

·专辑:人工智能与情报学·

## 机器学习视角的人工智能研究回顾及对图书情报学的影响\*

刘 浏<sup>1,2,3</sup> 王东波<sup>1,4</sup> 黄水清<sup>1,4</sup>

(1.南京农业大学领域知识关联研究中心 江苏南京 210095)

(2.江苏省数据工程与知识服务重点实验室(南京大学) 江苏南京 210023)

(3.南京大学信息管理学院 江苏南京 210023)

(4.南京农业大学信息科学技术学院 江苏南京 210095)

**摘要:**人工智能是近年来异常火热的话题,其对世界的影响和改变已经随处可见,这其中机器学习尤其是深度学习发挥着至关重要的作用。文章梳理了人工智能的发展脉络,阐述了机器学习在人工智能中所处的地位,对机器学习中重要和常见的模型和算法进行了简要介绍,其中包括近年最为重要的深度学习。人工智能的热潮对于图书情报来说是一个极佳的发展契机,文章也从多个角度分析了机器学习对图书情报学科产生的影响,展望了两者充分结合的发展前景和未来趋势,并以古文信息处理为例进一步阐述了机器学习方法在图书情报学研究中发挥的作用。

**关键词:**人工智能;机器学习;深度学习;图书馆学;情报学;古文信息处理

中图分类号:TP18;G250

文献标识码:A

DOI:10.11968/tsyqb.1003-6938.2017118

### Review of the Study about Machine Learning and Its Influence on Library Science

**Abstract** Artificial Intelligence is one of the most popular topics lately and it has been changing the way of our life. Machine learning is playing the most important role in all these effects. This study introduced the development of AI along with machine learning, deep learning and most of their models and methods. It is truly a great chance for LIS to catch up with such a wave of AI. The promising future of LIS topics with machine learning methods is widely analyzed and the ancient Chinese information processing is discussed as an example at last.

**Key words** artificial intelligence; machine learning; deep learning; science of library; science of information; information processing of Ancient Chinese

## 1 引言

近年来,深度学习获得了革命式的发展,在图像视频处理、语音识别、自然语言处理等领域的诸多研究问题中均取得了突破和成功<sup>[1]</sup>。这些发展对人类的生产生活产生了深远的影响,小到不易察觉的语音助手(如苹果的 Siri、微软的 Cortana 等),大到万众瞩目的 AlphaGo,人工智能这一专业性极强的研究课题成为人们茶余饭后的时髦话题,整个社会都在热切展望这次人工智能的“第三次浪潮”。以机器学习为主的人工智能的迅猛发展对图书情报领域的影响

也是深刻的,尤其是大数据时代下,文献、信息、数据呈爆发式增长,这使得传统的以个案、抽样为主的研究方法越来越难于获得可靠的结论,而机器学习方法凭借其在处理海量数据方面的先天优势以及不断提高的性能和效果,可以在图书情报学的未来发展中发挥重要作用。

## 2 人工智能的发展历程

简单地说,人工智能(Artificial Intelligence, AI)的目标就是制造出智能体(Intelligent Entities 或 Intelligent Agents),智能体能够感知周围环境,并做出

\* 本文系国家社会科学基金重大项目“基于《汉学引得丛刊》的典籍知识库构建及人文计算研究”(项目编号:15ZDB127)与南京农业大学人文社会科学基金项目“基于《汉学引得丛刊》的典籍知识库构建及人文计算研究”(项目编号:SKPT2016001)研究成果之一。

收稿日期:2017-12-15;责任编辑:魏志鹏

能够使其目标的成功率最大的行动<sup>[2]</sup>。

## 2.1 诞生前的发展

人类对于制造出智能机器人的想法早已有之,但人工智能作为一个正式的学科却十分年轻,最早的相关研究始于1943年,如McCulloch和Pitts<sup>[3]</sup>对于神经网络的逻辑计算能力的研究,被公认为现代人工智能领域最早的经典研究。这一时期人工智能研究的桂冠属于阿兰·图灵,其发表的《计算机与智能》<sup>[4]</sup>一文堪称杰作,其中最著名的就是“图灵测试”。图灵测试为人工智能带来的突破在于,研究者可以暂时不再深究人类的智能是什么,直接从行为主义的角度探索智能的机器,看哪些机器表现得像是具备智能。这在人工智能领域掀起一波浪潮,通过图灵测试成为人工智能研究者的目标。其后也有大量研究提出对其的批判或改进,其中最著名的要数塞尔的“中文屋”假设<sup>[5]</sup>,但是图灵测试的影响力至今仍未衰减。

## 2.2 早期发展的乐观

一般认为,1956年在达特茅斯学院举办的人工智能研讨会标志的人工智能的诞生,这是“人工智能(Artificial Intelligence)”一词第一次在正式场合被使用。在该会议上Newell和Simon提出的一个推理程序——逻辑理论家(Logic Theorist,LT)得到了最多的关注,该程序已经具备了推理能力,并在其后证明了罗素和怀特海的《数学原理》第二章中的38个定理(共52个)<sup>[6-7]</sup>。达特茅斯会议的重要性在于宣告人工智能成为了独立的学科。

人工智能前二十年的发展十分火热,其中最为常见的就是推理问题,如Newell等在LT基础上设计的通用问题解决程序<sup>[8]</sup>;Gelernter设计的几何定理证明程序<sup>[9]</sup>,其他类似的还有如问题回答和计划系统<sup>[10]</sup>,基于分解法则的一阶逻辑定理证明算法<sup>[11]</sup>等,著名的跳棋程序<sup>[12]</sup>也是在这样的背景下提出的。达特茅斯会议的发起者McCarthy在这一时期对人工智能的发展做出了突出的贡献,首先他提出的Lisp编程语言和分时操作对计算机发展产生了深远的影响。此外,他还设计了一个叫意见采纳者的程序<sup>[13]</sup>,具备获取知识的能力。与此同时,由Minsky带领的一批学者致力于研究在“微世界”中的人工智能,其

中最有影响力的要数“积木世界(blocks world)”,视觉问题<sup>[14]</sup>、学习问题<sup>[15]</sup>、自然语言理解<sup>[16]</sup>等一系列与该世界相关的问题得到关注。在同一时期,Rosenblatt对感知器相关的一系列研究<sup>[17,1]</sup>,也将神经网络研究推向新的高度。

## 2.3 中期发展的徘徊

早期人工智能的大量成功尝试给学界带来了一股乐观的气氛,然而“美好”的愿景没有如期而至。对困难的预计不足,使得人们对人工智能的发展瓶颈缺乏准备,以至于在诸多等问题上迟迟难以突破,在耗尽了政府的耐心和研究资金后,人工智能于1970年代初期堕入寒冬(AI winter)之中<sup>[2]</sup>。到了1970年代后半期,人工智能专家将目光转向了领域知识(domain knowledge)问题,如记忆的组织<sup>[19]</sup>、计划的理解等<sup>[20]</sup>。这一始于化学中分子结构判定问题<sup>[21]</sup>的研究模式,在自然语言理解上也得到了成功的应用<sup>[22]</sup>。

在1980年代初,专家系统<sup>[23]</sup>在商业上的成功,使得人工智能获得了新生。日本在同一时期建造“第五代计算机”的计划,刺激了美英等国对人工智能新的投入,短时间内人工智能又一次呈现出繁荣的景象。在神经网络方面,反向传播算法重新得到尝试,获得了广泛的关注<sup>[24]</sup>,随之兴起的联结主义(connectionist)方法开始经典的符号主义和逻辑主义方法分庭抗礼<sup>[25]</sup>。然而,产业上过高的期望并没有带来相应的成果,人工智能迅速又陷入低潮。

## 2.4 机器学习的新热潮

到了1990年代以后,随着计算机性能的不断提高,人工智能又迎来了一次深刻的转变,有数学依据的统计模型、大规模的训练语料、可重复的测试语料,这种在语音识别和机器翻译等领域的颇受欢迎的研究模式,这种融合了统计学、数学、信息论等各种领域知识的机器学习方法,逐渐成为主流,这使得人工智能真正成为一门严谨的、可靠的科学<sup>[2]</sup>。随着隐马尔可夫模型、神经网络、贝叶斯网络等各种模型方法的不断引入,机器学习方法的发展越发迅猛,尤其在模式识别、自然语言处理、数据挖掘等领域,更是成为中流砥柱。如今,以神经网络模型为基础的深度学习方法,正在引领新一轮的人工智能热潮,无数以机器学习为基础的应用和产品正在以显著或悄然

的方式改变着人类的生活,改变着全世界。

近年来话题度最高的人工智能应用非 AlphaGo 莫属,这个“围棋机器人”在接连战胜了李世石、柯洁等世界围棋冠军后,使得人工智能的话题传遍大街小巷。实际上,AlphaGo 就是一个机器学习的应用,其采用了最新的深度学习模型,从数以万计的围棋训练数据中“学习”了如何下围棋。类似的机器学习应用还有很多,如苹果、微软、谷歌等公司开发的各类语音助手,得益于近年来兴起的 LSTM-RNN 模型,谷歌的语音搜索、讯飞输入法等语音输入工具也是基于同样的技术完成的。另一个应用就是人脸识别技术,这其中卷积神经网络模型扮演了最重要的角色,使得机器能够学习到人脸的特征,从而对不同的人脸进行判别。除此之外,自动翻译工具如谷歌翻译等,也得益于机器学习在机器翻译中取得的长足进展。

机器学习正在而且将持续地改变人类对世界的认识,对于不论什么领域的研究者来说,了解、理解甚至跟随这样的发展趋势,对于研究者的成长,对于学科的发展,都是有利的。**图书情报学**这个与信息科学、计算机科学密切相关的学科,也理应积极迎接这样的发展潮流。用最新的**机器学习**的技术拓展自身的研究视野和研究跨度,将为学科带来值得期待的新发展。

### 3 机器学习概览

机器学习(统计学习、统计机器学习)是人工智能的一个重要分支,这种方法主要是一个根据已有数据建立数学模型,并根据模型对新的数据进行预测的过程<sup>[26]</sup>。而所谓的学习,就是计算机从数据中产生模型的过程<sup>[27]</sup>。随着机器学习的深入发展,训练集、模型和测试集三者构成了一种较为固定的研究范式,相关研究或是构建特征丰富的大规模训练集、测试集以提高已有模型效果或供以比较不同的模型,或者是调整或改进已有模型以期在原有训练集和测试集上获得更好的效果,而不同研究的效果一般通过正确率、召回率和调和平均值(F 值)来比较。这种研究范式的优势在于,可以比较清晰直观地比较出各种模型之间的优劣,而这也使得研究者不断致力于提高模型的性能,使其在特定测试集上获得

越来越好的效果。由于这种优势,机器学习不断保持着迅猛发展的态势,且随着计算机性能的不断提高,数据集不断变大,机器学习的效果也越来越好。尤其是以神经网络模型为基础的深度学习,在近年来获得了突破性的成功,使得机器学习成为人工智能中最炙手可热的研究领域。

#### 3.1 监督学习和无监督学习

机器学习根据不同的任务需求可以大致分为两大类:监督学习和无监督学习。(1)在监督学习中,会先对训练集(输入集)进行一定程度的标注得到相应的输出集,所谓标注就是指人工构建一个机器学习的“标准答案”,此时计算机从输入集和输出集之间的关系中学习得到相应的模型,之后就可以根据模型来对新出现的类似问题进行预测。根据输出集中数据的离散或连续的特点,**监督学习还可以细分为分类和回归两种**;(2)对于无监督学习来说,训练集没有经过任何标注,计算机要自动地从数据中挖掘出所需要的结果。与监督学习中的分类相对应的就是无监督学习中的聚类,两者之间的区别在于分类任务在计算机学习之前已有预先定义好的类别,而聚类任务没有预先定义的类别,需要根据数据的特点自行决定类别和类别数量,并进行相应的分类。

#### 3.2 朴素贝叶斯分类器

朴素贝叶斯分类器(Naive Bayes classifier)是机器学习中最经典、最简单的统计分类模型<sup>[28]</sup>。该模型以贝叶斯公式为基础,根据某类别条件下每个特征的条件概率和每个类别的先验概率,来求得某些特征条件下每个类别的条件概率,条件概率最大的类别就是模型所得的类别,以此来完成分类的任务。朴素贝叶斯的“**朴素**”之处在于,其假定不同特征之间相互独立,这也使得该模型特别适用于**文本分类**这样的问题<sup>[29]</sup>。

#### 3.3 决策树

决策树模型是一种用树形结构构造的分类模型。一个决策树由结点和有向边构成,分类时在决策树上的每一个内部结点上进行判断,判断的结果得到一个子结点,直到到达一个叶子节点,完成分类的任务<sup>[30-31]</sup>。决策树处理分类问题时,如果预测的结果都是离散的,此时也可以称为分类树,而当预测的结果是连



续的,决策树又可以化作回归树,而与之对应的分类与回归树模型<sup>[32]</sup>是应用最为广泛的决策树模型。

### 3.4 最大熵模型

最大熵模型是自然语言处理中最常见的模型之一,常用来处理序列化标注的问题<sup>[33]</sup>。该模型以最大熵原理为基础,认为满足一定约束条件的前提下,熵最大的模型是最好的模型。最大熵原理最早由 Jaynes 于 1957 年提出<sup>[34-35]</sup>,其要求模型先满足一定的约束条件,这些约束条件与特征密切相关。特征一般用特征函数的方式来表示,而特征函数的定义可以根据问题的需求自由地增加或者减少,这也是最大熵模型的一大优势。最大熵模型的学习过程,是根据训练语料对模型进行极大似然估计。最常见的优化算法有 GIS 算法<sup>[36]</sup>、改进的迭代尺度法<sup>[37]</sup>和拟牛顿法中的 L-BFGS 方法<sup>[38]</sup>等。最大熵模型在词性标注<sup>[39]</sup>、命名实体识别<sup>[40]</sup>、信息抽取<sup>[41]</sup>等问题上均有着较好的表现。

### 3.5 隐马尔可夫模型

隐马尔可夫模型是一种最经典的统计模型<sup>[42-44]</sup>。在自然语言处理中,隐马尔可夫模型一般用于标注问题,尤其在词性标注<sup>[45]</sup>、命名实体识别<sup>[46]</sup>、信息抽取<sup>[47]</sup>等问题上。隐马尔可夫模型是一个关于“双层的随机过程”的模型,其中一个随机过程无法直接观测得到(所谓的隐藏,Hidden),只能通过另一个可观测到的随机过程来间接获得。隐马尔可夫模型有两个重要的假设<sup>[48]</sup>:齐次马尔可夫假设和观测独立性假设。隐马尔可夫模型一般用于解决三种问题,其中解码问题最常被用于自然语言处理中的自动标注,该问题一般转化为一种求解最优序列的问题,一般使用维特比算法<sup>[49]</sup>来求解,可以极大地提高求解的效率。

### 3.6 条件随机场

条件随机场是一种条件概率模型,根据一组输入随机变量对相应的输出随机变量进行预测。条件随机场模型常用在序列化标注问题中,此时一般采用线性链形式的条件随机场<sup>[50]</sup>,将输出序列表示一个马尔可夫随机场(概率无向图),并将问题转换成由输入序列预测输出序列的问题。用线性条件随机场模型来描述序列化标注问题,与隐马尔可夫模型存在相似之处,其不同在于,隐马尔可夫模型中每个

标记(状态)只与其前一个标记(状态)有关,而线性链条件随机场中每个标记与其前后一个标记都有关,后者显然比前者能够表示更为丰富的关系。条件随机场一般也使用维特比算法来求解。条件随机场凭借其较强的学习能力,在词性标注<sup>[51]</sup>、命名实体识别<sup>[52-53]</sup>、关系识别<sup>[54]</sup>、浅层分析<sup>[55]</sup>等众多自然语言处理问题中均表现出色,是一个用途广泛且性能优化的机器学习模型。

### 3.7 支持向量机

支持向量机是一种二类分类模型<sup>[56]</sup>,是文本分类研究中,最常用的一种机器学习模型。其在文本向量空间中构造一个超平面,将不同类别的文本向量分割开来,以实现分类的效果。对于线性可分的文本空间来说,支持向量机通过间隔最优化的方法求解得到唯一的超平面<sup>[57]</sup>;而当线性不可分时,一般有两种策略:对于近似线性可分的文本空间,可以采用软间隔最大化方法<sup>[58]</sup>;对于线性不可分的文本空间而言,可以引入核方法<sup>[59]</sup>。支持向量机的求解可以看作一种凸二次规划的问题,一般采用序列最小优化这种启发式的算法来快速求解<sup>[60]</sup>。支持向量机是一种非常有效的分类方法,由于其面对线性可分和线性不可分问题时灵活变化的能力,使得其在文本分类研究中广受青睐<sup>[61-64]</sup>,而 Chang 和 Lin 构建的开源工具包——LIBSVM<sup>[65]</sup>更是促进了该方法的传播,使之成为机器学习中最常用的分类模型之一。

### 3.8 聚类方法

聚类是无监督学习中最主要的一种研究方法<sup>[66-67]</sup>,在机器学习中常被用在模式识别、图像分析、信息抽取等领域。其思想是将数据集中有相同特征的个体聚合成类。聚类并不是一种特定的模型或算法,而是一组具有共同目标和思路的算法的统称,一般可以将聚类分为原型聚类、密度聚类和层次聚类三种<sup>[27]</sup>。常见的原型聚类方法有 k 均值、学习向量量化、高斯混合聚类<sup>[68]</sup>;密度聚类方法有 DBSCAN<sup>[69]</sup>和 OPTICS<sup>[70]</sup>等;层次聚类的方法有 AGNES 算法和 DIANA 算法<sup>[71]</sup>。

## 4 深度学习的革新

深度学习(deep learning)是近年来最成功和最

受关注的机器学习方法<sup>[1]</sup>,其核心思路是使用多层的神经网络来表征数据,这样做的优势在于训练数据无需进行人工标注,而是直接通过表征学习的方式自动获取数据中的特征,因此深度学习可以灵活地用于监督学习、半监督学习和无监督学习问题中<sup>[72-73]</sup>。深度学习并不是一个特定的模型或算法,而是一系列具有共同特点的机器学习方法的统称,这种特点总得来说就是“深度”,这具体包括使用多层非线性处理单元表征数据,进行监督或无监督学习,多层的表征对应了不同的抽象层次,使用反向传播算法进行训练的优化<sup>[74]</sup>。

深度学习的思想始终伴随着人工智能的发展,在神经网络模型提出伊始,就已有相应的机器学习的尝试<sup>[75-76]</sup>。而到了20世纪80年代,当时正值人工智能第二次兴起,反向传播算法重新发挥活力,联结主义高歌猛进,神经网络模型又重新得到关注<sup>[77]</sup>。然而受制于当时计算机的性能,神经网络的训练时间过长,无法投入实际使用当中<sup>[78]</sup>,也并未获得太多关注,而随着人工智能冬天的来临,深度学习也随之陷入沉寂。到了20世纪90年代以后,人工智能又一次兴起,神经网络模型又一次被重新审视,而在早期的尝试中,以训练速度过慢为主的一系列问题仍然很难解决<sup>[79-80]</sup>,虽然仍有研究者在该方法上坚持探索<sup>[81]</sup>,但与同时期的SVM、HMM等方法相比,深度学习(神经网络)方法不论在速度上还是在效果上都要逊色不少。

然而随着计算机性能不断提高,大规模数据越来越普遍,一场属于深度学习(神经网络)的“文艺复兴”席卷而来。所谓的“深度学习革命”,源自Hinton等人于2006年提出的深度置信网络(Deep Belief Nets, DBN)的高效训练方法<sup>[82]</sup>,该方法表明深度学习在训练时间效率上已经不逊色于其他机器学习方法。随着研究的深入,深度学习逐渐崭露头角,尤其是在语音识别和图像识别领域,深度学习方法迅速地实现了追赶、反超到统治的过程。(1)在语音识别方面,深度神经网络在Hinton和Deng的合作下不断取得突破<sup>[83-84]</sup>,并解决了深度学习中梯度消失困难的长短时记忆网络(Long Short-Term Memory, LSTM)<sup>[85]</sup>,更是刮起了一阵旋风,使得LSTM-RNN方法在语音识别中占据了统治地位<sup>[86-87]</sup>,如苹果的

SIRI、谷歌的Now、微软的Cortana、亚马逊的Alexa等语音助手无一不是基于LSTM-RNN方法实现的<sup>[88]</sup>;(2)而在图像识别方面,则是卷积深度神经网络<sup>[89]</sup>(Convolutional deep neural networks, CNNs)在大放光彩,相关方法在短时间内(2011-2012)的数个图像识别评测会议上拔得头筹,并不断获得更好的效果<sup>[90-91]</sup>,将深度学习热潮继续向前推进,如今在OCR、人脸识别等应用中,CNN都是首选方法。近年家喻户晓的AlphaGo<sup>[92]</sup>,更是将人工智能和深度学习推向了一个高潮。

深度学习的另一项研究热点在自然语言处理,其中最为重要的就是词向量<sup>[93]</sup>(word embedding)概念的提出,以及word2vec对该概念的模型实现<sup>[94-95]</sup>。这种将文本中的词语分布式地表示在向量空间中的方法,可以直接将表征结果作为输入层用于RNN等深度神经网络中<sup>[96]</sup>,从而避免了传统方法过多地通过句法、语义等语言学知识来选择特征的过程。目前在句法成分分析<sup>[97]</sup>、情感分析<sup>[98]</sup>、关系抽取<sup>[99]</sup>、机器翻译<sup>[100]</sup>等领域,深度学习均发挥出不错的效果,其中著名的谷歌机器翻译使用了LSTM网络<sup>[101]</sup>。而在自然语言处理的传统领域如词性标注、命名实体识别、语义角色分析等问题上,深度学习方法也有着不俗的表现<sup>[102]</sup>。

## 5 机器学习在图书情报领域的应用前景

深度学习的热潮还在持续,对于图书情报学来说,抓住这样的时机,紧跟时代的步伐对于学科的发展将大有裨益。将传统的信息服务和信息处理手段与新兴的深度学习技术有效结合,将为学科带来许多值得期待的新思路、新理念。如文本信息的自动化、数字化服务,智能的专家知识问答系统,以文本、语音、图像为对象的情报自动采集和加工技术,基于大数据的人工智能决策系统,以语义内容分析为基础的科研成果评价等。深度学习正给世界带来深刻变革,图书情报学科有充分的理由在这场变革中迎来发展的机遇。以下试举几个实例,说明机器学习未来在图书情报领域的可能应用。

### 5.1 智能问答系统在图书馆中的应用

在图书馆由“馆藏为中心”向“读者为中心”转移的过程中,智能问答系统可以发挥更加重要和关键

的作用。智能问答系统的前身是基于关键词搜索的FAQ,所能回答的是简单事实型问题。人工智能时代的问答系统在提问形式、回答问题的深度方面有了长足的进步。随着机器学习技术的深入应用,智能问答系统在词库与知识更新速度、所使用的实体知识的类型等方面有了更大的提高。新的智能问答系统将对提高图书馆数字参考咨询的智能化和深度语义知识化起到更加促进的作用。首先,基于自然语言处理中的浅层句法分析和深层句法分析技术,图书馆智能问答系统可利用序列化的深度学习模型对提问进行精准的分类和多层面的子句拆分;其次,通过充分利用深度学习模型在海量结构化、半结构化和非结构化数据中自动构建特征并进行知识挖掘的特征,结合图书馆的馆藏和业务方面的数据,在一定量的人工干预下,可以构建深层备选答案的数据集,并对该数据集进行相关的分类;最后,未来图书馆智能问答系统的突出特征是支持并行计算,尤其是调用相应的自然语言处理模型面对海量答案的自动匹配过程中,并行计算不仅能确保答案查询的准确性而且能确保答案查询的全面性。深度学习技术支持下的图书馆智能问答系统,将成为与图书馆馆员一样的“百事通”,很大程度上代替图书馆馆员面向读者开展服务,且态度和蔼,百问不厌。

## 5.2 机器学习在文本信息处理中的应用

机器学习特别是深度学习技术,应用于各类文本的深入加工和处理,并籍此得到各个角度、多个层面的知识,将成为图书情报学未来研究的重要内容之一。首先,基于领域化的文本数据,构建领域化的词汇、术语和实体资源,结合隐马尔可夫模型、最大熵模型、条件随机场、支持向量机和深度学习的各种模型,实现对文本内容的分词、词性标记、实体识别、实体关系挖掘,进行完成对文本知识的组织;其次,通过浅层句法关系、深层次句法关系、语义自动分析、篇章结构的计算,从已经进行了知识组织的文本中挖掘出词汇层面、实体层面、句子层面、段落层面和篇章层面的多个层面的知识,作为构建知识库和知识本体的基础;最后,结合可视化和虚拟现实的技术,从历史的角度,对处理后的文本知识进行时间序列层面上的呈现,从情景的角度,结合相应的地理

和历史场景,对文本中的相关事件进行适度的还原,从而让文本知识真正活起来。图书情报机构原本是以单册文本作为收藏单元的,如此处理后则能以细颗粒度的方式向读者或用户呈现文本内容,及其内容所蕴含的知识以及知识与知识间的关联。同时,用户需求也可以利用上述机器学习方法采集获取。机器学习方法将助力图书情报机构进入基于文本内容和用户需求的自动化知识服务时代。

## 5.3 机器学习在学术评价中的应用

基于机器学习的相关技术,可以改善非结构化学术文献全文数据处理技术,从学术文献的全文数据中挖掘出更多的特征知识,并把所挖掘出来的知识融入到学术评价当中。譬如,以前费时费力的引用行为分析便可通过全文挖掘方式获得基础数据。以机器学习为技术基础的评价数据采集与加工方式将日益受到学术评价领域研究者的关注。

未来随着大数据和机器学习方法的快速发展,充分发挥机器学习在自然语言处理、图像自动识别与分析以及深度语义关联与挖掘上的强大功能,可实现对学术论文题录和全文的内容理解、评价分析,从而面向海量学术文献模拟人类专家在理解文献内容的基础上实现由机器充任的同行评议,摆脱目前高度依赖学术文献题录信息和引文数据的单一评价模式,达到学术论文的评价真正基于学术论文内容进行的目标,提升学术评价的科学性、规范性和智能性。

## 5.4 机器学习在信息服务方面的应用

机器学习对于用户的信息检索、信息利用行为可以进行深入的挖掘。在图书情报机构广泛的读者或用户使用行为数据基础上,机器学习大有作为。

首先,机器学习方法可以挖掘并分析用户的信息需求,以“推”方式向用户提供主动的、个性化的信息服务。未来的图书馆在充分发挥机器学习功效的基础上,可以成为用户手边的个人信息助理,随时随地提供精准的主动信息服务。

其次,在确保隐私的前提下,机器学习完全有能力将用户在图书情报机构的信息行为与个人的日常生活行为进行关联,并提炼、分析用户需求,使用户日常生活行为与图书馆的主动信息服务融为一体,将前述个人信息助理升级为个人生活助理。如生病



时推荐医院、医生、药品,出行时选择并预订交通工具,工程项目开展前推荐相关资料甚至形成基础性文档,等等。

### 5.5 机器学习对图书情报学人才培养的影响

在机器学习迅猛发展的大趋势下,一个全面、立体、贯穿智能元素的图书情报学教育体系,应当将培养精通机器学习技术的高端图书情报人才作为重要目标,设立与机器学习理论和技术相关的一系列课程,将此作为图书情报学教育和人才培养方面一个重要的主题,同时也是图书情报学学科建设与发展的重要保障。

随着大数据时代的推进,图书情报机构面对的数据信息、数据资源将越来越复杂,其种类、层次呈多样化和立体化,内容愈加异构化、非结构化,挖掘、利用的难度将越来越大,而机器学习大有用武之地,图书情报机构对机器学习技术需求的程度将越来越高。掌握先进的机器学习技术的专业人才,将成为当前环境下引领学科与行业发展的生力军,对于机器学习人才培养的看法和重视程度,也将影响图书情报学学科未来的发展方向和趋势。

## 6 机器学习视角下的古文信息处理

古籍是图书情报机构的重要馆藏资源,综合性的大中型图书馆一般都设有古籍部,古汉语文献的加工与处理是图书情报机构的一项重要工作内容。计算机自动化加工处理古汉语文本,即古文信息处理,是最能体现机器学习应用价值的研究主题之一,也是图书情报机构古籍工作未来的发展方向。古文信息处理就是以古代汉语文本为对象的信息处理,是文本信息处理中的组成部分,黄水清<sup>[103]</sup>将其定义为“借助信息技术手段对古代汉语文本的音、形、义进行处理和加工,并可在此基础上实现对古代汉语文本的深度挖掘与知识发现。”在提倡弘扬中华文化的新时代,以古汉语文本形式呈现的中国典籍在既有的历史价值之上又被赋予了民族文化复兴的现实重任,图书情报机构应该在民族文化复兴中承担应尽的义务,古籍业务工作迫切需要信息处理、文本挖掘、人文计算等多种有别于传统视角的新思路,而机器学习仍将在其中扮演重要的角色。

### 6.1 机器学习对古籍检索智能化的影响

机器学习在处理大规模文本语料时,具有速度快、精度高的先天优势,因此该方法将为以古籍文本为对象的全文检索、知识检索及智能问答系统带来巨大发展助力。

古籍文本检索系统在检索词选取、词汇标注、同义词控制、异文控制等方面工作量巨大。以代表了古籍文本手工检索工具最高水准的哈佛燕京学社《汉学引得丛刊》(以下简称《引得》)为例,其64种、81册《引得》涵盖了由周遍型字词引得、综合性关键词引得、书名(篇名)引得、图谱引得、人物姓名与字号引得等多种类型引得组成的引得体系,对典籍文本进行了全面、立体、充分地揭示。然而《引得》的编纂耗费了巨大的人力和物力,仅仅编纂周期就跨越了近20多年的时间。而且,由于人工方法在面对大规模文本工程时极易出错,其成果内容中难免存在瑕疵。机器学习方法的引入可以有效地解决这种人工编纂带来的固有困难,同时将引得形式的资源与机器学习中的全文检索、知识挖掘、智能问答技术相结合构造出一种高效的古籍智能检索系统,能够以内容全面、形式丰富的方式承载中华历史文化资源。

以机器学习为技术核心的古籍智能检索系统,将高效、精确、智能地为专业人员及普通读者提供差异化、个性化的检索结果。其检索方式也是多样的,可以是满足专业需求的关键词检索、实体检索、知识检索等,也可以是以历史知识普及为目的的智能问答系统、可视化检索系统、VR/AR体验系统等。

### 6.2 机器学习应用于典籍知识库的构建及推广服务

机器学习方法可以帮助构建超大规模、细颗粒度的典籍知识库,推动中华文化知识的回溯与传承。具体地说,可以通过机器学习中的序列化标注模型,准确地从海量古代典籍中挖掘出人名、地名、时间、机构、多义项词和句子中的异文等语义知识点,进而构建成为典籍知识库。进一步,还可以将知识库中的资源用具体的数字和文字结合图片、视频、VR/AR等多维形式呈现给读者,从历史的角度阐述中华文化的“历史渊源、发展脉络、基本走向”。

随着深度学习研究的推进,个性化推荐、智能语音问答、移动端自动推送等功能值得期待,这将使得

典籍知识库成为一个提供多维知识的智能资源库。在机器学习的推动下,这些丰富的、有吸引力的、智能的典籍知识库,将帮助图书馆在民族精神、传统文化的“移动化、社交化、智能化”的网络传播、弘扬和传续中起到关键性作用。

### 6.3 机器学习在辨别古籍真伪等研究中的潜力

机器学习中的文本分类模型,非常适用于解决古籍真伪辨别、写作年代判定、文体风格判断等传统古籍研究中经典难题。已有多项研究应用支持向量机等算法讨论了红楼梦前80回后40回的作者问题<sup>[104]</sup>,但更为经典的历史文献真伪、年代问题,具有更广阔的应用空间。如《尚书》部分章节及某些新发现竹简的真伪,《左传》成书时代,等等。机器学习在处理类似问题时,较之传统方法有着巨大的优势。基于机器学习方法的文本分类器能够更客观、准确地把握文本中的文体风格特征,以此为基础得出的古籍真伪、时代等方面的推论将为历史文献的研究提

供另辟蹊径,往往会有意想不到的创新性结论。

## 7 结语

人工智能和机器学习的大发展还将持续,这其中带来的重要的技术和方法对于图书情报学界来说是一个十分难得的机遇。两者之间有价值的结合点较多,每一个结合点都值得研究者深入探究和挖掘,而这样的结合对于图书情报学来说是有利的。如果能够充分地、有效地借鉴和利用人工智能的新思路和新方法,图书情报学可以获得全方位的、值得期待的新发展。而在古文信息处理方面,机器学习方法有着巨大的价值和潜力,不论是面向专业领域的古籍检索智能化,还是面向普及大众的中华文化知识传承,机器学习方法都能够发挥重要作用。目前以深度学习为主的机器学习方法在古文信息处理中还较少见,但可以期待古文信息处理在机器学习的引领下,将有一个更加智能化的前景和未来。

### 参考文献:

- [1] Lecun Y, Bengio Y, Hinton G. Deep learning[J]. Nature, 2015, 521(7553): 436-444.
- [2] Russell S J, Norvig P. Artificial Intelligence: A Modern Approach (3rd Edition) [M]. New Jersey: Pearson Education, Inc., 2010.
- [3] McCulloch W S, Pitts W. A logical calculus of the ideas immanent in nervous activity[J]. The bulletin of mathematical biophysics, 1943, 5(4): 115-133.
- [4] Turing A M. Computing machinery and intelligence[J]. Mind, 1950, 59(236): 433-460.
- [5] Searle J R. Minds, brains, and programs[J]. Behavioral and brain sciences, 1980, 3(3): 417-424.
- [6] Newell A, Simon H. The logic theory machine—A complex information processing system[J]. IRE Transactions on information theory, 1956, 2(3): 61-79.
- [7] Newell A, Shaw J C, Simon H A. Empirical explorations of the logic theory machine: a case study in heuristic [C]. Papers presented at the February 26-28, 1957, western joint computer conference: Techniques for reliability. ACM, 1957: 218-230.
- [8] Newell A, Simon H A. GPS, a program that simulates human thought[R]. RAND CORP SANTA MONICA CALIF, 1961.
- [9] Gelernter H. Realization of a geometry theorem proving machine[C]. IFIP Congress, 1959: 273-281.
- [10] Green C. Theorem proving by resolution as a basis for question-answering systems[J]. Machine intelligence, 1969(4): 183-205.
- [11] Robinson J A. A machine-oriented logic based on the resolution principle[J]. Journal of the ACM (JACM), 1965, 12(1): 23-41.
- [12] Samuel A L. Some studies in machine learning using the game of checkers. II—recent progress[J]. IBM Journal of research and development, 1967, 11(6): 601-617.
- [13] McCarthy J. Programs with common sense[M]. RLE and MIT Computation Center, 1960.
- [14] Huffman D. Impossible objects as nonsense sentences[J]. Machine intelligence, 1971(6): 295-324.
- [15] Winston P H. Learning Structural Descriptions From Examples[M]. Massachusetts Institute of Technology, 1970.
- [16] Winograd T. Understanding natural language[J]. Cognitive psychology, 1972, 3(1): 1-191.
- [17] Rosenblatt F. The perceptron: A probabilistic model for information storage and organization in the brain[J]. Psychological



- review, 1958, 65(6):386.
- [18] Rosenblatt F. Principles of neurodynamics[M]. Spartan Nooks, 1962.
- [19] Kolodner J L. Maintaining organization in a dynamic long-term memory[J]. Cognitive science, 1983, 7(4):243-280.
- [20] Wilensky R. Planning and understanding: A computational approach to human reasoning[M]. Addison-Wesley Pub. Co, 1983.
- [21] Buchanan B, Sutherland G, Feigenbaum E A. Heuristic DENDRAL: A program for generating explanatory hypotheses in organic chemistry[M]. Stanford University, 1968.
- [22] Wilensky R. Understanding Goal-Based Stories[M]. DBLP, 1978.
- [23] McDermott J R. 1: A rule-based configurer of computer systems[J]. Artificial intelligence, 1982, 19(1):39-88.
- [24] Rumelhart D E, Hinton G E, McClelland J L. A general framework for parallel distributed processing[J]. Parallel distributed processing: Explorations in the microstructure of cognition, 1986(1):45-76.
- [25] Smolensky P. On the proper treatment of connectionism[J]. Behavioral and brain sciences, 1988, 11(1):1-23.
- [26] 李航. 统计学习方法[M]. 北京: 清华大学出版社, 2012.
- [27] 周志华. 机器学习[M]. 北京: 清华大学出版社, 2016.
- [28] Ng A Y, Jordan M I. On discriminative vs. generative classifiers: A comparison of logistic regression and naive bayes[C]. Advances in neural information processing systems, 2002:841-848.
- [29] McCallum, Andrew, Kamal Nigam. A comparison of event models for naive bayes text classification[C]. Proc. AAAI-98 workshop on learning for text categorization, 1998:41-48.
- [30] Quinlan J R. Induction of decision trees[J]. Machine learning, 1986, 1(1):81-106.
- [31] Rokach L, Maimon O. Data mining with decision trees: theory and applications[EB/OL]. [2017-09-10]. [https://eric.univ-lyon2.fr/~ricco/tanagra/fichiers/fr\\_Tanagra\\_DM\\_with\\_Decision\\_Trees.pdf](https://eric.univ-lyon2.fr/~ricco/tanagra/fichiers/fr_Tanagra_DM_with_Decision_Trees.pdf).
- [32] Breiman L, Friedman J, Stone C J, et al. Classification and regression trees[M]. CRC press, 1984.
- [33] Berger A L, Pietra V J D, Pietra S A D. A maximum entropy approach to natural language processing[J]. Computational linguistics, 1996, 22(1):39-71.
- [34] Jaynes E T. Information Theory and Statistical Mechanics[J]. Physical Review, 1957, 106(4):620-630.
- [35] Jaynes E T. Information theory and statistical mechanics. II[J]. Physical review, 1957, 108(2):171.
- [36] Darroch J N, Ratcliff D. Generalized iterative scaling for log-linear models[J]. The annals of mathematical statistics, 1972:1470-1480.
- [37] Berger A. The improved iterative scaling algorithm: A gentle introduction[EB/OL]. [2017-09-10]. <http://luthuli.cs.uiuc.edu/~daf/courses/optimization/papers/berger-iis.pdf>.
- [38] Chen S F, Rosenfeld R. A Gaussian prior for smoothing maximum entropy models[M]. School of Computer Science, Carnegie Mellon University, 1999.
- [39] Ratnaparkhi A. A maximum entropy model for part-of-speech tagging[C]. Proceedings of the conference on empirical methods in natural language processing, 1996(1):133-142.
- [40] Borthwick A, Grishman R. A maximum entropy approach to named entity recognition[D]. New York University, Graduate School of Arts and Science, 1999.
- [41] McCallum A, Freitag D, Pereira F C N. Maximum Entropy Markov Models for Information Extraction and Segmentation[C]. Icml, 2000:591-598.
- [42] Baum L E, Petrie T. Statistical inference for probabilistic functions of finite state Markov chains[J]. The annals of mathematical statistics, 1966, 37(6):1554-1563.
- [43] Baum L E, Petrie T, Soules G, et al. A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains[J]. The annals of mathematical statistics, 1970, 41(1):164-171.
- [44] Baum L E. An Inequality and Associated Maximization Thechnique in Statistical Estimation for Probabilistic Functions of Markov Process[J]. Inequalities, 1972(3):1-8.
- [45] Cutting D, Kupiec J, Pedersen J, et al. A practical part-of-speech tagger[C]. Proceedings of the third conference on Applied natural language processing. Association for Computational Linguistics, 1992:133-140.

- [46] Zhou G D, Su J. Named entity recognition using an HMM-based chunk tagger[C]. proceedings of the 40th Annual Meeting on Association for Computational Linguistics. Association for Computational Linguistics, 2002: 473-480.
- [47] Seymore K, McCallum A, Rosenfeld R. Learning hidden Markov model structure for information extraction[C]. AAAI-99 workshop on machine learning for information extraction, 1999: 37-42.
- [48] Rabiner L, Juang B. An introduction to hidden Markov models[J]. IEEE ASSP Magazine, 1986, 3(1): 4-16.
- [49] Forney G D. The viterbi algorithm[J]. Proceedings of the IEEE, 1973, 61(3): 268-278.
- [50] Lafferty J, McCallum A, Pereira F C N. Conditional random fields: Probabilistic models for segmenting and labeling sequence data[EB/OL]. [2017-09-10]. [https://repository.upenn.edu/cgi/viewcontent.cgi?article=1162&context=cis\\_papers](https://repository.upenn.edu/cgi/viewcontent.cgi?article=1162&context=cis_papers).
- [51] 洪铭材, 张阔, 唐杰, 等. 基于条件随机场(CRFs)的中文词性标注方法[J]. 计算机科学, 2006, 33(10): 148-151.
- [52] McCallum A, Li W. Early results for named entity recognition with conditional random fields, feature induction and web-enhanced lexicons[C]. Proceedings of the seventh conference on Natural language learning at HLT-NAACL 2003-Volume 4. Association for Computational Linguistics, 2003: 188-191.
- [53] 周俊生, 戴新宇, 尹存燕, 等. 基于层叠条件随机场模型的中文机构名自动识别[J]. 电子学报, 2006, 34(5): 804-809.
- [54] Sutton C, McCallum A. An introduction to conditional random fields for relational learning[M]. Introduction to statistical relational learning. MIT Press, 2006.
- [55] Sha F, Pereira F. Shallow parsing with conditional random fields[C]. Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology-Volume 1. Association for Computational Linguistics, 2003: 134-141.
- [56] Cortes C, Vapnik V. Support-vector networks[J]. Machine learning, 1995, 20(3): 273-297.
- [57] Smola A J, Schölkopf B. A tutorial on support vector regression[J]. Statistics and computing, 2004, 14(3): 199-222.
- [58] Boser B E, Guyon I M, Vapnik V N. A training algorithm for optimal margin classifiers[C]. Proceedings of the fifth annual workshop on Computational learning theory. ACM, 1992: 144-152.
- [59] Schölkopf B, Smola A J. Learning with kernels: support vector machines, regularization, optimization, and beyond[M]. MIT press, 2002.
- [60] Zeng Z Q, Yu H B, Xu H R, et al. Fast training Support Vector Machines using parallel sequential minimal optimization[C]. Intelligent System and Knowledge Engineering, 2008. ISKE 2008. 3rd International Conference on. IEEE, 2008: 997-1001.
- [61] Joachims T. Text categorization with support vector machines: Learning with many relevant features[J]. Machine learning: ECML-98, 1998: 137-142.
- [62] Joachims T. Transductive inference for text classification using support vector machines[EB/OL]. [2017-09-10]. <http://www1.cs.columbia.edu/~dplewis/candidacy/joachims99transductive.pdf>.
- [63] Tong S, Koller D. Support vector machine active learning with applications to text classification[J]. Journal of machine learning research, 2001, 2(Nov): 45-66.
- [64] Joachims T. Learning to classify text using support vector machines: Methods, theory and algorithms[M]. Kluwer Academic Publishers, 2002.
- [65] Chang C C, Lin C J. LIBSVM: a library for support vector machines [J]. ACM transactions on intelligent systems and technology (TIST), 2011, 2(3): 27.
- [66] Jain A K, Dubes R C. Algorithms for clustering data[M]. Prentice-Hall, Inc., 1988.
- [67] Jain A K, Murty M N, Flynn P J. Data clustering: a review[J]. ACM computing surveys (CSUR), 1999, 31(3): 264-323.
- [68] Jain A K. Data clustering: 50 years beyond K-means[J]. Pattern recognition letters, 2010, 31(8): 651-666.
- [69] Ester M, Kriegel H P, Sander J, et al. A density-based algorithm for discovering clusters in large spatial databases with noise[C]. Kdd. 1996: 226-231.
- [70] Ankerst M, Breunig M M, Kriegel H P, et al. OPTICS: ordering points to identify the clustering structure[C]. ACM Sigmod record. ACM, 1999: 49-60.
- [71] Kaufman L, Rousseeuw P J. Finding groups in data: an introduction to cluster analysis[M]. John Wiley & Sons, 2009.
- [72] Bengio Y, Courville A, Vincent P. Representation learning: A review and new perspectives[J]. IEEE transactions on pattern

- analysis and machine intelligence, 2013, 35(8): 1798-1828.
- [73] Schmidhuber J. Deep learning in neural networks: An overview[J]. Neural networks, 2015(61): 85-117.
- [74] Deng L, Yu D. Deep learning: methods and applications[J]. Foundations and Trends? in Signal Processing, 2014, 7(3-4): 197-387.
- [75] Ivakhnenko A G, Lapa V G. Cybernetic predicting devices[R]. Purdue Univ Lafayette Ind School of Electrical Engineering, 1966.
- [76] Ivakhnenko A G. Polynomial theory of complex systems[J]. IEEE transactions on Systems, Man, and Cybernetics, 1971, 1(4): 364-378.
- [77] Fukushima K, Miyake S. Neocognitron: A self-organizing neural network model for a mechanism of visual pattern recognition[M]. Competition and cooperation in neural nets. Springer, Berlin, Heidelberg, 1982: 267-285.
- [78] LeCun Y, Boser B, Denker J S, et al. Backpropagation applied to handwritten zip code recognition[J]. Neural computation, 1989, 1(4): 541-551.
- [79] Hinton G E, Dayan P, Frey B J, et al. The "wake-sleep" algorithm for unsupervised neural networks[J]. Science, 1995, 268(5214): 1158.
- [80] Hochreiter S, Bengio Y, Frasconi P, et al. Gradient flow in recurrent nets: the difficulty of learning long-term dependencies [EB/OL]. [2017-09-10]. <http://www.bioinf.jku.at/publications/older/ch7.pdf>.
- [81] Waibel A, Hanazawa T, Hinton G, et al. Phoneme recognition using time-delay neural networks[M]. Readings in speech recognition, 1990: 393-404.
- [82] Hinton G E, Osindero S, Teh Y W. A fast learning algorithm for deep belief nets[J]. Neural computation, 2006, 18(7): 1527-1554.
- [83] Hinton G, Deng L, Yu D, et al. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups[J]. IEEE Signal Processing Magazine, 2012, 29(6): 82-97.
- [84] Deng L, Hinton G, Kingsbury B. New types of deep neural network learning for speech recognition and related applications: An overview[C]. Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on. IEEE, 2013: 8599-8603.
- [85] Hochreiter S, Schmidhuber J. Long short-term memory[J]. Neural computation, 1997, 9(8): 1735-1780.
- [86] Graves A, Mohamed A, Hinton G. Speech recognition with deep recurrent neural networks[C]. Acoustics, speech and signal processing (icassp), 2013 IEEE international conference on. IEEE, 2013: 6645-6649.
- [87] Graves A, Jaitly N. Towards end-to-end speech recognition with recurrent neural networks [C]. Proceedings of the 31st International Conference on Machine Learning (ICML-14), 2014: 1764-1772.
- [88] Schmidhuber J. Our Impact on the World's 5 Most Valuable Public Companies (2017) [EB/OL]. [2017-12-27]. <http://people.idsia.ch/~juergen/impact-on-most-valuable-companies.html>.
- [89] LeCun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.
- [90] Ciregan D, Meier U, Schmidhuber J. Multi-column deep neural networks for image classification[C]. Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on. IEEE, 2012: 3642-3649.
- [91] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[C]. Advances in neural information processing systems, 2012: 1097-1105.
- [92] Silver D, Huang A, Maddison C J, et al. Mastering the game of Go with deep neural networks and tree search[J]. Nature, 2016, 529(7587): 484-489.
- [93] Bengio Y, Ducharme R, Vincent P, et al. A neural probabilistic language model[J]. Journal of machine learning research, 2003, 3(Feb): 1137-1155.
- [94] Mikolov T, Chen K, Corrado G, et al. Efficient estimation of word representations in vector space[J]. arXiv preprint arXiv, 2013: 1301.3781.
- [95] Mikolov T, Sutskever I, Chen K, et al. Distributed representations of words and phrases and their compositionality[C]. Ad-



- vances in neural information processing systems, 2013:3111–3119.
- [96] Goldberg Y, Levy O. word2vec Explained: deriving Mikolov et al.'s negative-sampling word-embedding method [J]. arXiv preprint arXiv, 2014:1402.3722.
- [97] Socher R, Bauer J, Manning C D. Parsing with compositional vector grammars [C]. Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2013:455–465.
- [98] Socher R, Perelygin A, Wu J, et al. Recursive deep models for semantic compositionality over a sentiment treebank [C]. Proceedings of the 2013 conference on empirical methods in natural language processing, 2013:1631–1642.
- [99] Zeng D, Liu K, Lai S, et al. Relation Classification via Convolutional Deep Neural Network [C]. COLING, 2014:2335–2344.
- [100] Sutskever I, Vinyals O, Le Q V. Sequence to sequence learning with neural networks [C]. Advances in neural information processing systems, 2014:3104–3112.
- [101] Schuster M, Johnson M, Thorat N. Zero-shot translation with Google's multilingual neural machine translation system [J]. Google Research Blog, 2016.
- [102] Collobert R, Weston J. A unified architecture for natural language processing: Deep neural networks with multitask learning [C]. Proceedings of the 25th international conference on Machine learning. ACM, 2008:160–167.
- [103] 黄水清, 王东波. 古文信息处理研究的现状及趋势 [J]. 图书情报工作, 2017, 61(12):43–49.
- [104] 施建军. 基于支持向量机技术的《红楼梦》作者研究 [J]. 红楼梦学刊, 2011(5):35–52.

**作者简介:** 刘浏, 男, 南京大学信息管理学院博士生; 王东波, 男, 南京农业大学信息科学技术学院副教授, 硕士生导师; 黄水清, 男, 南京农业大学信息科学技术学院教授, 博士生导师。

(上接第 73 页)

- [13] Salton G, Allan J, Singhal A. Automatic Text Decomposition and Structuring [J]. Information Processing & Management, 1996, 32(2):127–138.
- [14] Allen J. Natural Language Understanding (2nd ed.) [M]. Redwood City, CA: The Benjamin/Cummings Publishing Co., 1995.
- [15] Chomsky N. Aspects of the Theory of Syntax [M]. Cambridge, MA: MIT Press, 1965.
- [16] Chomsky N. The Logical Structure of Linguistics Theory [M]. New York: Plenum Press, 1975.
- [17] Schank R C, K M Colby. Computer Models of Thought and Language [M]. San Francisco, CA: W.H. Freeman and company, 1973.
- [18] Schank R C. The Concept Analysis of Natural Language. Natural Language Processing (Edited by R. Rustin) [M]. New York: Algorithm Press, 1973.
- [19] 鲁川, 缙瑞隆, 董丽萍. 现代汉语基本句模 [J]. 世界汉语教学, 2000(4):11–24.
- [20] 鲁川. 汉语语法的意合网络 [M]. 北京: 商务印书馆, 2001.
- [21] 黄曾阳. HNC 理论概要 [J]. 中文信息学报, 1997, 11(4):11–20.
- [22] 黄曾阳. HNC (概念层次网络) 理论: 计算机理解语言研究的新思路 [M]. 北京: 清华大学出版社, 1998.
- [23] What is WordNet? [EB/OL]. [2017-08-10]. <https://wordnet.princeton.edu/>.
- [24] HowNet Knowledge Database [EB/OL]. [2017-08-11]. <http://www.keenage.com/>.
- [25] Berners-Lee T, Hendler J, Lassila, O. The Semantic Web [J]. Scientific American, 2001, 284(5):34–43.
- [26] Bizer C, Heath T, Berners-Lee T. Linked Data—The Story So Far [J]. International Journal on Semantic Web and Information Systems, 2009, 5(3):1–22.

**作者简介:** 叶鹰 (1962–), 男, 南京大学信息管理学院教授, 博士生导师, 研究方向: 定量信息分析、智能信息处理。