

Design of Question Answering System of Intelligent Customer Service Robot Based on Knowledge Base

Jie Lu, Shaobo Li

The Mechanical Engineering College, Guizhou University, Guiyang Guizhou
Email: 282102166@qq.com

Received: Nov. 1st, 2019; accepted: Nov. 14th, 2019; published: Nov. 21st, 2019

Abstract

Firstly, this paper analyzes the current situation of intelligent customer service robot, summarizes the characteristics of information retrieval, database query and Q&A system. It also compares the question answering system based on Web information retrieval, the question answering system based on knowledge base and the community question answering system. Secondly, it decomposes the knowledge base management and relationship architecture of Q&A system, and puts forward the specific learning process and implementation framework. Finally, the weather query application scenario is simulated. The results show that the question answering system of intelligent customer service robot based on knowledge has good practical application effect.

Keywords

Question Answering, Knowledge Base, Natural Language Processing, Resource Description Framework

基于知识库的智能客服机器人问答系统设计

陆 婕, 李少波

贵州大学机械工程学院, 贵州 贵阳
Email: 282102166@qq.com

收稿日期: 2019年11月1日; 录用日期: 2019年11月14日; 发布日期: 2019年11月21日

摘 要

本文首先对智能客服机器人现状进行分析, 总结了基于任务处理模式的信息检索、数据库查询、问答系

统特点, 并对基于Web信息检索的问答系统、基于知识库的问答系统和社区问答系统进行对比; 而后, 对知识库问答系统的知识库管理、关系架构进行分解, 提出了具体的学习过程和实现框架; 最后, 结合天气查询应用场景进行仿真。结果证明, 本文提出的基于知识库的智能客服机器人问答系统具有良好的实际应用效果。

关键词

问答系统, 知识库, 自然语言处理, 资源描述框架RDF

Copyright © 2019 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

本文对智能客服机器人现状和问答系统进行了分析和比对, 结合实际细化具体的学习过程和实现框架; 并结合天气查询应用场景进行仿真。

问答系统通过 Web 搜索或知识库链接等方式, 对用户问题进行自然语言理解和检索、潜在答案对比推荐, 并用准确、简洁的自然语言进行回复。作为信息检索系统的高级方式, 问答系统跟接近信息检索中的语义检索, 并在信息检索推荐结果基础上进行二次分析, 给出精准答案, 并通过指代消解和内容补全处理后展现给用户。问答系统针对特定领域的知识进行一问一答, 侧重于知识结构构建、知识融合和知识推理。

2. 问答系统简介

2.1. 问答系统的特点

在任务处理的具体模式中, 问答系统与信息检索、数据库查询具有相同点, 其中问答系统与信息检索均需要根据用户提出的问题, 在 Web 上进行答案信息检索, 问答系统与数据库查询均需要在数据库或知识库上进行答案信息查询, 但在具体的输入输出、信息获取过程、应用场景等方面存在不同[1]。三种主流问答系统的特点如表 1 所示。

Table 1. Question and answer system feature table based on task processing mode

表格 1. 基于任务处理模式的问答系统特点表

类别	输入	输出	数据解释	信息获取过程	适用场景
信息检索	关键字	文档或结构化数据	用户让搜索引擎明白搜索意图	依赖多种检索操作, 是回复驱动过程	简单信息获取, 网络上有大量可参考资料
数据库查询	结构化	数据聚合	用户需理解数据库模式和 SQL	依赖多次查询操作	问题规模小而集中, 存在少量语义异构信息
问答系统	自然语言	准确答案	由机器承担数据解释工作	依赖自然语言处理和知识库, 是问题驱动过程	多样化、非结构化信息, 需进行自动化语义理解

2.2. 问答系统的分类

根据问题答案的数据来源和回答方式的不同, 主流问题系统可分为以下三类。

1) 基于 Web 信息检索的问答系统(Web Question Answering, WebQA): WebQA 系统在搜索引擎基础上发而来, 在理解和分析用户问题意图后, 以搜索引擎为支撑, 在全网范围内搜索相关答案反馈用户, 典型系统包括 Ask Jeeves 和 AnswerBus 系统。

2) 基于知识库的问答系统(Knowledge Based Question Answering, KBQA): KBQA 系统通过与已有的知识库或数据库资源(如 Freebase、DBpedia、Yago、Zhishi.me)结合, 利用非机构化文本信息(如维基百科、百度百科), 通过信息抽取的方式提前有价值的信息, 并构建知识图谱作为问答系统的支撑, 结合知识推理等方法为用户提供更高层次语义理解的答案[2]。

3) 社区问答系统(Community Question Answering, CQA): CQA 系统也称为基于社交媒体的问答系统, 其大多数答案由网友提供, 通过检索社交媒体中心(如 Yahoo! Answers、百度百科、知乎)与用户提问语义相似的问题, 并将答案反馈给用户。

除了上述三类问答系统, 还包括其他特殊场景的问答系统, 如混合式问答系统(Hybrid QA)、多语言问答系统(Multilingual QA)、基于常见问题库的问答系统(Frequently Asked Question, FAQ)等。其中, KBQA 应用最为广泛, 它不仅实现了对复杂问题的语义理解, 还融合了多个知识库间的相关知识, 并对复杂问题进行知识推理。

目前, 一些基于搜索引擎的问答系统也结合了知识图谱的知识, 使用语义检索的方式从多种来源收集信息, 可以根据用户的问题进行一定的推理, 并将适合的答案返回给用户以提高搜索质量, 例如 Google 知识图谱和百度知识图谱等。

3. KBQA 系统设计

3.1. 知识库管理

知识库(Knowledge Base, KB)是用于相关领域知识的采集、整理及提取的特殊数据库[3]。知识库中的知识来源于相关领域的专家, 是求解问题所需领域知识的集合, 包括基本事实、规则和其他相关信息。

知识库的表示形式是一个对象模型(Object Model), 通常为本体, 包含类、子类和实体。不同于传统的数据库, 知识库中存放的知识蕴含特殊的知识表示, 其结构比数据库更复杂, 可以用来存放更多复杂语义表示的数据。知识库最早被应用于专家系统, 它是一种基于知识的系统, 包含表示客观世界事实的一系列知识及一个推理机(Inference Engine), 并依赖一定的规则和逻辑形式推理出一些新的事实。

3.2. 关系架构设计

KBQA 系统是目前应用最广泛的问答系统之一, 适用于人们生活的方方面面, 例如在医疗、银行、保险、零售等行业建立相应专业知识的问答系统(智能客服系统), 可以给用户提供更好的服务。知识库一般采用资源描述框架(Resource Description Framework, RDF)格式对其中的知识进行表示, 除此之外, 还包括维基百科等无结构化文本知识库。

典型的 KBQA 系统包含问句理解、答案信息抽取、答案排序和生成等核心系统[4]。结合 KBQA 典型结构, 本文对其系统各模块间进行细化, 具体的关系架构图如图 1 所示, 其主要模块包括问句分析(Question Analysis)、短语映射(Phrase Mapping)、消歧(Disambiguation)和查询构建(Query Construction)。

问句理解模块在提取问题实体后, 答案信息抽取模块通过在知识库中查询该实体得到以该实体节点为中心的知识库子图, 并依据某些规则或模板从提取到的子图中抽取相应的节点或边, 得到表征问题和候选答案特征的特征向量, 最后将候选答案的特征向量作为分类模型的输入, 通过模型输出的分值对候选答案进行筛选, 得出最终答案[5], 学习过程如图 2 所示。

1) 问句分析模块

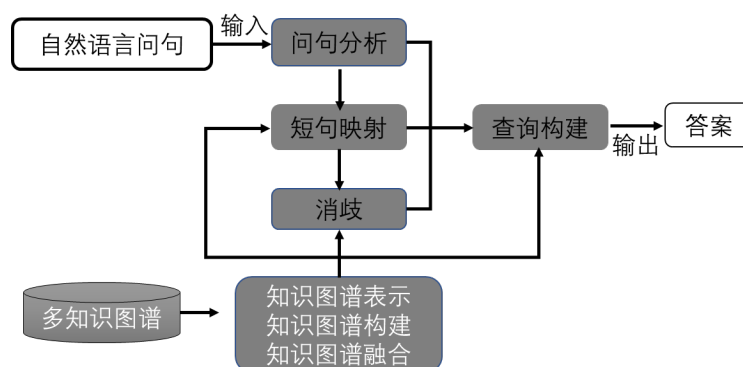


Figure 1. KBQA system relationship architecture diagram

图 1. KBQA 系统关系架构图

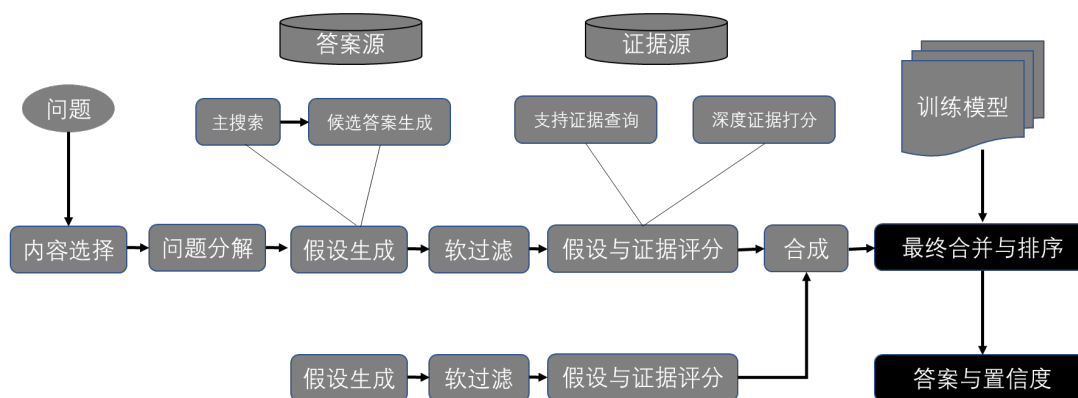


Figure 2. Learning process of the KBQA question answering system

图 2. KBQA 问答系统的学习过程

KBQA 系统中用到的问句分析技术属于自然语言处理范畴的任务, 并将自然语言转化成计算机可以理解的形式化语言的过程, 包括自动分词(对于中文)、词性标注、命名实体识别、指代消解、句法分析等任务。

2) 短语映射模块

短语映射模块负责将问题分析模块提取的信息词与知识库或知识图谱中的资源对应的标签映射连接起来。常用方法包括本体映射、同义词映射等, 并通过字符串相似度计算、语义相似度匹配(Sense-based Similarity Matcher)方法进行相似度计算。

3) 消歧模块

消歧模块又为候选答案排序(Rank)模块, 负责解决短语映射模块中出现的歧义问题, 以确保问句信息词和知识库实体(资源的标签)的无歧义映射。常用的方法包括基于字符串相似度的方法和基于属性和参数的判断方法。

4) 查询构建模块

查询构建模块融合前 3 个模块生成结果, 得到最终 SPARQL 查询语句, 并将查询结果返回给用户。查询 SPARQL 语句的方法分为模板、问题分析和机器学习三类。

3.3. 学习过程设计

结合广义知识库, 集中各种结构化知识和非结构化的文本语料和语言学知识, 是集理解、推理、学

习、交互功能于一体的强大问答系统, 其学习处理信息的过程通过模拟人的认知思考过程, 分为观察、推断、评估、决策 4 个阶段, 包含问题分解、假设生成、基于证据进行假设评估及排序等关键步骤[6]。

图 3 所示为 KBQA 问答系统的学习过程。首先, 通过分析问题的语义, 找出查询所需的依赖关系及查询的焦点; 然后, 根据查询线索生成候选答案, 并给出相关性的评分; 最后, 归并重复候选答案, 由候选答案评估算法做排序选出最终的答案[7]。

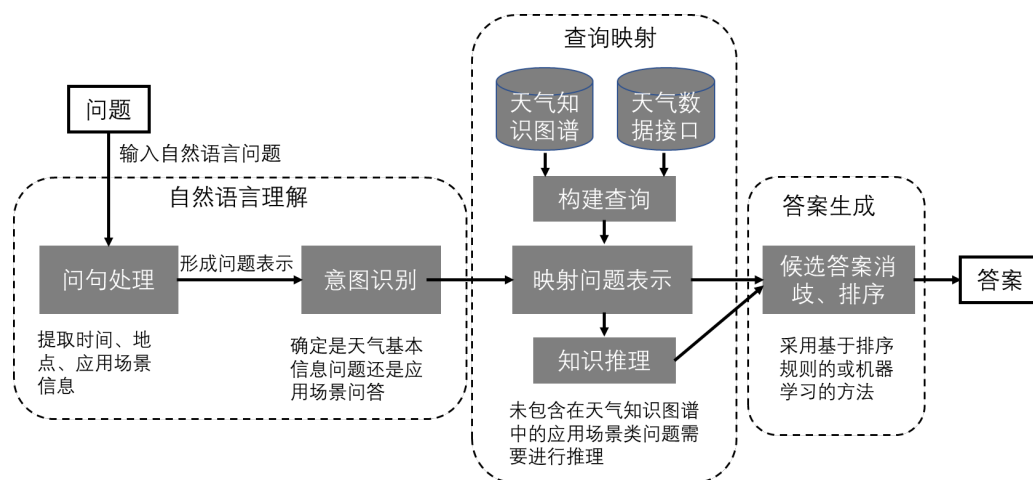


Figure 3. KBQA implementation architecture based on the weather Q&A application scenario
图 3. 基于天气问答应用场景的 KBQA 实现架构图

3.4. 实现架构设计

以天气领域应用场景为例, 根据用户输入的与天气相关的问题, 理解用户的问题意图, 从天气知识图谱数据中检索答案, 或加以一定的推理生成候选答案, 通过算法进行排序, 将最优答案反馈给用户, 包含自然语言理解、查询映射和答案生成三个核心模块。

1) 问句处理模块

该模块的主要任务是识别出问句中的天气信息词, 确定问句与天气问答相关, 然后提取与天气相关的应用场景词、地域词、时间节点词等。

2) 意图识别模块

确定是问天气基本属性类还是应用场景类问题。根据天气信息词确定咨询的是关于天气的哪一类型的信息, 根据是否有场景信息词确定问题属于哪一应用场景。

3) 映射问题表示

用户咨询的问句不一定直接对应知识图谱中的标准表示。例如, 知识图谱中存放的是气温字段, 而用户咨询的是温度, 因此要做词汇映射消歧, 并对映射天气服务接口与知识图谱中的标准进行表示。通过字符串相似度匹配和同义词表映射解决映射问题, 并进行必要的拆分和合并操作。

4) 构建查询

该模块通过对输入的问题进行处理, 将问题转化为知识图谱查询语言, 进而访问知识图谱, 通过检索获得答案。

5) 知识推理

如果问题问的是天气基本属性或知识图谱中定义的一些应用场景, 则可以从知识图谱中查找, 直接返回属性值[8]。如果询问的是未定义的天气应用场景类问题, 则需要通过推理获得答案。以明天是否需

要带伞为例, 需要构建的规则样例是: 天气状态为下雨则需要带伞, 否则无须带伞。

6) 候选答案消歧、排序

从知识图谱中查询到的答案可能不止一个, 需结合具体的领域进行对返回答案进行排序, 返回最优的答案。

7) 天气知识图谱

将整个天气问答系统看作本体, 该本体内部有多个用户查询意图, 意图之间也有层次关系, 即多级意图。如“天气查询”这个意图为一级意图, 而由这个一级意图可以延伸出多个和天气相关的话题, 即二级意图, 例如洗车咨询、晾晒咨询、防晒咨询、带伞咨询、穿衣咨询等。每个意图又有许多与之密切关联的属性和规则库中对应的规则, 例如, 在二级意图“防晒咨询”中, 一个天气对象名为 `weather`, 与其相关联的属性有:

- 天气对象的温度 `weather.temperature`;
- 天气对象的天气状况 `weather.condition`. 查询天气的时间属性 `time`。

另外, 通常还需要执行规则对意图加以约束。例如, 当最高气温高于 30 度, 天气状况为晴天, 且用户咨询天气时间在 10 点~16 点间, 是否需要做好防晒工作, 回答应为“是”。

定义好上述天气知识图谱的结构后, 天气知识图谱需要与问答系统中的其他模块产生交互, 以便生成最终的答案。天气知识图谱与其他模块交互示意图如图 4 所示。

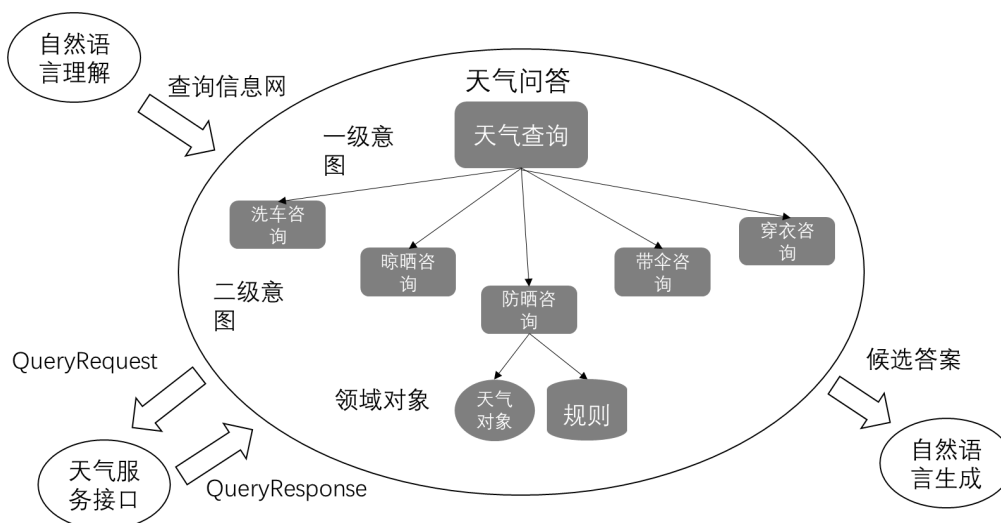


Figure 4. Schematic diagram of the weather knowledge map interacting with other modules
图 4. 天气知识图谱与其他模块交互示意图

4. 具体应用实现

具体实现过程中, 由自然语言理解模块进行查询信息词的提取(时间、地点、查询意图词), 并将提查询信息词输入天气知识图谱。并通过本体映射将用户自然语言与天气知识图谱标准定义相配套。天气知识图谱会根据用户的意图, 发送一个查询天气请求(query_Request), 并将查询响应(query_Response)返回给天气知识图谱; 天气知识图谱将意图所需天气信息及意图对应的规则输出给自然语言生成模块, 由自然语言生成模块生产候选答案并排序后, 反馈给用户。

查询输入有 4 个, 分别是问句信息词(info_Word)和用户意图(user_Intent)、外部天气数据(weather_Interface)、天气知识图谱(weather_KG)。

获得自然语言理解模块的输出后, 通过调用 `ontology_Mapping_NL()` 操作将自然语言理解模块输出的自然语言信息词映射到本体, 接着根据用户意图进行问题的分类, 根据所需天气信息产生一个 `query_Request` 请求, 发送给天气服务接口。天气服务接口会将相应的天气信息反馈给系统, 系统仍然需要进行一次 `ontology_Mapping_Ser()` 映射操作, 将天气信息映射到天气知识图谱的标准表示上, 而后结合意图规则, 推导出候选答案。

查询映射模块获得候选答案后, 将候选答案作为答案生成模块的输入。候选答案在该模块中要经过消歧、打分排序等操作, 然后系统将获得唯一的答案。最后, 通过 `transform_NL()` 将最终答案转化为用户可以理解的自然语言表示。

5. 总结

本文在分析了客服领域知识的特点后, 选择构建领域本体库知识库来结构化的存储客服聊天机器人需要的领域知识, 并结合本体构建了基于公共服务的业务知识库客服聊天机器人系统, 提升了聊天机器人的多轮对话能力。仿真实验证明, 以本体知识体系为基础的问答系统, 在融合智能搜索引擎, 并支持内外部多渠道应用后, 真正实现了知识积累有序化、知识应用智能化、知识管理统一化、知识展现个性化, 满足特定场景的多渠道知识问答应用需求。

基金项目

贵州省科技计划项目(黔科合人才[2015] 4011、黔科合平台人才[2016] 5103、黔科合平台人[2017] 5788)。

参考文献

- [1] 佚名. 智能机器人服务环境下传感数据映射及服务组合方法研究[J]. 机器人, 2019, 41(1): 30-39.
- [2] Peters, M.E., Neumann, M., Iyyer, M., *et al.* (2018) Deep Contextualized Word Representations. arXiv:1802.05365.
- [3] Devlin, J., Chang, M.W., Lee, K., *et al.* (2018) BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding. arXiv:1810.04805.
- [4] 佚名. 中文基本复合名词短语语义关系体系及知识库构建[J]. 中文信息学报, 2019, 33(4): 20-28.
- [5] Zhu, Y., Wan, J., Zhou, Z., *et al.* (2019) Triple-to-Text: Converting RDF Triples into High-Quality Natural Languages via Optimizing an Inverse KL Divergence. *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, Paris, 21-25 July 2019, 455-464. <https://doi.org/10.1145/3331184.3331232>
- [6] Abujabal, A., Yahya, M., Riedewald, M., *et al.* (2017) Automated Template Generation for Question Answering over Knowledge Graphs. *Proceedings of the 26th International Conference on World Wide Web*, Perth, 3-7 April 2017, 1191-1200. <https://doi.org/10.1145/3038912.3052583>
- [7] Chen, Z.Y., Chang, C.H., Chen, Y.P., *et al.* (2019) UHop: An Unrestricted-Hop Relation Extraction Framework for Knowledge-Based Question Answering. arXiv:1904.01246.
- [8] Lima, E., Shi, W., Liu, X., *et al.* (2019) Integrating Multi-level Tag Recommendation with External Knowledge Bases for Automatic Question Answering. *ACM Transactions on Internet Technology (TOIT)*, **19**, 1-22. <https://doi.org/10.1145/3319528>