

面向频谱大数据处理的机器学习方法

吴启晖 邱俊飞 丁国如

(解放军理工大学通信工程学院, 南京, 210007)

摘要: 随着移动互联网与物联网的迅猛发展, 个人无线设备的数量呈现指数级增长, 随之产生的海量频谱数据与日俱增, 频谱大数据的存在已成事实。同时, 频谱赤字也日益严峻。为提高频谱利用率, 有效的频谱大数据处理显得十分重要。本文从无线通信的角度, 首先给出了频谱大数据的定义并分析了它的基本特征; 然后总结了一些对于频谱大数据分析或利用颇具前景的机器学习方法, 如分布式和并行式学习、极速学习、核学习、深度学习、强化学习、博弈学习和迁移学习; 最后给出了几个开放性话题和研究趋势。

关键词: 大数据; 频谱大数据; 机器学习; 数据挖掘; 无线通信; 物联网

中图分类号: TP181 **文献标志码:** A

Machine Learning Methods for Big Spectrum Data Processing

Wu Qihui, Qiu Junfei, Ding Guoru

(College of Communications Engineering, PLA University of Science and Technology, Nanjing, 210007, China)

Abstract: With the rapid development of the mobile Internet and the Internet of Things, the number of personal wireless devices has grown exponentially, resulting in the increase of massive spectrum data. Therefore, the big spectrum data are literally formed. Meanwhile, the spectrum deficit is also increasingly precarious. Effective big spectrum data processing is significant in improving the spectrum utilization. Firstly, from a perspective of wireless communication, a definition of big spectrum data is presented and its characteristics are also analyzed. Then, promising machine learning methods to analyze and utilize the big spectrum data are summarized, such as, the distributed and parallel learning, extreme learning machine, kernel-based learning, deep learning, reinforcement learning, game learning, and transfer learning. Finally, several open issues and research trends are addressed.

Key words: big data; big spectrum data; machine learning; data mining; wireless communication; Internet of Things

引 言

当今, 来自不同领域的海量数据以前所未有的速度爆炸式地增长。据 IDC 报告^[1] 预测, 从 2005 年到 2020 年, 全球数据将以每两年翻倍的速度从 130 艾字节 (Exabytes) 增长到 40 000 艾字节 (1 艾字节 =

2^{16} 字节)。于是,大数据(Big data)的概念便应运而生,用于捕捉这种数据爆炸趋势所带来的深远意义。而且,大数据的巨大潜力也引领了一个新兴研究领域的产生,并受到了各行各业的广泛关注。谷歌、雅虎、微软、IBM、英特尔和华为等公司已经开始开发各式各样的大数据应用案例,像智能家居、智慧城市和物联网等^[2,3]。在过去几年,中国政府也资助了多项关于大数据的研究项目^[4],用于研制和改善分析、利用大数据的工具和技术。总之,大数据研究成为当下热潮之一。

大数据本身是一个抽象的概念,对于它的定义不一而足,典型的说法包括麦肯锡公司给出的是“超越典型数据库软件工具所能获取、存储、管理和分析的数据集”^[5],IBM和微软公司给出的3V“大量(Volume)、多样(Variety)和高速(Velocity)”模型^[6,7]。事实上,大数据的精确定义和特点依据场景而定。就无线通信领域来说,近年来,随着移动互联网与物联网的迅猛发展,个人无线设备的数量呈现指数级增长,人们对无线多媒体业务愈发青睐,据预测,在未来十年无线业务量将激增1 000倍^[8],而随之产生的频谱数据也将急剧增加,频谱大数据的存在已成事实。此外,由于频谱是无线信号赖以传输的媒介,是有限的自然资源,因此,随着无线业务量的持续增长,频谱赤字也将日益严峻。所以,为解决频谱危机,提高频谱资源的有效利用,需要对无线通信业务中产生的海量频谱数据进行有效地分析与处理。

机器学习技术^[9,10]伴随着计算能力的不断进步,在数据分析和知识发现方面已经扮演着至关重要的角色。因此,为解决频谱赤字危机这一问题,需要借助先进的机器学习方法,挖掘频谱大数据中隐含的有用信息,从而实现高效频谱利用。然而,从大量结构繁多的频谱数据中获取隐藏规律并非易事,传统的针对小规模数据集的机器学习算法很多已不再适用,所以,研究频谱大数据环境下的机器学习方法也成为关注的焦点。

本文首先给出了频谱大数据的定义,并从大量、多样、高速、真实和价值等5个方面,对频谱大数据的特征加以分析,然后,从频谱大数据分析和频谱大数据利用两个角度,系统地梳理了当前可用于频谱大数据处理的机器学习方法,并在最后提供了一些开放性的话题和研究趋势。

1 频谱大数据的定义和基本特征

频谱数据(Spectrum data)是指与无线通信环境相关的数据,主要包括以下几个方面^[11]:(1)在时域、空域和频域上的无线电频谱状态数据,如空闲或者繁忙、信号能量值以及信号特征等;(2)用户或者设备数据,如设备ID、设备容量、用户频谱需求和用户反馈等;(3)环境信息,如地形数据、水文气象数据等。而频谱大数据(Big spectrum data)作为大数据在无线或者无线电领域中的一种特定模式,是指无法使用传统的系统和工具进行处理和分析的海量复杂的频谱数据。特别地,图1用5个关键词来表征频谱大数据:大量(Volume)、多样(Variety)、高速(Velocity)、真实(Veracity)和价值(Value)。

1.1 频谱大数据的大规模特性

大量(Volume)无疑是大数据具备的一个基本属性^[12]。就频谱大数据而言,单以无线频谱状态数据为例,如图2所示,假设用1 Byte来表示一个 $100\text{ m} \times 100\text{ m}$ 的空间网格内100 kHz带宽在100 ms时隙长度中的频谱感知能量值,那么存储在 $100\text{ km} \times 100\text{ km}$ 空间范围内0~5 GHz频段的总的规模将达到^[11]

$$7\text{ days/week} \times 24\text{ hours/day} \times 3\,600\text{ seconds/hour} \times \frac{1\text{ s}}{100\text{ ms}} \times \frac{5\text{ GHz}}{100\text{ kHz}} \times \frac{100\text{ km} \times 100\text{ km}}{100\text{ m} \times 100\text{ m}} \times 1\text{ Byte} = 3.024 \times 10^{17}\text{ Byte} = 3.024 \times 10^5\text{ Terabyte(TB)} = 3.024 \times 10^2\text{ Petabyte(PB)} \quad (1)$$

频谱状态数据的规模将从时域、空域和频域3个维度不断增长,而且,倘若进一步考虑用户或设备数据和地理、水文、气象等环境参数,产生的数据量将会更大,甚至有可能从拍字节(Petabyte)变化到泽它字节(Zettabyte, 1 Zettabyte = 10^{21} Bytes)。对于海量频谱数据的分析与处理成为当前数据处理系统的一大挑战,需要分布式和并行式的处理方法来加以应对。

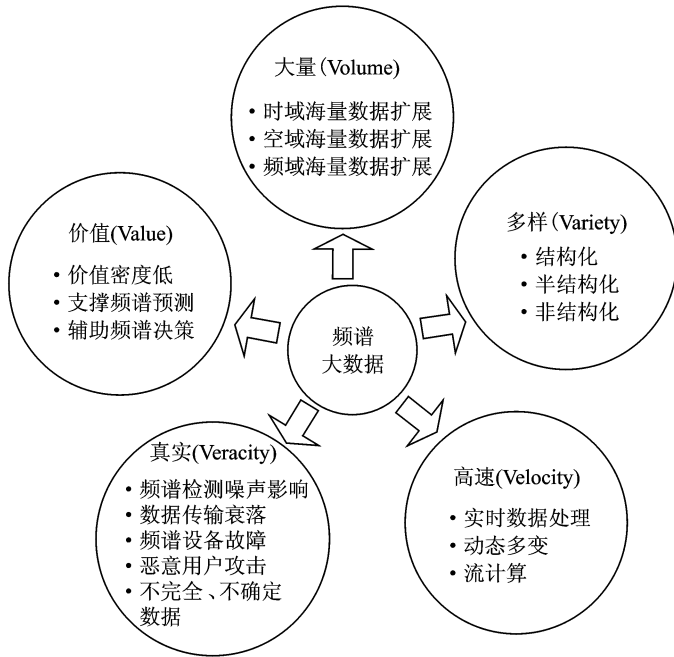


图 1 频谱大数据的基本特征

Fig. 1 Basic characteristics of big spectrum data

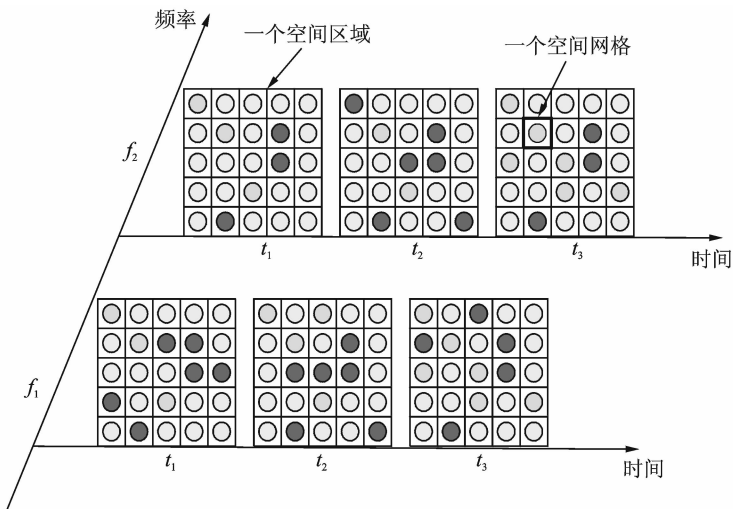


图 2 时-空-频多维空间中的频谱大数据

Fig. 2 Big spectrum data in time-space-frequency multi-dimensional space

1.2 频谱大数据的多样化特性

大数据的第 2 个属性是多样性 (Variety), 即数据往往来自于不同的数据源并具备不同的形式^[13]。对于频谱大数据而言, 多样化的频谱数据主要来源于频谱感知和地理频谱数据库, 可以从 4 个方面进行分类, 如图 3 所示。

从频谱数据的关联程度来看, 可分为直接、间接和综合频谱数据。直接频谱数据主要是指频谱占用

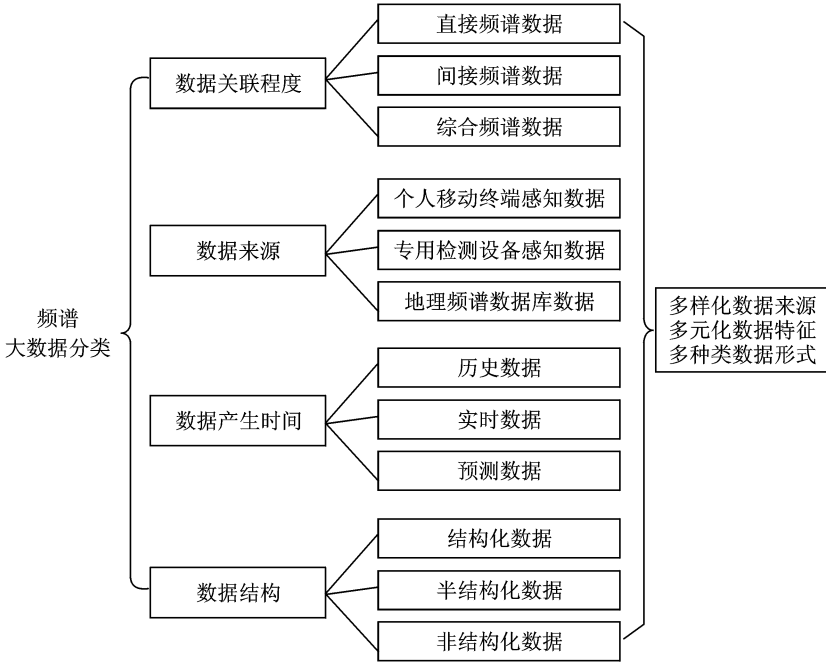


图3 频谱大数据分类

Fig. 3 Classification of big spectrum data

情况、频谱质量、频谱噪声和频谱干扰等。间接频谱数据包括时间信息、空间位置信息、地理信息、用频设备数据、用频规则与政策数据、作战频谱数据、干扰机位置、速度以及装备干扰关系等。综合频谱数据包括频谱态势图、演化视频、文本以及表格等。从频谱数据的来源来看,可以分为个人移动终端感知数据、专用检测设备感知数据和地理频谱数据库数据。从频谱数据的产生时间来看,可分为历史数据、实时数据和预测数据。从频谱数据的结构来看,可分为结构、半结构、非结构。多源的频谱大数据往往呈现出异构特质,对于传统的数据处理方法带来了难题,需要具备多领域处理功能的技术的产生。

1.3 频谱大数据的高速特性

大数据带来的另一挑战是它的速度(Velocity)^[14],即数据以极快的速度产生,需要进行实时处理,否则,数据将稍纵即逝或者时延的处理机制得到的结果价值很低甚至无用。针对频谱大数据,以无线网络为例,频谱数据分析是发现频谱共享机会的基本手段,其功能主要是从频谱检测获取的数据中分析推断频谱演化态势,为频谱决策提供信息支撑。由于频谱环境的动态变化,快速的频谱数据处理是有效频谱预测与决策的前提条件,图4给出了一个频谱数据分析与预测、处理与决策的示意图。

在 t_0 时刻对于获取的频谱数据进行处理,如果实际用时 T' 大于频谱状态改变时间 T ,由于频谱环境的时变特性,那么预测结果将不再具备价值意义,而且有可能是错误的判决,还会进一步影响到后面的频谱决策,导致占用冲突,降低频谱利用的有效性。因此,针对频谱数据处理中的时间敏感性,尤其在即将面临的频谱大数据环境下,快速实时处理和流计算机制很值得关注与研究。

1.4 频谱大数据的真实特性

真实性(Veracity)主要是指数据的质量。由于大数据来源多样,在数据采集、整理、存储和传输的过程中,难免出现模糊、错误、属性丢失等现象,影响到数据本身的可信度。数据不完全、不确定是大数据在真实性方面常见的现象^[4]。对于频谱大数据而言,数据的真实性问题主要体现在以下几个方面:

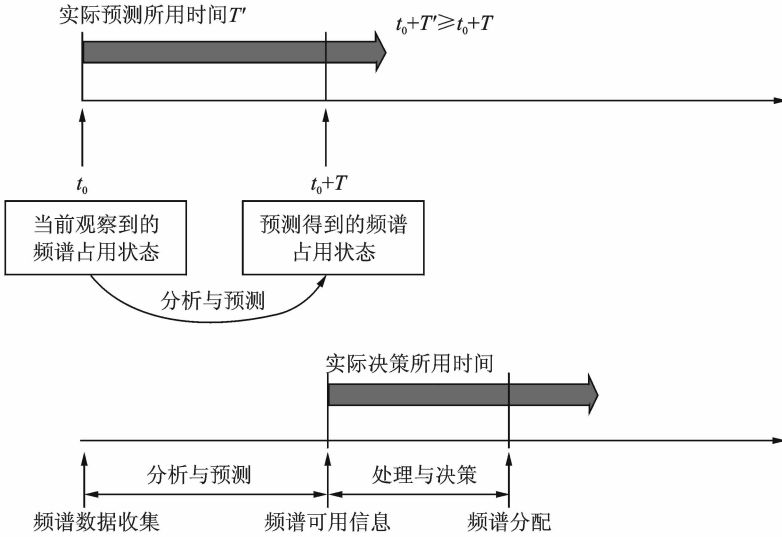


图 4 频谱数据的实时处理流程

Fig. 4 Flow of real-time spectrum data processing

(1) 频谱测量中固有的噪声影响; (2) 频谱监测设备的测量误差以及偶发的设备故障误差; (3) 异构频谱监测设备(测量能力、精度不同)引入的误差; (4) 恶意用户上传的虚假频谱测量数据。

以认知无线网络(Cognitive radio network)^[15,16]为例,对于频谱数据可能存在的真实性问题加以说明,如图 5 所示。对于数据融合中心来说,由于网络中存在恶意用户,因此,获取的频谱数据可能存在错误信息。另外,即使是诚实用户上传的数据,由于衰落、噪声的存在,也可能出现数据不完全、不确定的现象,这样的频谱数据质量问题会严重影响到融合中心的最终判决。因此,有效的数据修复和补全变得必要。此外,随着 5G 移动网络^[17-19]、物联网^[20,21]和互联网+等新兴技术的不断发展,频谱数据量相比于认知无线网络场景下会更多,频谱大数据的真实性问题也需要引起研究者们更多的关注。

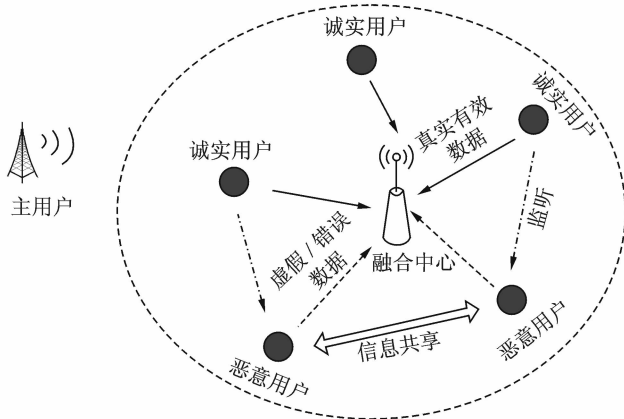


图 5 认知无线网络中频谱数据真实性问题描述

Fig. 5 Description of veracity of spectrum data in cognitive radio networks

1.5 频谱大数据的价值特性

事实上,探索各式各样的数据处理技术最终目的就是要从海量数据中提取有价值的信息,因此,价

值(Value)也成为大数据的显著特征之一^[3, 22]。然而,从大量低价值密度的数据中获取重要的价值信息是颇具挑战的工作。对于频谱大数据而言,低价值密度主要体现在3个方面:(1)由于空间相关性,局部范围内的频谱数据可能存在较大的冗余;(2)由于时间相关性,同一设备在短时间内的样本存在冗余;(3)由于频率相关性,相邻频率上的频谱数据存在冗余。因此,为有效获取频谱价值信息,更加智能的数据处理方法显得尤其紧迫。

总的来说,以上5个方面从不同角度分析了频谱大数据的基本特征,它所带来的既有前所未有的机遇,也有难以应对的挑战,这些都需要新的数据分析方法从频谱数据中获取关于无线频谱使用的知识,去提高频谱决策效用。机器学习(Machine learning)^[9, 10, 23]是使用计算机模拟技术对人类获取知识的学习过程进行研究的,并从中对现有知识进行创新,提升分析和解决问题,它最本质的目的就是大量的数据分析中汲取知识。因此,大数据技术的目标实现与机器学习的发展必然密不可分,机器学习与大数据也成为数字领域最热的研究趋势^[2]。因此,为解决频谱赤字危机,挖掘频谱数据中的价值信息来辅助频谱决策,机器学习技术是不可或缺的有效方法。然而,就频谱大数据而言,由于数据的规模大、种类多、变化快,部分数据不完全、不确定和价值密度低等特性,传统数据上的机器学习方法很多已不再适用。因此,研究频谱大数据环境下的机器学习技术也将成为未来的热点话题。

2 面向频谱大数据处理的机器学习技术

海量的频谱数据来源主要分为频谱感知数据和地理频谱数据库。针对不同的数据任务、数据类型和数据特点,需要运用特定的学习方法。本文将面向频谱大数据处理的机器学习技术分为两大类:一类是用于频谱大数据分析,从海量频谱数据中挖掘有效信息,发现潜在规律和获取统计特性;另一类是用于频谱大数据利用,基于数据分析中得到的频谱演化态势,执行有效的频谱决策。如图6所示,本文给出了频谱大数据环境下的机器学习技术框架,分为“数据来源层—数据分析层—数据利用层”。本文将从频谱大数据分析和频谱大数据利用两个方面梳理和总结现有研究中有效的机器学习技术。

2.1 面向频谱大数据分析的机器学习方法

针对频谱大数据分析,本文给出了4种各具特色的机器学习方法,包括分布式和并行式学习、极速学习、核学习和深度学习。

(1)分布式和并行式学习:针对海量频谱数据分析,分布式学习(Distributed learning)^[24]似乎是一个颇具前景的研究,因为不同于传统学习方法中需要将收集的数据存储于一个工作站中进行集中式处理,分布式学习通过将学习过程分配在不同的工作站以一种分布式的机制有效地实现学习算法的扩展。由于分布式学习避免了集中式处理带来的时间和能量开销,解决了大规模频谱数据存储于单一频谱设备带来的挑战,成为比传统学习技术更具优势的频谱大数据学习方法。在最近几年,分布式学习算法也陆续被提出,典型的包括判定规则(Decision rules)^[25]、堆积归纳(Stacked generalization)^[26]和分布式助推(Distributed boosting)^[27]。与分布式学习相似,另一个可用于扩展传统机器学习算法的学习技术便是并行学习(Parallel learning)^[28]。协同多核处理器和云计算平台的功能,分布并行式处理系统最近也慢慢出现,更多关于分布式和并行式学习的内容可参考文献[29]。

(2)极速学习:快速的机器学习技术是解决频谱大数据时间敏感性的有效方法,在线学习(Online learning)^[30, 31]作为一种即时学习方式,在大数据处理领域已得到广泛的研究。极速学习(Extreme learning machine, ELM)^[32, 33]是一种单隐层前馈神经网络学习算法,通过随机赋值单隐层神经网络的输入权值和偏差项,并通过一步计算即可解析求出网络的输出权值。相比于传统前馈神经网络训练算法需经多次迭代调整才可最终确定网络权值,极速学习具有学习速度更快,泛化能力更强,人为参与少等优势,成为了当下快速数据处理的优选学习方法,也逐步引起了人们越来越多的关注。基于极速学习的

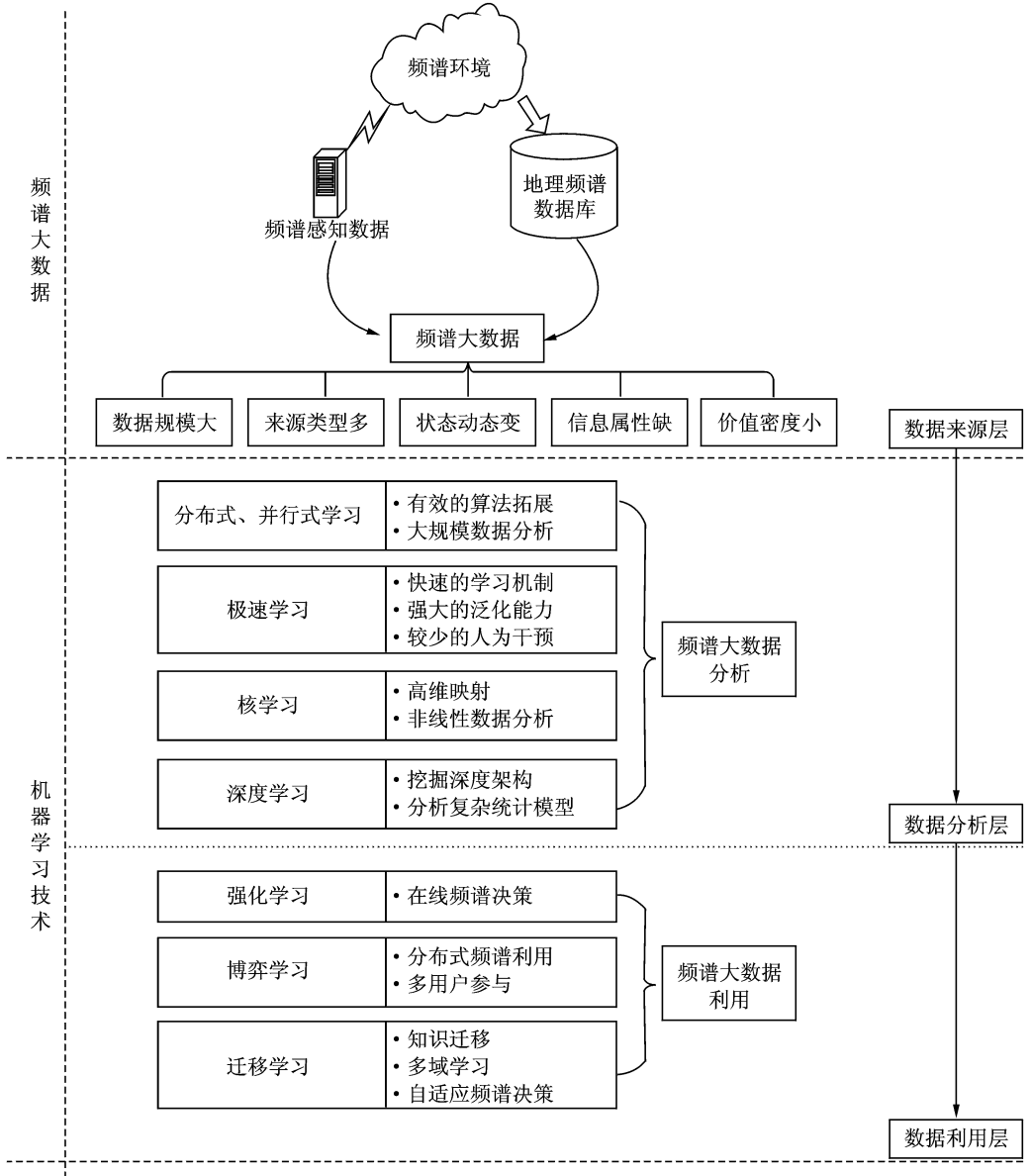


图6 面向频谱大数据处理的机器学习技术

Fig. 6 Machine learning techniques for big spectrum data processing

拓展算法也被陆续提出,例如,增量极速学习(Incremental ELM)^[34]、进化极速学习(Evolutionary ELM)^[35]、序数极速学习(Ordinal ELM)^[36]和均衡极速学习(Symmetric ELM)^[37],这些算法之间的关系在文献[38]中有详细的分析。

(3)核学习:频谱大数据由于来源多样,运用传统的线性学习理论方法往往存在局限性。针对这一挑战,核学习(Kernel-based learning)^[39, 40],一种新颖的可提升学习机计算能力的,并有效进行非线性数据分析的学习理论得到人们越来越多的关注。在核学习理论中,通过一个核函数将原始输入空间的数据映射到高维(甚至无限维)特征空间中,在特征空间中进行线性处理,这样使得复杂的学习问题变得易

于操作,大大增强了学习机的计算能力。针对现有线性协同频谱感知研究中存在的检测结果不准确这一问题,研究面向频谱大数据非线性统计处理的核学习理论,有助于提升非线性频谱数据融合和高维数据聚类等方面的性能^[41]。

(4)深度学习:深度学习(Deep learning)^[42]无疑成为机器学习领域最热的研究趋势之一。大多数传统的机器学习方法都是利用潜结构的学习架构,而深度学习则是利用监督或者非监督策略自主学习深度架构中的分层表示机制,它的优势在于可以捕获更加复杂的统计模型以实现自适应新领域^[2]。主流的深度学习方法有深度信念网(Deep belief networks)和卷积码神经网络(Convolutional neural networks)^[43, 44]。深度学习在最近几年已经引起了学术界乃至工业界的注意,并被成功应用到大数据处理系统中。如谷歌公司利用深度学习算法来处理谷歌翻译机、图像搜索引擎等产生的海量数据,微软公司的语言翻译机、IBM的智能计算机都成功利用深度学习技术来处理大数据^[2, 45]。因此,鉴于深度学习在大数据环境下的先进优势,针对频谱大数据分析而言,深度学习也将起到关键的作用。

2.2 面向频谱大数据利用的机器学习方法

针对频谱大数据利用,本文将分析可用于频谱智能决策的机器学习方法。

(1)强化学习:在频谱大数据利用过程中,往往存在网络状态动态变化、环境条件未知等因素,因此,需要先验样本进行训练的离线学习算法不再适用。而强化学习(Reinforcement learning)^[46, 47]采用“学习—决策—执行”的环路实现频谱的实时决策,优点在于:首先,它不依赖于先验的环境信息;其次,可实现边学边用;最后,具备较强的稳健性。因此,基于强化学习技术,利用频谱大数据进行实时信道调整、功率控制和网络选择等决策时具有一定的优势。

(2)博弈学习:博弈论目前已经成为无线通信中多用户分布式决策优化的重要方法之一^[48]。一方面,博弈学习(Game learning)方法非常适合于解决如频谱拍卖^[49]等直接类比成经济活动的宏观无线资源管理问题;另一方面,它也能很好地解决信道接入^[50]、网络选择^[51]等微观无线资源优化问题。下一代移动通信系统5G,将向着网络密集化、自组织的方向不断发展,分布式、多用户的特征越加明显,频谱数据量也将急剧增加。博弈学习在利用频谱大数据进行智能决策的作用也将愈显突出。

(3)迁移学习:许多传统的机器学习算法的一大假设就是训练和测试数据来源于同样的特征空间,具备同样的分布^[52]。然而,来自于频谱感知设备和地理频谱数据库的频谱大数据,具有结构、半结构和非结构的多样特性。这样的数据异构化特质使得传统机器学习的假设不再成立。为应对这一问题,迁移学习(Transfer learning)^[52, 53],一种可以从不同领域、不同任务和不同分布的数据中提取知识的新兴学习技术被提出。迁移学习的优势就在于它可以应用先前学到的知识来更快速地解决新遇到的问题,具备一定的多域学习能力^[54]。对于频谱大数据利用而言,迁移学习可以基于自身的知识迁移能力,很好地实现频谱的智能决策。

2.3 讨 论

本节系统总结了在频谱大数据环境下具备独特优势的机器学习技术。在频谱大数据分析方面,介绍了面向大规模频谱数据分析的分布式和并行式学习、面向快速频谱数据分析的极速学习、面向异构频谱数据分析的核学习和面向复杂统计模型频谱数据分析的深度学习。在频谱大数据利用方面,介绍了侧重于在动态不确定频谱环境下,从单用户视角利用频谱数据,寻求最优决策的强化学习方法;分析了侧重于在多用户频谱系统中,为实体寻求分布式决策的博弈学习方法;阐述了具备知识迁移,可实现自适应多域频谱数据利用的迁移学习方法。

其他的一些学习技术,如表示学习(Representation learning)^[55]、主动学习(Active learning)^[56]、子空间学习(Subspace learning)^[57]等,对于频谱大数据处理也有一定的应用价值。总的来说,面向频谱大数据处理的机器学习技术不再是传统学习中仅关注与某一算法的性能提升,将来需要的可能是多种学习

方法的联合使用,甚至是一个可供数据处理系统访问的学习算法库。

3 未来研究趋势

在过去几年,虽然运用机器学习来分析和处理大数据的研究已经取得一定进展,然而,在无线通信领域,为应对频谱赤字危机,利用机器学习技术挖掘频谱大数据中隐藏的宝贵信息来实现无线网络中有效的频谱管理依然面临许多挑战。这里,作者将给出一些开放性话题和可能的研究趋势,以供读者做进一步的探索。

(1)提起大数据,经常会将数据挖掘、云计算、知识发现、信号处理等技术同机器学习联系起来,一方面是因为它们在海量数据处理方面可以发挥重要的作用,另一方面也在于它们之间有很强的相关性。在频谱大数据环境下,未来的发展必将向着技术融合的方向不断前进,然而,如何将这些智能处理方法集成到机器学习理论中去将会是一个很值得研究的问题。

(2)5G 移动网络、物联网和互联网+等新兴技术的不断发展,使得无线业务量持续增长。在这些不同领域的频谱数据具有一定的共性特征,也有各自的鲜明特点,如何利用机器学习方法针对特定领域中的频谱大数据进行分析处理依然是一项挑战性的工作。

(3)频谱作为无线信号赖以传输的媒介,其中包含的海量频谱数据,对于管理者来说,将蕴含着巨大的商业价值。对于消费者来说,同时也涵盖着许多人的隐私。因此,如何在运用机器学习和数据挖掘技术进行频谱数据分析利用的同时,还能保护用户的隐私安全也将成为一个非常值得关注的问题。

4 结束语

有效的频谱数据处理对于频谱资源管理具有重要的意义,然而伴随着信息技术的飞速发展,海量规模的复杂频谱大数据的产生给传统的数据处理系统带来了极大挑战。本文主要分析了频谱大数据的特点,并从5个方面对频谱大数据进行了阐述。然后,从频谱大数据分析和频谱大数据利用两个角度,系统梳理了在频谱大数据环境下,具备先进优势的机器学习技术。最后给出了一些开放性话题和研究趋势。面向频谱大数据处理的机器学习方法这一领域的研究才刚刚开始,将更加智能、高效、便捷的学习技术应用于频谱数据处理依然有待在接下来的实际工作和研究中继续探索。

参考文献:

- [1] Gantz J, Reinsel D. The digital universe in 2020: Big data, bigger digital shadows, and biggest growth in the far east[M]. Massachusetts: IDC, 2012.
- [2] Chen X W, Lin X. Big data deep learning: Challenges and perspectives[J]. IEEE Access, 2014,2:514-525.
- [3] Hu H, Wen Y, Chua T S, et al. Toward scalable systems for big data analytics: A technology tutorial[J]. IEEE Access, 2014,2:652-687.
- [4] Ding W, Wu G, Zhu X, et al. Data mining with big data[J]. IEEE Transactions on Knowledge and Data Engineering, 2014, 26(1):97-107.
- [5] Manyika J, Chui M, Brown B, et al. Big data: The next frontier for innovation, competition, and productivity [M]. USA: McKinsey Global Institute, 2011.
- [6] Meijer E. The world according to LINQ[J]. Communications of the ACM, 2011,54(10):45-51.
- [7] Zikopoulos P, Eaton C. Understanding big data: Analytics for enterprise class Hadoop and streaming data[M]. New York: McGraw-Hill, 2011.
- [8] Wang C X, Haider F, Gao X, et al. Cellular architecture and key technologies for 5G wireless communication networks[J]. IEEE Communications Magazine, 2014,52(2):122-130.
- [9] Russell S, Norvig P. Artificial intelligence: A modern approach[M]. New Jersey: Prentice-Hall, 1995.
- [10] Mitchell T M. Machine learning[M]. New York: McGraw-Hill, 1997.
- [11] Ding G, Wu Q, Wang J, et al. Big spectrum data: The new resource for cognitive wireless networking[EB/OL]. <http://arxiv.org/abs/1404.6508>, 2014-04-23.

- [12] Mayer-Schonberger V, Cukier K. Big data: A revolution that will transform how we live, work, and think[M]. USA: Eamon Dolan/Houghton Mifflin Harcourt, 2013.
- [13] Chen M, Mao S, Liu Y. Big data: A survey[J]. *Mobile Networks and Applications*, 2014,19(2):171-209.
- [14] Che D, Safran M, Peng Z. From big data to big data mining: Challenges, issues, and opportunities[C]//18th International Conference on Database Systems for Advanced Applications (DASFAA). Bonghee Hong, Berlin, Germany: Springer Berlin Heidelberg, 2013: 1-15.
- [15] Liang Y C, Chen K C, Li G Y, et al. Cognitive radio networking and communications: An overview[J]. *IEEE Transactions on Vehicular Technology*, 2011,60(7):3386-3407.
- [16] Sengupta S, Subbalakshmi K P. Open research issues in multi-hop cognitive radio networks [J]. *IEEE Communications Magazine*, 2013,51(4):168-176.
- [17] Andrews J G, Buzzi S, Choi W, et al. What will 5G be? [J]. *IEEE Journal on Selected Areas in Communications*, 2014,32(6):1065-1082.
- [18] Li Q C, Niu H, Papathanassiou A T, et al. 5G network capacity: Key elements and technologies[J]. *IEEE Vehicular Technology Magazine*, 2014,9(1):71-78.
- [19] Jungnickel V, Manolakis K, Zirwas W, et al. The role of small cells, coordinated multipoint, and massive MIMO in 5G[J]. *IEEE Communications Magazine*, 2014,52(5):44-51.
- [20] Atzori L, Iera A, Morabito G. The internet of things: A survey[J]. *Computer Networks*, 2010,54(15):2787-2805.
- [21] Tsai C W, Lai C F, Chiang M C, et al. Data mining for internet of things: A survey[J]. *IEEE Communications Surveys and Tutorials*, 2014,16(1):77-97.
- [22] Gantz J, Reinsel D. Extracting value from chaos [M]. Hopkinton: EMC, 2011.
- [23] Rudin C, Wagstaff K L. Machine learning for science and society[J]. *Machine Learning*, 2013,95(1):1-9.
- [24] Peteiro-Barral D, Guijarro-Berdiñas B. A survey of methods for distributed machine learning[J]. *Progress in Artificial Intelligence*, 2013,2(1):1-11.
- [25] Chen H, Li T, Luo C, et al. A rough set-based method for updating decision rules on attribute values' coarsening and refining[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2014,26(12):2886-2899.
- [26] Chen J, Wang C, Wang R. Using stacked generalization to combine SVMs in magnitude and shape feature spaces for classification of hyperspectral data[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2009,47(7):2193-2205.
- [27] Sarnovsky M, Vronc M. Distributed boosting algorithm for classification of text documents[C]//IEEE 12th International Symposium on Applied Machine Intelligence and Informatics (SAMII). Herl'any, Slovakia: IEEE Computer Society Press, 2014:217-220.
- [28] Langford J. Parallel machine learning on big data[J]. *XRDS: Crossroads, The ACM Magazine for Students*, 2012,19(1):60-62.
- [29] Bekkerman R, Bilenko M, Langford J. Scaling up machine learning: parallel and distributed approaches[M]. Oxford: Cambridge University Press, 2011.
- [30] Slavakis K, Kim S J, Mateos G, et al. Stochastic approximation vis-a-vis online learning for big data analytics[J]. *IEEE Signal Processing Magazine*, 2014,31(6):124-129.
- [31] Wang J, Zhao P, Hoi S C H, et al. Online feature selection and its applications[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2014,26(3):698-710.
- [32] Huang G B, Zhu Q Y, Siew C K. Extreme learning machine: Theory and applications[J]. *Neurocomputing*, 2006,70:489-501.
- [33] Huang G B, Wang D H, Lan Y. Extreme learning machines: A survey[J]. *International Journal of Machine Learning and Cybernetics*, 2011,2(2):107-122.
- [34] Huang G B, Chen L, Siew C K. Universal approximation using incremental constructive feedforward networks with random hidden nodes[J]. *IEEE Transactions on Neural Networks*, 2006,17(4):879-892.
- [35] Zhu Q Y, Qin A K, Suganthan P N, et al. Evolutionary extreme learning machine[J]. *Pattern Recognition*, 2005,38(10):1759-1763.
- [36] Deng W Y, Zheng Q H, Lian S, et al. Ordinal extreme learning machine[J]. *Neurocomputing*, 2010,74(13):447-456.
- [37] Liu X, Li P, Gao C. Symmetric extreme learning machine[J]. *Neural Computing and Applications*, 2013,22(3/4):551-558.
- [38] Ding S, Xu X, Nie R. Extreme learning machine and its applications[J]. *Neural Computing and Applications*, 2014,25(3/4):549-556.
- [39] Li C, Georgiopoulos M. A unifying framework for typical multi-task multiple kernel learning problems[J]. *IEEE Transac-*

tions on Neural Networks and Learning Systems, 2014,25(7):1287-1297.

- [40] Montavon G, Braun M, Krueger T, et al. Analyzing local structure in kernel-based learning: Explanation, complexity, and reliability assessment[J]. IEEE Signal Processing Magazine, 2013,30(4):62-74.
- [41] Ding G, Wu Q, Yao Y D, et al. Kernel-based learning for statistical signal processing in cognitive radio networks: Theoretical foundations, example applications, and future directions[J]. IEEE Signal Processing Magazine, 2013,30(4):126-136.
- [42] Arel I, Rose D C, Karnowski T P. Deep machine learning—a new frontier in artificial intelligence research[J]. IEEE Computational Intelligence Magazine, 2010,5(4):13-18.
- [43] Callet P L, Viard-Gaudin C, Barba D. A convolutional neural network approach for objective video quality assessment[J]. IEEE Transactions on Neural Networks, 2006,17(5):1316-1327.
- [44] Yu D, Deng L. Deep learning and its applications to signal and information processing[J]. IEEE Signal Processing Magazine, 2011,28(1):145-154.
- [45] Jones N. Computer science: The learning machines[J]. Nature, 2014,505(7482):146-148.
- [46] Vassiliades V, Cleanthous A, Christodoulou C. Multiagent reinforcement learning: Spiking and nonspiking agents in the iterated Prisoner's Dilemma [J]. IEEE Transactions on Neural Networks, 2011,22(4):639-653.
- [47] Hu Y, Yang G, Bo A. Multiagent reinforcement learning with unshared value functions[J]. IEEE Transactions on Cybernetics, 2014,45(4):647-662.
- [48] Xu Y, Anpalagan A, Wu Q, et al. Decision-theoretic distributed channel selection for opportunistic spectrum access: Strategies, challenges and solutions[J]. IEEE Communications Surveys and Tutorials, 2013,15(4):1689-1713.
- [49] Gao L, Wang X, Xu Y, et al. Spectrum trading in cognitive radio networks: A contract-theoretic modeling approach[J]. IEEE Journal on Selected Areas in Communications, 2011,29(4):843-855.
- [50] Xu Y, Wu Q, Shen L, et al. Opportunistic spectrum access with spatial reuse: Graphical game and uncoupled learning solutions[J]. IEEE Transactions on Wireless Communications, 2013,12(10):4814-4826.
- [51] Elias J, Martignon F, Chen L, et al. Joint operator pricing and network selection game in cognitive radio networks: Equilibrium, system dynamics and price of anarchy[J]. IEEE Transactions on Vehicular Technology, 2013,62(9):4576-4589.
- [52] Pan S J, Yang Q. A survey on transfer learning[J]. IEEE Transactions on Knowledge and Data Engineering, 2010,22(10):1345-1359.
- [53] Xiang E W, Cao B, Hu D H, et al. Bridging domains using world wide knowledge for transfer learning[J]. IEEE Transactions on Knowledge and Data Engineering, 2010,22(6):770-783.
- [54] Wang P, Domeniconi C, Hu J. Cross-domain text classification using Wikipedia[J]. IEEE Intelligent Informatics Bulletin, 2008,9(1):5-17.
- [55] Bengio Y, Courville A, Vincent P. Representation learning: A review and new perspectives[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012,35(8):1798-1828.
- [56] Fu Y, Li B, Zhu X, et al. Active learning without knowing individual instance labels: A pairwise label homogeneity query approach[J]. IEEE Transactions on Knowledge and Data Engineering, 2014,26(4):808-822.
- [57] Mardani M, Mateos G, Giannakis G B. Subspace learning and imputation for streaming big data matrices and tensors[J]. IEEE Transaction on Signal Processing, 2015,63(10):2663-2677.

作者简介:



吴启晖(1970-),男,教授,博士生导师,研究方向:认知无线电、频谱管理以及信号处理, E-mail: wuqihui2014@sina.com。



邱俊飞(1990-),男,通信作者,硕士研究生,研究方向:机器学习、大数据分析以及无线网络, E-mail: junfeiqiu@163.com。



丁国如(1986-),男,讲师,研究方向:认知无线电、统计信号处理、机器学习以及无线网络大数据分析。