

# 问答系统研究综述

吴灵慧

**摘要** 目前,问答系统已经广泛应用到社会各个方面,其作为整个计算机领域重点的研究领域,引起了人们的高度重视。文章对问答系统的基础概念、发展历史、关键技术及未来的发展趋势进行简要阐述,旨在为问答系统今后发展打下良好基础。

**关键词** 问答系统;信息检索;答案抽取;问题分析

**中图分类号** TP3

**文献标识码** A

**文章编号** 1674-6708(2019)230-0147-02

DOI:10.16607/j.cnki.1674-6708.2019.05.069

随着互联网的迅猛发展和计算机的广泛普及,人们一直寻找迅速并准确地获取信息的方法。在过去20年里,各类搜索引擎产品不断出现和发展,迎合了用户的信息检索需求。然而当数据量不断增加,海量数据资源汇集到一起时,搜索引擎的缺点逐渐开始暴露。传统的搜索引擎只能以检索关键词作为输入,这就要求用户必须用数量有限的词来准确表达查询需求,对用户的信息检索能力要求较高,这对于绝大多数普通用户来说并不友好。

此外,传统搜索引擎返回的结果通常也不尽如人意,仍需用户自己在搜索引擎系统返回的网页列表中查找具体的信息,这就容易使用户的体验大打折扣。基于上述存在的问题,问答系统应运而生,与传统搜索引擎不同,问答系统可以针对任意以自然语言形式提出的问题返回给用户确切的答案。由此可见,问答系统能更好地表达和满足用户的要求。本文对问答系统的发展历史、关键技术及发展趋势进行简要介绍,旨在帮助读者对问答系统有全面的认识和了解。

## 1 问答系统发展历史

问答系统的发展历史已有70余年,早在1950年,计算机科学之父艾伦图灵便提出了使用自然语言进行人机交互的相关技术,进而使得问答系统的诞生成为了现实。

在20世纪60年代,借助人工智能的相关技术,第一批问答系统出现在人们的视眼中,典型的代表系统有BASEBALL、LUNAR和ELIZA<sup>[1]</sup>。Baseball允许用户对篮球比赛的相关内容进行提问,并返回相应的结果;Lunar能够对岩石样本分析实验的相关内容进行问答;ELIZA系统主要用于精神疾病患者的恢复过程,能够实现简单的交流过程。这3个系统都在各自的领域中取得了巨大的成功,但这些问

答系统只能接受特定形式的自然语言问句形式,无法实现开放领域的问答。

在20世纪70年代和80年代期间,随着计算机语言学理论的深入研究及发展,问答系统也逐渐向着更为复杂的领域发展。这个时期诞生了大量的项目,其中就包含了著名的Berkeley Unix Consultant(UC),然而UC系统提供的对话实例并不能完美应用到现实对话场合中。

20世纪90年代以后,问答系统进入了开放领域和基于自由文本的新时期,在此期间,国外开发出一些相对成熟的问答系统。1993年,麻省理工学院(MIT)的人工智能实验室开发出了世界上第一个名为START的问答系统。START可以回答一些有关地理、文化、科技、历史等方面的简单问题。例如:对于问题“What's the capital of China?”START将会回答“The capital of China is Beijing.”。密歇根大学于2002年开发的AnswerBus<sup>[2]</sup>是个多语种的自动问答系统,它可以回答以英语、法语、和葡萄牙语等语言提出的问题。

在文本信息检索会议(TREC, Text Retrieval Conference)将QA track引入到会议主题后,新型问答系统的研究发展得到了进一步促进,问答系统研究自此进入了开放域研究的时代<sup>[3]</sup>。现如今,越来越多的高校和公司都参与到了有关问答系统技术的研究中,包括人们熟知的苹果、三星、Google、阿里巴巴、剑桥大学等,使得问答系统的发展得到了进一步推进。同时,凭借着信息技术的发展,问答系统的研究者们总结提炼了大量的常问问题与回答数据,为问答系统技术的进一步发展打下了坚实的基础。其中典型的代表为2005年出现的社区问答(community based question answering, CQA<sup>[4]</sup>)数据,如Yahoo! Answer。

作者简介:吴灵慧,启东市汇龙中学。

## 2 问答系统关键技术

问答系统关键技术主要包括问题分析、信息检索、答案抽取 3 个部分。本章节将对这 3 个部分进行简要介绍。

### 2.1 问题分析

在问答系统工作时，首要的任务便是对用户提出的问题进行分析。问题分析包括类型分析、主旨识别、问句指代及语法分析等工作。问答系统的首要工作是判定问题的类别，通常问题的类别也是答案的类别，所以这个过程对问答系统来说必不可少。问答系统规则分类器根据问题的特征对用户提出的问题进行分类，之后对问题中的关键词进行提取；关键词的提取是问题分析的核心步骤，问答系统往往会对关键词进行扩展来更好的理解用户的查询意图；最后问答系统将对问题的表达形式进行重新组织，以便于计算机进行答案查找。

### 2.2 信息检索

信息检索是问答系统与普通搜索引擎之间的桥梁。信息检索的步骤包括文档检索、段落检索和富信息索引。信息检索过程的核心内容是将问题分析过程中提取到的关键词在后台文档库中进行匹配，以找到最相关的文档。信息检索模块的输入一般是几个关键词的组合，同时根据语种的不同，问答系统还需要对英文关键词进行词根操作，对中文关键词进行分词处理等。信息检索步骤的关键在于确定文档相关性的权重和对文档进行合理排序，这样才能保证向用户呈现最相关的信息。

### 2.3 答案抽取

答案抽取是问答系统的最后一步。答案抽取的主要目的是从文档信息中获得用户想要的答案，以满足用户的需求。答案抽取一般分为两个步骤，第一步为生成候选答案集合，第二步为提取答案。生成候选答案集合即使用在文档中匹配得到的短语构成候选答案集，提取答案即在所得候选答案集的基础上进行选择，最后将最准确的答案返回给用户。在这个过程中，答案抽取一般会通过基于表层特征的答案提取方法，选择出现频率最高的信息作为最终答案，或者采用关系抽取答案、模式匹配抽取答案和统计模型抽取答案等方法。

## 3 问答系统发展趋势

问答系统作为人工智能技术的有效评价手段，目前已有 60 年的研究历史。多年来，基于 Wikipedia 等高质量且会动态更新的开发资源建立起来的知识资源库日趋成熟，包含的知识也越来越多，再加上基于统计机器学习的自然语言处理技术和知识推理技术有了极大的发展，这两方面的进步分别为智能问答系统的发展奠定了资源基础和技术

基础。问答技术的发展从限定领域向开放领域，从单轮问答向多轮对话，从单个数据向多个数据，从浅层语义分析向深度逻辑推理不断推进<sup>[5]</sup>。

但是问答系统仍然存在一些急需解决的问题，主要有知识库的融合、目前存在信息孤岛的问题；自然理解能力偏弱，还需要进一步提高等。为解决这些问题，问答系统将会有以下发展趋势。

1) 网络中充满着大量为特定领域建立的知识资源库，这些资源库在各自领域发挥了很好的作用。未来的趋势是开放领域的多领域覆盖智能问答系统，所以要把全部异构的知识源统一起来，满足用户的统一查询需求。

2) 我们目前建立的知识资源库中的知识大多是事实性知识，缺乏常识性知识。常识性知识在人的推理中具有及其重要的作用，人们正在研究如何将常识性知识融入到智能问答系统中来。

3) 随着深度神经网络技术在各个领域的成功应用，人们希望深度神经网络算法能够在问答系统领域代替传统的方法。随着深度学习技术的逐步发展和计算机硬件性能的进一步提升，能够学习数据深层特征的深度神经网络技术将成为未来智能问答系统的绝对主流技术。

随着相关技术的飞速发展，自动问答系统在未来将会有越来越多的应用实例，其效果也会得到突破。伴随着 IBM Watson、Apple Siri 等实际应用的落地，我们更有信心看到这一技术在不远的未来得到更广泛的应用。

## 4 结论

本文主要对问答系统的发展历史、关键技术及发展趋势 3 个方面进行了阐述。就目前而言，问答系统暂时还未达到我们预期的目标，问答技术也只是在起步阶段，并且问答系统准确率也有待提高，相信在不久的将来，问答系统将会取得重大的突破，为人们的工作和生活提供更多的便利。

### 参考文献

- [1] 李舟军, 李水华. 基于Web的问答系统综述[J]. 计算机科学, 2017(6): 1-7.
- [2] 郑实福, 刘挺, 秦兵, 等. 自动问答综述[J]. 中文信息学报, 2002(6): 46-52.
- [3] 毛先领, 李晓明. 问答系统研究综述[J]. 计算机科学与探索, 2012, 6(3): 193-207.
- [4] 宋万里, 卜磊. 智能答疑系统的设计与实现[J]. 电脑知识与技术, 2017, 13(36): 163-165.
- [5] 冯升. 聊天机器人问答系统现状与发展[J]. 机器人技术与应用, 2016(4): 34-36.