

# 大数据与人机对话:语必在言的集合里选取

邹晓辉<sup>1,3</sup> 王肖群<sup>2</sup> 邹顺鹏<sup>3\*</sup>

<sup>1</sup>(中美赛尔研究中心学科组 北京 100871)

<sup>2</sup>(北京大学教师教学发展中心 北京 100871)

<sup>3</sup>(中美赛尔研究中心高教组 北京 100083)

**摘要** 人们学习说话的教育过程就像是在语言的超级棋盘暨形式化字符组成的单词矩阵中做各种选择。提出一种必然涵盖所有言语(话语或术语)的新方法:构建语言的集合暨字符棋盘或词表;通过人机交互和协作,生成大量的言语(话语或术语)大数据,涵盖代表知识本体的各种话语或术语;通过机器学习和人机交互过程,比较、查询或重复使用这些话语或术语。结果表明,话语或术语的选取都可通过双语或多语转换以多种方式自动查询。该方法不仅可用于创建大数据与人机对话的环境平台,还可用于智能化文本分析和知识模块精加工,从而搭起大数据与知识大生产的桥梁。

**关键词** 大数据 软件 人机对话 机器学习

**中图分类号** TP3 **文献标识码** A **DOI**:10.3969/j.issn.1000-386x.2019.09.033

## BIG DATA AND HUMAN-MACHINE DIALOGUE: SPEECH MUST BE SELECTED IN THE SINGLE SET OF LANGUAGES

Zou Xiaohui<sup>1,3</sup> Wang Xiaoqun<sup>2</sup> Zou Shunpeng<sup>3\*</sup>

<sup>1</sup>(Subject Group, Sino-American Searle Research Center, Beijing 100871, China)

<sup>2</sup>(Peking University Teacher Teaching Development Center, Beijing 100871, China)

<sup>3</sup>(Higher Education Group, Sino-American Searle Research Center, Beijing 100083, China)

**Abstract** The educational process of learning to speak is like making choices in a word matrix of formal characters or a language's super chessboard. This paper introduced a new approach that must cover all speech (discourse or terminology). We constructed a single set of languages, that is, a character or vocabulary matrix. Through human-computer interaction and collaboration, a large number of big data of speech (discourse or terminology) were generated, covering all kinds of discourse or terminology representing knowledge ontology. We compared, queried or reused these words or terms through machine learning and human-computer interaction. The results show the choice of discourse or terminology can be automatically queried in a variety of ways through bilingual or multilingual conversion. This method and can be used not only to create a platform for big data and human-machine dialogue environment, but also for intelligent text analysis and knowledge module finishing, so as to bridge the gap between big data and knowledge big production.

**Keywords** Big data Software Man-machine dialogue Machine learning

## 0 引言

本文的目的是介绍一种必须涵盖所有言语(话语

或术语)的新方法。因此,人们学习说话的教育过程就像在语言的超级棋盘或形式化字符组成的单词矩阵中做出各种选择。

从背景知识来看,最接近大数据技术应用前沿和

教育教学前沿的前沿科技可从最典型的应用实例<sup>[1-3]</sup>开始,这是普通教师和学生的捷径。我们不能从学术前沿的最新探索开始<sup>[4-9]</sup>(否则大多数教师和学生将撤退),围棋软件击败人类的职业国际围棋选手(甚至横扫欧洲和世界冠军)仅依靠基于大数据的新一代人工智能,涵盖了基于统计的机器学习<sup>[10]</sup>和基于神经网络的深度学习<sup>[11]</sup>。对于普通人甚至大多数专家来说这些都是不可想象的。如何将这些大数据(前沿技术)与教育(日常教学)联系起来(制作出原创成果),对许多大学教师来说是一个巨大的挑战。因此,本文从最简单的算术二进制数和语言(仅以0和1两个单音节符号作为基本成员)开始,回到图灵机<sup>[12]</sup>、图灵测试<sup>[13]</sup>和塞尔中文屋<sup>[14-15]</sup>,及最新的间接形式化方法和间接计算模型<sup>[16]</sup>。新近发现并且宣布的形式化理解模型的一系列基本原理<sup>[17]</sup>,探讨适合不同学科知识背景的两大类形式化方略<sup>[18]</sup>,是另一种方便的大数据处理方式(相当于流行的基于统计的机器学习),基于神经网络的深度学习,可以在这个干净的云平台上发挥更好的作用(例如,做大数据采集、存储和分类计算或统计,都更容易)。最重要的是,教师和学生能够尽早开始参与“教育、管理、学习和应用”结合的社会化系统工程<sup>[19]</sup>。比较围棋和数字化字符棋即双字棋的异同,这是本文选择实现其目标的方式。

## 1 大前提与大趋势

首先构建语言集合即字符棋盘或形式词汇表;然后,通过人机交互协作,生成大量的言语(话语或术语)大数据,涵盖代表知识本体的话语和术语;最后,通过机器学习和人机交互过程,比较、查询或重用这些话语或术语。它们共同遵循的基本原理和形式化理解模型,这是其大前提。上述三个步骤涉及的则是其一系列具体的小前提。

### 1.1 形式化理解模型

图1为三种形式化理解模型及各自的特点。模型A是在两个相同的图灵机之间建立全等关系,其统一的输出输入基础是二进制数及其变体均遵循序位逻辑而可物理实现的测序定位法则。模型B从数字系统扩展到符号系统(广义文本的基本单位),它们仍在两个相同的图灵机(数字计算机)之间运行,所不同的仅仅是广义文本的符号系统虽然也只是人机之间的转换,但是,其所遵循的规则增加了用户协议,至少涵盖

两套形式信息和内容信息,在同意和不同意的理解方式之间由于主体的介入而存在差异。模型C只是对汉字系统进行间接正式化理解,可视为前者的收敛模型。必须特别说明的是三类模型都只是形式化理解模型(仅仅涉及棋法、棋盘和棋子的测序和定位、计算和统计,不涉及对“是什么”和“为什么”的解释,而只涉及怎么做暨形式化理解模型作为人机操作工具如何发挥其作用的特定角色含义,一句话,物理与数理一致)。

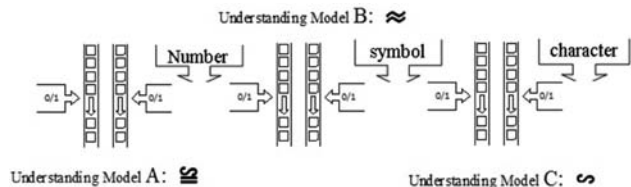


图1 ABC三类形式化理解模型

第一组选择是由一个用户(通常是专家)与系统交互过程中设计的(不排除基于统计的机器学习和基于神经网络的深度学习也介入其中)。随后可由机器识别进行相应的重复选择(其关键是:如何明确告诉机器如何为用户做出特定的选择)。机器可以进行基于统计的学习、分析和形式化理解和基于神经网络的深度学习。也就是说,可开发基于统计的机器学习软件以及基于神经网络的深度学习软件植入其中的相应部分,基本前提是拥有大数据和相应的计算资源及硬盘存储空间。该形式化理解模型有别于以前的现有信息处理范式,可以看作是大数据介入之后的新一代人工智能的形式化解释。为此,有必要回顾信息处理、人工智能和大数据以及它们之间的相互关系及其发展变化趋势<sup>[20]</sup>。

### 1.2 信息处理、人工智能和大数据的趋势分析

以下是以论文发表的主题信息处理、人工智能和大数据三方面做的三类可检验的调查结果及分析。

(1) 第一类可验证的结果及其分析。图2为信息处理(IP)主题研究的发展变化趋势。从标题包含信息处理这一主题词语而发表论文的数量及其发展变化趋势,以及每年发表的论文数量,可以看出1960年是一个起点,而2016年则是一个具有很大波动起伏的转折点。它究竟是怎么回事呢?如果孤立来看,是很难解释得通的。尽管如此,人们还是可从图2所示IP趋势及其对应的年份和发表的代表论著的研读来做判断。如果没有人工智能(AI)和大数据(BD)的相继发展,那么,人们也就只能根据图2所示IP趋势做与其相关的一系列研究了。但事实上,不仅有人工智能(AI),而且还有大数据(BD)的发展和相继伴随。

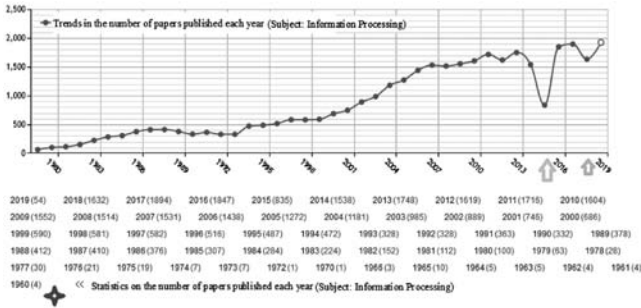


图2 信息处理(IP)发展变化的趋势

(2) 第二类可验证的结果及其分析。图3为人工智能(AI)主题研究的发展变化趋势。从标题包含人工智能这一主题词语而发表论文的数量及其发展变化趋势,以及每年发表的论文数量,可以看出1978年是一个起点,2016年则是一个十分明显而巨大的转折点。图3与图2的表现有巨大的反差。基于此,我们发现以往人工智能(AI)主题研究的发展变化曲线长期走低的可视化表现;还发现以往信息处理(IP)主题研究的发展变化曲线的一路逐渐高涨的可视化表现(虽然也有些许波动)。但是,令人费解的是:为什么IP曲线和AI曲线都在2016年不约而同地出现了大转折点?

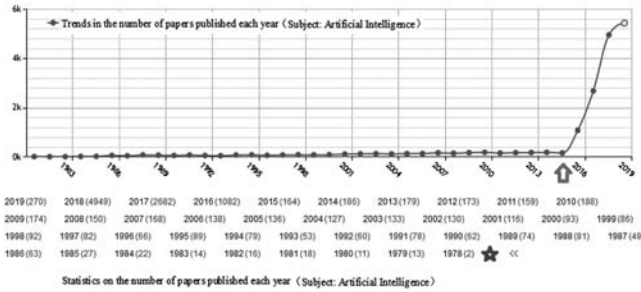


图3 人工智能(AI)发展变化的趋势

(3) 第三类可验证的结果及其分析。图4为大数据(BD)主题研究的发展变化趋势。从标题包含大数据这一主题词语而发表论文的数量及其发展变化趋势,以及每年发表的论文数量,可以看出1985年是一个起点,2012至2016年出现了几个转折点。此时我们把BD曲线与前面的IP曲线和AI曲线放在一起观察,可以发现2016年是一个大转折点,它在IP曲线、AI曲线和BD曲线的走势十分接近。

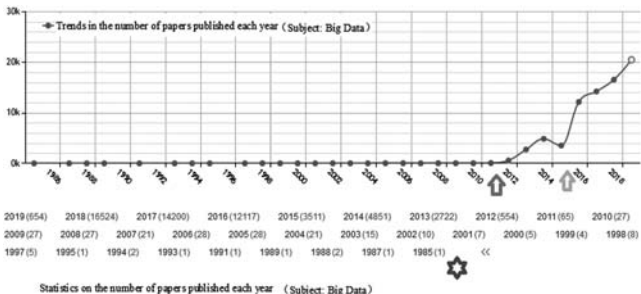


图4 大数据(BD)发展变化的趋势

### 1.3 大前提与大趋势的关系

只要认真观察IP曲线、AI曲线和BD曲线的走势变化就可发现它们的异同。回顾信息处理(IP)和人工智能(AI)及大数据(BD)这三大科技领域的发展历程,就不难发现相应的这三条发展变化曲线蕴含的深刻意义。

形式化理解模型与上述三方面的发展变化趋势有怎样的相互关系,是以往的科学范式及其配套的形式化技术忽略的。新的科学范式及其配套的形式化技术虽然已经公开,但是,其推广和普及还需要时间。

## 2 三个步骤的聚集

从图5可以看出,美国信息交换标准码(ASCII)是基于小字符集,而国际统一编码(Unicode)覆盖了采用统一字符编码标准(GB)的大字符集,对双字节字符进行编码,仅限于狭义的文本范围。本研究中的中国标准(Z)指广义文本(字符、公式、图形、表格、声音、图像、立体、活体)。因此,信息处理的新旧融合标准(Z-ASCII)是中美融合标准,它涵盖所有类型的狭义和广义的文本及其(直接和间接)形式化表达的元素。构建全球语言定位系统(GLPS)的广义文本语境是其最终目标,含跨&多学科(领域或行业)的全球知识定位系统(GKPS)。这就为形式化理解模型及其实际应用奠定了坚实的基础(特别是在人机交互智能系统的背景下)。

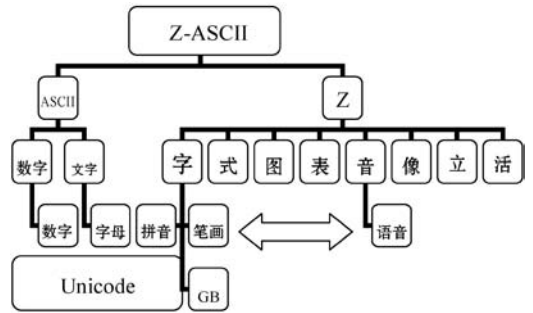


图5 信息处理新旧标准(Z-ASCII)<sup>[21]</sup>

用以下三个步骤的聚集可衔接GLPS与大小前提:

(1) 基于孪生图灵机(具有自动学习能力和形式化理解能力的新型智能机)的发现和发明,构建一对数字棋盘。例如:字符的智能化文本分析示例,英语的词形式组合的智能化文本分析示例。由此可显示人性和人机的双重形式化方略的协同机制。

这不仅揭示了语言学和语言哲学通过间接形式化途径如何进入语言科学殿堂的短程线,而且还发现了信息学和信息哲学通过间接形式化途径如何进入信息科学殿堂的短程线,甚至也发现了教育学和教育哲学通过间接形式化途径如何进入教育科学殿堂的短程线。

(2) 基于形式化理解模型而生成的大数据与过去

旧范式获得方式生成的大数据,形成了鲜明对比。基于数字棋盘产生的大数据和基于网络爬虫所获取的大数据,虽然都是大数据,但是存在可控性的区别。例如,虽然都是基于美国标准信息交换代码的技术来调用字母符号,但是却不能直接调用笔画符号,更不能同时直接调用广义文本的所有符号(Z 却可以)。

可以说同样是大数据,其效率却大不相同。例如:谷歌的围棋软件和深度学习的结合带来了快速发展;数字棋盘和文字棋盘结合构成的孪生图灵机即可发挥协同智能系统的作用。由于逻辑、数学和物理三方面存在秩序和位置的一致性,如{真,假}、{0,1}和{开,关},因此,它们只是纯形式的计算、统计和变换,棋盘、棋谱和棋法不受周期和概率的控制,棋理也一样。这对于人脑和电脑(好比硬件)用的软件、知识和语言而言,是一个更好的灵感。通过实际示例介绍棋理,可以发现人脑擅长的棋谱应用与电脑擅长的机器学习都是独一无二的。

(3) 机器的批处理和人机交互的启发式,在基于宽度的搜索和基于深度的搜索以及基于遗传算法的搜索中,都具有其自身特征。人机结合的意义和选择,仅在术语处理和学习方面有所不同(涉及:分析、比较、查询和重用)。受限于在间接形式化的数字和文本的双列表,人机双脑的协同操作,可实现人机协作甚至协同即有针对性的双脑大协作。

大数据与教育之间的关系反映了计算机第一步的优势;第二步反映人机交替的优势;第三步则展示了人机之间的(高度)分工与合作的双重优势。由此,才发现了其巨大的潜力。

### 3 棋盘示例

图 6 为可间接计算的双字棋盘示例。



图 6 可间接计算的双字棋盘示例(英文版的)

从图 6 可以看见,文本是 2017 年 2 月 19 日星期日在 AAAS 科学年会上塞尔研究中心主任研究员邹晓辉在展览厅(Hynes 会议中心)展示的论文方法的纯文本部分。文本框下方的数字和字符即词形式。

用户选出的术语是在双棋盘上调用的字符组合。示例如表 1 所示。

表 1 在此三对知识模块由六组术语构成

ID	代表人物	知识模块加工后提取的示例
1	Aristotle 亚里士多德	Formal logic based on language 基于语言的形式逻辑
2	Frege 弗雷格	Mathematical logic based on arithmetic 基于算术的数理逻辑
3	Turing 图灵	Strong artificial intelligence view 强人工智能的观点
4	Searle 塞尔	Weak artificial intelligence view 弱人工智能的观点
5	Saussure 索绪尔	General linguistic view 普通语言学的观点
6	Chomsky 乔姆斯基	Formal linguistic view 形式语言学的观点

将表 1 与图 6 结合可以看出两者的内在联系。图 6 英语文本的中文翻译如下:

基于数字和文本的双矩阵方法涉及以下步骤:

首先,亚里士多德基于语言的形式逻辑和弗雷格基于算术的数学逻辑,为基于序和位的广义双语逻辑奠定了基础。同时,图灵基于数字计算的强人工智能观点和塞尔基于自然语言的弱人工智能观点,为基础数字和文本双矩阵之间的联动函数关系奠定了基础。然后,在常识和跨学科、跨领域和跨行业的专业知识基础上建立另类双语,索绪尔普通语言观和乔姆斯基形式语言观结合,在英汉狭义双语的基础上,再结合广义双语,建立三类双语的广义翻译系统的基础。其特点是基于知识本体的广义翻译,与联动函数关系及序位逻辑结构一道,共同构成了人机双脑协作的基础。

图 7 为中文双字棋盘可间接计算的《弃》示例,可以看出,仅在汉字棋盘上选出“终、不、迷、胡”四个字即可起到画龙点睛的理解效果(在此内容与形式结合得很好)。



图 7 中文双字棋盘可间接计算的《弃》示例

图8为示例《存》,就这首诗而言,一旦找出“已、真、待、巧”四个字,再说明要重构的“基因”不是生物的,而是文化的,即形式化广义文本的基因,几乎瞬间就可表达其特定的含义和用意。



图8 中文双字棋盘可间接计算的《存》示例

图7和图8的中文双字棋盘及其蕴涵的序位逻辑和联动函数结合广义翻译可将知识棋谱和原创棋魂代表的知识模块精加工示例与英文双字棋盘的术语即知识模块的示例揭示出语言棋盘蕴涵的深刻棋理。棋理在此具通用性,如二进制的围棋与多进制的双字棋。这样的广义文本的棋理可作为统一参照系统理解广义语言。将汉字版的双字棋盘作为统一参照系统,不仅可以从原文中得到更好的支持,而且,是智能化文本分析和知识模块精加工的工具。话语或术语的选取都可通过双语或多语转换以多种方式一系列的双字棋盘后台自动查询。

## 4 结 语

无论是双音节及多音节的语(汉语的字组)还是混音节的词语(不限于汉语的语),都是单音节的言(汉语的字)构成的超级棋盘或矩阵里被用户选用的(某种组合),它们都是形式化的<sup>[22-25]</sup>。

其意义在于该方法及结果不仅可用于创建大数据与人机对话环境,而且,还可用于智能化文本分析和知识模块精加工,从而,搭起大数据与知识大生产的桥梁。例如,基于课堂特定学科领域的知识本体建构过程中产生的大数据,不仅可通过传统技术处理,还可通过双字棋盘这样的创新技术处理。基于术语知识本体的智能系统可直接用于知识大生产。

形式化理解模型及其蕴含的三大原理(序位结构遵循的双语逻辑、联动函数遵循的P进制原理和三类双语示例蕴含的基于知识本体的广义翻译原理)和双字棋盘蕴涵的三类孪生图灵机原理,为基于洁净大数据协同智能系统的研究与应用奠定了基础。具体表现是:自然语言处理和形式化理解,专家知识获取和形式

化表达,软件编程建模和形式化模拟,都可得到序位结构(逻辑)和联动函数(数学)在理解(知)上的支持,还可得到人工(标准化技术)和人类(个性化艺术)在操作(行)上的支持。

可视化的模式识别可先建模再模拟,软件工程和模式识别的许多问题及其解决方案也蕴涵在其中。

这种有益效果表明:过去语言、知识和软件三种工程技术,不仅可揭示其连贯的机理,而且还可获得意料之外情理之中的创新发展(语言、知识和软件的三大系统工程技术的发明创造,及信息、智能和数据三大科学原理的发现创新)。简而言之,它是对智能系统研究(科学)及其应用(技术)进一步发展的基础。

新方法有效的原因在于它充分发挥了人机合作的优点。其具体表现是:首先,不仅顶级人类专家的创造力已获得了优越的环境,而且教师和学生也获得了相同的环境。其次,不仅可以使间接形式化方法和间接计算模型,还可建构并使用英语和其他语言的双字棋盘。最后,不仅间接形式化方法和字符间接计算模型很受欢迎,而且图形图像和各种类型的媒体也会流行。

这三方面的结合意味着在解决大数据和教育中的有趣问题时,不仅可讨论最新解决方案、科学的结果和方法,还可讨论人机之间分工和合作的可能性。这是最吸引人的。人机交互的优势互补和优化互动才是信息处理、人工智能和大数据发展之后最具竞争力的综合技术创新,即智能系统研究及其应用。由于教育最重要的作用是启迪智慧,而基于洁净大数据的智能系统的研究和应用则在教育、管理、学习和应用结合的社会化系统工程方面,与语言、知识、软件和硬件结合的形式化系统工程遥相呼应,且具有独特性。

## 参 考 文 献

- [1] 田渊栋. 阿法狗围棋系统的简要分析[J]. 自动化学报, 2016, 42(5):671-675.
- [2] Dong X., Wu J S, Zhou L. Demystifying AlphaGo Zero as AlphaGo GAN[EB]. arXiv:1711.09091,2017.
- [3] Granter S R, Beck A H, Papke D J J. AlphaGo, deep learning, and the future of the human microscopist[J]. Archives of Pathology & Laboratory Medicine, 2017, 141(5): 619.
- [4] Chouard T. The Go Files: AI computer clinches victory against Go champion[OL]. Nature, 2016-03-12. <https://www.nature.com/news/the-go-files-ai-computer-clinches-victory-against-go-champion-1.19553>.
- [5] Chouard T. The Go Files: AI computer wraps up 4-1 victory against human champion[OL]. Nature, 2016-03-15. <https://www.nature.com/news/the-go-files-ai-computer->

wraps-up-4-1-victory-against-human-champion-1.19575.

- [6] Silver D, Schrittwieser J, Simonyan K, et al. Mastering the game of Go without human knowledge[J]. *Nature*, 2017, 550(7676):354-359.
- [7] Gibney E. Self-taught AI is best yet at strategy game Go [OL]. *Nature*, 2017-10-18. <https://www.nature.com/news/self-taught-ai-is-best-yet-at-strategy-game-go-1.22858>.
- [8] Zou S, Zou X. Understanding: How to resolve ambiguity [C]//International Conference on Intelligence Science. Springer, Cham, 2017:333-343.
- [9] Hua W, Zou S, Zou X, et al. Using two formal strategies to eliminate ambiguity in poetry text[C]//International Conference on Intelligence Science. Springer, Cham, 2018.
- [10] Mitra J, Shen K K, Ghose S, et al. Statistical machine learning to identify traumatic brain injury(TBI) from structural disconnections of white matter networks[J]. *NeuroImage*, 2016, 129:247-259.
- [11] Holden D, Saito J, Komura T. A deep learning framework for character motion synthesis and editing[J]. *ACM Transactions on Graphics*, 2016, 35(4):1-11.
- [12] Gulcehre C, Chandar S, Cho K, et al. Dynamic neural Turing machine with continuous and discrete addressing schemes [J]. *Neural Computation*, 2018, 30(4):1-28.
- [13] Cerf V G. Turing test 2[J]. *Communications of the ACM*, 2018, 61(5):5-5.
- [14] Rapaport W J. Searle's experiments with thought[J]. *Philosophy of Science*, 1986, 53(2):271-279.
- [15] Mper R I. The logic of searle's chinese room argument[J]. *Minds & Machines*, 2006, 16(2):163-183.
- [16] 邹晓辉, 邹顺鹏. 间接计算模型和间接形式化方法[J]. *软件*, 2011, 32(5):1-5.
- [17] 邹晓辉, 邹顺鹏. 双语信息处理方法及原理[J]. *计算机应用与软件*, 2015, 32(11):69-76, 102.
- [18] 邹晓辉, 邹顺鹏. 两大类形式化方略[J]. *计算机应用与软件*, 2013, 30(9):194-199.
- [19] 邹晓辉, 邹顺鹏. 当代中国大学新使命:基于汉语思维与双语处理的文化遗产和创新[J]. *南京理工大学学报(社会科学版)*, 2012, 25(5):106-113.
- [20] Zou X H, Zou S P, Ke L J. Fundamental law of information: Proved by both numbers and characters in conjugate matrices[J]. *Proceedings*, 2017, 1(3):60.
- [21] Zou S P, Zou X H. Ecological characteristics of information and its scientific research[J]. *Proceedings*, 2017, 1(3):59.
- [22] 邹晓辉, 邹顺鹏. 软件工程学科何以独特——形式化方法的双重路径[J]. *软件*, 2011, 32(7):1-14.
- [23] 邹晓辉, 邹顺鹏. 言和语的关系——在语言和言语的区分基础上进一步[C]//第四届中西语言哲学国际研讨会, 2012.
- [24] 邹晓辉. 智能教育协同创新:学术引领与产业驱动的数字教育进程[J]. *数字教育*, 2017(6):1-5.

## (上接第155页)

方法,和传统 GBDT 相比较能达到较优的预测精度, Xgboost 最低均方根误差可达到 1.850 3, LightGBM 最低均方根误差可达到 1.608。同时,经过三个预测模型中特征重要度的排序,识别出影响租金最关键的特征是面积因素和地段因素。本文的不足之处在于 Xgboost 和 LightGBM 虽然能够得到较好的预测精度,但是由于 Xgboost 和 LightGBM 都是基于启发式算法,寻找的解为局部最优并非全局最优。另外,后续研究也可以引入 Filter、Wrapper 等特征选择方法,以进一步提升预测精度。

## 参 考 文 献

- [1] 刘洪玉. 什么因素阻碍了租房市场健康发展[J]. *人民论坛*, 2017(24):88-90.
- [2] 施继元, 李涛, 李婧骅. 国外住房租赁管理经验及对我国的启示[J]. *软科学*, 2013, 27(1):31-36.
- [3] 黄燕芬, 王淳熙, 张超, 等. 建立我国住房租赁市场发展的长效机制——以“租购同权”促“租售并举”[J]. *价格理论与实践*, 2017(10):19-23.
- [4] Adair A S, Berry J N, McGreal W S. Hedonic modelling, housing submarkets and residential valuation[J]. *Journal of property Research*, 1996, 13(1):67-83.
- [5] 吴雪婷. 限购政策对住房市场价格与租金的影响研究[D]. 上海:华东师范大学, 2016.
- [6] 郑文娟. 中国城市住房价格与住房租金的影响因素及相互关系研究[D]. 杭州:浙江大学, 2011.
- [7] 汪业辉. 住房租赁市场租金的影响因素分析[D]. 武汉:华中科技大学, 2014.
- [8] 袁野. 城市住房租赁合同的租金管制[J]. *河北法学*, 2018, 36(10):124-137.
- [9] 贺灵. 基于特征价格模型的城市住宅租赁价格影响因素及空间差异研究[D]. 重庆:重庆大学, 2017.
- [10] 刘洪杰. 基于特征价格模型的社区商业租金微观影响因素实证研究[D]. 杭州:浙江大学, 2010.
- [11] 黄丽君. 基于 Hedonic 模型的长租公寓租金影响因素的实证研究[D]. 南昌:江西师范大学, 2018.
- [12] 李武斌. 西安市主城区住房租金空间分异及其影响因素[J]. *西部人居环境学刊*, 2015(4):77-81.
- [13] Friedman J H. Greedy function approximation: A gradient boosting machine [J]. *Annals of Statistics*, 2001, 29(5):1189-1232.
- [14] Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System[C]//ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2016:785-794.
- [15] Ke G L, Meng Q, Finley T, et al. LightGBM: A Highly Efficient Gradient Boosting Decision Tree [C]//Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, 4-9 December 2017, Long Beach, CA, USA.