

基于表示学习的知识库问答研究进展与展望

刘康¹ 张元哲¹ 纪国良¹ 来斯惟¹ 赵军¹

摘要 面向知识库的问答 (Question answering over knowledge base, KBQA) 是问答系统的重要组成部分。近些年, 随着以深度学习为代表的表示学习技术在多个领域的成功应用, 许多研究者开始着手研究基于表示学习的知识库问答技术。其基本假设是把知识库问答看做是一个语义匹配的过程。通过表示学习知识库以及用户问题的语义表示, 将知识库中的实体、关系以及问句文本转换为一个低维语义空间中的数值向量, 在此基础上, 利用数值计算, 直接匹配与用户问句语义最相似的答案。从目前的结果看, 基于表示学习的知识库问答系统在性能上已经超过传统知识库问答方法。本文将对现有基于表示学习的知识库问答的研究进展进行综述, 包括知识库表示学习和问句 (文本) 表示学习的代表性工作, 同时对于其中存在难点以及仍存在的研究问题进行分析和讨论。

关键词 知识库问答, 深度学习, 表示学习, 语义分析

引用格式 刘康, 张元哲, 纪国良, 来斯惟, 赵军. 基于表示学习的知识库问答研究进展与展望. 自动化学报, 2016, 42(6): 807–818

DOI 10.16383/j.aas.2016.c150674

Representation Learning for Question Answering over Knowledge Base: An Overview

LIU Kang¹ ZHANG Yuan-Zhe¹ JI Guo-Liang¹ LAI Si-Wei¹ ZHAO Jun¹

Abstract Question answering over knowledge base (KBQA) is an important direction for the research of question answering. Recently, with the drastic development of deep learning, researchers and developers have paid more attentions to KBQA from this angle. They regarded this problem as a task of semantic matching. The semantics of knowledge base and users' questions are learned through representation learning under the framework of deep learning. The entities and relations in knowledge base and the texts in questions could be represented as numerical vectors. Then, the answer could be figured out through similarity computation between the vectors of knowledge base and the vectors of the given question. From reported results, KBQA based on representation learning has obtained the best performance. This paper introduces the mainstream methods in this area. It further induces the typical approaches of representation learning on knowledge base and texts (questions), respectively. Finally, the current research challenges are discussed.

Key words Question answering over knowledge base (KBQA), deep learning, representation learning, semantic analysis

Citation Liu Kang, Zhang Yuan-Zhe, Ji Guo-Liang, Lai Si-Wei, Zhao Jun. Representation learning for question answering over knowledge base: an overview. *Acta Automatica Sinica*, 2016, 42(6): 807–818

随着人们对于信息精准化的需求越来越高, 传统以关键词匹配和文档排序为基本特点的搜索引擎急需一场革命。2011 年, 美国华盛顿大学图灵实验室的 Etzioni 教授在 *Nature* 上发表题为 “Search needs a shake-up”^[1] 一文指出, 问答系统是下一代搜索引擎的基本形态。区别于传统基于关键词匹配

的搜索模式, 问答系统最主要的特点在于: 1) 用户输入是自然语言的问句; 2) 返回答案不再是文档排序的形态, 而是直接给出用户所需要的答案。这需要对用户自然语言问句进行深度理解, 同时对网页中的目标文本进行细致的语义分析, 从中抽取知识, 并根据用户的问题准确匹配、推理相对应的答案。

为了达到这一目标, 近些年, 无论是学术界或工业界, 研究者们逐步把注意力投向知识图谱或知识库。其目标是把互联网文本内容组织成为以实体为基本语义单元 (节点) 的图结构, 其中图上的边表示实体之间语义关系。通过构建知识库, 可以从源头上分析网络文本中所蕴含的语义知识。目前, 互联网中已经有一些可以获取的大规模知识库, 例如 DBpedia^[2]、Freebase^[3]、YAGO^[4] 等。这些知识库多是以 “实体–关系–实体” 三元组 ((实体 1, 关系,

收稿日期 2015-11-02 录用日期 2016-05-03

Manuscript received November 2, 2015; accepted May 3, 2016
国家重点基础研究发展计划 (973 计划) (2014CB340503), 国家自然科学基金 (61533018), “CCF–腾讯” 犀牛鸟基金资助

Supported by National Basic Research Program of China (973 Program) (2014CB340503), National Natural Science Foundation of China (61533018), and CCF-Tencent Open Research Fund

本文责任编辑 贾珈

Recommended by Associate Editor JIA Jia

1. 中国科学院自动化研究所模式识别国家重点实验室 北京 100190

1. National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing 100190

实体 2), 简称为三元组) 为基本单元所组成的图结构. 基于这样的结构化的知识, 问答系统的任务就是要根据用户问题的语义直接在知识库上查找、推理出相匹配的答案, 这一任务也称之为面向知识库的问答系统或知识库问答 (Question answering over knowledge base, KBQA).

要完成在结构化数据上的查询、匹配、推理等操作, 最有效的方式是利用结构化的查询语句, 例如: SQL、SPARQL 等. 然而, 这些语句通常是由专家编写, 普通用户很难掌握并正确运用. 对其来说, 自然语言仍然是最自然的交互方式. 因此, 如何把用户的自然语言问句转化为结构化的查询语句便是知识库问答的核心所在, 其关键是对自然语言问句进行语义理解 (如图 1 所示). 目前, 主流方法是通过语义分析, 将用户的自然语言问句转化成结构化的语义表示, 例如 λ 范式^[5] 和 DCS-Tree^[6].

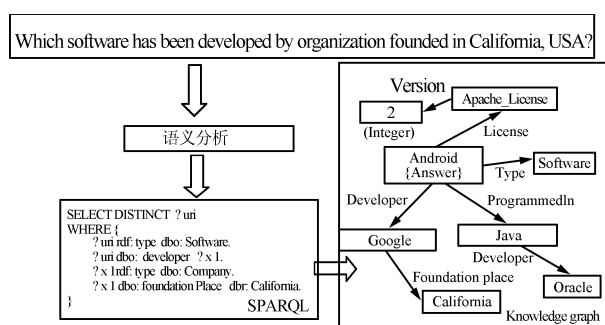


图 1 知识库问答过程

Fig. 1 The process of KBQA

但是, 这一处理范式仍然是基于符号逻辑的, 缺乏灵活性, 在分析问句语义过程中, 易受到符号间语义鸿沟影响. 同时从自然语言问句到结构化语义表达需要多步操作, 多步间的误差传递对于问答的准

确度也有很大的影响. 近年来, 深度学习技术以及相关研究飞速发展, 在很多领域都取得了突破, 例如图像、视频、语音. 在自然语言处理领域也逐步开始具有广泛的应用. 其优势在于通过学习能够捕获文本 (词、短语、句子、段落以及篇章) 的语义信息, 把目标文本投射到低维的语义空间中, 这使得传统自然语言处理过程中很多语义鸿沟的现象通过低维空间中向量间数值计算得以一定程度的改善或解决. 因此越来越多的研究者开始研究深度学习技术在自然语言处理问题中的应用, 例如情感分析^[7]、机器翻译^[8]、句法分析^[9]等. 知识库问答系统也不例外, 已有相关的研究工作包括文献 [10–12]. 与传统基于符号的知识库问答方法相比, 基于表示学习的知识库问答方法更具鲁棒性, 其在效果上已经逐步甚至超过传统方法, 如图 2 所示. 这些方法的基本假设是把知识库问答看作是一个语义匹配的过程. 通过表示学习, 我们能够将用户的自然语言问题转换为一个低维空间中的数值向量 (分布式语义表示), 同时知识库中的实体、概念、类别以及关系也能够表示成为同一语义空间的数值向量. 那么传统知识库问答任务就可以看成问句语义向量与知识库中实体、边的语义向量相似度计算的过程. 图 3 给出基于表示学习的知识库问答示意图, 其中方格表示学习到的语义表示.

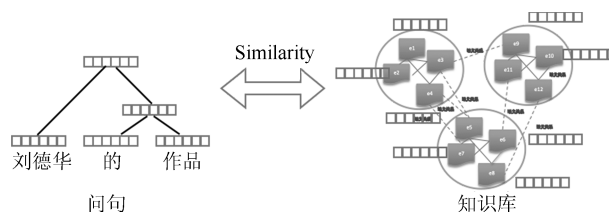


图 3 基于表示学习的知识库问答示意图

Fig. 3 Representation learning based KBQA

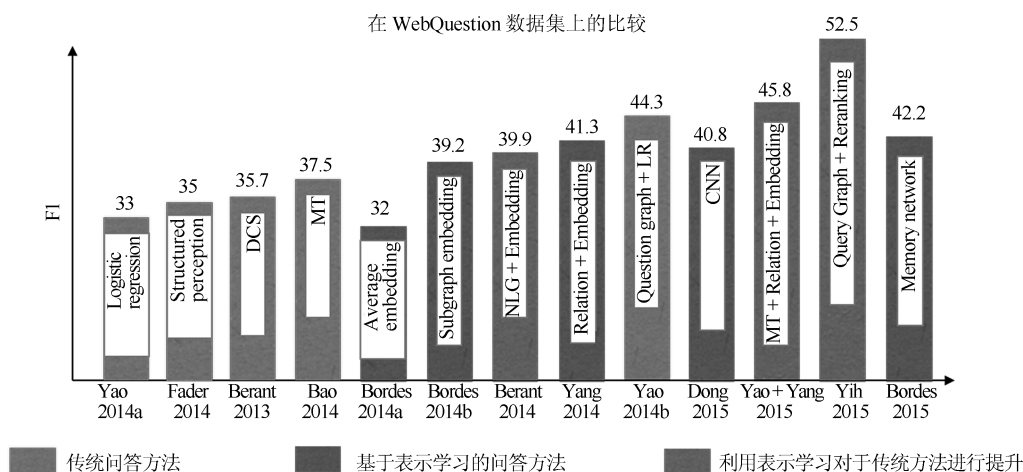


图 2 基于表示学习的知识库问答方法与传统方法的性能比较

Fig. 2 The comparisons between representation learning based KBQA and traditional KBQA

然而, 构建一个基于表示学习的知识库问答系统并不是一件容易的事情, 要完成这一目标, 我们首先要回答三方面问题:

1) 如何学习知识库的分布式表示? 如何将知识库中的语义单元, 包括节点(实体、类别)、边(关系), 表示成为语义空间中的数值向量.

2) 针对用户的问句, 我们应该如何通过深度学习学习问句的语义表示.

3) 基于学习到的问句和知识库的语义表示, 如何自动学习知识库语义表示和问句语义表示之间的关联, 学习它们间的映射关系, 从而构建基于表示学习的知识库问答系统.

本文将围绕这三个问题, 对于现有成果进行综述, 介绍其中的代表性方法, 同时探讨针对这一问题未来可能存在的研究问题与热点. 具体章节安排如下: 第 1 节介绍知识库表示学习的主要方法; 第 2 节介绍问句(文本)表示学习的常用模型和方法; 第 3 节介绍基于表示学习的知识库问答现有工作以及性能比较; 第 4 节可能存在的研究问题和未来的研究热点; 最后是结论.

1 知识库的表示学习

知识库表示学习的目标是通过知识库建模, 将知识库中的实体、类别以及关系等语义单元表示成为数值空间中的向量或矩阵. 向量中的每一维的数值表示该语义单元在某一语义维度上的投影. 由于实体和关系的数值表示是根据整个知识库得到, 因此这种数值表示方法包含更加全面的信息, 使得知识库能够很方便地应用到其他学习任务中. 根据学习方法, 已有知识库表示学习的方法主要分为两大类: 1) 基于张量分解的方法; 2) 基于映射的方法. 下面进行详细介绍.

1.1 基于张量分解的知识库表示学习

张量分解的方法以 RESCAL 系统^[13-17] 为主要代表. 图 4 是 RESCAL 的原理图. 它的核心思想是将整个知识图谱编码为一个三维张量(其中知识库包含的三元组对应值为 1, 其他为 0), 由这个张量分解出一个核心张量和一个因子矩阵, 核心张量中每个二维矩阵切片代表一种关系, 因子矩阵中每一行代表一个实体. 由核心张量和因子矩阵还原的结果被看作对应三元组成立的概率, 如果概率大于某个阈值, 则对应三元组正确; 否则, 不正确.

RESCAL 的目标函数为: $X_k \approx AR_kA^T$, 这里 A 是一个 $n \times r$ 矩阵, 其行向量是对于每一个实体的表示, R_k 是 $r \times r$ 矩阵, 表示知识库中第 k 个语义关系所定义的映射矩阵, X_k 表示在知识库中的实际观测. RESCAL 通过张量分解, 能够在编码实体和关系过程中综合整个知识库的信息, 它的主要缺

点是当关系数目较多时, 张量的维度很高, 分解过程计算量较大. 因此, 对于 Freebase 这类关系数目众多而又非常稀疏的大规模知识库效果不佳.

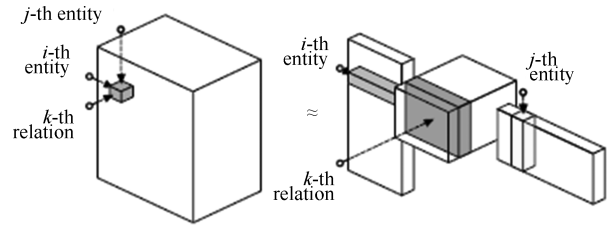


图 4 RESCAL 系统原理^[15]

Fig. 4 RESCAL system architecture^[15]

1.2 基于映射的知识库表示学习

为了解决上述张量分解方法在大规模知识库表示学习过程中学习效率低的问题, 很多研究者转向对于知识库中的基本语义单元: 三元组进行独立建模. 这类方法通常将知识库中的三元组表示为 (h, r, t) , h 表示头实体, r 表示关系, t 表示尾实体, 它们的向量分别表示为 $\mathbf{h}, \mathbf{r}, \mathbf{t}$. 其假设 $\mathbf{h}$ 和 $\mathbf{t}$ 经过某种与 $\mathbf{r}$ 相关的映射后得到向量应该相似或者相等的. 为了刻画这一过程, 通常定义基于三元组的能量函数为 $f_r(\mathbf{h}, \mathbf{t})$, 则学习的目标函数为

$$L = \sum_{\mathbf{h}, \mathbf{t}} \sum_{\mathbf{h}', \mathbf{t}'} \max(0, f_r(\mathbf{h}, \mathbf{t}) + \gamma - f_r(\mathbf{h}', \mathbf{t}')) \quad (1)$$

其中, $\mathbf{h}', \mathbf{t}'$ 表示随机产生的三元组(错误的或不存在的三元组)中头实体和尾实体对应向量. 其学习过程是保证知识库中出现的三元组获得较高的目标函数值, 同时, 对于未出现的三元组进行惩罚. 根据对语义关系 $\mathbf{r}$ 的表示方式, 已有方法大致可以分为映射矩阵和映射向量两类, 我们分别对其进行介绍.

1.2.1 关系 $\mathbf{r}$ 表示为矩阵

Structured 模型^[18] 为每种关系定义两个矩阵 M_{r1} 和 M_{r2} , 其能量函数为 $f_r(\mathbf{h}, \mathbf{t}) = \|M_{r1}\mathbf{h} - M_{r2}\mathbf{t}\|_1$. 该模型用两个分离的矩阵表示关系, 不能很好地捕获关系与实体之间的联系. Unstructured 模型^[19] 是 Structured 模型的特例, 它令 $\mathbf{r} = 0$, 能量函数为 $f_r(\mathbf{h}, \mathbf{t}) = \|\mathbf{h} - \mathbf{t}\|_2^2$. 模型的缺点是没有考虑不同关系的影响.

Semantic matching energy (SME) 模型^[20] 为了克服参数过多的问题, 该模型将实体和关系都用向量表示, 所有三元组共享 SME 中的参数. SME 使用多维矩阵运算捕获实体和关系之间的联系, 它首先对 $\mathbf{h}, \mathbf{r}$ 进行线性运算得到向量 $\mathbf{v}_{hr}$, 然后再对 $\mathbf{r}, \mathbf{t}$ 做线性运算得到向量 $\mathbf{v}_{rt}$, 能量函数为 $f_r(\mathbf{h}, \mathbf{t}) = \mathbf{v}_{hr}^T \mathbf{v}_{rt}$, 其中线性运算的权重可以是矩阵, 也可以是三维张量. SME 具有很强的学习能力,

参数较少,但是由于需要对 $\mathbf{h}, \mathbf{r}$ 和 $\mathbf{r}, \mathbf{t}$ 分别进行线性运算,因此计算量大。

Latent factor (LF) 模型^[21-22] 将实体用向量表示,关系用矩阵 M_r 表示。LF 把关系看作实体间的二阶关联,定义能量函数 $f_r(\mathbf{h}, \mathbf{t}) = \mathbf{h}^T M_r \mathbf{t}$ 。这个模型使得实体和关系之间产生了很好的交互,实体和关系之间的联系得到充分的体现。

Single layer (SL) 模型^[23] 用神经网络刻画映射过程,其为每个三元组定义一个单层非线性神经网络作为能量函数,以实体向量为输入层,关系矩阵为网络的权重参数。同样,Socher 等^[23] 对于 SL 模型进行改进,提出 Neural tensor network (NTN) 模型。其在 SL 中加入关系与实体的二阶非线性操作,增强实体与关系的交互性。SL 模型是 NTN 模型的特例,当 NTN 中三维张量均为 0 时,NTN 退化为 SL。NTN 是目前表达能力最强的模型,适合学习稠密的知识库。它的主要缺点是参数太多,计算量大,不适合比较稀疏的知识库。

1.2.2 关系 $\mathbf{r}$ 表示为向量

这一类方法通常将关系 $\mathbf{r}$ 表示成向量,对于三元组的建模看作从头部实体到尾部实体的翻译。最早提出的模型是 TransE^[24],其势能函数定义为

$$f_r(\mathbf{h}, \mathbf{t}) = \|\mathbf{h} + \mathbf{r} - \mathbf{t}\|_2^2 \quad (2)$$

如图 5 所示,三元组中的实体和关系在同一语义空间内成加性关系。

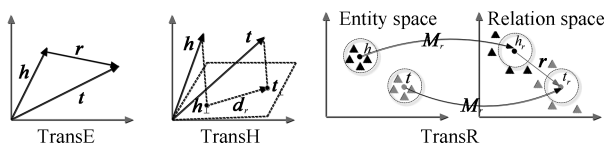


图 5 TransE、TransH 和 TransR^[24-26]

Fig. 5 TransE, TransH and TransR^[24-26]

1) 1-to- N , N -to-1 和 N -to- N 的语义关系

TransE 是一种计算效率很高、预测性能非常好的模型。对于“1-to-1”关系类型,这一模型通常能够很好的建模。但是对于“1-to- N ”、“ N -to-1”和“ N -to- N ”等关系类型存在不足。例如,我们用 $\mathbf{r}$ 表示“性别”这一语义关系。在知识库中,我们可以有(张三, 性别, 女)、(王五, 性别, 男),也可以有(李四, 性别, 女)。因此,“性别”是“ N -to- N ”关系类型的。基于 TransE 模型,在学习知识库的表示时,会得到:

$$\begin{aligned} \mathbf{h}_1 + \mathbf{r} &= \mathbf{t}_1 \\ \mathbf{h}_2 + \mathbf{r} &= \mathbf{t}_1 \end{aligned} \Rightarrow \mathbf{h}_1 = \mathbf{h}_2 \quad (3)$$

得到的学习结果 $\mathbf{h}_1 = \mathbf{h}_2$ 表示两个不同的实体的语义表示是相同的,这显然是错误的。为了解决这一问

题,许多研究者提出了一系列改进方法^[25-28]。其核心思想是根据语义关系类型,设定映射函数,将三元组的头实体和尾实体映射到与目标关系相关的语义空间中,在这一空间中再利用 TransE 中的加性模型学习知识库的表示。

TransH 模型^[25] 认为 TransE 中的翻译过程应当在关系所属的超平面上进行。如图 5 所示,它首先将头部、尾部实体向量投影到关系所在超平面,然后在超平面上完成翻译过程。该模型能够使同一个实体在不同的关系中扮演不同的角色,并且在 1-to- N , N -to-1 和 N -to- N 的关系类型上较 TransE 有较好的预测效果。

TransR 模型^[26] 也基于 TransE,它在不同维度的空间中对于实体和关系建模。因此,如图 5 所示,TransR 为每个关系定义一个矩阵 M_r ,用于将实体向量转换到关系所属空间中,然后在关系向量空间中完成翻译。TransR 较 TransE 和 TransH 有较好的推广性能。但是由于为每种关系定义了一个矩阵,因此它有更多的参数和更大的计算量,给扩展到大规模知识库的运用带来了困难。CTransR^[26] 是根据关系在不同实体对中表现出不同含义的现象,对 TransR 的一种改进。它将同一个关系的实体对聚类成若干类别,在每一个类别中,单独学习一个关系的向量表示,这样可以大大减小检索学习的参数规模。

TransD 模型^[27] 是对 TransR 的一种改进,其认为 TransR 中固定大小的转换矩阵 M_r 应由实体-关系对动态确定。同时,该模型考虑实体和关系的多类别性。除此之外,在计算过程中,TransD 用向量运算代替了 TransR 中的矩阵运算,大大减少了计算量。

除此之外,还有一些模型在 TransE 的基础上进行了改进 PTransE 模型^[28] 在 TransE 的基础上增加实体之间的路径信息,显式地增加了推理知识,相对 TransE 模型,在效果上有较大提升。SSE (Semantically smooth knowledge graph embedding) 模型^[29] 在 TransE 的基础上增加了语义平滑的约束,认为同类型的实体的向量表示应该在空间中位置更加靠近。该模型有较好的学习效果,但是时间复杂度较高,不易扩展到大规模知识库。

2) 基于分布的知识库表示学习

KG2E (Knowledge graphs with Gaussian embedding) 模型^[30] 在 TransE 的基础上提出一种基于分布的表示学习方法,使用基于高斯嵌入的方法在多维高斯分布空间中学习知识库中实体和关系的表示。不同于 TransE 以及其改进模型,KG2E 将知识库中的实体、类别、关系都约定服从高斯分布。如 $P_h = N(\mu_h, \Sigma_h)$ 以及 $P_r = N(\mu_r, \Sigma_r)$, 这里 h 表示实体, r 表示关系, N 表示高斯分布, μ 是高斯分布的均值,表示学习到的表示向量, Σ 协方差

矩阵, 表示该实体或关系的不确定性. 学习的过程如同 TransE 一样采用加性模型, 其设定三元组中头实体和尾实体之间的差 $h - t$ 同样服从高斯分布 $P_{h-t} = N(\mu_h - \mu_t, \Sigma_h + \Sigma_t)$, 则目标势能函数通过 KL 距离 $KL(P_{h-t}, P_r)$ 来衡量 $h - t$ 和 r 之间的距离, 保证在知识库中出现的三元组中 $h - t$ 和 r 的 $KL(P_{h-t}, P_r)$ 越小越好. 同时, 其假设知识库中的实体、关系均有不确定性 (频率、不同类型的关系均会引发不确定性). 通过引入协方差矩阵, 该模型能够对于知识库中实体和关系的不确定性进行建模, 尤其对于 1-to- N 和 N -to-1 的关系具有很好的学习效果.

2 问句表示学习

问句的表示学习是通过统计学习自动获取问句 (文本) 的语义表示. 德国数学家弗雷格 (Gottlob Frege) 在 1892 年就曾提出: 一段话的语义由其各组成部分的语义以及它们之间的组合方法所确定^[31]. 现有的问句语义表示也通常以该思路为基础, 通过语义组合的方式获得. 常用的组合语义组合函数, 如线性加权、矩阵乘法、张量乘法等, 在文献 [32] 中有详细的总结. 近年来基于神经网络的语义组合技术为文档表示带来了新的思路. 从神经网络的结构上看, 主要可以分为三种方式: 递归神经网络、循环神经网络和卷积神经网络.

2.1 递归神经网络

递归神经网络 (Recursive neural network) 的结构如图 6, 其核心为通过一个树形结构, 从词开始逐步合成各短语的语义, 最后得到整句话的语义.

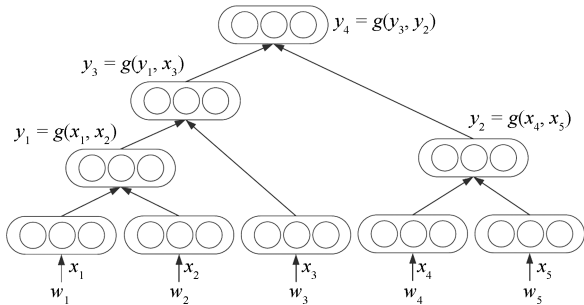


图 6 递归神经网络结构图

Fig. 6 Recursive neural network architecture

递归神经网络使用的树形结构一般为二叉树, 在某些特殊情况下 (如依存句法分析树^[33]) 也使用多叉树. 本文主要从树的构建方式和子节点到父节点的组合函数, 这两方面介绍递归神经网络.

树形结构有两种方式生成: 1) 使用句法分析器构建句法树^[7, 34]; 2) 使用贪心方法选择重建误差最小的相邻子树, 逐层合并^[35]. 这两种方法各有优劣,

使用句法分析器的方法可以保证生成的树形结构是一棵句法树, 树中各个节点均对应句子中的短语, 通过网络合并生成的各个节点的语义表示也对应各短语的语义. 使用贪心方法构建树形结构则可以通过自动挖掘大量数据中的规律, 无监督地完成这一过程, 但是树中的各个节点不能保证有实际的句法成分.

子节点到父节点的组合函数 $y = f(a, b)$ 主要有三种:

1) 句法组合. 这种方式下, 子节点的表示为向量 a, b , 父节点可以通过矩阵运算得到:

$$y = \phi(H[a; b]) \quad (4)$$

其中, ϕ 为非线性的激活函数, 权重矩阵 H 可能固定^[36], 也可能根据子树对应的句法结构不同, 而选用不同的矩阵^[37]. 该方法一般用于句法分析中.

2) 矩阵向量法^[38]. 在这种表示下, 每个节点由两部分表示组成, 一个矩阵和一个向量, 对于 A, a 子节点和 B, b 子节点, 其组合函数为

$$y = \phi(H[Ba; Ab]) \quad (5)$$

$$Y = W_M \begin{bmatrix} A \\ B \end{bmatrix}$$

其中, $W_M \in \mathbf{R}^{|a| \times |2a|}$, 保证父节点对应的语义变换矩阵 $Y \in \mathbf{R}^{|a| \times |a|}$, 与子节点的 A, B 矩阵维度一致. 使用这种方法, 每个词均有一个语义变换矩阵, 对于否定词等对句法结构另一部分有类似影响的词而言, 普通的句法组合方式没法很好地对其建模, 而这种矩阵向量表示则可以解决这一问题. Socher 等将该方法用于关系分类中^[39].

3) 张量组合. 张量组合方式使用张量中的每一个矩阵, 将子节点组合生成父节点表示中的一维.

$$y = \phi([a; b]^T W^{[1:d]} [a; b] + H[a; b]) \quad (6)$$

其中, $W^{[1:d]}$ 表示张量 W 中的第 1~ d 个切片矩阵. 不同的切片用于生成父节点 y 中不同的维度. 该方法是句法组合方法的泛化形式, 有更强的语义组合能力, Socher 等将其用于情感分析任务中^[23].

递归神经网络在构建文本表示时, 其精度依赖于文本树的精度. 无论使用哪种构建方式, 哪种组合函数, 构建文本树均需要至少 $O(n^2)$ 的时间复杂度, 其中, n 表示句子的长度. 当模型在处理长句子或者文档时, 所花费的时间往往是不可接受的. 更进一步地, 在做文档表示时, 两个句子之间的关系不一定能构成树形结构. 因此递归神经网络在大量句子级任务中表现出色, 但可能不适合构建长句子或者文档级别的语义.

2.2 循环神经网络

循环神经网络 (Recurrent neural network) 由 Elman 在 1990 年首次提出^[40]. 该模型的核心是通过循环方式逐个输入文本中的各个词, 并维护一个隐藏层, 保留所有的上文信息.

循环神经网络是递归神经网络的一个特例, 可以认为它对应的是一棵任何一个非叶结点的右子树均为叶结点的树. 这种特殊结构使得循环神经网络具有两个特点: 1) 由于固定了网络结构, 模型只需在 $O(n)$ 时间内即可构建文本的语义. 这使得循环神经网络可以更高效地对文本的语义进行建模. 2) 从网络结构上看, 循环神经网络的层数非常深, 句子中有几个词, 网络就有几层. 因此, 使用传统方法训练循环神经网络时, 会遇到梯度衰减或梯度爆炸的问题, 这需要模型使用更特别的方法来实现优化过程^[41-42].

在循环神经网络中, 当模型输入所有的词之后, 最后一个词对应的隐藏层代表了整个文本的语义.

$$h(i) = f(e(w_i) + Wh(i-1)) \quad (7)$$

优化方式上, 循环神经网络与其他网络结构也略有差异. 在普通的神经网络中, 反向传播算法可以利用导数的链式法则直接推算得到. 但是在循环神经网络中, 由于其隐藏层到下一个隐藏层的权重矩阵 W 是复用的, 直接对 W 矩阵求导非常困难. 循环神经网络最朴素的优化方式为沿时间反向传播技术 (Back propagation through time, BPTT). 该方法首先将网络展开成图 7 的形式, 对于每一个标注样本, 模型通过普通网络的反向传播技术对隐藏层逐个更新, 并反复更新其中的权重矩阵 W . 由于梯度衰减的问题, 使用 BPTT 优化循环神经网络时, 只传播固定的层数 (比如 5 层). 为了解决梯度衰减问题, Hochreiter 等在 1997 年提出了 LSTM (Long short-term memory) 模型^[43]. 该模型引入了记忆单元, 可以保存长距离信息, 是循环神经网络的一种常用的优化方案. LSTM 模型在传统循环网络的一个隐藏层节点上加入了三个门, 分别为输入门、输出门和遗忘门, 这三个门可以有选择地将远距离信息无衰减地传递下去. LSTM 的具体实现公式如下:

$$\begin{aligned} i_t &= \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i) \\ f_t &= \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f) \\ c_t &= f_t \odot c_{t-1} + i_t \odot \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \\ o_t &= \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_{t-1} + b_o) \\ h_t &= o_t \odot \tanh(c_t) \end{aligned} \quad (8)$$

无论采用哪种优化方式, 循环神经网络的语义都会偏向文本中靠后的词. 因此, 循环神经网络很少

直接用来表示整个文本的语义. 但由于其能有效表示上下文信息, 因此被广泛用于序列标注任务.

2.3 卷积神经网络

卷积神经网络 (Convolutional neural network, CNN) 最早由 Fukushima 在 1980 年提出^[44], 此后, LeCun 等对其做了重要改进^[45].

卷积神经网络的结构如图 8, 其核心是局部感知和权值共享. 在一般的前馈神经网络中, 隐藏层的每个节点都与输入层的各个节点有全连接; 而在卷积神经网络中, 隐藏层的每个节点只与输入层的一个固定大小的区域 (win 个词) 有连接. 从固定区域到隐藏层的这个子网络, 对于输入层的所有区域是权值共享的. 输入层到隐藏层的公式, 形式化为

$$\begin{aligned} x_i &= [e(w_{i-\lfloor win/2 \rfloor}); \cdots; e(w_i); \cdots; e(w_{i+\lfloor win/2 \rfloor})] \\ h_i^{(1)} &= \tanh(wx_i + b) \end{aligned} \quad (9)$$

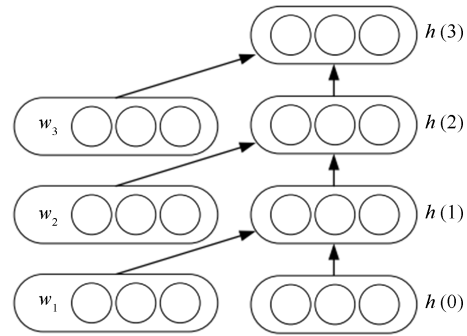


图 7 循环神经网络模型结构图

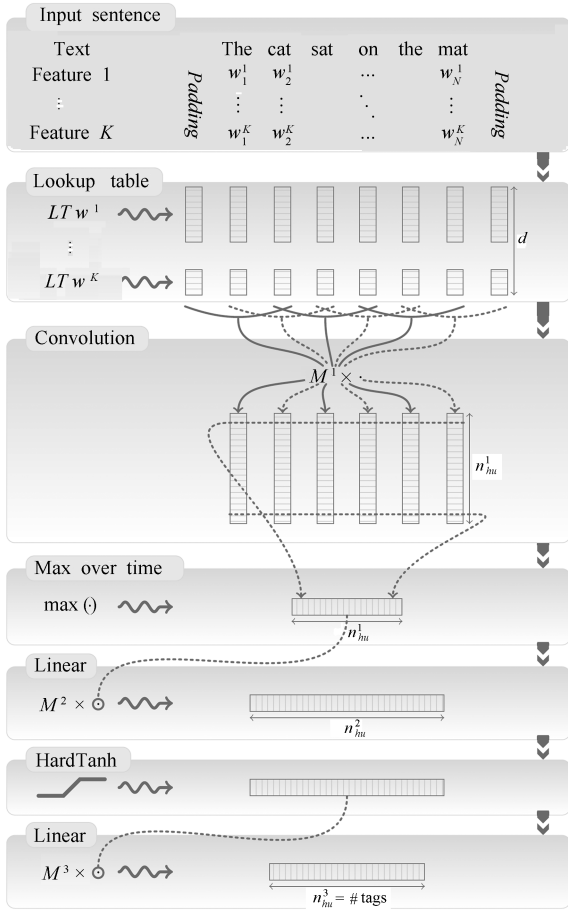
Fig. 7 Recurrent neural network architecture

在得到若干个隐藏层之后, 卷积神经网络通常会采用池化 (Pooling) 技术, 将不定长度的隐藏层压缩到固定长度的隐藏层中. 常用的有均值池化 (Average pooling) 和最大池化 (Max pooling)^[46]. 最大池化的公式为

$$h_i^{(2)} = \max_{1 \leq i \leq n} h_i^{(1)} \quad (10)$$

卷积神经网络通过其卷积核, 可以对文本中的每个部分的局部信息进行建模; 通过其池化层, 可以从各个局部信息中整合出全文语义, 模型的整体复杂度为 $O(n)$.

卷积神经网络应用非常广泛. 在自然语言领域, Collobert 等首次将其用于处理语义角色标注任务, 有效提升了系统的性能^[46]. 2014 年, Kalchbrenner 等与 Kim 分别发表了利用卷积神经网络做文本分类的论文^[47-48]. Zeng 等提出使用卷积神经网络做关系分类任务, 取得了一定的成功^[49].

图8 卷积神经网络模型结构图^[46]Fig. 8 Convolutional neural network architecture^[46]

3 基于表示学习的知识库问答

针对知识库问答, 近年来已经有一些研究者利用深度学习, 将表示学习应用其中. 这些方法的核心是把自然语言问句和知识库中的资源都映射到同一个低维向量空间中, 这样就可以将问句和答案(三元组)都用一个向量来表示, 知识库问答问题就被转化为求解向量相似度的问题.

Bordes 等^[10] 首先将基于词向量(Word embedding)的表示学习方法应用于知识库问答. 他将问句以及知识库中的三元组都转换为低维空间中的向量, 然后计算余弦相似度找出问句最有可能对应的答案三元组. 更具体地分为三个步骤:

1) 问句的向量表示为 $f(q) = V^T \phi(q)$, V 是词向量表示矩阵, $\phi(q)$ 表示哪些词在问句中出现过, 即问句中所有的词的向量直接相加.

2) 三元组的实体和关系的向量相加得到一个答案向量表示 $g(t) = W^T \psi(t)$, 其中 W 为知识库中实体和关系的向量表示矩阵, $\psi(t)$ 表示哪些实体和关系在三元组中出现了.

3) 最后计算这两个向量的相似度 $S(q, t) = f(q)^T g(t)$.

这种方法需要获得大量的问句-答案三元组对来训练, 以得到向量词典 V 和 W . 为了获得充足的训练语料, 其利用一系列人工设定的模板对已有的 Reberb^[50] 三元组进行扩展, 生成自然语言问句, 以弱监督的方式获取大量的训练数据. 例如, 已有三元组 (s, p, o) , 可以将 o 设为答案, 得到问句 “What does s p ?”. 而获取负样本的方法是随机破坏已有问句-答案三元组对中的三元组的元素. 训练目标是使得正样本的相似度得分大于负样本的得分加上一个间隔 0.1. 即:

$$\forall i, \forall t' \neq t_i, f(q_i)^T g(t_i) > 0.1 + f(q_i)^T g(t') \quad (11)$$

所以训练的损失函数如下:

$$L = [0.1 - f(q_i)^T g(t_i) + f(q_i)^T g(t')]_+ \quad (12)$$

该方法采用随机梯度下降法进行训练, 每一步更新 V 和 W . 与此同时, 利用 Paraphrasing 的语料进行多任务训练, 使得相似问句的向量更加相似, 以达到更好的训练效果. 这项工作在 Reverb 数据集上取得了不错的效果, F1 值达到 73%.

然而, 这一方法对于问句和知识库的语义分析十分粗糙, 仅仅是基于词、实体、关系的语义表示的简单求和. Bordes 等在文献 [11] 中对其进行了改进, 其基本假设是: 在答案端加入更多信息, 会提升问答的效果. 答案的表示可以分成三种: 1) 答案实体的向量表示; 2) 答案的路径(Path)的向量表示(和前面的工作用三元组表示的方法一样, 直接相加); 3) 和答案直接相关的实体和关系的向量表示, 这被称为子图向量表示(Subgraph embedding), 如图 9 所示.

同样地, 问句和答案的相似度表示如下:

$$S(q, a) = f(q)^T g(a) \quad (13)$$

其中, 问句的表示为 $f(q) = W^T \phi(q)$, 和前面的工作一样; 答案的表示为 $g(a) = W^T \varphi(a)$. $\varphi(a)$ 可以为上述的三种不同的表示方式. W 是向量表示矩阵, 自然语言的词汇以及知识库中的实体和关系都在这个表中. 这项工作的训练数据获取方式以及训练方法和文献 [10] 一样, 不同的是三元组是从 Freebase 中得到的. 在 WebQuestions 上的实验结果表明, 这种 Subgraph embedding 的方法是有效的, 其性能超过文献 [10].

Yih 等^[51] 把知识库问答转换成两个问题, 一个是找到问句中的实体和知识库中实体的对应; 另一个是问句中自然语言描述和知识库中语义关系的对应. 找到实体和关系后, 就可以从知识库中找到其指向的答案实体. 和已有基于表示学习的知识库问答方法不同之处在于, Yih 在进行上述两种匹配时都采用 CNN 来处理自然语言问句.

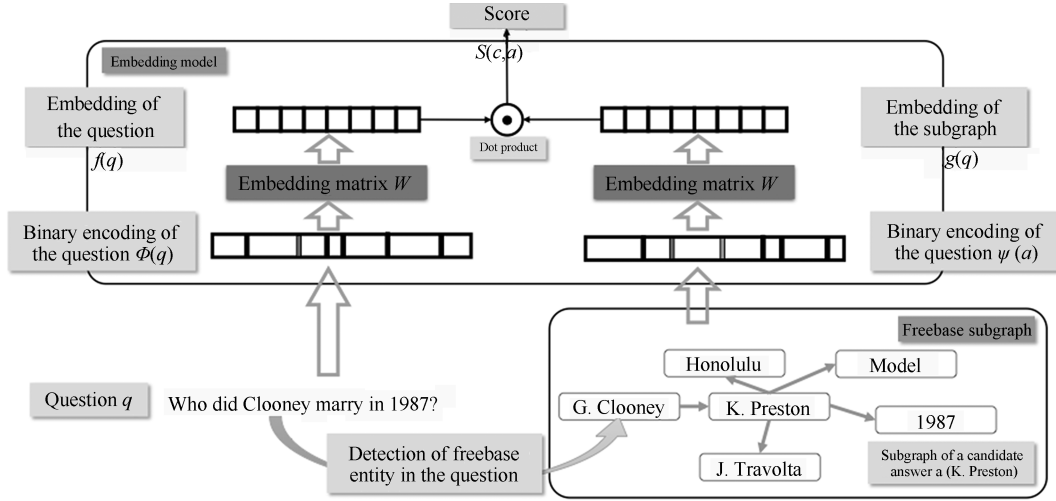
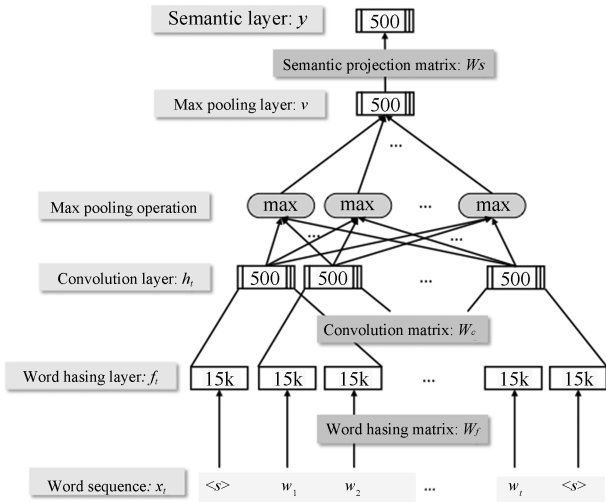
图 9 Subgraph embedding 模型^[11]Fig. 9 Subgraph embedding model^[11]

图 10 是基于 CNN 模型的知识库问答具体细节. 问句中的词经过一个哈希、卷积、最大池化和语义映射后成为一个向量表示. 相应的知识库中实体和关系用一个向量表示. 该工作的训练数据来自 Paralex^[52], 训练的目标是最大化后验概率 (此处仅给出 Relation pattern 匹配的目标函数, 实体匹配类似):

$$P(R^+|Q) = \frac{\exp(\gamma \cdot \cos(y_{R^+}, y_Q))}{\sum_{R'} \exp(\gamma \cdot \cos(y_{R'}, y_Q))} \quad (14)$$

在 Reverb 数据集上, 这项工作取得了比 Paralex^[52] 更好的效果, F 值为 0.57, 但是不如 Bordes 等^[10].

图 10 处理问句的 CNN 模型^[51]Fig. 10 CNN model used to process question^[51]

Dong 等^[12] 的思想和 Bordes 等^[11] 的 Subgraph embedding 相似, 同样是考虑了答案的更

多信息. 具体地, 答案类型、得到答案的路径以及答案周围的实体和关系这三种特征向量分别和问句向量做相似度计算, 最终的相似度为这三种相似度之和. 该方法在问句端的处理上使用了三个不同参数的 CNN 模型, 称为 Multi-column CNN, 图 11 为模型的具体细节.

$$S(q, a) = f_1(q)^T g_1(a) + f_2(q)^T g_2(a) + f_3(q)^T g_3(a) \quad (15)$$

其中 $f_1(q)^T g_1(a)$ 表示基于得到答案路径的相似度, $f_2(q)^T g_2(a)$ 表示基于答案周围实体和关系的相似度, $f_3(q)^T g_3(a)$ 表示基于答案类型的相似度.

训练方面, 仍然采用基于排序的方法, 损失函数为

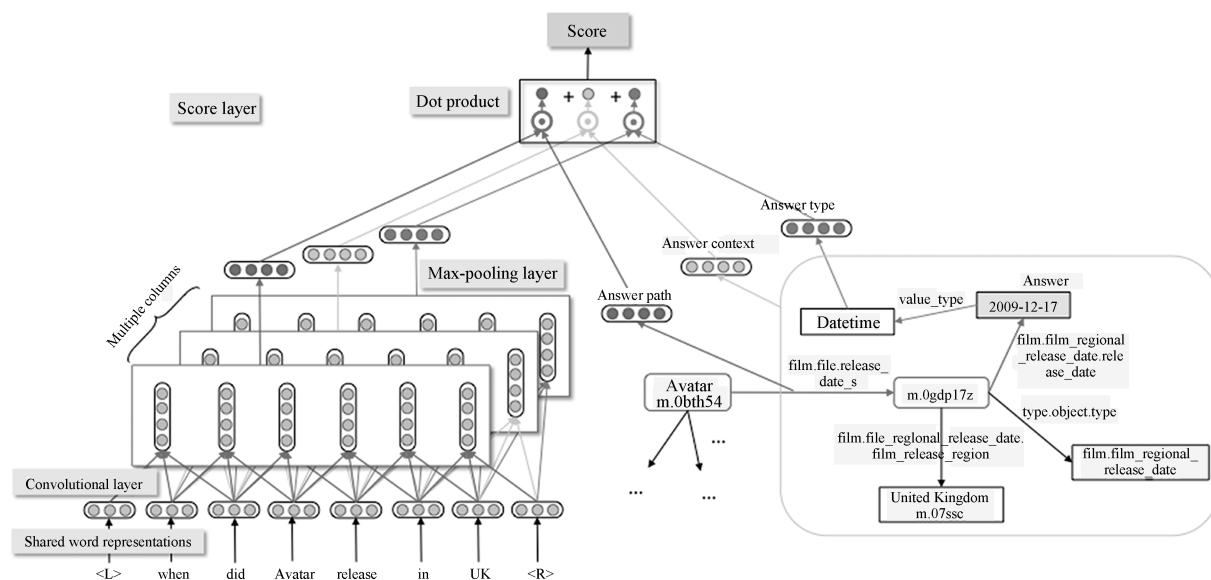
$$l(q, a, a') = [m - S(q, a) + S(q, a')]_+ \quad (16)$$

这项工作在 WebQuestions 上取得了比文献 [11, 53–54] 更好的效果, F 值为 40.8%.

从上面的几个代表性的工作来看, 基于表示学习的知识库问答并不能简单采用通用的表示学习方法对于文本和知识库的语义表示进行学习, 这样往往得不到好的效果, 需要对于问句的语义进行细致化的分析, 从多个角度考虑问句语义与知识库中实体、关系的匹配度, 例如答案类型、上下文相关性、语义关系匹配度等.

4 难点与挑战

由以上章节我们可以看出, 基于表示学习的知识库问答主要涉及知识库表示学习、文本表示学习以及基于表示学习的问答系统构建三方面任务. 这一类方法试图将传统基于符号的语义分析的知识库问答看做基于语义表示的语义匹配任务, 利用端到

图 11 Multi-column CNN 模型^[12]Fig. 11 Multi-column CNN model^[12]

端的思路解决问答问题, 从而省略中间步骤, 完全从数据中学习文本到知识之间的映射关系. 由于其采用基于学习到的语义向量进行语义匹配, 缺乏显式的问句语义解析, 相对于传统基于符号的问答方法, 能够有效地提高问答结果的召回率, 然而其准确率较以往方法存在不足. 在这三方面还存在如下难点和挑战, 亟待解决.

4.1 知识库表示学习的难点与挑战

目前, 知识库的表示学习方法主要有张量分解, 基于映射的模型等, 这些方法均具有较强的学习能力. 但是, 1) 在知识库中, 语义关系之间通常有一定的关联关系, 这种关系常常用于知识的推理或隐含知识的挖掘. 例如: “父亲”的“父亲”是“祖父”. 目前的知识库表示学习方法主要针对知识库中单个三元组进行建模, 对语义关系之间的关联关系考虑较少. Lin 等^[26] 已经将关系间的关系约束加入知识库的表示学习当中, 取得一定的效果. 但是如何将已有的常识知识 (如推理规则) 加入知识库表示学习过程中, 仍然是一个难点问题. 2) 当前的知识库表示学习方法主要集中在知识库内部的实体、关系的表示学习, 对于知识库未包含的实体、关系等语义单元, 不具有学习能力. 一个可能的解决途径是建立知识库和文本的联合表示学习模型. 这样可以从文本中挖掘知识库中尚不存在的实体和关系, 对知识库中的知识进行补充. 3) 知识库通常包含有噪声, 但是当前的方法均不考虑噪声的影响. 4) 在实际问答场景下, 单一知识库很难覆盖用户的问题, 在很多情况下, 需要综合多个异构知识库的知识回答当前问题. 然而, 目前的知识库表示学习方法都集中在单个知

识库上, 对于多知识库的表示学习的研究较少. 在学习过程中, 如何建立异构知识库间的实体对齐、关系对齐是一个尚待研究的问题.

4.2 文本表示学习的难点与挑战

现阶段文本表示学习主要有两个难点: 组合语义通用性和文本表示的时间效率.

1) 组合语义的通用性. 现有组合语义的方式, 无论是递归、循环还是卷积网络, 都采用一致的策略对文本元素进行组合. 比如递归神经网络会将句法结构上的左右两部分组合成一个整体, 对于所有的句法结构都采取同样的策略; 循环网络中, 对文本中的每一个词, 都会采用相同的策略对该词与其上下文进行语义组合; 卷积网络中, 对于每一段连续的短语片段, 也均采用同样的方式组合. 然而在一般的文本中, 这类简单的组合方式并不能完全表达语义. 例如, 句子中的不同成分的语素需要不同的组合方式, 比如形容词 + 名词合成的语义可能与两个名词合成名词短语的方式差异较大. 在这方面, Socher 等^[7] 在递归神经网络中尝试了对不同句法结构采用不同的组合函数; Zhao 等基于 LSTM 的思想提出了 AdaSent 模型^[55], 对不同语素的多样化组合进行了尝试. Luong 等^[56] 已经把 Attention-based neural network 应用于机器翻译, 并且取得了不错的效果. 其核心目标也是为了解决文本语义的组合问题.

2) 时间效率问题. 比如谷歌在网页去重时, 使用的是非常简单的 SimHash, 然而 SimHash 对同义词的建模较弱, 在短文本时, 对语义的建模并不理想. 但是这种方法运行效率比此前提到的递归、循环、卷积网络要高. 特别是递归神经网络, 时间复杂度与文

本长度的平方呈正比,无法投入到问答系统等大规模数据的使用场景下.因此寻找一种高效且有效的文本表示方法也是研究人员关注的难点和热点之一.

总体来看,文本表示目前处于简单模型不能很好地捕获各种语义,而复杂模型在提升语义捕获能力的同时,往往使模型变得非常复杂,难以实用.因此,无论是寻找有效的语义组合方式,更精确地捕获语义,还是寻找高效可用的文本表示方法,都将是下一步研究方向与热点.

4.3 基于表示学习的问答系统构建的难点与挑战

目前,基于表示学习的问答系统都是通过高质量已标注的问题-答案建立联合学习模型,同时训练知识库和文本的语义表示.但是仍然存在一些问题.1) 资源问题.表示学习的方法依赖大量的训练语料,而目前获取高质量的问题-答案对仍然是个瓶颈,如果进行人工标注代价昂贵. Bordes 等^[1]提出了一些模板利用已有三元组来生成问句,用较小的代价生成了大量的问题-答案对,但是相应的问句质量并不能保证,而且问句同质化严重.在训练资源上的提高空间仍然很大.2) 已有的基于表示学习的知识库问答方法多是针对简单问题,例如单关系问题设计的,对于复杂问题的回答能力尚且不足,例如有限制条件的问题(What did Obama do before he was elected president?)和聚合问题(When's the last time the steelers won the superbowl?).如何利用表示学习的方法解决复杂问题值得继续关注.3) 目前,大部分工作中所使用的知识库通常包含两类,一类是高质量的人工知识库,例如 Freebase;一类是通过自动抽取技术得到的开放知识库,例如 Reverb.后者尽管不如前者精准,但是知识覆盖程度更大.同时每一类中知识库的结构也存在差异性.综合不同知识库一起回答问题能够发挥各自知识库的优势.已经有传统的方法对此作出研究^[57],尚缺乏从表示学习的思路研究多领域异构知识库的问答方法.4) 在知识库问答过程中需要知识推理技术对于知识库中的未知知识进行预测与推理.传统推理方法基于谓词逻辑的推理策略,存在覆盖度低、推理速度慢的问题.如何利用表示学习自动学习推理规则,使其使用大规模知识库问答场景是一个值得研究的难点问题.

5 结论

本文对于基于表示学习的知识库问答方法进行详细的综述.特别是在知识库的表示学习、问句(文本)的表示学习以及基于表示学习的知识库问答系统构建等三方面介绍了目前主流的方法及其特点.同时,本文对于这三方面的主流方法存在的不足以及今后可能的研究难点和热点问题进行讨论.知识

库问答本身就是一个十分具有挑战性的问题,如今在深度学习的研究热潮中,如何利用基于深度学习的表示学习技术,构建面向知识库的深度问答系统,是一个极具挑战的研究课题.希望本文能够为这一领域的研究者提供一定的参考和启发.

References

- 1 Etzioni O. Search needs a shake-up. *Nature*, 2011, **476**(7358): 25–26
- 2 Lehmann J, Isele R, Jakob M, Jentzsch A, Kontokostas D, Mendes P N, Hellmann S, Morsey M, van Kleef P, Auer S, Bizer C. Dbpedia — a large-scale, multilingual knowledge base extracted from Wikipedia. *Semantic Web*, 2015, **6**(2): 167–195
- 3 Bollacker K, Evans C, Paritosh P, Sturge T, Taylor J. Freebase: a collaboratively created graph database for structuring human knowledge. In: *Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data*. Vancouver, Canada: ACM, 2008. 1247–1250
- 4 Suchanek F M, Kasneci G, Weikum G. Yago: a core of semantic knowledge. In: *Proceedings of the 16th International Conference on World Wide Web*. Alberta, Canada: ACM, 2007. 697–706
- 5 Kwiatkowski T, Zettlemoyer L, Goldwater S, Steedman M. Lexical generalization in CCG grammar induction for semantic parsing. In: *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. Scotland, UK: Association for Computational Linguistics, 2011. 1512–1523
- 6 Liang P, Jordan M I, Klein D. Learning dependency-based compositional semantics. *Computational Linguistics*, 2013, **39**(2): 389–446
- 7 Socher R, Perelygin A, Wu J Y, Chuang J, Manning C D, Ng A Y, Potts C. Recursive deep models for semantic compositionality over a sentiment treebank. In: *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*. Seattle, USA: Association for Computational Linguistics, 2013. 1631–1642
- 8 Cho K, Van Merriënboer B, Gülçehre C, Bahdanau D, Bougares F, Schwenk H, Bengio Y. Learning phrase representations using RNN encoder-decoder for statistical machine translation. In: *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*. Doha, Qatar: Association for Computational Linguistics, 2014. 1724–1734
- 9 Socher R, Manning C D, Ng A Y. Learning continuous phrase representations and syntactic parsing with recursive neural networks. In: *Proceedings of the NIPS-2010 Deep Learning and Unsupervised Feature Learning Workshop*. Vancouver, Canada, 2010. 1–9
- 10 Bordes A, Weston J, Usunier N. Open question answering with weakly supervised embedding models. *Machine Learning and Knowledge Discovery in Databases*. Berlin Heidelberg: Springer, 2014. 165–180
- 11 Bordes A, Chopra S, Weston J. Question answering with subgraph embeddings. In: *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*. Doha, Qatar: Association for Computational Linguistics, 2014. 615–620
- 12 Dong L, Wei F R, Zhou M, Xu K. Question answering over freebase with multi-column convolutional neural networks. In: *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International*

- Joint Conference on Natural Language Processing. Beijing, China: ACL, 2015. 260–269
- 13 Nickel M, Tresp V, Kriegel H P. A three-way model for collective learning on multi-relational data. In: Proceedings of the 28th International Conference on Machine Learning. Washington, USA, 2011. 809–816
 - 14 Nickel M, Tresp V, Kriegel H P. Factorizing YAGO: scalable machine learning for linked data. In: Proceedings of the 21st International Conference on World Wide Web. Lyon, France: ACM, 2012. 271–280
 - 15 Nickel M, Tresp V. Logistic tensor factorization for multi-relational data. In: Proceedings of the 30th International Conference on Machine Learning. Atlanta, USA, 2013.
 - 16 Nickel M, Tresp V. Tensor factorization for multi-relational learning. *Machine Learning and Knowledge Discovery in Databases*. Berlin Heidelberg: Springer, 2013. 617–621
 - 17 Nickel M, Murphy K, Tresp V, Gabrilovich E. A review of relational machine learning for knowledge graphs. *Proceedings of the IEEE*, 2016, **104**(1): 11–33
 - 18 Bordes A, Weston J, Collobert R, Bengio Y. Learning structured embeddings of knowledge bases. In: Proceedings of the 25th AAAI Conference on Artificial Intelligence. San Francisco, USA: AAAI, 2011. 301–306
 - 19 Bordes A, Glorot X, Weston J, Bengio Y. Joint learning of words and meaning representations for open-text semantic parsing. In: Proceedings of the 15th International Conference on Artificial Intelligence and Statistics. Cadiz, Spain, 2012. 127–135
 - 20 Bordes A, Glorot X, Weston J, Bengio Y. A semantic matching energy function for learning with multi-relational data. *Machine Learning*, 2014, **94**(2): 233–259
 - 21 Jenatton R, Le Roux N, Bordes A, Obozinski G. A latent factor model for highly multi-relational data. In: Proceedings of Advances in Neural Information Processing Systems 25. Lake Tahoe, Nevada, United States: Curran Associates, Inc., 2012. 3167–3175
 - 22 Sutskever I, Salakhutdinov R, Joshua B T. Modelling relational data using Bayesian clustered tensor factorization. In: Proceedings of Advances in Neural Information Processing Systems 22. Vancouver, Canada, 2009. 1821–1828
 - 23 Socher R, Chen D Q, Manning C D, Ng A Y. Reasoning with neural tensor networks for knowledge base completion. In: Proceedings of Advances in Neural Information Processing Systems 26. Stateline, USA: MIT Press, 2013. 926–934
 - 24 Bordes A, Usunier N, Garcia-Durán A, Weston J, Yakhnenko O. Translating embeddings for modeling multi-relational data. In: Proceedings of Advances in Neural Information Processing Systems 26. Stateline, USA: MIT Press, 2013. 2787–2795
 - 25 Wang Z, Zhang J W, Feng J L, Chen Z. Knowledge graph embedding by translating on hyperplanes. In: Proceedings of the 28th AAAI Conference on Artificial Intelligence. Québec, Canada: AAAI, 2014. 1112–1119
 - 26 Lin Y, Zhang J, Liu Z, Sun M, Liu Y, Zhu X. Learning entity and relation embeddings for knowledge graph completion. In: Proceedings of the 29th AAAI Conference on Artificial Intelligence. Austin, USA: AAAI, 2015. 2181–2187
 - 27 Ji G L, He S Z, Xu L H, Liu K, Zhao J. Knowledge graph embedding via dynamic mapping matrix. In: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics. Beijing, China: Association for Computational Linguistics, 2015. 687–696
 - 28 Lin Y K, Liu Z Y, Luan H B, Sun M S, Rao S W, Liu S. Modeling relation paths for representation learning of knowledge bases. In: Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. Lisbon, Portugal: Association for Computational Linguistics, 2015. 705–714
 - 29 Guo S, Wang Q, Wang B, Wang L H, Guo L. Semantically smooth knowledge graph embedding. In: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing. Beijing, China: Association for Computational Linguistics, 2015. 84–94
 - 30 He S Z, Liu K, Ji G L, Zhao J. Learning to represent knowledge graphs with Gaussian embedding. In: Proceedings of the 24th ACM International Conference on Information and Knowledge Management. Melbourne, Australia: ACM, 2015. 623–632.
 - 31 Frege G. Jber sinn und bedeutung. *Wittgenstein Studien*, 1892, **100**: 25–50
 - 32 Hermann K M. Distributed Representations for Compositional Semantics [Ph. D. dissertation], University of Oxford, 2014
 - 33 Socher R, Karpathy A, Le Q V, Manning C D, Ng A Y. Grounded compositional semantics for finding and describing images with sentences. *Transactions of the Association for Computational Linguistics*, 2014, **2**: 207–218
 - 34 Socher R, Huang E H, Pennin J, Manning C D, Ng A Y. Dynamic pooling and unfolding recursive autoencoders for paraphrase detection. In: Proceedings of Advances in Neural Information Processing Systems 24. Granada, Spain, 2011. 801–809
 - 35 Socher R, Pennington J, Huang E H, Ng A Y, Manning C D. Semi-supervised recursive autoencoders for predicting sentiment distributions. In: Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing. Scotland, UK: Association for Computational Linguistics, 2011. 151–161
 - 36 Socher R, Lin C C Y, Ng A Y, Manning C D. Parsing natural scenes and natural language with recursive neural networks. In: Proceedings of the 28th International Conference on Machine Learning. Bellevue, WA, USA, 2011. 129–136
 - 37 Socher R, Bauer J, Manning C D, Ng A Y. Parsing with compositional vector grammars. In: Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics. Sofia, Bulgaria: Association for Computational Linguistics, 2013. 455–465
 - 38 Mitchell J, Lapata M. Composition in distributional models of semantics. *Cognitive Science*, 2010, **34**(8): 1388–1429
 - 39 Socher R, Huval B, Manning C D, Ng A Y. Semantic compositionality through recursive matrix-vector spaces. In: Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning. Jeju Island, Korea: Association for Computational Linguistics, 2012. 1201–1211
 - 40 Elman J L. Finding structure in time. *Cognitive Science*, 1990, **14**(2): 179–211
 - 41 Bengio Y, Simard P, Frasconi P. Learning long-term dependencies with gradient descent is difficult. *IEEE Transactions on Neural Networks*, 1994, **5**(2): 157–166
 - 42 Hochreiter S. The vanishing gradient problem during learning recurrent neural nets and problem solutions. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 1998, **6**(2): 107–116
 - 43 Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Computation*, 1997, **9**(8): 1735–1780

- 44 Fukushima K. Neocognitron: a self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, 1980, **36**(4): 193–202
- 45 LeCun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 1998, **86**(11): 2278–2324
- 46 Collobert R, Weston J, Bottou L, Karlen M, Kavukcuoglu K, Kuksa P. Natural language processing (almost) from scratch. *The Journal of Machine Learning Research*, 2011, **12**: 2493–2537
- 47 Kalchbrenner N, Grefenstette E, Blunsom P. A convolutional neural network for modelling sentences. In: *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*. Baltimore, USA: Association for Computational Linguistics, 2014. 655–665
- 48 Kim Y. Convolutional neural networks for sentence classification. In: *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*. Doha, Qatar: Association for Computational Linguistics, 2014. 1746–1751
- 49 Zeng D H, Liu K, Lai S W, Zhou G Y, Zhao J. Relation classification via convolutional deep neural network. In: *Proceedings of the 25th International Conference on Computational Linguistics*. Dublin, Ireland: Association for Computational Linguistics, 2014. 2335–2344
- 50 Fader A, Soderland S, Etzioni O. Identifying relations for open information extraction. In: *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*. Scotland, UK: Association for Computational Linguistics, 2011. 1535–1545
- 51 Yih W T, He X D, Meek C. Semantic parsing for single-relation question answering. In: *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistic*. Baltimore, USA: Association for Computational Linguistic, 2014. 643–648
- 52 Fader A, Zettlemoyer L S, Etzioni O. Paraphrase-driven learning for open question answering. In: *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*. Sofia, Bulgaria: Association for Computational Linguistics, 2013. 1608–1618
- 53 Berant J, Liang P. Semantic parsing via paraphrasing. In: *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistic*. Baltimore, USA: Association for Computational Linguistic, 2014. 1415–1425
- 54 Yao X C, Van Durme B. Information extraction over structured data: question answering with freebase. In: *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistic*. Baltimore, USA: Association for Computational Linguistic, 2014. 956–966
- 55 Zhao H, Lu Z D, Poupard P. Self-adaptive hierarchical sentence model. In: *Proceedings of the 24th International Conference on Artificial Intelligence*. Buenos Aires, Argentina: AAAI, 2015. 4069–4076
- 56 Luong M T, Pham H, Manning C D. Effective approaches to attention-based neural machine translation. In: *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*. Lisbon, Portugal: ACL, 2015. 1412–1421
- 57 Fader A, Zettlemoyer L, Etzioni O. Open question answering over curated and extracted knowledge bases. In: *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York, USA: ACM, 2014. 1156–1165



刘 康 中国科学院自动化研究所副研究员. 主要研究方向为问答系统, 观点挖掘, 自然语言处理. 本文通信作者.

E-mail: kliu@nlpr.ia.ac.cn

(LIU Kang Associate professor at the Institute of Automation, Chinese Academy of Sciences. His research interest covers question answering, opinion mining, and natural language processing. Corresponding author of this paper.)



张元哲 中国科学院自动化研究所博士研究生. 主要研究方向为问答系统和自然语言处理.

E-mail: yzzhang@nlpr.ia.ac.cn

(ZHANG Yuan-Zhe Ph.D. candidate at the Institute of Automation, Chinese Academy of Sciences. His research interest covers question answering and natural language processing.)



纪国良 中国科学院自动化研究所博士研究生. 主要研究方向为知识工程和自然语言处理.

E-mail: guoliang.ji@nlpr.ia.ac.cn

(JI Guo-Liang Ph.D. candidate at the Institute of Automation, Chinese Academy of Sciences. His research interest covers knowledge engineering and natural language processing.)



来斯惟 中国科学院自动化研究所博士研究生. 主要研究方向为表示学习和自然语言处理.

E-mail: swlai@nlpr.ia.ac.cn

(LAI Si-Wei Ph.D. candidate at the Institute of Automation, Chinese Academy of Sciences. His research interest covers representation learning and natural language processing.)



赵 军 中国科学院自动化研究所研究员. 主要研究方向为信息检索, 信息提取, 网络挖掘, 问答系统.

E-mail: jzhao@nlpr.ia.ac.cn

(ZHAO Jun Professor at the Institute of Automation, Chinese Academy of Sciences. His research interest covers information retrieval, information extraction, web mining and question answering.)