

知识图谱在问答系统中的应用综述

刘俐婷, 师文轩, 程书芝

(南开大学软件学院, 天津 300350)

摘要: 随着互联网文档万维网向数据万维网的转型, 各大搜索引擎公司纷纷构建知识图谱。为更好地理解知识图谱的内涵及其利用价值, 对相关领域的研究进行汇总和梳理。首先, 介绍问答系统的起源发展和知识图谱的概论。然后, 分别从概念层次关系学习、命名实体识别与实体链接、事实知识学习和事件知识学习四个方面的信息抽取技术介绍知识图谱构建技术。最后, 将知识图谱应用到问答系统中, 得出知识图谱的构建有望有效地扩大知识库的规模、给问答系统的进一步研究提供新方向的结论。然而, 知识图谱的发展机遇与挑战并存, 仍需进一步研究与积累。

关键词: 计算机应用; 知识图谱; 综述; 问答系统; 信息抽取; 深度学习

中图分类号: TP391.1 **文献标识码:** A **文章编号:** 1674-2850(2017)10-1052-16

Review of the application of knowledge graph in question answering system

LIU Liting, SHI Wenxuan, CHENG Shuzhi

(College of Software, Nankai University, Tianjin 300350, China)

Abstract: With the transition of the Internet from the world wide document web to the world wide data web, major search engine companies turn to build the knowledge graph. This paper summarizes the researches in the field of knowledge graph in order to make a better understanding of the connotation of knowledge graph and its utilization value. Firstly, the origin and development of question answering system and the generality of knowledge graph are introduced. Then, the techniques of building knowledge graph by information extraction techniques are presented. Information extraction techniques are divided into four aspects, including concept hierarchical relationship learning, named entity recognition and entity linking, factual knowledge learning and event knowledge learning. Finally, knowledge graph is applied to question answering system, and it can be concluded that the knowledge graph is expected to expand the scale of knowledge base effectively and provide a new orientation for the further study of question answering system. However, the development of knowledge graph has both opportunities and challenges, which calls for further researches to gain a deeper understanding on knowledge graph.

Key words: computer application; knowledge graph; review; question answering system; information extraction; deep learning

0 引言

问答系统是信息检索的一种高级形式, 以其简单、准确的互动方式深受人类青睐, 在人工智能的大时代下, 问答系统的智能化也成为研究热点。

基金项目: 高等学校博士学科点专项科研基金 (20130031120042)

作者简介: 刘俐婷 (1993—), 女, 硕士研究生, 主要研究方向: 数据挖掘、机器学习等

通信联系人: 师文轩, 副教授, 主要研究方向: 数据挖掘、移动计算、软件工程等. E-mail: shiwx@nankai.edu.cn

第一个问答系统诞生于 20 世纪 60 年代, 一般认为是 WEIZENBAUM^[1]设计实现的“ELIZA”——一款聊天机器人, 通过反问方式引导说话, 用于对精神病人的心理治疗。在知识库中, 采用模式匹配的方式寻找答案。

20 世纪 70 年代, 一些专家系统的出现标志着问答系统中新的里程碑, 如 WOODS^[2]设计的 LUNAR 系统, 是一个专门回答有关阿波罗登月返回的月球岩石样本地质分析问题的自然语言信息检索系统; WINOGRAD^[3]建立的一个用自然语言指挥机器人动作的人机对话系统等。由于专家系统专业范围的严格限制, 使特定领域发展的成果惊人, 也成为了通用领域发展的桎梏。

20 世纪 80 年代, LENAT 等^[4]发起的 Cyc 研究项目, 将问答系统带入“知识原则”的时代, 在知识库的基础上进行自然语言的理解、学习及问答研究。20 世纪 90 年代, 国内的许多研究者也开始知识库的建设研究工作^[5]。

20 世纪 90 年代, 麻省理工学院 (Massachusetts Institute of Technology, MIT) 人工智能实验室开发出一个优秀的 Start (<http://www.ai.mit.edu/projects/infolab/start.html>) 问答系统, 包含能给出快速且准确回答的 2 个数据库, 同时也包含 1 个搜索引擎检索数据库外的知识数据库。

2014 年以来, 互联网开始了新的转型。由于 Linking Open Data 等项目的开展, 互联网悄然从仅包含网页之间超链接的文档万维网转变为包含大量描述实体间丰富关系的数据万维网。如今的几大主流搜索引擎公司 Google、百度和搜狗等借着互联网转型的东风, 开展了知识图谱的构建研究。因此, 对于问答系统的发展, 本研究更看好知识图谱作为知识库的新突破是如何应用到问答系统中的。

科学知识图谱^[6]将复杂的科学知识领域通过数据挖掘、信息处理、知识计量和图形绘制, 以可视化的方式显示科学知识的发展进程与结构关系, 揭示科学知识及其活动规律, 展现知识结构关系与演进规律。构建知识图谱, 突破知识库发展瓶颈成为这场变革的重中之重。本文从构建知识图谱的技术开始综述, 进而展望其在问答系统中的应用。

1 知识图谱概览

知识图谱的概念于 2012 年 5 月由 Google 率先提出, 目的是将搜索结果进行知识系统化, 让每一个关键字都拥有一个完整的知识体系, 从而提高搜索质量。Google 的辛格博士曾说: “The world is not made of strings, but is made of things”, 知识图谱旨在描绘真实世界中的各种实体或概念, 因此, 也使检索真正实现了基于内容的查询。

知识图谱的概念模型是构建知识图谱的基础。1) 概念: concepts (classes/types), 具有相同属性的一组对象; 2) 实例: instances (entities/objects), 属于某个概念中的一个对象; 3) 关系: taxonomy knowledge (hyponymy/hierarchy/subclassof), 分类知识; 4) 属性: properties (attributes), 对象拥有的特征, 属于事实知识共同构成知识图谱的概念模型。有观点认为知识图谱表示的概念模型和逻辑基础可看作本体, 即哲学中的本体^[7]。通过语义网络 (目前已定义网页信息的扩展) 的形式表达出来, 从而使人类和电脑能更好地合作。综上, 明晰知识图谱的原理, 利用数据万维网提供的共享知识和重用知识的平台, 借鉴互联网信息内容的语义描述的基础、人工智能的知识表示 (例如符号主义) 方式, 实现对客观世界从字符串描述到结构化语义描述, 进而对客观世界的知识映射。图 1 为知识图谱的结构示意图。

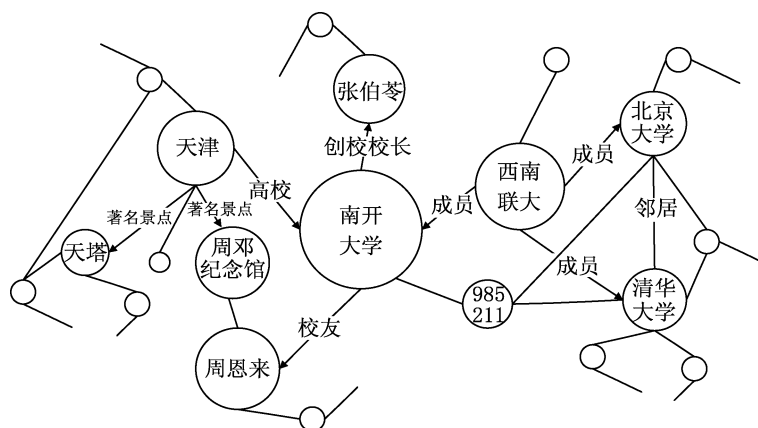


图1 知识图谱结构示意图

Fig. 1 Structure diagram of the knowledge graph

2 知识图谱构建技术

构建知识图谱最重要的两点是数据和信息抽取。

为满足对话搜索和复杂问答的新要求，知识图谱中不仅需要包括大量高质量的常识性知识数据，还需要及时补充和更新。数据的更新通常会选择如下几个方式。

1) **Wikipedia**: 现在公认最大的在线百科全书，其文本为半结构化形式，在百科类站点中结构化程度最为规整，并采用协同编辑实时更新与修正。

2) **Freebase**: 类似 Wikipedia 允许创作共享的百科类数据源，其完全采用结构化形式，规模远大于 Wikipedia。在 Google、百度、搜狗自建的知识图谱中也占有举足轻重的地位。

3) **YAGO**: 通用语义数据集。一个大型的语义数据集，整理、规范和结合来自维基百科、WordNet 和 GeoNames 的数据资源。

4) 搜索日志、点评网站、社交网站等一些特定领域的知识库、众包反馈等。

知识图谱的构造与信息抽取也密不可分。信息抽取是从自然语言文本中抽取指定类型的实体、关系、事件等事实信息，并形成结构化数据输出的文本处理技术，目的在于使信息更加机器可读化，即将文本中的信息进行结构化处理，以统一形式将信息集成起来形成巨大的数据库。

信息抽取^[8]的研究也有很长的历史。1987年至1997年，美国国防高级研究计划局(Defense Advanced Research Projects Agency, DARPA)资助了七届消息理解研讨会(Message Understanding Conference, MUC)，该会议特点是对信息抽取的评测中，在线与离线文本数量呈几何级增长，并且推动了20世纪80年代末的蓬勃发展。

1999年至2008年，由美国国家标准与技术研究所(National Institute of Standards and Technology, NIST)主办的自动内容抽取评测会议(Automatic Content Extraction, ACE)，旨在开发支持3种不同来源的语言文本(普通文本、由自动语音识别得到的文本、由光学字符识别得到的文本)内容的自动抽取技术。由于ACE评测不针对具体的领域场景，而是对系统跨文档处理能力进行评测，采用基于漏报、误报为基础的评价体系，进而把信息抽取技术研究引入至新的高度。

从2009年沿用至今的国际文本分析会议(Text Analysis Conference Knowledge Base Population, TAC-KBP)子任务，其评测会议一直推动着相关领域的研究与发展。

2.1 概念层次关系学习

抽取的主要思路是确定类与子类的关系, 考虑不同维度的多维分类问题。以基于启发式的规则和基于分布式的规则为例, 前者可以获得较高质量的候选对概念, 但覆盖率有限, 使用维基百科和大规模网页库也只是起到部分改善的作用; 后者不需要预知可能具有上下位关系的词对, 而是通过利用概念的分布特征来计算上下位关系, 但准确率低。因此, 两种方式各有利弊。

2.1.1 基于启发式规则的方法

启发式技术=启发式扫描+启发式监控, 重点在于更新特征值识别技术、解决单一特征码比对的缺陷。该技术的核心在于反编译: 计算机软件反向工程 (reverse engineering), 推导出他人的软件产品所使用的思路、原理、结构、算法、处理过程、运行方法甚至是源代码。探索了解反编译后的结果, 学习并产生启发。知识图谱的构建思路描述如下。

1) 从大规模资源中, 根据上下位概念的陈述模式, 找出可能具有上下位关系的概念对。

2) 层次树 (一种语义描述) 归纳。例如:

①Wikipedia 中学习的思路^[9]是

a. 将 Wikipedia 中的 category 当作概念。

b. 利用 Wikipedia 中条目下所属的分类, 初始化上下位关系。

其优点是大规模的概念和概念层次关系生成自动化; 缺点是概念覆盖不全、存在噪声等, 通常需要采用过滤、迭代、引用外部资源的方式进行改善。

②结合 Wikipedia 的学习, 如 YAGO 利用 WordNet 中良好的上下位关系定义概念层次结构, 联合使用 Wikipedia 概念作为叶子, 进行分类挂载。

③在非 Wikipedia 资源网页中学习, 大规模概念结构的学习, 概念结构这类数据库也存在不少, 常被应用的有 Freebase、Probase. 其中, 作为 Bing 的法宝, Probase 有自己独特的优势。

a. 自动学习规模大, 有 16 亿网页。

b. 概念的层次结构规模大, 可达 200 多万。

c. 概率概念结构在关系抽取上优势明显 (Probase 中的数据, 就像人脑中的思想, 不是简单的非黑即白, 会量化不确定性)。

d. 没有子节点的概念可认为是实例。

建立一系列符合解决实际问题的启发式规则, 也就形成了启发式算法。20 世纪 40 年代已经提出启发式算法, 50 年代发展鼎盛, “贪婪算法和局部搜索” 受到关注, 到 60 年代发现了不足: 速度快, 但不能保证解的质量, 而且对大规模的问题收敛速度慢。在构建小型知识图谱时, 仍有很多基于启发式规则^[10]建立选择的实例, 然而随着数据以惊人的速度增长, 该方法显得力不从心。但人们并没有停止研究启发式算法的步伐, 如今已有模拟退火算法 (simulated annealing algorithm)、人工神经网络 (artificial neural network)、禁忌搜索 (tabu search)、演化算法 (evolutionary algorithm)、蚁群算法 (ant algorithms) 等。

2.1.2 基于分布式规则的方法

该方法通常在一个给定的实体集合和大规模文档集合中, 利用词或实体的分布相似性, 自主建立实体集合的概念结构。基本思路^[11-12]如下。

1) 获取实体的各种分布特征, 计算其间关系。

2) 确定概念层次结构的特征度量方法。

3) 构造分类 (Taxonomy) 结构。例如:

①基于概念结构度量的分类学习, 其特点为

a. 使用大量对找出有益于上下位关系特征的概念 (如: 上下文特征、术语共现、语法从属特征、重复定义等), 给每个概念对定义信息函数。

b. 采用增量层次聚类。增量式是针对数据变化的部分增量地更新挖掘结果, 这样若能充分利用前一次挖掘结果, 挖掘效率就会提高。层次聚类不再是一个单一聚类, 而产生了一个聚类层次。层次聚类算法就是产生了一个嵌套聚类的层次, 在执行步骤 t 时, 在 $t-1$ 步的聚类情况下产生一个新聚类。

c. 概念结构学习转化为多标准优化问题。

②基于隐变量的分类学习, 其特点为

a. 从公众分类法 (folksonomy, 随互联网发展出现的一种新的分类方法, 典型的代表网站是 Del.icio.us, 由公众为信息贴加标签) 中建立实体集合的概念结构。

b. 使用话题模型计算标签间的关系。

c. 4 种计算分类的标签距离度量方法 (tag-divergence, hypernym-divergence, merging-divergence, keep-divergence)。

③根据信念传播 (belief propagation, BP) 算法特点, 学习局部间的联系, 得到全局最优解, 既可减少运算时间, 又能提高分类质量。基于信念传播的分类结构学习^[13]的优点为

a. 不需要用 pattern 找上下位概念对。

b. 同时考虑上下位关系、兄弟节点关系和整个树结构。

c. 非增量的分类归纳的概念模型。

基于特征分布进行分类的研究很多, 该方法对于非均衡语料、特别是稀有类别的判断存在显著的优势。例如: ZHENG 等^[14]提出的显示近似最优组合正/负特征, 徐燕等^[15]提出的在较少类别中出现的词条具有较好的区分类别能力等, 但该方法应用于构建知识图谱计算上下位的关系中, 准确率较低、效果不佳。

2.2 命名实体识别与实体链接

本节主要从命名实体识别和实体链接的角度构建知识图谱。

2.2.1 命名实体识别

首先, 明确命名实体识别的任务: 待处理文本中, 识别出 7 类命名实体 (人名、机构名、地名、时间、日期、货币和百分比); 其中包含 2 个子任务——实体边界识别和确定实体类别。

根据任务需求与识别特点形成的识别方法:

1) 命名实体的内部构成和外部语言环境具有的特征;

2) 命名实体所在的上下文特征和实体内部特征;

3) 相应类别实体识别模型;

4) 计算特征权重的序列标注工具, 如最大熵马尔可夫模型 (max entropy Markov model, MEMM)、隐马尔可夫模型 (hidden Markov model, HMM)、条件随机场 (conditional random fields, CRF)。

采用以上识别方法的结果于 2006 年在国际会议上得到了评定, 识别英语的最好结果为准准确率 95%、召回率 92%, 该结果还算理想, 已经具备了相当程度的大规模文本处理能力。而识别汉语最好的结果 (准确率, 召回率) 对于人名、地名、机构名分别为 (66%, 92%)、(89%, 91%)、(89%, 88%)。

第一代的实体识别中, 汉语的识别效果并不理想, 在实际应用中, 效果性能甚至还会大打折扣。回

顾第一代, 训练语料的规模和类别数, 一方面限制了识别的性能, 由于训练语料网页信息的不规范和噪音干扰, 性能剧烈下降; 另一方面也限制了类别数, 使训练语料不能在实际应用中达到理想的效果。

开放域实体抽取的提出使实体识别有了新的发展方向: 其可识别实体类型更多、粒度更细, 甚至可识别随着时间更新的实体概念。开放域实体识别既不限定实体类别, 也不限定目标文本。给定某一类别的实体实例, 就能从网页中抽取同一类别的其他实体实例。从几个实例分析并总结其主要方法。

1) 利用查询日志 (query log) 进行开放域实体抽取:

①以上下文为种子实例, 分析得到模板, 进而找到其他同类别实例;

②构造候选向量、种子上下文向量, 计算相似度。

2) 利用网页文档 (web page): 若处理列表型网页, 种子与目标实体有相同的网页结构。以此为推动, 总结网页信息的开放域实体抽取 3 个模块。

①爬取模块 (fetcher): 种子送到搜索引擎, 抓取网页作为语料。

②抽取模块 (extractor): 单个网页学习模板, 进而抽取候选实例。

③排序模块 (ranker): 利用种子、网页、模板、候选构造一个图, 综合考虑网页 (一个网页包含的高质量模板越多, 该网页质量越高, 反之亦然) 和模板 (一个模板抽取出的正确实例越多, 该模板的质量越高, 反之亦然) 的质量, 使用随机游走 (random walk) 算法为候选打分并排序。

3) 多数据源融合进行实体抽取: 分别对不同数据源选取不同特征进行实例扩展, 对结果进行融合, 不同数据选取不同的模板和特征, 同时计算不同特征候选的置信度。实体抽取的数据源倾向于多样化、多源化发展。找到种子词与目标词之间的联系 (相同或类似的上下文), 由种子词生成模板, 从模板中找到更多同类实体。

开放域实体抽取的思想不再是单纯的召回, 而是进行实例扩展。针对不同数据源的特点设计方法, 其针对性、灵活性更强; 通常包括模板抽取和计算实例候选置信度两个模块, 两部分迭代进行、相互依赖。以无监督的方法为主, 但缺少公认的数据集和相关评测。随着研究的深入, 新的问题也凸显出来, 比如命名实体的歧义。几种基于聚类的实体消歧如下。

1) 以实体对象为聚类中心, 所有实体指称项按其指向的目标实体进行聚类。

2) 每一个实体指称项对应到一个单独的类别。

①词袋模型:

a. 利用待消歧实体周边的词来构造向量;

b. 利用向量的余弦相似度, 聚类到最相近的实体类下。

②社会网络模型:

a. 每个人都有自己的社会关系, 物以类聚、人以群分;

b. 利用社会网络分析技术, 根据不同人的社会化关联信息所表现出来的网页链接特征, 计算节点间的拓扑距离, 实现人名聚类消歧。

③百科网站模型:

a. 页面超链接关系;

b. 条目间的关联关系反映语义相关度;

c. 用实体上下文的条目对实体进行向量表示;

d. 利用条目间的相关度计算指称项间的相似度, 以解决数据稀疏的问题。

④多源异构知识: 挖掘和集成多源异构知识, 解决单一知识源覆盖度有限的问题, 提高消歧性能。

多源（不同类型、不同结构、不同领域的数据来源的融合）包括：

- a. Wikipedia，捕捉概念间的语义关联；
- b. WordNet，捕捉词语间的语言学关联；
- c. Web 网页库，捕捉命名实体间的社会化关联。

异构^[16]：语义图表示框架。同时捕捉显示语义知识（语义图的边：所有知识源中概念间的显示语义关联）和结构化语义知识（语义图的结构：概念间的隐藏语义关联）。存在一个传递性计算原则：若邻居概念与另一个概念存在语义关联，则这个概念也与另一个概念存在语义关联。

对于命名实体识别抽取还面临着众多的挑战，面向开放域信息抽取方法的准确率和召回率与面向封闭域信息抽取存在一定差距。面向开放域的信息抽取算法的研究还处于起步的阶段，要实现算法由量到质的飞跃，还需要进一步研究与积累。

2.2.2 实体链接

实体链接^[17]是给定实体指称项和它所在的文本，将其链接到对应知识库的正确实体对象的操作。包含主要步骤为候选实体的发现和候选实体的链接，如图2所示。

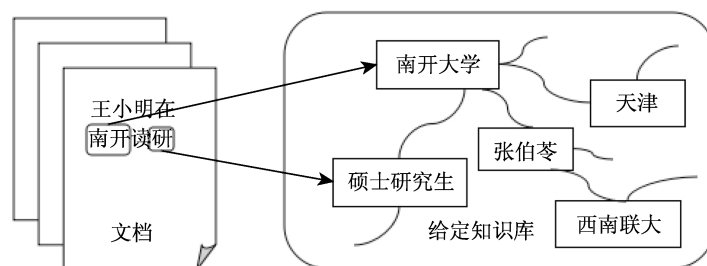


图2 实体链接结构示意图

Fig. 2 Structure diagram of the entity linking

步骤1是通过给定实体指称项，链接系统根据知识、规则等信息找到实体指称项的候选实体。Wikipedia的信息和上下文信息可作为数据源。

1) Wikipedia:

- ①Wikipedia中锚文本的超链接关系；
- ②Wikipedia的消歧页面^[18]；
- ③Wikipedia中的重定向页面来获取候选实体。

2) 上下文：通过定位全称来解决缩略语指称项的强歧义性。

步骤2则是系统根据指称项与候选实体间的相似度等特征，解决实体消歧、共指消解的问题，选择实体指称项的目标实体。

其发展经历了单一实体链接向协同实体链接发展的过程，目的是更有效地挖掘实体指称项信息，更准确地计算实体指称项与实体概念之间的相似度。

1) 单一实体链接：

①基于词袋子模型计算相似度，用词袋子向量形式表示上下文文本（实体指称项、候选实体），计算向量间的夹角确定相似度，选择相似度最大的候选实体进行链接。

②加入候选实体的类别特征，为改善候选实体文本内容太短小导致相似度计算不准确的问题，以Wikipedia中的超链接为训练数据，以文本相似度、指称项文本中词与候选实体类别的共现信息为特征，

训练 SVM 分类器对候选实体进行选择。

③加入候选实体的流行度等特征, 综合考虑候选实体的背景知识和先验信息。

2) 协同实体链接: 既要计算实体指称项与目标实体的语义相似度, 又要计算目标实体间的语义相似度, 因此采用排序学习 (Pairwise) 算法优化策略。以下介绍几种协同链接的具体方法。

①基于图的协同链接: 由于 Pairwise 策略只考虑两两实体间的关系, 结果并非全局最优。全局考虑目标实体间的语义关联, 采用图方法构成两种关系指称项图 (referent graph)。

a. 指称项与实体间的关系, 该指称项文本与实体文本的相似度由传统的向量空间模型 (vector space model, VSM) 得到。

b. 实体间的语义关系, 利用目标实体间的链接关系计算实体间的语义相关度。

例如, HAN 等^[19]构建包含指称-实体、实体-实体关系图, 提出有效的协同推理算法。HOFFART 等^[20]在图的基础上, 利用边加权计算稠密子图, 得到映射效果。

②基于深度学习的方法: 传统方法中, 由于没有考虑各个概念间的内在联系, 计算待消歧实体上下文和目标实体语义相似度的方法可扩展性之差。协同协作的方法中, 提出利用深度学习的方法自动联合学习实体和文档, 进而完成实体链接任务。

③跨语言实体链接: 由于传统方法要先翻译成目标语言, 过程中会产生错误而传递风险, 需要大量句子级平行的双语训练预料。双语隐含主题模型, 将实体指称项与候选实体映射到同一个主题空间, 给定一种语言的实体指称项及其所在的上下文, 将其链接到另外一种语言的知识库中。例如, WANG 等^[21]提出用链接因子图模型挖掘百度百科与 Wikipedia 的实体关联, 效果良好。

④结构化数据实体链接: 结构化数据没有上下文, 消歧利用实体的流行度和实体共现。

⑤社交数据实体链接: 社交媒体 (Twitter) 作为重要信息来源。由于社交媒体上下文短、不规范, 因此需要综合考虑用户信息和交互信息。

2.3 事实知识学习

事实知识的抽取包括两种: 实体-属性-值、实体-关系-实体。根据实体间共享事实知识而形成彼此之间的联系, 从而构建知识图谱。

需要注意的是, 网络文本的信息结构有 3 种: 置信度高、规模小但缺乏个性化属性信息的结构化数据, 置信度较高、规模较大、有个性的信息但形式多样、含有噪音的半结构化数据和置信度低、复杂多样但规模大的纯文本数据。由于结构化和半结构化数据的规范化, 可用于网页结构进行信息块的识别、模板的学习以及属性值的抽取, 从而实现属性和关系抽取。由于属性也可被看作一种名词性关系, 就可以将属性抽取问题也抽象为一种关系抽取问题。因此, 非结构化文本的实体关系抽取成为本节的重点, 以下分别从传统关系抽取、开放域关系抽取和开放域关系发现进行具体方法的介绍。

2.3.1 传统关系抽取

目前, 传统关系抽取主要采用统计机器学习的方法, 将关系实例转换成高维空间中的特征向量或直接用离散结构表示, 在标注语料库上训练生成分类模型, 然后再识别实体间关系。

1) 基于特征向量方法: 获取各种有效的词法、句法、语义等特征, 并有效地集成起来, 产生描述实体语义关系的各种局部特征和简单的全局特征。其中, 特征选取从自由文本及其句法结构中抽取各种表面特征及结构化特征。优点是实用、计算速度快, 缺点是挖掘有效性有限、再提高不易。

2) 基于核函数方法: 需要人工提供大量标注语料作为训练集的有监督学习方法。优点是有效挖掘结

构化信息特征，缺点是计算速度缓慢、实用性差。其中，核代表卷积树核，用两个句法树之间的公共子树的数目来衡量相似度。

①标准的卷积树核函数 (convolution tree kernel, CTK)，在计算两个子树相似度时，只考虑子树本身，不考虑子树上下文信息。

②上下文相关卷积树核函数 (context-sensitive convolution tree kernel, CS-CTK)，在计算子树相似度时，同时考虑子树的祖先信息，并对不同祖先的子树相似度进行加权平均。

3) 基于神经网络的方法：设计合理的网络结构，捕捉更多信息，提升关系抽取的准确度。该方法属于一种深度监督学习下的机器学习模型 (深度学习)，通过训练关系分类器来判别句子中实体之间的语义关系。可规避传统特征抽取需要的自然语言处理 (natural language processing, NLP) 预处理和人工设计特征产生的错误。

不同的网络结构捕捉不同的信息：

①递归神经网络 (recurrent neural network, RNN)，构建过程更多考虑句子的句法结构，需要依赖复杂的句法分析工具。

②卷积神经网络 (convolutional neural network, CNN)，通过卷积操作完成句子级信息的捕获，无需复杂的 NLP 工具。

2.3.2 开放域关系抽取

开放域关系抽取的特点是不限定关系类别和目标文本，也因此产生相应的难题，获取训练语料也是其中之一。

1) 针对网页，采用一种 TextRunner (传统信息抽取系统专注于在小而均匀的语料中抽取少量精确狭窄预先要求的信息，采用全新的开放信息抽取方式可实现在无人工输入的情况下，抽取大量相关元组集合) 的方式：

- ①用户输入特定的谓词和论元；
- ②搜索引擎返回与用户输入相关的句子；
- ③用 TextRunner 抽取谓词论元三元组。

2) 针对在 Wikipedia 文本中抽取关系信息，采用一种机阅读 (machine reading)^[22] 的方式，其具体方法如下：

- ①Infoboxes (维基的信息框) 抽取关系信息；
- ②维基条目文本中进行回标，产生训练语料。

3) 从纽约时报 (NYT) 中抽取 Freebase 的关系类别，比较多个数据库发现，Freebase 的实体类别和关系类别最为丰富：

- ①人工标注训练集不可行，需要寻求无监督或弱监督的关系抽取方法；
- ②将弱监督关系抽取看作多示例问题；
- ③利用分段卷积网络的自动学习特性，设计分段最大池化层，更好地保留句子的结构化信息。

4) 基于细粒度实体类型特征发现的弱监督关系抽取，强化粗粒度实体类型特征对特定关系类别的指示作用。其具体方法为

①细粒度实体类型，将实体类型分类看成文本分类问题，利用搜索引擎扩展实体指称，再用预先训练的分类器，分类扩展实体指称文本。

②3种融合方法, 替换、扩展和选择。

2.3.3 开放域关系发现

关系发现在知识图谱中也称为链接预测: 利用知识图谱中现有的知识推断未知的知识; 知识图谱本身就蕴含推理模式。

从知识表示的角度, 关系发现可以分为符号逻辑表示和分布式表示。符号逻辑表示表达能力强、理解度高, 还可提供精确的结果。其关键技术是事实的抽取、规则的学习和事实在规则上的推理。其中, 逻辑推理规则学习包括统计关系路径的共现学习霍恩子句表示的推理规则, 以及实体间在图中的链接特征学习关系分类器, 以得到路径与关系的推理规则。

但是随着知识库的规模越来越大, 逻辑表示很难高效地扩展到大规模知识库上, 而分布式表示学习可用线性参数来表示指数数量的区域, 将知识库中的实体和关系表示为低维空间的对象(向量)及其操作(空间转换), 能蕴含其在知识库中的性质, 特点为每个实体和关系的表示是通过优化整个知识库的目标函数编码得到的, 凝练了整个知识库的信息, 能够进行全面的知识推理, 拥有一个丰富的相似度空间; 是一个统计学习过程, 事物间的隐式规律蕴含在对象表示中, 反映在推理过程中。

根据分布式知识表示方式的特点, 知识库表示学习的方法有以下几种:

1) 张量分解法^[23], 将知识图谱编码为一个三维邻接张量, 然后分解为一个核心张量与因子矩阵的乘积, 分解后的结果作为对应三元组存在与否的概率。

2) 基于翻译的模型^[24], 将关系看作头部实体到尾部实体的翻译。

3) 基于神经网络的能量函数, 为三元组打分, 通过惩罚错误的三元组完成学习过程。

4) 基于稀疏矩阵的知识库补全方法^[25], 用稀疏矩阵替换转换矩阵。在使用稀疏矩阵的前提下, 可根据不同关系的复杂度调节转换矩阵的自由度, 这种做法对数据有很强的适应性。

5) 基于高斯分布表示学习的补全方法^[26], 针对符号不确定性的问题, 实体用粒度大小表示, 关系用指示性强弱表示; 针对实体不确定性的问题, 包含事实越多, 含义不确定性越小; 针对关系不确定性的问题, 关系越复杂, 含义不确定性越大。用多维高斯分布表示, 均值向量表示该符号的位置及含义, 协方差矩阵表示该符号的多样性及不确定性。

6) 基于动态映射矩阵的补全方法^[27], 通过对事实中的实体和关系构造的动态映射矩阵, 表示知识的多样性, 主要解决“一对多、多对一、多对多”类型的关系。

前三种方法中, 所有关系的转换矩阵均有相同的自由度和自由变量数目, 且头部和尾部实体共享同一个转换矩阵。然而, 知识库中关系存在: 异质性, 关系链接很多实体对(复杂关系)及关系链接很少的实体对(简单关系); 不平衡性, 关系链接很多头部(尾部)实体及关系链接很少尾部(头部)实体。

不同复杂度的关系应当使用不同表达能力的模型学习, 头部与尾部实体应分开学习; 使用不同稀疏度的矩阵学习不同复杂性的关系。分布式表示和逻辑表示融合的混合方法, 融合分布表示概率图模型。将表示学习和概率推理融合^[28]: 将知识图谱中的元素进行向量化的表示; 用马尔可夫逻辑网学习推理规则; 在事实推理中, 将量化中得到的候选元组的相似度值作为随机游走的状态转移概率注入逻辑推理中, 减少搜索空间、提高效率。

一些具体操作的例子值得借鉴。对结构化或半结构化文档: 如 COHEN 等^[29]提出的方法 WR、KOSALA 等^[30]提出的树自动机、TENIER 等^[31]提出的利用最小子树、ZHAI 等^[32]基于树匹配的部分匹配方法、ÁLVAREZ 等^[33]利用聚类技术以及 HONG 等^[34]和 TANG 等^[35]基于机器学习两阶段语义标注的方法、iASA 语义标注方法。对纯文本自由结构的文档: MeatAnnot 系统、Armadillo 方法、KnowItAll 方法、PAN-KOW 方法、基于

Cyc 的自动语义标注系统、KIM 平台、iOkra 框架等。其中，均包含一些自动或半自动的知识抽取技术，且英文抽取技术已经较为成熟且效果良好，由于中文自身的特点具有一定的挑战，中文抽取技术仍具有很大的研究空间。

2.4 事件知识学习

事件抽取是针对特定领域的事件进行相关信息的抽取。一个事件包括时间、地点、原因、实体行为等多个因素。与关系抽取相比，另外还需要一个事件触发词和更多的论点个数，需要预测的目标也不再是两个概念间的语义关系，而是触发词及事件类型、事件元素及其扮演的角色。

事件抽取的典型方法如下。

1) 基于模式匹配的方法：

①平面模式主要基于词袋等字符串特征；

②结构模式更多地考虑句子的结构信息，融入句法分析特征。

存在的缺点有：领域相关度高导致可扩展性差，需要大量的人工标注，耗时耗力。模式匹配的一般过程如图 3 所示。

2) 基于机器学习的方法（从特征和结构两个角度）：

基于特征的方法是将事件抽取看成一个多分类的问题。

①利用句子级信息，从句子中提取特征，利用最大熵、朴素贝叶斯和支持向量机等模型去完成事件抽取，代表方法有基于动态最大池化技术的卷积神经网络（dynamic multi-pooling convolutional neural networks, DMCNN）^[36]。

②4 部分组成：嵌入词学习（word-embedding learning）、词本身特征代表（lexical-level feature representation）、句子上下文特征提取（sentence-level feature extraction）和论点分类输出（argument classifier output）。

③传统的 CNN 利用最大池化技术，选取最大值代表每个特征映射，但事件抽取任务中一句话存在 2 个或多个事件。提出动态最大池化技术，根据触发词和论点候选的位置，在不损失最大值的前提下，获取更多的有用信息。

④利用篇章级信息^[37]，在句子级信息基础上更多地考虑篇章级信息和丰富的背景知识（跨文档信息、跨语言信息、跨文本事件信息、跨实体信息）。代表方法有基于概率软逻辑推理^[38]、可利用全局信息（例如事件之间的相关性）和更精确的局部信息（隐含的局部信息）。

a. 局部模块，基于局部信息做出初步分类，会考虑细粒度实体类别等精准信息。

b. 全局模块从语料中收集全局信息，学习到事件与话题间、事件与事件间的共现信息。

c. 综合以上信息进行全局推理。

通常，按照触发器检测（trigger detection）、论点分类（argument classification）、属性分类（attributes classification）和上报分类（reportability classification）的顺序进行多分类。然而将事件抽取分成多个步骤会导致错误由上而下传递，没有考虑触发词和时间元素间的相互影响。

基于结构的方法：将事件抽取成一个最优结构预测。

①看作依存树结构预测问题。

②自定义联合结构的预测问题，触发词和事件元素联合预测。一个词存在歧义，同时考虑事件元素和事件触发词的标注会提升最终效果。

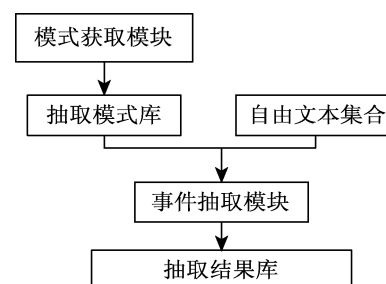


图 3 模式匹配过程示意图

Fig. 3 Pattern matching process diagram

基于事件的抽取也是自然语言处理领域中重要的研究方向之一, 抽取较高级的任务信息是热点也是难点。该思路的提出是为实现自动将非结构化的文本结构化, 方便进一步利用信息。事件抽取对于突发事件^[39]、灾难性事件^[40]的研究相对较多且准确, 还擅长进一步追踪分析。因此, 提高新类型事件的抽取能力是新的研究挑战与目标。

3 知识图谱应用问答系统

问答系统^[41]的发展经历了长期的探索与改进, 现有的问答系统分为聊天机器人、基于知识库的问答系统、问答式检索系统和基于自由文本的问答系统。其中, 前三者均受制于知识库规模的大小, 被知识库覆盖的知识, 问答系统能高效而准确地给出答案, 反之性能骤减为零。而后者因为避免建立大规模知识库而被看好, Cymfony 公司的 Textract 系统成功代表着问答系统的发展方向。

知识图谱的诞生使知识库的规模有量的提升。据不完全统计, 目前 Google Graph 包含了 5 亿实体和 35 亿条事实, 同时百度和搜狗推出的知识图谱规模仅略小于 Google^[42]。同时技术上, 若采用深度分析逻辑化的模式, 需要至少 500 条规则覆盖所有模式, 而应用知识图谱则是可以开发的。所谓知识图谱就是数据的管理手段, 这种技术能弥补机器的缺失: 机器对语言认知和概念认知的巨大障碍, 即缺失的类似人脑中海量而有组织的知识体系。当然, 知识图谱也非万能的, 对于在线问询, 也存在语调识别失败以及机器自我学习能力制约的问题。知识图谱的发展也是机遇与挑战并存。

从基于关键词匹配、信息抽取并基于浅层语义分析的 IR-based QA 到依赖于网民贡献的问答系统, 再到知识库问答系统。问答系统一般包括 3 个主要组成部分: 问题分析、信息检索和答案抽取。一般问答系统流程图 4 所示。

知识库规模的发展成为问答系统发展的瓶颈, 而互联网的转型、知识图谱的诞生逐渐成为突破瓶颈的可能。先进行问句语义解析, 再进行语义表示, 最后进行语义匹配查询得到推理的过程。

将深度学习的算法模型应用于知识图谱的构建, 可以使知识图谱信息更加丰富, 表示更多的信息。进而在知识图谱中将问句分析进行最优子图的匹配, 以实现问什么答什么的智能问答系统。采用知识图谱的方式查询相比传统简单, 基于字符串匹配搜索引擎, 让搜索结果与查询内容更加紧实。同时, 知识图谱的构建源自海量数据的整合处理, 也避免了跨领域问题查询偏差, 即在知识图谱这张知识网编织成功的基础上, 通过各种逻辑算法分析处理自然语言, 实现抓住最符合问题的关键匹配知识图谱, 从而找出答案, 实现智能问答。

问答系统实现案例中, 最为成功的当属 IBM 于 2011 年研发的超级计算机“沃森”, 在美国的知识竞赛节目《危险边缘》中表现出色, 成功战胜人类选手。“沃森”采用深度开放域技术, 其逻辑架构如图 5 所示。

仔细分析比赛现场可发现, “沃森”虽有一定排除干扰信息的能力, 抓住问题的要点回答, 但不能及时获取现场选手错误答案信息进行更新反应, 排除错误。由此可知, 知识图谱是问答系统的重要支撑, 而且仅有全面而强大的知识库也不足够, 还需要其他技术的支撑才能使问答系统更智能。自然语言理解 (natural language understanding) 技术使问答系统具有更强的理解能力, 甚至一些双关、暗喻的知识还

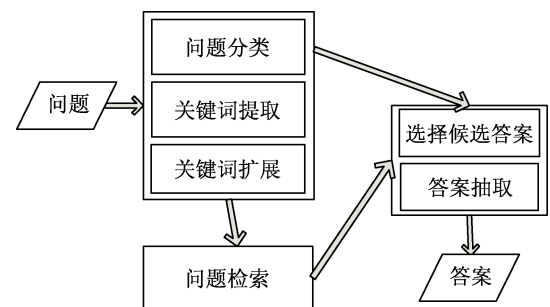


图 4 一般问答系统流程图示意图

Fig. 4 Process diagram of the general question answering system

具有交互的能力。动态分析各类假设和问题、精细的个性化分析能力、优化问题解答都是“沃森”成功的关键。IBM 超级电脑“沃森”的中国团队专家、IBM 中国研究院资深经理潘越接受采访时说到“沃森”系统几个特点：

1) 沃森的领域知识库包括百科全书、字典、地理类娱乐类的专题数据库、新闻报道、经典著作等，存储大小约为 70 GB。

2) 依赖上下文去除歧义。例如，考察上下文中提到的人、事件、地点等其他相关对象，从多个证据的角度解释这一问题。即使在一个算法中去除歧义失败，其他上百个算法会从其他方面给出答案是否正确的证据，这就削弱了对单个算法去除歧义能力的依赖。

3) 对于自然语言的处理是“沃森”的优势所在。

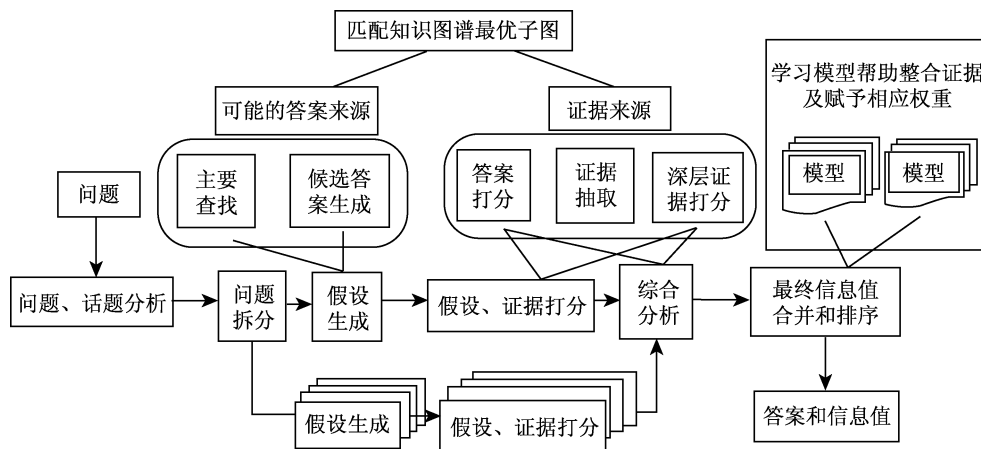


图5 深度开放域“沃森”问答系统逻辑架构图

Fig. 5 Logical architecture diagram of “Watson” question answering system in depth opening domain

4 结论与展望

将深度学习应用于知识图谱中是近年来研究的热潮，研究也刚刚起步。深度学习在很多领域虽已有显著成效而且有一定的学习能力，但不能完全自主学习；在自然语言的处理上虽已有了飞跃发展，但还存在不擅长的领域，未达到完美状态。同时，深度学习在知识库问答的应用上也是一项重要技术，同样存在很多的挑战。知识图谱应用的问答系统中存在两大难点：一是问句的解析挂载查询，查询单关系类型问题困难较少、易优化，但查询者实际应用中往往不是简单的单关系查询，因此难度提升。将问题分解，由于搜集并审议了各子问题的相关证据，最终答案的分值应该比未分解时高。即使不需要分解问题，也会因分解提高整体的置信度。二是知识图谱的构建，知识图谱的研究越来越倾向于多源异构，这就需要处理上的变化。日益剧增的海量数据，需要一种能不断更新智能学习的算法，才能实现一张知识面广且准确的知识图谱。尽管知识图谱的发展使得问答系统有了新的突破，但对于问答系统来说，各个技术、各个算法的完美结合与相互配合才是技术发展和市场应用的关键。

[参考文献] (References)

[1] WEIZENBAUM J. ELIZA—a computer program for the study of natural language communication between man and machine[J]. Communications of the ACM, 1966, 9(1): 36-45.

[2] WOODS W A. Lunar rocks in natural English: explorations in natural language question answering[J]. Linguistic Structures Processing, 1977, 5: 521-569.

- [3] WINOGRAD T. Understanding natural language[J]. *Cognitive Psychology*, 1972, 3(1): 1-191.
- [4] LENAT D B, GUHA R V, PITTMAN K, et al. Cyc: toward programs with common sense[J]. *Communications of the ACM*, 1990, 33(8): 30-49.
- [5] 王树西. 问答系统: 核心技术、发展趋势[J]. *计算机工程与应用*, 2005, 41 (18): 1-3.
WANG S X. Question answering system: core technology, application[J]. *Computer Engineering and Applications*, 2005, 41(18): 1-3. (in Chinese)
- [6] 曹树金, 吴育冰, 韦景竹, 等. 知识图谱研究的脉络、流派与趋势——基于 SSCI 与 CSSCI 期刊论文的计量与可视化[J]. *中国图书馆学报*, 2015, 41 (5): 16-34.
CAO S J, WU Y B, WEI J Z, et al. History, schools and trend in knowledge map: investigation and visualization based on SSCI and CSSCI[J]. *Journal of Library Science in China*, 2015, 41(5): 16-34. (in Chinese)
- [7] 刘岍, 李杨, 段宏, 等. 知识图谱构建技术综述[J]. *计算机研究与发展*, 2016, 53 (3): 582-600.
LIU Q, LI Y, DUAN H, et al. Knowledge graph construction techniques[J]. *Journal of Computer Research and Development*, 2016, 53(3): 582-600. (in Chinese)
- [8] 李保利, 陈玉忠, 俞士汶. 信息抽取研究综述[J]. *计算机工程与应用*, 2003, 39 (10): 1-5.
LI B L, CHEN Y Z, YU S W. Research on information extraction: a survey[J]. *Computer Engineering and Applications*, 2003, 39(10): 1-5. (in Chinese)
- [9] SHIRAKAWA M, NAKAYAMA K, HARA T, et al. Concept vector extraction from Wikipedia category network[C]//*Proceedings of the 3rd International Conference on Ubiquitous Information Management and Communication*. New York: ACM, 2009: 71-79.
- [10] 廉成洋, 毛宇光, 黄玉明. 基于启发式规则的 Web 信息抽取技术研究[J]. *计算机技术与发展*, 2009, 19 (8): 4-7.
LIAN C Y, MAO Y G, HUANG Y M. Web information extraction technology research based on heuristic rules[J]. *Computer Technology and Development*, 2009, 19(8): 4-7. (in Chinese)
- [11] 唐杰, 梁邦勇, 李涓子, 等. 语义 Web 中的本体自动映射[J]. *计算机学报*, 2006, 29 (11): 1956-1976.
TANG J, LIANG B Y, LI J Z, et al. Automatic ontology mapping in semantic Web[J]. *Chinese Journal of Computers*, 2006, 29(11): 1956-1976. (in Chinese)
- [12] 靖红芳, 王斌, 杨雅辉, 等. 基于类别分布的特征选择框架[J]. *计算机研究与发展*, 2008, 46 (9): 1586-1593.
JING H F, WANG B, YANG Y H, et al. Category distribution-based feature selection framework[J]. *Journal of Computer Research and Development*, 2008, 46(9): 1586-1593. (in Chinese)
- [13] 刘洁晶, 张建光. 信念传播算法在分类模型中的应用[J]. *福建电脑*, 2012, 28 (10): 6, 12.
LIU J J, ZHANG J G. Belief propagation algorithm in the application of the classification model[J]. *Fujian Computer*, 2012, 28(10): 6, 12. (in Chinese)
- [14] ZHENG Z H, WU X Y, SRIHARI R. Feature selection for text categorization on imbalanced data[J]. *SIGKDD Explorations Newsletter*, 2004, 6(1): 80-89.
- [15] 徐燕, 李锦涛, 王斌, 等. 不平衡数据集上文本分类的特征选择研究[J]. *计算机研究与发展*, 2007, 44 (增 2): 58-62.
XU Y, LI J T, WANG B, et al. A study of feature selection for text categorization on imbalanced data[J]. *Journal of Computer Research and Development*, 2007, 44(Suppl. 2): 58-62. (in Chinese)
- [16] 肖芳. 异构系统中实体识别研究[J]. *自动化与信息工程*, 2009, 30 (3): 35-37.
XIAO F. Research on entity identification in heterogeneous systems[J]. *Automation & Information Engineering*, 2009, 30(3): 35-37. (in Chinese)
- [17] 陆伟, 武川. 实体链接研究综述[J]. *情报学报*, 2015 (1): 105-112.
LU W, WU C. Literature review on entity linking[J]. *Journal of the China Society for Scientific and Technical Information*, 2015(1): 105-112. (in Chinese)
- [18] HAN X, ZHAO J. 2009 Named entity disambiguation by leveraging Wikipedia semantic knowledge[C]//*Proceeding of the*

- 18th ACM Conference on Information and Knowledge Management. New York: ACM, 2009: 215-224.
- [19] HAN X, SUN L, ZHAO J. Collective entity linking in web text: a graph-based method[C]//Proceeding of the International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM Press, 2011: 765-774.
- [20] HOFFART J, YOSEF M A, BORDINO I, et al. Robust disambiguation of named entities in text[C]//Proceedings of the Conference on Empirical Methods in Natural Language Processing. Stroudsburg: Association for Computational Linguistics, 2011: 782-792.
- [21] WANG Z C, LI J Z, WANG Z G, et al. Cross-lingual knowledge linking across Wiki knowledge bases[C]//Proceedings of the 21st International Conference on World Wide Web. New York: ACM, 2012: 459-468.
- [22] POON H. Markov logic for machine reading[D]. Seattle: University of Washington, 2011.
- [23] NICKEL M, TRESP V. Tensor factorization for multi-relational learning[M]//BLOCKEEL H, KERSTING K, NIJSSEN S, et al. Machine Learning and Knowledge Discovery in Databases: European Conference. Heidelberg: Springer, 2013: 617-621.
- [24] WANG Z, ZHANG J W, FENG J L, et al. Knowledge graph embedding by translating on hyperplanes[R]. Québec: AAAI Conference on Artificial Intelligence, 2014.
- [25] JI G L, LIU K, HE S Z, et al. Knowledge graph completion with adaptive sparse transfer matrix[EB/OL]. [2016-02-21]. <https://aaai.org/ocs/index.php/AAAI/AAAI16/paper/view/11982>.
- [26] HE S, LIU K, JI G, et al. Learning to represent knowledge graphs with Gaussian embedding[C]//Proceedings of the 24th ACM International on Conference on Information and Knowledge Management. New York: ACM, 2015: 623-632.
- [27] JI G L, HE S Z, XU L H, et al. Knowledge graph embedding via dynamic mapping matrix[C]//Proceedings of 53rd Annual Meeting of the Association for Computational Linguistics and 7th International Joint Conference on Natural Language Processing. Stroudsburg: Association for Computational Linguistics, 2015: 687-696.
- [28] WEI Z Y, ZHAO J, LIU K, et al. Large-scale knowledge base completion: inferring via grounding network sampling over selected instances[C]//Proceedings of the 24th ACM Conference on Information and Knowledge Management. New York: ACM, 2015: 1331-1340.
- [29] COHEN W, HURST M, JENSEN L. A flexible learning system for wrapping tables and lists in HTML documents[C]//Proceedings of the 11th International World Wide Web Conference. New York: ACM, 2002: 232-241.
- [30] KOSALA R, BLOCKEEL H, BRUYNOOGHE M, et al. Information extraction from structured documents using k -testable tree automaton inference[J]. Data & Knowledge Engineering, 2006, 58(2): 129-158.
- [31] TENIER S, TOUSSAINT Y, NAPOLI A, et al. Instantiation of relations for semantic annotation[C]//Proceedings of IEEE/WIC/ACM International Conference on Web Intelligence. Washington DC: IEEE Computer Society, 2006: 463-472.
- [32] ZHAI Y H, LIU B. Structured data extraction from the Web based on partial tree alignment[J]. IEEE Trans. on Knowledge and Data Engineering, 2006, 18(12): 1614-1628.
- [33] ÁLVAREZ M, PAN A, RAPOSO J, et al. Extracting lists of data records from semi-structured Web pages[J]. Data & Knowledge Engineering, 2008, 64(2): 491-509.
- [34] HONG M C, TANG J, LI J Z. Semantic annotation using horizontal and vertical contexts[C]//Proceedings of the 1st Asian Semantic Web Conference. Berlin: Springer, 2006: 58-64.
- [35] TANG J, LI J Z, LU H J, et al. iASA: learning to annotate the semantic Web[J]. Journal on Data Semantic IV, 2005, 3740(4): 110-145.
- [36] CHEN Y, XU L, LIU K, et al. Event extraction via dynamic multi-pooling convolutional neural networks[C]//Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing. Stroudsburg: Association for Computational Linguistics, 2015: 167-176.
- [37] LIAO S, GRISHMAN R. Using document level cross-Event inference to improve event extraction[C]//Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics. Stroudsburg: Association for Computational Linguistics, 2010: 789-797.

- [38] HE S, LIU K, ZHANG Y, et al. Question answering over linked data using first-order logic[C]//Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing. Stroudsburg: Association for Computational Linguistics, 2014: 1092-1103.
- [39] 陈湧. 基于知识元的突发事件案例信息抽取及检索[D]. 大连: 大连理工大学, 2014.
CHEN Y. Emergency case information extraction and retrieval based on knowledge element[D]. Dalian: Dalian University of Technology, 2014. (in Chinese)
- [40] 钟涛, 陈群秀. 基于 Web 主题性信息检索的灾难性事件信息抽取系统[C]//第七届中文信息处理国际会议. 武汉: 中国中文信息学会, 2007: 321-327.
ZHONG T, CHEN Q X. A calamity event information extraction system in topical Web[C]//7th International Conference on Chinese Computing. Wuhan: Chinese Information Processing Society of China, 2007: 321-327. (in Chinese)
- [41] 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.
MAO X L, LI X M. A survey on question and answering systems[J]. Journal of Frontiers of Computer Science & Technology, 2012, 6(3): 193-207. (in Chinese)
- [42] 36 大数据. 知识图谱技术原理介绍[EB/OL]. [2016-01-09]. <http://mt.sohu.com/20160109/n434003191.shtml>.
36 Big Data. Introduction to the principle of knowledge graph technology[EB/OL]. [2016-01-09]. <http://mt.sohu.com/20160109/n434003191.shtml>. (in Chinese)