

开放域问答技术研究进展

张志昌^{1,2}, 张宇¹, 刘挺¹, 李生¹

(1. 哈尔滨工业大学计算机学院信息检索研究室, 黑龙江哈尔滨 150001;
2. 西北师范大学数学与信息科学学院, 甘肃兰州 730070)

摘要: 问答技术是信息检索和自然语言理解中的研究热点. 本文对开放域问答技术在近年来的研究进展进行了总结. 针对不同类型的问题, 比较了问答系统中问题分析、文档和句段检索、答案抽取各个部分不同实现方法的特点, 讨论了这些技术中存在的不足. 最后, 对问答技术未来的研究趋势进行了分析与展望.

关键词: 开放域问答; 问题分析; 相关文档检索; 句段检索; 答案抽取

中图分类号: TP391 **文献标识码:** A **文章编号:** 0372-2112 (2009) 05-1058-12

Advances in Open-Domain Question Answering

ZHANG Zhi-chang^{1,2}, ZHANG Yu¹, LIU Ting¹, LI Sheng¹

(1. IR Laboratory of Computer Science and Technology School, Harbin Institute of Technology, Harbin, Heilongjiang 150001, China;
2. Mathematic and Information Science School, Northwest Normal University, Lanzhou, Gansu 730070, China)

Abstract: Question answering is one of the research hotspots in information retrieval and natural language understanding. This paper summarizes the up-to-date research advances in open-domain question answering, compares the different approaches in question analysis, document and passage retrieval, answer extraction according to different question types. The shortcomings of state-of-art techniques are also discussed, and further research directions for question answering are analyzed and prospected at last.

Key words: open domain question answering; question analysis; relevant document retrieval; passage retrieval; answer extraction

1 引言

随着信息技术发展, 人们可用的数据资源呈爆炸式增长, 用户需要有强大的检索工具从海量数据中找到所需信息. 以搜索引擎为代表的各种检索系统, 在给人们带来极大方便的同时, 表现出不少缺点: 系统返回给用户的是排好序的一批文档链接, 用户需要在其中浏览定位真正的有用信息; 用户需要自己构造查询词, 而构造的好坏将严重影响检索的性能.

开放域问答 (Question Answering, QA) 研究的目的是提供更自然的信息访问交互界面, 特点是: 用户可以用自然语言提问的形式提出查询需求, 系统从各种数据资源中自动找出准确的答案; 用户的提问不限于某个特定的领域或应用^[1].

从 20 世纪 60 年代开始, 问答技术的研究涉及了以自然语言为前端的数据库系统、限定域对话系统、阅读

理解系统、基于 FAQ 的问答系统等^[2]. 随着检索技术的发展, 基于海量文档集的开放域问答技术研究逐渐成为热点. 从 1999 年开始, 在 TREC 问答评测推动下, 相关研究得到了全球众多学者的重视.

2 主要的研究机构与解决的问题类型

对 2001 年以后发表在 ACL、SIGIR、COLING、NAACL、EACL、HLT、IJCNLP、EMNLP 等会议上关于问答的文章按照研究机构进行统计, 数量较多的一些机构如表 1.

鉴于这些会议的代表性, 可以看出问答技术主要的研究机构. 国内研究中, 复旦大学、中科院自动化所、计算所、北京大学、清华大学、哈工大、山西大学、沈阳航空工业学院等单位均有涉及. 复旦大学、哈工大还多次参与 TREC QA 评测, 取得了不错的成绩.

受 TREC 等问答评测的影响, 现有研究对问答系统所回答的问题主要包含以下类型:

表 1 国外主要的研究机构和代表人物

机构	代表人物	论文数
LCC 公司、德克萨斯州立大学达拉斯分校的 HLTRI	Dan Moldovan, Sanda Harabagiu	9
IBM 沃森研究中心	Jennifer Chu-Carroll, John Prager	9
新加坡国立大学计算机系	Tat-Seng Chua	8
马里兰大学信息学院	Jimmy Lin	7
南加州大学的 ISI	Eduard Hovy	7
CMU 的 LTI	Eric Nyberg, Jamie Callan	4
滑铁卢大学计算机学院	Charles L. A. Clarke	4
微软研究院	Eric Brill	4
德国萨尔大学计算语言学系	Dietrich Klakow	3
纽约州立大学 Albany 分校 ILS	Tomek Strzalkowski	3
MIT 的 AI 实验室	Boris Katz	2
密执安大学计算机系	Joyce Y. Chai	2

(1) 事实型 (factoid) 和列举型 (list) 问题. 询问实体或实体的属性, 答案是一个或多个命名实体或短语;

(2) 定义型 (definition) 问题. 询问对某个目标的定义, 答案是对目标在不同侧面的定义说明.

(3) 复杂问题. “复杂”本身存在歧义, 但基本特点是: 答案包含多个句子或句子片断的详细描述. 现有研究所涉及的复杂问题包括: (a) why 型问题, 询问某个话题的原因或解释等; (b) how-to 型问题, 询问某个动作或操作的过程描述; (c) 基于情景 (scenario) 的问题, 首先给出一段关于某个情景的描述, 然后针对情景中的多个实体, 提问它们间的关系及支持关系的证据, 本文称之为“关系型”问题.

另外, 问答技术研究也逐渐拓展到交互式问答 (Interactive QA) 方面, 即用户和系统以问答形式进行交互, 用户所提问题处于对话上下文中.

3 问答技术研究进展

3.1 问答系统一般的体系结构

MITRE 的 Hirschman 等^[1]对问答系统结构进行了总结. 典型的问答系统一般包含有: 问题分析、文档和句段 (passage) 检索, 答案抽取和生成. 在该体系结构下, 系统接受用户的问题并给出反应的一般过程如图 1:

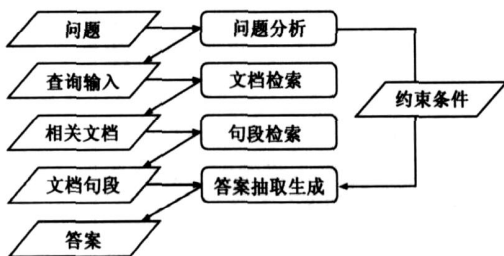


图 1 开放域问答系统一般的体系结构

该过程在不同的系统实现中会有变化, 但从整体

看, 体系结构在近几年仍没有太大变化^[3].

处理过程中的每一步都可以分成若干个模块. LCC 的 Moldovan 等^[4]对该流程中各模块对系统整体性能的影响进行了分析. 不同系统的实现方法不同, 但这些分析仍然可以反映影响系统整体性能的主要因素.

3.2 事实型与列举型问题的处理

3.2.1 问题分析

(1) 问题分类. 面对新问题, 系统首先要明确问题对预期答案的语义约束条件, 如“哪个国家人口最多?”是提问国家名. 现有问题分类的方法主要有: (a) 经验规则的方法^[5,6]. 这些规则是通过问题表层或者深层分析, 从中抽取出的公共分类模式. (b) 统计机器学习的方法, 表 2 对这些方法及采用的特征等进行对比.

(2) 预期答案与问句词之间的关系获取. 不同的答案提取算法, 需要不同的答案上下文约束条件. 在预期答案与问句关键词的关系方面, 现有的表示方法有:

(a) 简单共现关系. 如 Cymfony 公司的 Srihari 等^[13], 答案约束条件包括答案与问句关键词之间的共现关系和命名实体类型; (b) 句法依存关系. 如 LCC 公司 Harabagiu 等^[14]把问句表示成预期答案与问句词以及问句词相互之间的句法依存关系; (c) 浅层语义关系. 新加坡国立大学的 Sun 等^[15]对问句进行浅层语义分析, 将问题表示成核心词与语义角色的集合.

表 2 基于统计的问题分类方法

作者	算法/工具	测试语料和分类体系	特征	性能
Li 和 Roth ^[7]	SNoW	TREC 10 问题; 6 大类, 50 小类	词, 词性, 词义, NE, 类相关词	大类 92.5%; 小类 85%
Zhang 和 Lee ^[8]	SVM	同上	词	大类 86%; 小类 80%
Moschitti 等 ^[9]	SVM/句法树核、词线性核	同上	词, 句法结构, 语义角色	大类 91.8%
Suzuki 等 ^[10]	SVM/HDAG 核	四层类别; NTCIR-QAC 和 CRL-QA 问题	句法结构, 词, 词义, NE, Chunk	第四层: 74.9%
Wei Li ^[11]	语言模型	693 个 TREC 评测问题; 7 类	词	81.54%
Lin 和 Lu ^[12]	Bigram 语言模型	NTCIR-5 CLQA 中文问题; 6 类	词, 焦点词, 从 web 挖掘的类特征词	无焦点词问题: 54%

3.2.2 相关文档检索

为了减少答案抽取所需处理的数据量, 需要从海量文档集中检索出可能包含答案的相关文档. 需要解决的问题有: 检索模型的选择、需要检出的文档数量、查询输入的构造与反馈技术的应用.

(1) 检索模型的选择. 在 ad hoc 信息检索评测中, 统计语言模型经常要比其它模型表现出更好的性能. Moldovan 与 Harabagiu 的研究认为布尔模型可能更适合于问答系统^[4,14], 主要是布尔模型更容易构造反馈结构, 可以根据上次检索结果的多少, 调整关键词数量来影响下次的检索结果数量. 但参加 TRECQA 评测的大部分系统都采用统计语言模型或向量空间模型或直接用开源工具 Lemur 或 Lucene 来实现检索.

(2) 需要返回的文档数量. Hovy 等^[16]的实验发现, 在检索结果的前 1000 个文档句中, 有 8% 的问题仍然找不到答案. 因此, 检索应该返回尽量多的文档, 以确保包含答案的文档在检索结果中.

Thompson 等^[17]验证了文档检索的质量和问答系统整体性能之间的关系: 包含答案的文档越多, 则该答案越容易被抽取出来. 但另一方面, 返回太多的检索结果, 会导致后续的答案抽取性能下降^[18]. 原因是返回的结果越多, 其中包含的不相关文档也就越多, 这反而影响了答案的抽取.

(3) 查询输入的构造. 根据问句构造检索的查询输入有很多方法, 最简单的是将问句中非停用词作为查询关键词, 缺点是: 有些不包含关键词的文档却包含答案. 因此, 需要进行关键词扩展等处理.

(a) 词干提取 (stemming). 阿姆斯特丹大学的 Monz^[19]对传统文档检索中的词干提取在问答中的效果进行研究, 发现除了个别问题类型 (如定义型) 之外, 应用词干提取能稳定提高文档检索的准确率. 而 MIT 人工智能实验室 Bilotti 等^[20]的研究表明, 在问答中应用词干提取, 很多情况下会降低检索的性能. 可以看出, 采用不同的检索模型时, 同样的查询构造会有不同的效果.

(b) 伪相关反馈. Monz^[19]的研究表明, 实现查询扩展的伪相关反馈技术, 在问答系统中会降低检索性能.

(c) 同义词和语义相关词扩展. 这些方法有:

基于知识库的扩展. 如 Harabagiu 等^[14]和 Hovy 等^[16]利用问句词在 WordNet 中的上位词和同义词来扩展查询. IBM 的 Ittycheriah 等^[21]从在线百科全书中对问句词的解释信息中抽取关键词作为扩展. Moldovan 等^[22]从 WordNet 中获取与查询词语义相关的词汇链 (lexical chain) 作为扩展.

基于检索结果的扩展. 有研究从 Web 上获取词的语义相关词作为扩展. 如 Sun 等^[15]把问句作为查询输入, 在 Google 检索结果的前 50 个检索摘要中, 提取和原始问句关键词有依存关系的词作为扩展词.

基于统计机器学习方法的扩展. 从“问句-查询输入”对的训练集中学习查询扩展模型. 用到的模型有:

噪音通道模型. 如 Monz^[19]等在一个问题 Q 所有可

能的查询构造集中, 寻找通过噪音通道产生 Q 的最可能的查询构造与扩展.

基于训练数据的简单统计与验证模型. 哥伦比亚大学的 Agichtein 等^[23]在 FAQ 语料中统计不同疑问词及疑问短语的各种答案构造形式, 如对“what is a”开始的问题, 答案表述形式有“refers to”, “to describe”等. 在这些答案形式中提取各种候选查询扩展, 然后利用“问题-答案对”语料, 对这些候选形式进行验证, 对准确率超过阈值的扩展形式存入知识库.

统计机器翻译模型. 如 Google 公司的 Riezler 等, 在 100 万个 FAQ“问答对”语料中学习基于短语的“问题到答案”的翻译模型, 从而获得不同短语在给定问题上上下文时的各种“翻译”形式作为扩展^[24].

3.2.3 相关文档句段检索

在检索到的相关文档中, 把可能包含答案的文档句段提取出来, 以进一步减少答案抽取所需处理的内容.

(1) 句段的切分. 句段检索时, 首要的问题是如何将一个文档切成多个句段. 有些将文档句段定义为一个句子, 如 Hovy 等^[25]和 MITRE 公司的 Light^[26]; 有些定义为 n 个句子, 如 Alicante 大学的 Vicedo 等^[27]; 还有些定义为一个连续的词序列, 边界可以在一个句子的中间, 如 Clarke 等^[28].

(2) 句段检索方法. 最简单的是计算句段和问题之间匹配词的个数, 作为句段的排序权值^[26]. 还有些是计算问题和句段之间的余弦相似度, 如 Alicante 大学的算法^[27]. 滑铁卢大学 MultiText 系统^[28]的算法不仅考虑匹配词的个数及在句段中的 idf 权值, 也考虑匹配词在句段中的相邻距离, 即匹配词在句段中的密度.

MIT 人工智能实验室的 Tellex 等^[29]对各种句段检索算法进行了对比, 性能好的方法其共同点是: 句段排序都考虑匹配词在句段中的分布密度, 即匹配词之间间隔的紧密程度.

(3) 句段检索的文档范围. 参与 TREC 评测的系统中, 大部分都是先从海量数据集中检索出可能包含答案的文档, 然后在少量的文档中进行句段检索. 但有一些系统, 是直接在整个原始数据集中进行句段检索.

舍费尔德大学的 Roberts 等^[30]研究表明, 综合来看, 在语料库上首先进行文档检索, 然后将排序靠前的 n 篇文档切成句段, 再对这些句段进行句段检索重排序, 则排序靠前的句段准确率较高.

3.2.4 答案抽取与排序

在检索出可能包含答案的句段之后, 要将满足约束条件的答案从这些句段中抽取出来返回给用户.

对于简单事实型问题, 一般是用某个算法, 从一批

排好序的文档句段中抽取出多个候选答案,然后对这些候选进行排序,排第一的被作为答案返回。

对于列举型问题,许多研究都是使用和事实型问题相同的答案抽取和排序方法,只是在返回结果时,将排序分值超过一定阈值的多个候选答案作为最终结果^[3]。也有研究利用各种在线知识库^[31]和网页中的表格来抽取答案^[32]。

总结已有的事实型问题答案抽取方法,可归类为:

(1) 基于词袋模型的简单匹配。在句段中抽取和预期答案类型一致的命名实体作为候选答案^[13]。对候选答案排序时,综合其所在句段的排序和在所有句段中出现的次数作为分值。Light 等^[26]分析了这种方法的性能,在 TREC 评测语料上,这种方法至少可以保证 10 ~ 24 % 的问题能被正确回答。该方法的问题之一是一个句段中,可能包含多个同样类型的命名实体。

(2) 基于表层模式匹配。基本思想:答案和问句关键词之间总有特定的表层关系。因此,算法不使用太多深层的语言处理,而是从句段中抽取出满足模式的候选答案。模式构造方式有:

(a) 手工构造方式。特点是实现较简单,准确率高,但费时费力。如 InsightSoft-M 公司的 Soubbotin 等^[33]使用该方法,在 TREC 2001 评测中是性能最好的系统。

(b) 自动学习方式。训练时,从文档句段中提取出包含问句关键词和答案的子串,将其中的关键词和答案替换成变量,则得到了一个候选规则。然后,在测试数据中,用这些候选规则抽取答案,将答案准确率高于一定阈值的选为最终的规则。Hovy 等^[34]用搜索引擎来获取训练和测试数据,利用自举方法学习表层模式。

(3) 基于句法结构比较。基本思想是:候选句子中的候选答案与问句词之间的句法结构,和在问题中疑问短语与问句句之间的句法结构如果能良好匹配,则该候选答案很可能是正确答案。比较的方法有:

(a) 句法结构模式的匹配。答案选择过程是:首先找到匹配问题的问题模式,然后根据对应该问题模式的答案句法结构模式,从候选句中抽取满足该模式的答案。

模式的构造方式分两类:手工构造,如萨尔大学的 Kaisser 等^[35];从训练预料中自动抽取,如萨尔大学的 Shen 等^[36]。

(b) 句法结构的严格匹配。纽约州立大学 Albany 分校的 Wu 等^[37]对问题所有依存关系三元组(核心词,依存关系,修饰词)和候选句依存关系三元组进行匹配,匹配最多的候选句被选择为答案句,再从该句中将满足依存关系约束的命名实体抽取出来作为答案。

(c) 句法结构的统计匹配。前面所述的两种匹配方法,共同之处是基于这样的假设:正确答案在问题和答

案句中都有同样的句法关系。但实际上,这种假设很多时候并不成立,这些方法尽管可以获得较高的准确率,但是召回率却受限。

一种改进方法:设法从训练语料中学习“问题-答案句”的句法结构是正确匹配的概率。候选句即使和问题的结构不同,也可能有最大概率。用到的模型有:

统计机器翻译模型。新加坡国立大学 Cui 等^[38]用统计翻译模型来计算问题和答案句两个句法结构的相似度。句法结构被分解成包含多个句法关系的路径。两个句法关系之间的翻译概率用它们在训练问答对中的互信息来表示。该方法存在的问题是:没有区分不同路径的重要性;没有考虑问句词在候选答案句中的形态变化和同义词。

最大熵模型。针对 Cui 等方法的不足,萨尔大学的 Shen 等^[39]在三个方面进行了改进:计算两个关系路径的相关性时,不是叠加两组关系所有对齐方式的翻译概率,而是寻找两组关系间的最可能对齐;用最大熵模型来对不同路径赋予不同权值;在候选答案句中寻找和问句关键词对应的短语时,采用近似的问题短语映射方法,而不是词的严格匹配。

噪音通道模型。如南加州 ISI 的 Echiabi 等^[40]提出的方法。训练时,对答案句语法树进行裁剪,得到一个词、语法标识、答案的序列。在问题-序列对集合上训练得到模型。测试时,对所有的答案候选句语法树作同样的裁剪和标记,利用模型对裁剪和标记后的序列进行排序。

(4) 基于海量数据的冗余特性。基本思想是:在海量的文档集中,总会存在和问题描述方式相近的答案句子;而且一般来说,正确答案与问句关键词在检索结果中出现的次数要相对更多。因此,可以使用浅层语言技术辅助简单统计方法就可确定答案。Web 是海量数据的最好实例,从近几年参与 TREC 问答评测的系统看,几乎都会利用 Web,同时会明显提高系统性能。

Clarke 等^[41]选择 Web 作为补充辅助数据源,将候选答案在 Web 检索结果中出现的次数作为答案排序的一项因素。意大利 ITC 的 Magnini 等^[42]用候选答案在搜索引擎检索结果中出现的次数作为排序分值。

(5) 基于答案的逻辑推理验证。这些研究主要在美国 LCC 公司的 Moldovan 和 Harabagiu 带领的研究组,他们的系统也在历次 TREC 问答评测中取得了最好性能。

基本思想是:从相关文档句段中抽取出多个满足问题预期答案类型的命名实体作为候选答案,根据它们在文档句段和 Web 上出现的次数等因素对它们进行初始排序;然后,利用逻辑推理的方法验证这些候选答案的正确性,验证正确则为最终答案^[43]。

Moldovan 等^[44]介绍了逻辑推理工具 COGEX。

COGEX 把问句和句段表示成一阶逻辑表达式,应用世界公理知识和语言公理知识,在问题和候选句之间对候选答案进行合一推理,若成功,则该候选答案正确。

(6) 基于多特征的统计机器学习方法. 利用特定类型的单一特征总会存在不足,一种方法是将多种特征结合起来,发挥它们在不同条件下各自的优势。

这类系统的经典之一是 IBM 的统计问答系统^[21],用最大熵模型的概率来选择答案,特征包括问题和候选句间匹配词的分值、候选句中是否包含期望的命名实体类型、候选句是否匹配某词汇模式等多种特征。

利用这种方法时,正确答案与非正确答案在训练集上的不平衡,会影响模型的性能. 针对此问题,有研究对训练数据进行预处理,如哈工大的李鹏等^[45]。

3.3 定义型问题的处理

3.3.1 问题分析

已知问题为定义型的基础上,有研究对定义目标进行细分,如 IBM 的 PIQUANT 系统^[46]将目标分为“人物”、“机构”、“事物”等;复旦大学的 FDUQA 系统^[47]应用规则,将定义目标细分成 7 个类别. 对不同类型的定义目标,可以在不同的数据源中去抽取定义。

3.3.2 相关文档检索

(1) 查询扩展. 定义型问题包含很少的限定信息,因而使查询输入构造更困难. 如问题“*What is the vagus nerve?*”,可用的关键词只有目标词“*vagus nerve*”. 但如果只用目标词作查询,检索会返回太多文档。

因此,在定义型问题的文档检索中需要查询扩展. 如舍费尔德大学的 Gaizauskas 等^[18]从 WordNet 中查找对定义目标词的注释,从大英百科全书中查找含有定义目标词的句子,从 Web 上检索包含满足模式且包含定义目标词的句子,然后从这些注释或句子中选择关键词作为对定义目标词的扩展。

(2) 检索模型与返回文档的数量. 一般都是选择一个传统的文档检索模型^[48,49]. 对于检索要返回的文档数量,有些研究是利用全部检索结果^[50],也有些是用检索的前 n 篇文档^[48]。

3.3.3 答案抽取和生成

从检索到的文档中抽取、生成定义一般包含这几步:(1) 候选定义句抽取:从文档中抽取对目标进行描述定义的句子;(2) 排序:对句子的重要程度进行区分;(3) 冗余信息去除:去除和已有句子在内容上重复的句子,以及对目标定义无意义的部分;(4) 定义生成:将没有冗余信息的候选句子组织、综合成定义描述。

(1) 候选定义句子抽取. 抽取时,许多系统都利用规则,而规则的设计是利用各种表征定义句子的特征(称为定义型语言构造, *definitional linguistic constructs*). 这些特征包括:词-词性表层模式;句法树定义构造模式;同位语 (*Appositives*);系动词 (*Copulas*);定语从句 (*Relative Clauses*);谓词和动词短语。

规则的构造既有手工方式(如 Harabagiu 等^[51]),也有自动归纳学习方式(如 Cui 等^[50]). 某句子如果能和规则严格匹配,则该句子被选择为候选句,这种匹配方式称为“硬匹配”,其缺点是:候选句和模式规则之间即使有非常微小的差异,句子也会被判定为不匹配。

Cui 等^[50]提出一种“软模式匹配”方法,从训练样例中自动学习定义句子模式. 该方法不是把匹配看作二元分类问题,而是利用基于概率的模式匹配模型,来计算候选定义句和模式间的匹配度。

(2) 候选定义句子排序. 排序时,一般是首先给定义目标构造一个质心向量 (*centroid vector*) 或用户描述文件 (*profile*),然后依据各个候选句和该质心向量的相似度大小对候选句排序。

构造质心向量或描述文件时,词的来源有:(a) 外部资源中对定义目标词的解释^[48];(b) 全部候选定义句. 在这些句子中选择和定义目标共现次数超过阈值的词^[52]. 也有研究利用外部资源中的解释,对质心向量中词的权值进行调整处理^[53]. 与质心向量相似度的计算有 VSM 模型,也有统计语言模型方法^[52]。

Cui 等^[50]对参加 TREC 评测的各种定义型问答系统采用的语言构造和候选句子统计排序技术进行了总结,见表 3。

表 3 参加 TREC 评测的部分定义型问答系统采用的技术总结

TREC System	Definitional Linguistic Constructs					Statistical Ranking	
	Surface Pattern	Patterns on Parsing Trees	Appositives and Copulas	Relative Clauses	Predicates and Verb Phrases	Centroid Vector or Profile	Mining External Definitions
NUS ^[50]	×		×	×		×	
BBN ^[48]	×	×	×	×	×	×	×
Columbia ^[54]		×	×		×	×	
LCC ^[51]	×		×				
MIT ^[49]	×		×	×	×	×	×
IBM PIQUANT ^[46]	×		×	×		×	×
Sheffield ^[55]	×		×	×		×	×

(3) 候选定义句子中冗余信息过滤与去除

(a) 冗余句子过滤. 大多是采用多文档文摘的方法, 主要有: 基于词的句子过滤, 包括词覆盖方法和余弦相似度方法; 基于 n-gram 覆盖的句子过滤. Cui 等^[50]发现, n-gram 覆盖过滤方法, 在去冗余句子效果上要优于词覆盖方法, 而词覆盖方法要好于余弦相似度的方法.

(b) 冗余短语去除. 在候选定义句中, 可能会有一些与目标定义无关的短语. 采用的去除方法也是多文档文摘中已有的方法^[50, 55].

3.4 复杂问题的处理

3.4.1 问题分析

复杂问题的分析一方面是为后续答案抽取提供约束条件, 另一方面是对问题进行分解、简化.

(1) 问题细分类. 内梅亨大学的 Verberne 等对 why 型问题也进行了细分类, 以利于答案的抽取^[56].

(2) 问题分解. 将一个复杂问题分解为多个简单的问题, 对这些简单问题的答案句进行合并综合, 结果就是原始复杂问题的答案. 现有的问题分解研究主要在 LCC 公司, 分解方法有^[57]:

(a) 基于语法和浅层语义的分解. 复杂问题中往往包含有连接短语、连续的列举, 或嵌套子问题, 因此可以在语法语义分析的基础上, 利用规则对连接短语、列举、嵌套等进行分解.

(b) 基于随机步模型的分解. 用已有问题中的“关系对”作为查询去检索答案句, 然后用检索到的答案句中的其它关系, 利用规则来构造新问题, 这些新问题成为对原始复杂问题的分解; 反复进行该过程的迭代, 直至达到一定随机步数阈值.

3.4.2 相关文档检索

对于提问关于实体间关系的问题, 大部分研究都只以实体名称作为查询关键词^[3]. 答案句中的实体名称, 不一定和问题中的实体名完全匹配, 如问题中是“John McCain”, 而在句子中是“Senator McCain”. 因此, 滑铁卢大学的 MacKinnon 等利用 Wikipedia 中的链接锚文本, 对所指向文章标题中的实体名进行扩展^[58].

检索时, 现有研究都是用传统成熟的检索模型和工具, 如 BM25 模型、统计语言模型等. 但也有研究对特殊类型问题的文档检索结果进行重排序, 如布莱顿大学的 Yin 对 how-to 类型问题的检索结果, 判断文档属于“过程描述”的概率并排序^[59].

3.4.3 答案抽取

(1) 关系型复杂问题的答案抽取

(a) 候选答案句抽取与排序. 在检索阶段获得的前 n 篇文档中, 将包含查询词的句子提取出来作为候选答

案句. 对这些句子排序时, 简单方法是按照包含关键词的多少进行排序. 但也有研究利用机器学习方法, 如密执安大学的 Zhang 等^[60]采用逻辑回归模型对句子分类, 按照概率大小进行排序.

(b) 答案的综合生成. 对包含查询关键词的句子, 用多文档文摘方法综合压缩, 如马里兰大学 Madnani 等^[61]; 还有研究复杂问题分解而产生多个简单问题, 对这些简单问题的答案句, 采用多文档文摘方法进行综合、重排序, 得到最终答案, 如 LCC 公司的 Lacatusu 等^[62].

(2) why 型复杂问题的答案抽取

奈梅亨大学的 Verberne 等^[56]提出基于修辞结构理论的 why 型问题答案抽取方法. 答案抽取过程是: 首先在文档集中识别对应问题话题的文本段, 把和该文本段有特定修辞关系的文本段抽取出来作为答案.

3.5 交互式问答的处理

3.5.1 问题分析

在交互问答中, 用户问题会形成一个问题流. 对问题流中的单个问题, 除了上面所述的各种处理外, 还需要针对对话特点的处理:

(1) 问题上下文的关联 (relevancy) 识别. 判断用户新的提问是否和前一问题属于相同话题. AT&T 的 Yang 等^[63]利用问题是否包含代词、省略等特征, 用决策树方法, 进行问题关联的自动识别.

(2) 问题上下文的语篇处理. 在属于同一话题的多个连续问题所形成的语篇中, 会存在指代、引用等. 密执安大学的 Sun 等^[64]采用中心理论进行问题上下文语篇的指代消解、引用的识别.

(3) 用户提问意向 (intent) 的获取. 交互过程中, 用户所提的问题不仅反映其信息需求, 同时也体现用户对系统已有返回结果的认可意向. 密执安州立大学的 Chai 等^[65]针对新问题来分类判别用户对系统返回结果的意向. 可以将用户意向作为一种隐性反馈, 用于答案抽取策略的调整.

3.5.2 相关文档检索

交互式问答的已有研究主要着重于对话接口上, 而对系统采用的检索模型、检索排序等的研究很少, 大部分是采用和一般问答相同的检索方法.

Sun 等^[64]利用中心理论, 根据已有问题序列, 对当前新问题中的指代、引用等进行识别, 将识别结果作为当前问题的查询扩展. 约克大学的 De Boni 等^[66]认为, 如果问题存在表达的省略或指代, 则可以利用前面问题的相关文档, 而不需要重新检索.

3.5.3 答案抽取与生成

交互式问答中系统和用户之间有“对话”, 这种对

话并不只是“用户问”和“系统答”的过程.系统也可以主动提问或提示,通过协商来明晰用户的真正需求.

现有研究目前主要着重于对话接口的设计.而答案抽取技术一般采用非交互式技术.已有的对话接口方面的研究工作包括:

(1)聊天机器人形式的对话接口设计.通过良好的对话设计,使用户和系统的对话流畅自然.如约克大学的 Quarteroni 等^[67]用 AIML 解释器来实现对话接口管理.

(2)基于隐式反馈的答案正确性自动评估.通过用户后续的反应,对输出结果进行自动评估,然后根据评估结果对后续检索、答案抽取进行改进.如 Chai 等^[68].

(3)“建议”和“提示”问题的生成和输出.系统在给出答案的同时,也给出一些用户在后续可能会提的问题,则会显著提高用户的满意度.LCC 的 Hickl 等^[69]在给出答案的同时,从已构造好的“问/答对”库中选择出一批问题,利用 CRF 模型判断这些问题是否与用户问题相关,然后将相关问题作为给用户的“建议”输出.

3.6 中文开放域问答技术研究现状

中文开放域问答技术研究,所解决的问题类型、系统体系结构设计,基本是参考英文问答系统现有研究.

(1)问题分类.哈工大文勘等^[70]利用问句依存句法结构作为一种特征,训练贝叶斯分类器分类.南京理工大学的张亮^[72]在贝叶斯模型基础上,加入了疑问词和扩展特征项特征,改进了性能.沈阳航院孙景广等^[71]使用知网作为语义资源,利用最大熵模型进行分类,取得比贝叶斯模型更好的性能.

(2)检索查询构造.台湾中研院的 Tang 等^[73],利用 CRF 模型从人工的问题查询构造中学习自动的查询构造模型,提高了检索性能.

(3)答案抽取

(a)事实型问题答案抽取.台湾大学的林川杰^[74]采用简单模式匹配方法来抽取答案.自动化所的吴友政等^[75,76]提出一种无监督的学习算法,从 Web 上抽取中文问答的答案模式.王树西^[77]提出一种模式推理模型及推理算法,应用到人物关系问答中.孙昂等^[78]综合利用 SVM 和最大熵模型,采用多种特征,进行有指导的答案抽取.

(b)定义型问题答案抽取.相关研究极少,较早的是 BBN 公司的 Peng 等^[79]将人工构造表层模式的方法和基于深层语言分析的模式自动学习方法相结合.

在复杂问题和交互式问答方面,目前尚没有见到中文方面的研究.

总体来看,中文问答研究采用的技术路线、模型和英文问答研究没有太大差别,主要在实现上针对中文特点进行改进.所用技术中,规则方法和机器学习方法

都有研究,但由于国内缺少标准问答评测集,因此对不同方法进行比较较困难.

在 NTCIR-5 和 NTCIR-6 的跨语言问答评测中,设立了繁体中文事实型问答评测,最好的系统能正确回答 52.7% 的问题^[80],但和参与 TREC 评测的英文问答系统相比,仍有很大差距.总结产生差距的可能原因如下:

(a)在自然语言处理技术方面,中文总体上性能偏低,影响了答案的正确抽取.

(b)相对英文问答技术研究,中文方面的研究不够深入.如问题和文本句子在逻辑表示和推理方面,LCC 的研究已经比较成熟.但在中文方面,类似研究极少.

(c)中文问答研究可用的数据集偏少.除了 NTCIR 问答评测语料,其它公开的数据集很少.利用统计机器学习方法时,训练数据的规模是影响性能的关键因素之一.

(d)国内没有与问答相关的评测,来激励和带动中文问答研究,同时涉及中文问答研究的机构也相对较少.

4 现有方法中有待解决的问题

4.1 问题分析

(1)问题分类体系不统一.目前,问题分类体系没有一个统一的标准,各个系统都是根据自己的处理算法来定义类别,导致训练数据很难共享.

(2)训练数据不平衡和多种分类问题.训练数据在问题类别上数量的不平衡,导致分类器在不同类别上有不同的性能.另外,分类方法均未考虑问题属于多类别的情况,如“巴黎有多大?”,答案可以是面积、人口数量等.

(3)缺乏对问题上下文的语义表示和处理模型的研究.用户提问都有上下文环境,因此问题中会有省略、隐含等情况.如“美国的总统是谁?”,不同时期的总统可能不同.在问题上下文的理论模型的建立和表示、用户提问行为分析等方面,尚需要更多的研究.

在交互式问答中,存在很多需要解决的问题,如问题流中的语篇处理与普通篇章语篇处理的关系的研究,用户模型、交互模型表示方法,交互界面的设计等.

(4)缺乏在真实应用条件下的用户输入处理.现有问题分析研究大多是使用技术评测的问题作为训练和测试语料.但在实际应用中,用户输入千差万别,表述的语法和拼写也不一定规范.因此需要针对真实应用下的用户输入处理,进行自动提示和更正、抽取用户真实需求等容错性处理.

4.2 文档和句段检索

现有问答系统中的检索大部分是利用传统文档检索的模型和工具.但是,问答系统和一般文档检索任务

在需求上有本质的不同:文档检索任务,考虑的是文章主题和查询主题之间整体的相似度;而问答系统中,答案很可能在文章局部位置偶尔出现,问题话题和文章的整体主题可能并无紧密关系。

由于任务不同,在传统信息检索任务中表现好的检索模型,用在问答系统中不一定有更好的性能,还需要继续对已有的检索模型进行公平比较。

在查询扩展方面,尽管已经有了各种扩展方法的研究,但是在这些方法之间进行横向的性能比较的研究很少,绝大部分研究都是考虑扩展对检索准确率的影响,而极少考虑对检索时间效率的影响。

4.3 答案抽取与生成

(1) 关于事实型与列举型问题答案抽取

(a) 自然语言处理技术的性能限制。如命名实体识别,两个主要的问题是:命名实体的类别体系大多是采用 MUC 定义的标准,类别的划分较粗;而答案抽取需要更细的类别划分,如江河、城市、药品、电影等。命名实体识别技术总体性能难以达到要求。另外如句法分析的整体性能不尽人意,影响了对系统性能提高的帮助。

(b) 答案置信度评价方法不足。游澜等^[81]利用最大熵模型对答案进行置信度评估。逻辑推理答案验证更有说服力,但这类系统的实现非常复杂。

许多系统利用 Web 去验证、排序候选答案,前提是认为正确答案和问句词在 Web 检索结果中的出现次数相对更高。但实际上,出现次数和正确性之间并非必然关系。Web 冗余数据的应用还需要有更合理的模型。

(c) 阈值选取缺乏理论指导。对于列举型问题答案抽取,在确定返回结果的个数时,多数系统都使用阈值方法,而且对所有问题都选择同样的阈值,缺乏对答案完备性的量化模型。而目前的阈值选择主要靠实验,并没有一个理论依托,因此很难保证在测试时能达到最优效果。

(2) 关于定义型问题的答案抽取

(a) 缺乏对定义信息所在文档的可信度判断。总结现有系统对候选定义句子进行排序的算法,不足之处是:主要看和模式以及质心向量的相似度排序,而没有考察句子所在文档本身的质量。

(b) 缺乏对抽取到的定义在宏观角度的完备性度量。和列举型问题的答案抽取类似,在所抽取定义的准备性方面,一方面是现有研究不区分定义目标类型,都采用同样的方法来抽取不同目标的定义,另一方面是缺乏定义是否完备的检验标准。实际上,不同的目标有其不同的完备性要求,如对一个人物和一种昆虫的定义描述,就应该有不同的组成要素。因此,不仅需要进行单个信息片断是否属于定义的微观度量,也需要结

合目标本体,判断定义在不同侧面是否已完备的宏观度量。

(c) 缺乏对最终定义的逻辑一致性和可读性判断。在生成定义时,现有研究几乎都不考虑最终所产生的定义的流畅与逻辑一致性。在目前的 TREC 评测中,也不考虑系统返回定义的可读性。因此,目前的方法离应用的需求仍然有很大差距。

(3) 关于复杂型问题的与交互问答中的答案抽取

现有关系型复杂问题的研究,问题是提问关于实体间的关系及证据。复杂问题种类较多,除了关系型、how-to 型、why 型问题之外,还有询问关于“是/否”的逻辑判断问题和选择型问题等,而目前对这些问题鲜有研究。

复杂问题的难点在于对预期答案的约束条件很难用词之间的浅层关系来表述,而需要用多个概念的深层语义关系来表达。已有方法在把这种深层语义约束分解成多个简单问题时,一方面分解过程本身是经验性的,另外也缺乏对复杂问题语义模型的分析 and 建立。

why 型问题的研究利用了修辞结构理论^[57]。但在目前还没有一个公开自动的篇章修辞结构分析工具。

交互式问答是问答技术的重要发展方向,已有研究中,对用户输入问题的处理(类似于“听”)研究较多,但对系统输出问题的处理(类似于“问”)研究极少。

总的来看,复杂问题和交互式问答解决方法的研

5 总结与展望

本文从开放域问答的问题分析、文档与句段检索、答案抽取生成等方面总结了近几年的主要研究成果,分析比较了在事实型、列举型、定义型、复杂交互型这几种问题类型上各种技术方法的特点和优劣,讨论了现有方法有待解决的问题。

总体上,对事实型问题的处理技术研究已较深入。从 TREC 评测看,最好的系统能正确回答约 70% 的简单问题^[3]。但对于复杂问题,已有的研究仍不够广泛深入。问答技术研究的长期目标是给不同应用水平的用户提供精准的在线信息访问。基于这一目标,不断深入的研究工作为该领域提出了更多需要解决的问题。总结现有技术的不足,需要在以下方面继续进行深入研究:

(1) 问答系统的真实应用模式研究

问答技术成为研究热点,以 TREC 为代表的各种评测无疑起了极大的促进作用。评测总会有终点,市场的真正需求才是技术得以进步的巨大动力。但是目前还没有哪个开放域问答系统能真正被用户大量使用。显然,评测中所设计的问题类型和模式,还不能完全代表

真实应用需要.因此,寻找使问答系统真正为用户精准信息需求提供服务的模式,是需要深入研究的问题.

(2) 以用户为中心的交互模型及界面的研究

现有答案抽取技术研究的焦点还主要在文档和问题本身的内容上,基本不考虑用户一端的使用背景.问答系统需要能给不同知识背景的用户提供个性化答案.根据用户行为对用户进行层次类别划分、用户兴趣和意图的获取等,都是需要研究的问题.

(3) 复杂问题的分析和回答方法研究

问答技术研究解决的问题也逐渐扩展到复杂问题上但是对复杂问题的分析仍不够深入.在复杂问题的分解模型、结合多文档文摘技术的答案综合方法、篇章内修辞关系的建立等方面,还需更多研究.

(4) 异构数据源的检索和集成方法及检索模型研究

Web 上存在各种不同形式的数据,包括新闻网页、blog、论坛和 FAQ 库,以及百科全书、结构化的数据库等.如何将这些异构数据进行一致的索引和集成,从不同侧面进行答案抽取和正确性验证,综合利用它们各自的优势,是问答系统,特别是面向企业的问答系统在数据源方面需要研究的重要问题.

(5) 答案置信度、完备性与一致性判断方法研究

抽取答案时要区分候选答案所在数据源的可信度差别,在给用户返回答案的同时,系统应该对答案有一定水平的置信度.同时,需要完善对答案完备性的量化指标评估,而不是仅仅依靠经验估计;利用语言生成技术,给用户返回易于理解、逻辑一致、可读性好的答案,而不是信息片段的简单堆砌.

(6) 问答系统自动评测方法研究

现有问答技术的评测中需要很多人工手段,评测任务与真正的应用尚有差别,只是考虑系统的准确率和召回率,不考虑系统的时间性能、可靠性等因素.因此,涉及多维度的自动评测方法也是需要深入研究的重点.

问答的难点之一是用户问题和文档在叙述形式上的鸿沟,该问题的解决包含了许多子问题,如文本蕴含、复述等.在这些子问题上,国内目前的研究较少.

最后,限于篇幅,问答的其他一些方向,如限定域问答、跨语言问答、基于社区论坛的回答等方面的研究在近几年取得的其他成果,本文未有论及,但也都是重要的研究方向.

参考文献:

[1] Hirschman L, Gaizauskas R. Natural language question answering: the view from Here[J]. Natural Language Engineering,

2001,7(4):275-300.

- [2] 吴友政,赵军,等.问答式检索技术及评测研究综述[J].中文信息学报,2005,19(3):1-13.
Wu Youzheng, Zhao Jun, et al. Research on question answering & evaluation: a survey[J]. Journal of Chinese Information Processing, 2005, 19(3):1-13. (in Chinese)
- [3] Dang H-T, et al. Overview of the TREC 2007 question answering track[A]. Proc of the 16th Text REtrieval Conference[C]. Gaithersburg: NIST Special Publication, 2007. 44-60.
- [4] Moldovan D, et al. Performance issues and error analysis in an open-domain question answering system[J]. ACM Trans on Information Systems, 2003, 21(2):133-154.
- [5] Hovy E, et al. Toward semantics-based answer pinpointing [A]. Proc of the HLT Conference (HLT2001) [C]. San Diego, USA: Association for Computational Linguistics, 2001. 1-7.
- [6] Brill E, et al. An analysis of the AskMSR question-answering system[A]. Proc of 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP) [C]. Philadelphia, USA: Association for Computational Linguistics, 2002. 257-264.
- [7] Li X, Roth D. Learning question classifiers: the role of semantic information [J]. Natural Language Engineering, 2004, 1(1):229-249.
- [8] Zhang D, Lee W-S. Question classification using support vector machines[A]. Proc of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval[C]. Toronto, Canada: ACM, 2003. 26-32.
- [9] Moschitti A, et al. Exploiting syntactic and shallow semantic kernels for question/answer classification [A]. Proc of the 45th Annual Meeting of the Association of Computational Linguistics[C]. Prague, Czech Republic: Association for Computational Linguistics, 2007. 776-783.
- [10] Suzuki J, et al. Question classification using HDAG kernel [A]. ACL 2003 Workshop on Multilingual Summarization and Question Answering[C]. Morristown, USA: Association for Computational Linguistics, 2003. 61-68.
- [11] Li W. Question classification using language modeling[R]. CIIR Technical Report, 2002.
- [12] Lin S-J, Lu W-H. Learning question focus and semantically related features from web search results for Chinese question classification [A]. Proc of AIRS 2006, the 3rd Asia Information Retrieval Symposium [C]. Singapore: Springer, 2006. 284-196.
- [13] Srijari R, Li W. A question answering system supported by information extraction [A]. Proc of the Sixth Conference on Applied Natural Language Processing [C]. Seattle, Washington: Morgan Kaufmann Publishers Inc, 2000. 166-172.
- [14] Harabagiu S, et al. Falcon: boosting knowledge for answer en-

- gins [A]. Proc of the Ninth Text REtrieval Conference [C]. Gaithersburg: NIST Special Publication, 2001. 479-488.
- [15] Sun R-X, et al. Using syntactic and semantic relation analysis in question answering [A]. Proc of the 14th Text REtrieval Conference [C]. Gaithersburg: NIST Special, 2005. 611-629.
- [16] Hovy E, et al. Question answering in Webclopedia [A]. Proc of the Ninth Text REtrieval Conference [C]. Gaithersburg, USA: NIST Special Publication, 2001. 655-664.
- [17] Thompson K C, et al. The effect of document retrieval quality on factoid question answering performance [A]. Proc of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval [C]. Sheffield, UK: ACM, 2004. 574-575.
- [18] Gaizauskas R, et al. The University of Sheffield's TREC 2003 Q&A experiments [A]. Proc of the 12th Text REtrieval Conference (TREC 2003) [C]. Gaithersburg: NIST Special Publication, 2003. 782-790.
- [19] Monz C. Document retrieval in the context of question answering [A]. Proc of the 25th European Conference on Information Retrieval Research (ECIR-03) [C], Lecture Notes in Computer Science 2633, Springer, 2003. 571-579.
- [20] Bilotti M W, et al. What works better for question answering: stemming or morphological query expansion [A]. Proc of the SIGIR Workshop on Information Retrieval for Question Answering (IR4QA) [C]. Sheffield, UK: ACM, 2004. 1-7.
- [21] Ittycheriah A, et al. IBM's statistical question answering system- TREC-10 [A]. Proc of the 10th Text REtrieval Conference [C]. Gaithersburg: NIST Special Publication, 2002. 258-264.
- [22] Moldovan D, Novischi A. Lexical chains for question answering [A]. Proc of the 19th International Conference on Computational Linguistics [C]. Taipei: Association for Computational Linguistics, 2002. 1-7.
- [23] Agichtein E, et al. Learning to find answers to questions on the Web [J]. ACM Trans on Internet Technology (TOIT), 2003, 4 (2): 129-162.
- [24] Riezler S, et al. Statistical machine translation for query expansion in answer retrieval [A]. Proc of the 45th Annual Meeting of the Association of Computational Linguistics [C]. Prague, Czech Republic: Association for Computational Linguistics, 2007. 464 - 471.
- [25] Hovy E, et al. The use of external knowledge in factoid QA [A]. Proc of the 10th Text REtrieval Conference [C]. Gaithersburg: NIST Special Publication, 2002. 644-652.
- [26] Light M, et al. Analysis for elucidating current question answering technology [J]. Natural Language Engineering, 2001, 7(4): 325-342.
- [27] Vicedo J L, Ferrández A. University of Alicante at TREC-10 [A]. Proc of the 10th Text REtrieval Conference [C]. Gaithersburg: NIST Special Publication, 2002. 630-637.
- [28] Clarke C, et al. Question answering by passage selection [A]. Proc of the Ninth Text REtrieval Conference [C]. Gaithersburg: NIST Special Publication, 2001. 673-654.
- [29] Tellex S, et al. Quantitative evaluation of passage retrieval algorithms for question answering [A]. Proc of the Twenty-Sixth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval [C]. Toronto, Canada: ACM, 2003. 41-47.
- [30] Roberts I, Gaizauskas R. Evaluating passage retrieval approaches for question answering [A]. Proc of 26th European Conference on Information Retrieval (ECIR '04) [C]. Sunderland, UK: Springer-Verlag, 2004. 72-84.
- [31] Moldovan D, Clark C, et al. Lymba's PowerAnswer 4 in TREC 2007 [A]. Proc of the Sixteenth Text REtrieval Conference [C]. Gaithersburg: NIST Special Publication, 2007. 277-285.
- [32] Yang H, Chuan T-S. Web-based list question answering [A]. Proc of the 20th International Conference on Computational Linguistics [C]. Geneva, Switzerland: Association for Computational Linguistics, 2004. 1277-1283.
- [33] Martin M. Soubbotin, Sergei M. Soubbotin. Patterns of potential answer expressions as clues to the right answers [A]. Proc of the 10th Text REtrieval Conference [C]. Gaithersburg: NIST Special Publication, 2002. 293-302.
- [34] Ravichandran D, Hovy E. Learning surface text patterns for a question answering system [A]. Proc of the 40th Annual Meeting of the Association for Computational Linguistics [C]. Pennsylvania: Association for Computational Linguistics, 2002. 41-47.
- [35] Kaiser M, Becker T. Question answering by searching large corpora with linguistic methods [A]. Proc of the 13th Text Retrieval Conference [C]. Gaithersburg: NIST Special Publication, 2005. 535-544.
- [36] Shen D, Kruijff G-J M, et al. Exploring syntactic relation patterns for question answering [A]. Proc of Second International Joint Conference on Natural Language Processing [C]. Jeju, Korea: Springer, 2005. 507-518.
- [37] Wu M, et al. ILQUA-An IE-Driven question answering system [A]. Proc of the 14th Text REtrieval Conference [C]. Gaithersburg: NIST Special Publication, 2005. 77-83.
- [38] Cui H, et al. National university of Singapore at the TREC 13 question answering [A]. Proc of the 13th Text Retrieval Conference [C]. Gaithersburg: NIST Special Publication, 2004. 34-41.
- [39] Shen D, Klakow D. Exploring correlation of dependency relation paths for answer extraction [A]. Proc of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the ACL [C]. Sydney: Association for Com-

- putational Linguistics, 2006. 889-896.
- [40] Echihabi A, Marcu D. A noisy-channel approach to question answering [A]. Proc of the 41st Annual Meeting of the Association for Computational Linguistics [C]. Sapporo, Japan: Association for Computational Linguistics, 2003. 16-23.
- [41] Clarke C L A, et al. Exploiting redundancy in question answering[A]. Proc of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval[C]. New Orleans: ACM, 2001. 358-365.
- [42] Magnini B, et al. Is it the right answer? exploiting web redundancy for answer validation [A]. Proc of the 40th Annual Meeting of the Association for Computational Linguistics [C]. Philadelphia: Association for Computational Linguistics, 2002. 425-432.
- [43] Harabagiu S, et al. Employing two question answering systems in TREC-2005[A]. Proc of the 14th Text REtrieval Conference[C]. Gaithersburg: NIST Special Publication: 2005. 94-103.
- [44] Moldovan D, et al. Cogex: a logic prover for question answering[A]. Proc of HLT-NAACL 2003[C]. Edmonton: Association for Computational Linguistics, 2003. 87-93.
- [45] 李鹏, 王晓龙, 等. 一种基于混合策略的失衡数据集分类方法[J]. 电子学报, 2007, 35(11): 2161-2165.
Li Peng, Wang Xiaolong, et al. A classification method for imbalance data set based on hybrid strategy[J]. Acta Electronica Sinica, 2007, 35(11): 2161-2165. (in Chinese)
- [46] Chur-Carroll J, et al. IBM's PIQUANT in TREC 2004 [A]. Proc of the 13th Text REtrieval Conference [C]. Gaithersburg: NIST Special Publication, 2005. 403-410.
- [47] Wu L-D, et al. FDUQA on TREC 2005 QA Track [A]. Proc of the 13th Text REtrieval Conference [C]. Gaithersburg: NIST Special Publication, 2005. 459-468.
- [48] Xu J, et al. Evaluation of an extraction-based approach to answering definitional questions [A]. Proc of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval [C]. Sheffield, UK: ACM, 2004. 418-424.
- [49] Katz B, et al. Answering multiple questions on a topic from heterogeneous resources [A]. Proc of the 13th Text REtrieval Conference [C]. Gaithersburg: NIST Special Publication, 2005. 488-497.
- [50] Cui H, et al. Soft pattern matching models for definitional question answering[J]. ACM Trans on Information Systems (TOIS), 2007, 25(2): Article 8.
- [51] Harabagiu S, et al. Answer mining by combining extraction techniques with abductive reasoning [A]. Proc of the 12th Text REtrieval Conference [C]. Gaithersburg: NIST Special Publication, 2004. 375-382.
- [52] Chen Y, Zhou M, et al. Reranking answers for definitional QA using language modeling [A]. Proc of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the ACL [C]. Sydney: Association for Computational Linguistics, 2006. 1081 - 1088.
- [53] Zhang Z-Z, et al. Answering definition questions using web knowledge bases [A]. Proc of Second International Joint Conference [C]. Jeju Island, Korea, 2005. 498-506.
- [54] Blair-Goldensohn S, et al. A hybrid approach for QA track definitional questions [A]. Proc of the 12th Text REtrieval Conference [C]. Gaithersburg: NIST Special Publication, 2004. 185-192.
- [55] Gaizauskas R, et al. The University of Sheffield's TREC 2004 Q&A experiments [A]. Proc of the 13th Text REtrieval Conference [C]. Gaithersburg: NIST Special Publication, 2005. 758-764.
- [56] Verberne S, et al. Discourse-based answering of why-questions [J]. Traitement Automatique Des Langues, Special Issue on Computational Approaches to Discourse and Document Processing, 2007, (47): 21-41.
- [57] Harabagiu S, Lacatusu F, et al. Answering complex questions with random walk models [A]. Proc of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval [C], Seattle: ACM, 2006. 220-227.
- [58] MacKinnon I, Vechtomova O. Complex interactive question answering enhanced with Wikipedia [A]. Proc of the 16th Text Retrieval Conference [C]. Gaithersburg: NIST Special Publication, 2007. 290-294.
- [59] Yin L. A two-stage approach to retrieving answers for how-to questions [A]. Proc of the 11th conference of the European Chapter of the ACL, Student Research Workshop [C]. Trento, Italy: Association for Computational Linguistics, 2006. 63-70.
- [60] Zhang C, et al. Michigan State University at the 2007 TREC ciQA evaluation [A]. Proc of the Sixteenth Text REtrieval Conference [C]. Gaithersburg: NIST Special Publications, 2008. 501-508.
- [61] Madnani N, et al. TREC 2007 ciQA Task: University of Maryland[A]. Proc of the Sixteenth Text REtrieval Conference [C]. Gaithersburg: NIST Special Publications, 2008. 661-666.
- [62] Lacatusu F, Hickl A, et al. Impact of question decomposition on the quality of answer summaries [A]. Proc of the Seventh Language Resources and Evaluation Conference [C]. Genoa, Italy, 2006.
- [63] Yang F, et al. A data driven approach to relevancy recognition for contextual question answering[A]. Proc of the IQA Workshop at HLT-NAACL 2006 [C]. New York City, USA: Association for Computational Linguistics, 2006. 33 - 40.

- [64] Sun M, Chai J. Discourse processing for context question answering based on linguistic knowledge knowledge-based systems [J]. Special Issues on Intelligent User Interfaces, 2007, 20(6): 511-526.
- [65] Chai J, Zhang C, et al. Towards conversational QA: automatic identification of problematic situations and user intent [A]. Proc of the COLING/ ACL 2006 Main Conference Poster Sessions [C]. Sydney: Association for Computational Linguistics, 2006. 57 - 64.
- [66] De Boni M, Manandhar S. Implementing clarification dialogue in open-domain question answering [J]. Natural Language Engineering, 2005, 11(4): 343 - 361.
- [67] Quarteroni S, Manandhar S. A chatbot-based interactive question answering System [A]. Proc of the 11th Workshop on the Semantics and Pragmatics of Dialogue [C]. Trento, Italy: University of Trento, 2007. 83 - 90.
- [68] Chai J, Baldwin T, et al. Automated performance assessment in interactive QA [A]. Proc of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval [C]. Seattle: ACM, 2006. 631-632.
- [69] Hickl A, Harabagiu S. Enhanced interactive question-answering with conditional random fields [A]. Proc of the IQA Workshop at HL T-NAACL [C]. New York City, USA: Association for Computational Linguistics, 2006. 25 - 32.
- [70] 文勳, 张宇, 等. 基于句法结构分析的中文问题分类[J]. 中文信息学报, 2006, 20(2): 33-39.
Wen Xu, Zhang Yu, et al. Syntactic Structure Parsing Based Chinese Question Classification [J]. Journal of Chinese Information Processing, 2006, 20(2): 33-39. (in Chinese)
- [71] 孙景广, 蔡东风, 等. 基于知网的中文问题自动分类[J]. 中文信息学报, 2007, 21(1): 90-95.
Sun Jingguang, Cai Dongfeng, et al. HowNet Based Chinese Question Automatic Classification [J]. Journal of Chinese Information Processing, 2007, 21(1): 90-95. (in Chinese)
- [72] 张亮. 面向开放域的中文问答系统问句处理相关技术研究[D]. 南京: 南京理工大学, 2005.
- [73] Tang H, et al. Query term selection strategies for web-based chinese factoid question answering [A]. Proc of the 11th Conference on Artificial Intelligence and Applications (TAAI 2006) [C]. Kaohsiung, Taiwan: KUAS, 2006.
- [74] 林川杰. 中文开放领域自动问答系统之研究[D]. 台北: 台湾大学, 2004.
- [75] 吴友政, 赵军, 等. 基于无监督学习的问答模式抽取技术[J]. 中文信息学报, 2007, 21(2): 69-76.
Wu Youzheng, Zhao Jun, et al. Unsupervised Answer Pattern Acquisition [J]. Journal of Chinese Information Processing, 2007, 21(2): 69-76. (in Chinese)
- [76] Wu Y-Z, et al. Learning unsupervised svm classifier for answer selection in web question answering [A]. Proc of the 2007 Joint Conference on Empirical Methods in Natural Language [C]. Prague, Czech Republic: Association for Computational Linguistics, 2007. 33 - 41.
- [77] 王树西. 基于文本模式推理的问答系统研究[D]. 北京: 中国科学院计算技术研究所, 2005.
- [78] 孙昂, 江铭虎, 等. 基于句法分析和答案分类的中文问答系统[J]. 电子学报, 2008, 36(5): 833-839.
Sun Ang, Jiang Minghu, et al. Chinese question answering based on syntax analysis and answer classification [J]. Acta Electronica Sinica, 2008, 36(5): 833-839. (in Chinese)
- [79] Peng F, et al. Combining deep linguistics analysis and surface pattern learning: a hybrid approach to chinese definitional question answering [A]. Proc of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing [C]. Vancouver, Canada: Association for Computational Linguistics, 2005. 307 - 314.
- [80] Sasaki Y, et al. Overview of the NTCIR-6 cross-lingual question answering (CLQA) task [A]. Proc of NTCIR-6 Workshop Meeting [C]. Tokyo, Japan: NII, 2005. 153-163.
- [81] 游澜, 周雅倩, 等. 基于最大熵模型的 QA 系统置信度评分算法[J]. 软件学报, 2005, 16(8): 1407-1414.
You Lan, Zhou Yaqian. A maximum entropy model based confidence scoring algorithm for QA [J]. Journal of Software, 2005, 16(8): 1407-1414. (in Chinese)

作者简介:



张志昌 男, 1976 年生于甘肃天水. 哈尔滨工业大学计算机科学与技术学院博士研究生. 研究方向为问答系统、自然语言处理.

Email: pangzhang@gmail.com



张宇 男, 1972 年生于黑龙江大庆. 哈尔滨工业大学计算机科学与技术学院副教授、硕士生导师. 研究方向为信息检索、自然语言处理.