了快速的进步,它的性能迅速提升。当前限制深度神经网络性能进一步提升的一个可能性是网络容量。因此,一个可能的解决方案是增加网络容量[51]。但是这个方案面临着两个困难:一方面,大网络的训练难度可能成倍增长;另一方面,如何增加网络容量还不明确。因此,应该更加谨慎地增加网络容量,提升网络能力。而增量学习则为逐步的、增量的改善网络提供了一种可能性,当前已经有一些相关的工作对这种可能性展开了一定的研究[39,52-54]。因此在模型方面来说,面向深度学习的增量学习也是合理的。

5.3 声图文融合的多模态增量学习

基于智能交互的增量学习系统通过多模态交互进行增量学习。由此看来,增量学习的内容也应当是多模态的。

智能系统通过多模态交互进行增量学习,反过来,增量学习的结果也会提升多模态交互的性能。

听觉、视觉和文字是智能交互系统感知外界环境信息最主要的 3 种形式。通过声图文融合的增量学习方式,可以使智能交互系统逐步全面地适应不断变化的外界环境。

5.4 知识条目和识别能力的增量学习

现在的大部分研究工作更加关注于独立的视觉概念的识别或是单纯知识条目的增加构建。但实际生活中不同的概念之间具有或隐性或显性的关系,物体也拥有不同的属性。这些概念和属性可以构成关于交互物体、交互者和外界环境的知识条目。人类可以基于这些额外的关系或属性信息学习到更多的知识。智能交互系统也应该利用这些信息进行更全面的学习,对周围环境或任务目标得到一个更加全面的认识。

另一个值得关注的方面是,智能交互系统应该能够系统并有效地组织已学习到的知识。文献[55]指出,将小规模的信息加入到已经组织好的大规模信息中是人类感知,学习,和组织信息等过程中十分重要的部分。因此,智能交互系统应该拥有一个合理的学习机制,并可以自动在学习到的知识间建立合理有效的联系。

6 结束语

目前,增量学习在智能交互、物体识别等许多方面都得到了广泛的研究,由于应用环境远比训练环境更加复杂多变,它更加注重于解决自动学习,改善应用效果的问题。这说明智能交互系统从实验环境逐渐开始走向真实的应用场景。

由于不同任务关注方面各不相同,大多数研究工作都无法完全满足增量学习的定义。但真实场景的复杂多变是单一任务目标无法模拟的,若要智能交互系统真正走向现实,需要综合解决增量学习 4 个方面的问题,这是增量学习算法本身的发展趋势。

同时也应该结合不同的任务,实现适用于不同场景、不同侧重点的智能增量学习系统。根据任务本身设计不同的策略实现个性化的应用。这是从应用场景来看的增量学习发展趋势。

当这些发展趋势真正变为现实的时候,智能交互系统有望真正走进人类社会,为我们的日常生活带来更多帮助,安全、便捷和高效地辅助我们完成更多任务。

参考文献:

[1]ERNST M O, BÜLTHOFF H H. Merging the senses into a robust percept[J]. Trends in cognitive sciences, 2004, 8(4): 162-169.

[2]CORRADINI A, MEHTA M, BERNSEN N O, et al. Multimodal input fusion in human-computer interaction[J]. NATO Science Series Sub Series III Computer and Systems Sciences, 2005, 198: 223.

[3]NODA K, ARIE H, SUGA Y, et al. Multimodal integration learning of robot behavior using deep neural networks[J]. Robotics and autonomous systems, 2014, 62(6): 721-736.

[4]MERIÇLI C, KLEE S D, PAPARIAN J, et al. An interactive approach for situated task specification through verbal instructions[C]//Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems. Paris, France: International Foundation for Autonomous Agents and Multiagent Systems, 2014: 1069-1076.

[5]CANTRELL R, BENTON J, TALAMADUPULA K, et al. Tell me when and why to do it! Run-time planner model updates via natural language instruction[C]//Proceedings of the 2012 IEEE International Conference on Human-Robot Interaction. Boston, MA: IEEE, 2012: 471-478.

[6]THOMASON J, ZHANG S Q, MOONEY R, et al. Learning to interpret natural language commands through human-robot dialog[C]//Proceedings of the 24th international conference on Artificial Intelligence. Buenos Aires, Argentina: AAAI Press, 2015.

[7]EBERHARD K M, NICHOLSON H, SANDRA K, et al. The Indiana "Cooperative Remote Search Task"(CReST) corpus[C]//Proceedings of the 2010 International Conference on Language Resources and Evaluation. Valletta, Malta: LREC, 2010.

[8]LOWE D G. Distinctive image features from scale-invariant keypoints[J]. International journal of computer vision, 2004, 60(2): 91-110.

[9]MORISSET B, RUSU R B, SUNDARESAN A, et al. Leaving flatland: toward real-time 3D navigation[C]//Proceedings of the 2009 IEEE International Conference on Robotics and Automation. Kobe: IEEE, 2009: 3786-3793.

[10]HINTERSTOISSER S, HOLZER S, CAGNIART C, et al. Multimodal templates for real-time detection of texture-less objects in heavily cluttered scenes[C]//Proceedings of the 2011 IEEE International Conference on Computer Vision. Barcelona: IEEE, 2011: 858-865.

[11]WANG Anran, LU Jiwen, CAI Jianfei, et al. Large-margin multi-modal deep learning for RGB-D object recognition[J]. IEEE transactions on multimedia, 2015, 17(11): 1887-1898.

[12]LECUN Y, BOSER B, DENKER J S, et al. Backpropagation applied to handwritten zip code recognition[J]. Neural computation, 1989, 1(4): 541-551.

[13]THOMASON J, SINAPOV J, SVETLIK M, et al. Learning multi-modal grounded linguistic semantics by playing I spy[C]//Proceedings of the 25th International Joint Conference on Artificial Intelligence. New York, 2016.

[14]LIU C S, CHAI J Y. Learning to mediate perceptual differences in situated human-robot dialogue[C]//Proceedings of the Twenty-Ninth American Association Conference on Artificial Intelligence. Austin, Texas: AAAI Press, 2015: 2288-2294.

[15]PARDE N, HAIR A, PAPAKOSTAS M, et al. Grounding the meaning of words through vision and interactive gameplay[J]. Proceedings of the 24th International Conference on Artificial Intelligence. Buenos Aires, Argentina: AAAI Press, 2015.

[16]MATUSZEK C, FITZGERALD N, ZETTLEMOYER L, et al. A joint model of language and perception for grounded attribute learning[C]//Proceedings of the 29th International Conference on Machine Learning. Edinburgh, Scotland, 2012.

[17]赵鹏, 陈浩, 刘慧婷, 等. 一种基于图的多模态随机游走重排序算法[J]. 哈尔滨工程大学学报, 2016, 37(10): 1387-1393.
ZHAO Peng, CHEN Hao, LIU Huiting, et al. A multimodal graph-based re-ranking through random walk algrithm[J]. Journal of Harbin Engineering University, 2016, 37(10): 1387-1393.

[18]段喜萍, 刘家锋, 王建华, 等. 多模态特征联合稀疏表示的视频目标跟踪[J]. 哈尔滨工程大学学报, 2015, 36(12): 1609-1613.
DUAN Xiping, LIU Jiafeng, WANG Jianhua, et al. Visual target tracking via multi-cue joint sparse representation[J]. Journal of Harbin Engineering University, 2015, 36(12): 1609-1613.

[19]FISHER J W, DARRELL T. Signal level fusion for multimodal perceptual user interface[C]//Proceedings of the 2001 Workshop on Perceptive User Interfaces. New York, NY, USA: ACM, 2001: 1-7.

[20]JOHNSTON M, BANGALORE S. Finite-state multimodal parsing and understanding[C]//Proceedings of the 18th conference on Computational linguistics. Saarbrücken, Germany: ACM, 2000: 369-375.

[21]BETTERIDGE J, CARLSON A, HONG S A, et al. Toward never ending language learning[C]//Proceedings of the American Association for Artificial Intelligence. 2009: 1-2.

[22]CHERNOVA S, THOMAZ A L. Robot learning from human teachers[M]. San Rafael, CA, USA: IEEE, 2014.

[23]MATUSZEK C, BO L F, ZETTLEMOYER L, et al. Learning from unscripted deictic gesture and language for hu-

man-robot interactions[C]//Proceedings of the 28th American Association Conference on Artificial Intelligence. Québec City, Québec, Canada: AAAI Press, 2014: 2556-2563.

[24] CUAYÁHUITL H, DETHLEFS N. Dialogue systems using online learning: beyond empirical methods[C]//Proceedings of the NAACL-HLT Workshop on Future Directions and Needs in the Spoken Dialog Community: Tools and Data. Montreal, Canada: Association for Computational Linguistics, 2012: 7-8.

[25] 顾海巍, 樊绍巍, 金明河, 等. 基于灵巧手触觉信息的未知物体类人探索策略[J]. 哈尔滨工程大学学报, 2016, 37(10): 1400-1407.
GU Haiwei, FAN Shaowei, JIN Minghe, et al. An anthropomorphic exploration strategy of unknown object based on haptic information of dexterous robot hand[J]. Journal of Harbin Engineering University, 2016, 37(10): 1400-1407.

[26] KEIZER S, FOSTER M E, WANG Z R, et al. Machine learning for social multiparty human-robot interaction[J]. ACM transactions on interactive intelligent systems (TIIS), 2014, 4(3): 14.

[27] BOHUS D, SAW C W, HORVITZ E. Directions robot: In-the-wild experiences and lessons learned[C]//Proceedings of the 2014 International Conference on Autonomous Agents and Multi-agent Systems. Richland, SC, 2014: 637-644.

[28] KRAUSE E A, ZILLICH M, WILLIAMS T E, et al. Learning to recognize novel objects in one shot through human-robot interactions in natural language dialogues[C]//Proceedings of the 28th American Association Conference on Artificial Intelligence. Québec City, Québec, Canada: AAAI Press, 2014: 2796-2802.

[29] MENSINK T, VERBEEK J J, PERRONNIN F, et al. Distance-based image classification: generalizing to new classes at near-zero cost[J]. IEEE transactions on pattern analysis and machine intelligence, 2013, 35(11): 2624-2637.

[30] IBA W, WOGULIS J, LANGLEY P A T. Trading off simplicity and coverage in incremental concept learning[C]//Proceedings of the Fifth International Conference on Machine Learning. Ann Arbor: University of Michigan, 1988: 73.

[31] GROSSBERG S. Nonlinear neural networks: Principles, mechanisms, and architectures[J]. Neural networks, 1988, 1(1): 17-61.

[32] POLIKAR R, UPDA L, UPDA S S, et al. Learn++: An incremental learning algorithm for supervised neural networks[J]. IEEE transactions on systems, man, and cybernetics, part C (Applications and reviews), 2001, 31(4): 497-508.

[33] 贾刚, 王宗义. 混合迁移学习方法在医学图像检索中的应用[J]. 哈尔滨工程大学学报, 2015, 36(7): 938-942.
JIA Gang, WANG Zongyi. The application of mixed migration learning in medical image retrieval[J]. Journal of Harbin Engineering University, 2015, 36(7): 938-942.

[34] RÜPING S. Incremental learning with support vector machines[C]//Proceedings of the 2011 IEEE International Conference on Data Mining. Washington, DC, USA: IEEE, 2001: 641.

[35] CAUWENBERGHS G, POGGIO T. Incremental and decremental support vector machine learning[C]//Proceedings of the 13th International Conference on Advances in neural information processing systems. Cambridge, MA, USA: MIT Press, 2000, 13: 409.

[36] JORDAN M I, JACOBS R A. Hierarchical mixtures of experts and the EM algorithm[J]. Neural computation, 1994, 6(2): 181-214.

[37] WANG E H C, KUH A. A smart algorithm for incremental learning[C]//Proceedings of the 1992 IEEE International Joint Conference on Neural Networks. Baltimore: IEEE, 1992, 3: 121-126.

[38] ENGELBRECHT A P, CLOETE I. Incremental learning using sensitivity analysis[C]//Proceedings of the 1999 International Joint Conference on Neural Networks. Washington DC: IEEE, 1999.

[39] ZHANG B T. An incremental learning algorithm that optimizes network size and sample size in one trial[C]//Proceedings of the 1994 IEEE World Congress on Computational Intelligence. Orlando, FL, USA: IEEE, 1994, 1: 215-220.

[40] LI F F, FERGUS R, PERONA P. One-shot learning of object categories[J]. IEEE transactions on pattern analysis and machine intelligence, 2006, 28(4): 594-611.

[41] TOMMASI T, ORABONA F, CAPUTO B. Learning categories from few examples with multi model knowledge transfer[J]. IEEE transactions on pattern analysis and machine intelligence, 2014, 36(5): 928-941.

[42] LAMPERT C H, NICKISCH H, HARMELING S. Learning to detect unseen object classes by between-class attribute transfer[C]//Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition. Miami, FL: IEEE, 2009: 951-958.

[43] KUZBORSKIJ I, ORABONA F, CAPUTO B. From N to N+1: Multiclass transfer incremental learning[C]//Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition. Portland, OR: IEEE, 2013: 3358-3365.

[44] RISTIN M, GUILLAUMIN M, GALL J, et al. Incremental

learning of NCM forests for large-scale image classification [C]//Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus, OH: IEEE, 2014: 3654-3661.

[45] DA Qing, YU Yang, ZHOU Zhihua. Learning with augmented class by exploiting unlabeled data[C]//Proceedings of the 28th American Association Conference on Artificial Intelligence. Québec, Canada: AAAI Press, 2014: 1760-1766.

[46] CARPENTER G A, GROSSBERG S, REYNOLDS J H. ARTMAP: Supervised real-time learning and classification of nonstationary data by a self-organizing neural network [J]. Neural networks, 1991, 4(5): 565-588.

[47] VIJAYAKUMAR S, OGAWA H. RKHS-based functional analysis for exact incremental learning[J]. Neurocomputing, 1999, 29(1/2/3): 85-113.

[48] KARASUYAMA M, TAKEUCHI I. Multiple incremental decremental learning of support vector machines[J]. IEEE transactions on neural networks archive, 2010, 21(7): 1048-1059.

[49] GRETTON A, DESOBRY F. On-line one-class support vector machines. an application to signal segmentation [C]//Proceedings of the 2003 IEEE International Conference on Acoustics, Speech, and Signal Processing. Hong Kong, China: IEEE, 2003.

[50] LASKOV P, GEHL C, KRÜGER S, et al. Incremental support vector learning: Analysis, implementation and applications[J]. The Journal of machine learning research archive, 2006, 7: 1909-1936.

[51] XIAO Tianjun, ZHANG Jiaxing, YANG Kuiyuan, et al. Error-driven incremental learning in deep convolutional neural network for large-scale image classification[C]// Proceedings of the 22nd ACM international conference on Multimedia. New York, NY: ACM, 2014: 177-186.

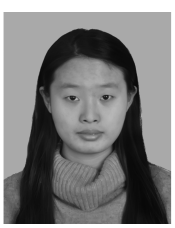
[52] LOMONACO V, MALTONI D. Comparing incremental learning strategies for convolutional neural networks [M]//SCHWENKER F, ABBAS H, EL GAYAR N, et al, eds. Artificial Neural Networks in Pattern Recognition. ANNPR 2016. Lecture Notes in Computer Science. Cham: Springer, 2016.

[53] GRIPPO L. Convergent on-line algorithms for supervised learning in neural networks[J]. IEEE transactions on neural networks, 2000, 11(6): 1284-1299.

[54] FU Limin, HSU H H, PRINCIPE J C. Incremental backpropagation learning networks[J]. IEEE transactions on neural networks, 1996, 7(3): 757-761.

[55] GOBET F, LANE P C R, CROKER S, et al. Chunking mechanisms in human learning[J]. Trends in cognitive sciences, 2001, 5(6): 236-243.

作者简介:



李雪,女,1992 年生,硕士研究生,主要研究方向为智能信息处理与机器学习。

蒋树强,男,1977 年生,博士生导师,主要研究方向为图像/视频等多媒体信息的分析、理解与检索技术。IEEE 和 CCF 高级会员,发表学术论文 100 余篇,授权专利 10 项。