

# 基于知识图谱的多轮对话技术研究综述

杨阳<sup>1,2</sup>, 盛胜利<sup>3</sup>, 奚雪峰<sup>1,2</sup>

1. 苏州科技大学 电子与信息工程学院, 江苏 苏州 215009;
2. 苏州市虚拟现实智能交互及应用重点实验室(苏州科技大学), 江苏 苏州 215009;
3. 数据分析实验室(德州理工大学), 德克萨斯州 拉伯克市 79409)

**摘要:**随着自然语言技术的不断进步与发展,人机交互取得了跨越式的进步。然而,目前人机交互系统往往都是用户与机器双方在特定的应用场景下设计完成的,在开放域下进行难度较大的多轮对话效果差强人意。而知识图谱作为实现对话系统的重要工具之一,其被证明在多轮对话任务中是有效的。该文从基于知识图谱的多轮对话技术总结了多轮对话中使用的相关技术,其中基于知识图谱的多轮对话模型包括 TransE、TransH、TransR 和 TransD 等,以及涉及到基于知识图谱的多轮对话相关数据集及评价标准。最后提出了基于知识图谱的多轮对话技术当前面临的挑战并进行了总结。

**关键词:**知识图谱;多轮对话;人机交互;自然语言处理;对话系统

中图分类号:TP18

文献标识码:A

文章编号:1673-629X(2023)04-0027-07

doi:10.3969/j.issn.1673-629X.2023.04.004

## Recovery of Multi-turn Dialogue Based on Knowledge Graph

YANG Yang<sup>1,2</sup>, SHENG Sheng-li<sup>3</sup>, XI Xue-feng<sup>1,2</sup>

1. School of Electronic and Information Engineering, Suzhou University of Science and Technology, Suzhou 215009, China;
2. Suzhou Key Laboratory of Virtual Reality Intelligent Interaction and Application Technology, Suzhou University of Science and Technology, Suzhou 215009, China;
3. Data Analytics Lab, Soochow University (Texas Tech University), Lubbock 79409, USA)

**Abstract:** With the continuous progress and development of natural language technology, human-computer interaction has made a great progress. However, at present, human-computer interaction systems are often designed by both users and machines in specific application scenarios, and the effect of difficult multi-turn dialogue in open domain is not satisfactory. As one of the important tools to implement dialogue system, knowledge graph has been proved to be effective in multi-turn dialogue tasks. We summarize the related technologies used in multi-turn dialogue from the multi-turn dialogue technology based on knowledge graph. The multi-turn dialogue model based on knowledge graph includes TransE, TransH, TransR and TransD, as well as the related data sets and evaluation standards of multi-turn dialogue based on knowledge graph. Finally, we put forward the current challenges of multi-turn dialogue technology based on knowledge graph and make a summary.

**Key words:** knowledge graph; multi-turn dialogue; human-computer interaction; natural language processing; dialogue system

## 0 引言

近年来,随着自然语言处理技术和人工智能的蓬勃发展,从备受关注的人机对战到贯穿人们日常生活的无人驾驶、智能家居等,人工智能正在逐渐地改变人们的日常生活<sup>[1]</sup>。例如,通过 WIFI 等接入方式与一些智能家居设备相连并进行控制操作<sup>[2-3]</sup>。得益于深

度学习技术的不断完善与发展,加上使用海量的大数据和云计算的不断普及,自然语言处理技术和语音识别技术的判断准确率得到了明显的提高。

对话系统得益于自然语言技术的不断成熟,其意在能够精准猜测出用户的意图,并且快速准确地回答问题。当然对话不可能只是一问一答,现实中对话一

收稿日期:2022-06-17

修回日期:2022-10-19

基金项目:国家自然科学基金项目(61876217,62176175);江苏省“六大人才高峰”高层次人才项目资助(XYDXX-086)

作者简介:杨阳(1996-),男,硕士研究生,CCF会员(D3192G),研究方向为自然语言处理、知识图谱;盛胜利(1966-),男,教授,博士,研究方向为自然语言处理;通信作者:奚雪峰(1978-),男,副教授,博士,CCF高级会员(16888S),研究方向为自然语言处理。

定是多问多答,并且是连贯的、合乎逻辑的。然而目前多轮对话仍然存在诸多问题,例如缺乏相对应的知识推理能力,多轮对话的能力。在实际的人机对话过程中,必须进行多方考虑,目前部分机器学习的实际应用案例以及相关的技术研究还仅仅处在一个“低级智能”阶段,为了使机器的回复更精准、更高效、更有趣,从“低级智能”向“高级智能”阶段进一步发展,因此许多研究人员努力将知识图谱技术应用到多轮对话中。基于知识图谱的多轮对话是结合实体和实体间的关系,将知识三元组和原始对话上下文数据融合在一起,更好地实现多轮对话。

## 1 问题定义

根据当前的应用场景(包括开放域和特定域),目前的对话系统主要分为单轮对话系统和多轮对话系统。

单轮对话的定义公式为:  $\gamma = g(q)$  ;

多轮对话的定义公式为:  $\gamma = g(q | c)$  。

以上公式中,  $q$  表示查询语句;  $\gamma$  表示答复语句;  $c$  表示历史上下文对话信息;  $g(*)$  表示从查询语句中对应答复的数据库中经过筛选匹配后给出最合适数据的结果。

## 2 多轮对话中使用的相关技术

随着自然语言处理技术的不断成熟与发展,当前对自然语言处理的研究侧重点已经转移到深度学习,深度学习是加深了层的深度神经网络,是研究人员模仿人类大脑的工作运转模式,通过计算模拟出人工神经元以此来组成人工神经网络,解决一系列复杂问题。多轮对话中的主要技术有词向量技术、知识图谱等。

### 2.1 词向量技术

词向量技术为人类自然语言和计算机语言处理提供了一座桥梁,让原本零散的、孤立的、稀疏的词序列演变成成为密集的语义向量,为人工神经网络更好地理解人类自然语言提供了有力的保障。现实中自然语言符号往往采用高维度向量表示,为了方便计算机读取和处理数据,需要采用相对应的技术手段将这些零散的文字进行向量化,即转为词向量。目前采用的方法主要有 One-Hot 编码和词的分布式表达法。

One-Hot 编码主要采用的是将文章中所有出现的词汇融合成一个字典,其目的是为每一个字典中出现的词设定一定的维度并确定为词向量。该向量的要求为词对应的位置上为“1”,即每一个词在指定的并相同的部分为 1,其他部分均为 0。例如,在一个包含 10 000 个词的词典,该词典里包含了所有的交通工具,

例如“高铁”“地铁”,那么这两个词可以表示为:“高铁”:[0,0,0,0,0,0,0,1,0,0],“地铁”:[0,0,0,1,0,0,0,0,0,0]。但由于自然语言字词符号等数量庞大,因此相应的词表也会非常大,从而造成了资源的极大浪费。

鉴于 One-Hot 编码的缺点及其发展的局限性,研究人员在表示词向量的维度上做出了很多改进。词的分布式表达法是目前广泛用于自然语言处理(Natural Language Processing, NLP)任务中的一种新范式。该方法最初在 1986 年由 David E Rumelhart、Geoffrey E Hinton、Ronald J Williams 提出,它克服了 One-Hot 编码的缺点,并具有很优势。其旨在以固定长度、连续和密集的特征向量来表示单词。此外, Mikolov 等人<sup>[4-5]</sup>提出了一种使用神经网络学习分布式词向量表示(Word2Vec)的模型架构,这种技术目前主要用于自然语言技术中捕捉语义和语法之间的关系。其他的分布式单词表示法有语义分析法<sup>[6]</sup>和潜在 Dirichlet 分配法<sup>[7]</sup>。

### 2.2 知识图谱技术

由于在多轮对话过程中,多轮对话需要深度结合历史对话信息,确保对话上下文具有连贯性,针对提问者提出来的问题,从答复语料库中选取匹配相似度最相关的回复。为了解决开放域多轮对话中存在的问题,研究人员建议将有关对话的先验信息表示为图,旨在建立更加连贯的对话。基于知识图谱<sup>[8]</sup>的多轮对话是给定与当前对话相关的知识三元组,以当前用户语句和对话历史信息作为输入,以合适的回答作为输出。

在基于知识图谱的多轮对话中,其主要采用的是 End-to-end<sup>[9]</sup>的生成模型来构建系统模型。在系统建模时,主要工作为对知识图谱、历史对话和用户当前语句进行建模,与此同时针对这三个部分的建模信息进行相对应的解码,以产生相对应的回复。

知识图谱的技术架构如图 1 所示。

## 3 基于知识图谱的多轮对话模型

### 3.1 预训练模型

深度上下文语言模型(Deep Context Model, DCM)被证明在语言表达方面是有效的,在一系列的自然语言处理模型中取得了最先进的结果。例如语言模型嵌入<sup>[10]</sup>、广义自回归预训练<sup>[11]</sup>、鲁棒优化伯特预训练方法<sup>[12]</sup>和 AL-BERT<sup>[13]</sup>。通过提供细粒度的上下文信息,这些预训练的模型既可以作为编码器轻松应用于下游模型,也可以用于微调,这些 DCM 在表示特定领域语料库中的上下文化信息方面仍然受到限制,因为它们通常在通用的语料库中进行训练<sup>[14-15]</sup>。

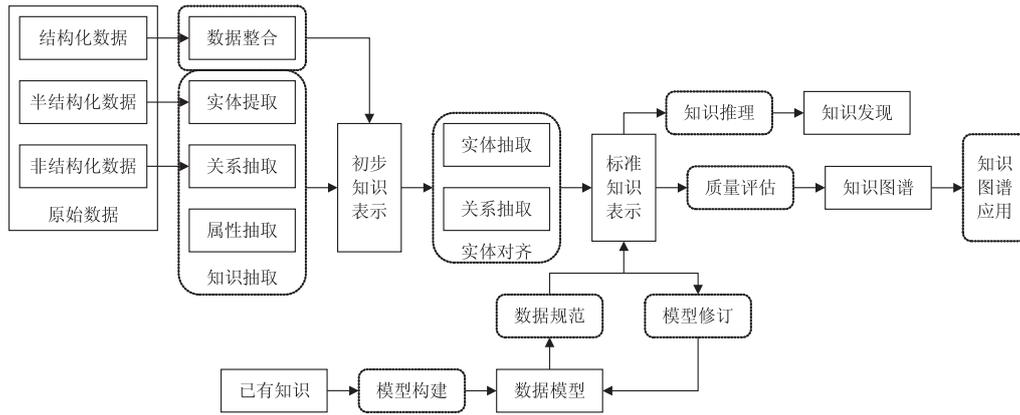


图 1 知识图谱的技术架构

### 3.2 多轮对话建模

开发对话系统意味着训练机器使用自然语言与人类对话。目前已经设计了许多对话系统,主要分为两种类型的架构:一种连接所有上下文语句<sup>[16-18]</sup>,另一种分离并聚合上下文语句<sup>[19-24]</sup>。

### 3.3 基线模型

#### 3.3.1 基于后验生成的模型

Liu 等人<sup>[25]</sup>在 2017 年提出了采用先验分布来逼近后验分布的方法,其原理为从候选的知识图谱库中选择出最合适的知识三元组,生成相对应的回复语句。先验分布是根据用户的提问来选择知识,后验概率主要是根据回复和用户输入选择知识。模型由四个部分组成,包含上下文编码器、知识编码器、知识管理器和解码器。上下文编码器是将用户当前语句和历史对话信息作为输入,记为  $I$ ,输出的为上下文向量,记为  $i$ 。知识编码器是与当前产生的所有对话产生的知识图谱和产生的回复作为输入,最终输出结果为每一个知识的三元组向量和回复向量。知识管理器是以上下文编码器和知识编码器的输出结果作为输入,输出经过大量筛选后的知识向量。同理,解码器是以上下文编码器和知识管理器的输出结果作为输入,最后生成回复。

#### 3.3.2 基于注意力机制的后验生成的模型

Bahdanau 等人<sup>[26]</sup>在 2014 年在机器翻译任务中引入注意力机制,大幅度地提升了机器翻译的性能。于是后期研究人员将注意力机制广泛应用于自然语言处理的各个领域。在基于序列到序列的生成任务中,可以根据当前解码状态去选择最相关的原始端信息,来提升解码效果,同时该方法也符合人类思考的思维习惯:关注重点信息,选择性忽视不重要的信息。可以将注意力机制分为三个部分,分别为查询语句  $q$ , 键项  $k = \{k_1, k_2, \dots, k_n\}$ , 值项  $v = \{v_1, v_2, \dots, v_n\}$ , 注意力机制的计算公式如公式 1 ~ 公式 3 所示。

$$e_i = \text{score}(q, k_i) \quad (1)$$

$$a_i = \frac{\exp(e_i)}{\sum_{j=1}^n \exp(e_j)} \quad (2)$$

$$c = \sum_{i=1}^n a_i v_i \quad (3)$$

目前注意力机制采取的得分函数有若干种计算方式。Luong 等<sup>[27]</sup>研究人员提出了三种注意力机制计算方法,如公式 4 所示。

$$\text{score}(q, k_i) = \begin{cases} q^T k_i & \text{dot} \\ q^T W_a k_i & \text{general} \\ V_a^T \tanh(W_a [q; k_i]) & \text{concat} \end{cases} \quad (4)$$

其中, dot 计算公式中要求有相同维度的  $q$  和  $k_i$ , 而 general 计算公式主要通过可变参数  $W_a$  使得  $q$  和  $k_i$  的维度保持不同, concat 的计算方式主要采用的是将  $q$  和  $k_i$  拼接联系起来,通过使用一个单层感知机和可变参数  $V_a$  计算得出最后的分数。Vaswani 等<sup>[28]</sup>研究人员提出一种新的计算手段-放缩点积,计算公式如公式 5 所示。

$$\text{score}(q, k_i) = \frac{q^T k_i}{\sqrt{n}} \quad (5)$$

其中  $n$  和  $k_i$  的维度,此外公式中加上  $\sqrt{n}$  这个放缩因子,是考虑到当  $k_i$  的维度很大时,随之而来  $q$  和  $k_i$  点积的结果也会变得很大,会造成 softmax 函数得到的梯度非常小,非常不利于反向传播,最终导致很难学习好网络参数。因此,考虑加入缩放因子这一策略,在一定程度上可以缓解这个问题。

可以清楚地发现:基于后验生成的模型和基于注意力机制的后验生成的模型无明显差别,其唯一存在的区别就是在解码时加入注意力机制这一手段。使用解码时上一时刻状态  $s_{t-1}$  以及上下文编码器中的  $h_i$  使用注意力机制可以得到的结果为  $c_t$ , 然后根据  $c_t$  以及知识管理器的输出  $k$  和  $s_{t-1}$ , 结果产生下一时刻的解码状态  $s_t$ , 最后判断当前时刻解码状态  $s_t$  以及上下文

向量  $x$  输出当前时刻解码单词, 计算过程如公式 6 ~ 公式 10 所示。

$$e_i = V_a^T \tanh(W_a [q; h_i]) \quad (6)$$

$$a_i = \frac{\exp(e_i)}{\sum_{j=1}^n \exp(e_j)} \quad (7)$$

$$c = \sum_{i=1}^n a_i h_i \quad (8)$$

$$s_i = \text{GRU}([y_{i-1}; \bar{k}], s_{i-1}, x) \quad (9)$$

$$Y_i = \text{softmax}(s_i, x) \quad (10)$$

### 3.4 层次上下文建模的后验生成模型

#### 3.4.1 基于层次上下文建模的后验生成模型

上述讨论的基于后验生成基线模型中, 将历史对话上下文信息拼接成一句话, 最后对拼接的句子建模。但是存在的问题是: 上下文信息不可能是单独的一句话, 一般是由多句话组成。并且拼接的句子语句长度较长, 这种建模方式会产生历史上下文对话中最前面的几句话被忽略, 导致上下文信息缺失, 向量信息不充分。为了解决此问题, 研究人员采用了层次上下文建模方式。其主要原理为: 先使用一个编码器建模上下文中每一句话的信息, 再使用另一个编码器建模所有的句子信息, 最后生成上下文信息。可以发现: 基于层次上下文建模的后验生成的模型和基于后验生成的模型无明显差别, 各模块基本相同, 其唯一存在的区别就是编码器不同。一个双向的门控循环单元 (Gate Recurrent Unit, GRU) 和一个单向的 GRU 组合构成基于层次上下文建模的后验生成的模型的编码器。例如给定上下文  $X = \{u_1, u_2, \dots, u_n\}$ , 其中  $u_i$  表示上下文  $X$  中的第  $i$  句话;  $U_i = \{w_{i1}, w_{i2}, \dots, w_{in}\}$ , 其中  $w_{in}$  表示  $U_i$  中的第  $j$  个词。Utterance Encoder 是一个双向 GRU, 主要功能是将历史对话上下文中每一句话  $\{u_1, u_2, \dots, u_n\}$  编码形成句子向量  $\{u_1, u_2, \dots, u_n\}$ 。Context Encoder 是一个 GRU, 主要功能是将句子向量编码成上下文向量  $x$ , 计算步骤如公式 11 ~ 公式 12 所示。

$$u_i = [\vec{h}_{iT}; \overleftarrow{h}_{i1}] = [\text{GRU}(w_{iT}, \vec{h}_{iT-1}); \text{GRU}(w_{i1}, \overleftarrow{h}_{i2})] \quad (11)$$

$$x = s_n = \text{GRU}(u_n, s_{n-1}) \quad (12)$$

#### 3.4.2 融合层次上下文建模和注意力机制的后验生成模型

在多轮对话系统中, 解码需要充分考虑上下文信息。即解码每一个词时, 需要提前判断历史对话中哪些话需要考虑进来, 并且是重要的。其次, 需要充分考虑历史对话中哪些词对解码的意义至关重要。融合层次上下文建模和注意力机制的后验生成模型和基于层次上下文建模的后验生成模型无明显差别, 其唯一存

在的区别在于上下文编码器和注意力使用机制有所不同。

上下文编码器主要采用层次上下文方法进行建模, 其模型结构和基于层次上下文建模的后验生成模型的上下文编码器大体相同。需要注意的是, 模型中不仅将注意力机制用于 Utterance Encoder 的输出, 而且也将该技术运用于 Context Encoder 的输出。Utterance Encoder 的输出表示为输出每个词的状态, 将其定义为词级别的注意力机制 (Word Attention)。Context Encoder 的输出表示为上下文中每句话的状态, 将其定义为句子级别的注意力机制 (Sentence Attention)。在使用中需要注意词级别的注意力机制和句子级别的注意力机制的相同之处以及不同之处, 并做好区分。例如, 给定上下文  $X = \{u_1, u_2, \dots, u_n\}$ , Utterance Encoder 的功能是将历史对话中每一句话  $U_i$  编码生成词向量  $H_i = \{h_{i1}, h_{i2}, \dots, h_{in}\}$  和对应句子向量  $u_i, u_i$  计算公式如公式 11 所示。Context Encoder 的功能是将句子向量  $C = \{c_1, c_2, \dots, c_n\}$  编码生成  $\{s_1, s_2, \dots, s_n\}$ 。在解码时使用上一时刻状态  $s_{i-1}$  并分别对所有的词向量和句子向量使用注意力机制, 最后得到的结果为  $c_i$ , 计算步骤如公式 13 ~ 公式 15 所示。

$$c_{i1} = \text{Attention}(s_i, H) \quad (13)$$

$$c_{i2} = \text{Attention}(s_i, C) \quad (14)$$

$$c_i = \text{MLP}([c_{i1}; c_{i2}]) \quad (15)$$

其中, MLP 表示全连接层,  $s_i$  表示当前解码状态,  $H$  表示上下文所有词的隐藏状态。最后由  $c_i$ 、知识管理器的输出  $\bar{k}$  和  $s_{i-1}$  得到  $s_i$ , 计算步骤如公式 9 所示。

### 3.5 基于 TransD 后验生成模型

在基于知识图谱的多轮对话中, 需要将知识图谱建模并且融入到多轮对话系统中。基于注意力机制的后验生成的模型的主要原理为将历史对话中与当前对话相关的知识三元组相互连接组成一句话, 然后将连接后的知识图谱使用双向 GRU 进行建模。该建模方式存在的缺点为没有充分利用实体间的关系, 导致知识图谱的建模信息不完整。为了解决该问题, 研究人员引入知识表示学习, 在知识图谱量化时, 也充分考虑了各个实体和关系之间的语义关系。

知识图谱主要包含实体及其对应关系的图结构网络, 图中的每一条边都表示为一个知识三元组, 即头实体、关系和尾实体。两个实体之间通过特定的关系相连, 每个实体可以和若干个实体相连, 每个关系也可以出现在若干个不同的三元组中。Bordes 等人<sup>[29]</sup>最早提出了 TransE 模型, 其主要采用的方法是将实体和关系映射到向量空间。同时研究人员借鉴 word2vec 的平移不变性的特性, 可以将知识图谱中的关系看作是

实体间的某种平移不变量。即针对一个已经建立好的知识图谱三元组  $(h, r, t)$ , 其中  $h$  表示头实体,  $r$  表示关系,  $(h, r, t)$  表示尾实体, TransE 的作用原理主要是将每个三元组  $(h, r, t)$  映射到向量空间, 使得  $h + r$  的值和  $t$  尽可能靠近, 如图 2 所示。

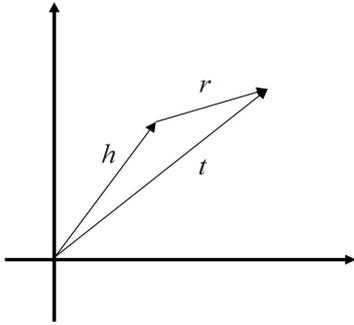


图 2 TransE 模型

同时在 TransE 模型训练过程中, 可以构造出大量的负样本, 要求负样本中的  $h' + r'$  和  $t'$  尽可能背离。具体损失函数公式如公式 16 所示:

$$L = \sum_{(h,r,t) \in S} \sum_{(h',r',t')} [\gamma + d(h+r,t) - d(h'+r',t')] \quad (16)$$

TransE 模型中参数较少, 计算复杂程度低, 可以很好解决一对一的问题, 但是无法解决一对多和多对一的问题。为了解决该问题, Wang 等人<sup>[30]</sup> 提出 TransH 模型, TransH 主要原理为针对每一组关系都定义一个超平面  $W_r$  和关系向量  $d_r$ , 主要采用的方法为将头实体和尾实体通过超平面映射到关系平面上, 这样可以将关系平面中头实体向量加关系向量接近于尾实体向量, 如公式 17 所示。

$$f(h, t) = h - w_r^T h w_r + d_r - t + w_r^T t w_r \quad (17)$$

TransH 使得不同的实体在不同的关系中可以用不同的向量来表示, 同时也可以使得不同的实体在同一关系中, 使用关系向量的映射, 得到不同的向量表示, 如图 3 所示。因此 TransH 相对于 TransE, 可以很好地解决一对多和多对一的情况。

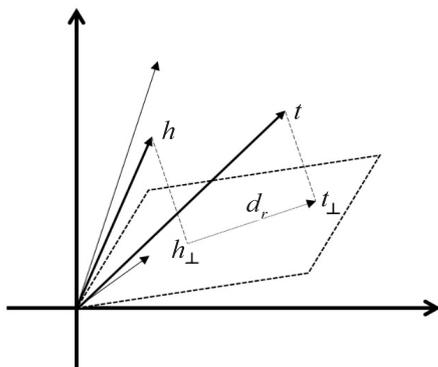


图 3 TransH 模型

TransE 和 TransH 的应用场景均为假设实体和关

系运用于同一向量空间中, 但是实体和关系表示为不同的对象, 可能运用于不同的向量空间中, 这就会导致在不同的向量空间中无法表示的可能性。为了解决该问题, Lin 等人<sup>[31]</sup> 提出 TransR 模型。TransR 主要原理为一个实体可以包含多种属性, 不同的关系需要考虑实体属性也不同, 因此不同的关系对应映射到不同的语义空间。在 TransR 模型中, 实体和关系均有不同的向量维度, 针对每一个关系都定义一个语义空间  $M_r$ , 它采用的是将实体映射到关系空间。然后在关系空间内, 使头实体加关系尽可能接近尾实体, 如公式 18 ~ 公式 19 所示。

$$h_r \equiv h M_r, t_r \equiv t M_r \quad (18)$$

$$f_r(h, t) \equiv h_r + r - t_r \quad (19)$$

在 TransE、TransH 和 TransR 模型中, 都是假设每种模型中每种关系仅有一层含义, 但是同一种关系可能存在多层含义。例如, relationship 可以表示人和某个团体之间的关系, 也可以表示人和某个国家之间的关系。在 TransR 模型(见图 4)中, 针对同一个关系, 头实体和尾实体可共享同一投影矩阵, 但头实体和尾实体所包含的对应属性可能存在明显的差异。为了解决该问题, Ji 等人<sup>[32]</sup> 提出 TransD 模型, TransD 主要原理基于动态矩阵生成, 对应的生成映射矩阵由实体向量和关系向量两部分组成。这样生成的优势为在同一种关系下在不同的头实体和尾实体所对应的映射矩阵完全独立且都不相同, 如公式 20 ~ 公式 22 所示。对于每一个三元组  $(h, r, t)$  都分别对应于两个向量, 即  $h, h_p, r, r_p, t, t_p$ , 一个可以用来表示它们的具体含义, 另一个用于构造出相对应的映射矩阵,  $I$  表示单位矩阵。TransD 模型中映射矩阵包含实体向量和关系向量, 因此大幅度提高了 TransD 计算复杂度, 并远小于 TransR。

$$M_{rh} = r_p h_p^T + I, M_{rt} = r_p t_p^T + I \quad (20)$$

$$h_{\perp} = M_{rh} h, t_{\perp} = M_{rt} t \quad (21)$$

$$f_r(h, t) = h_{\perp} - t_{\perp} \quad (22)$$

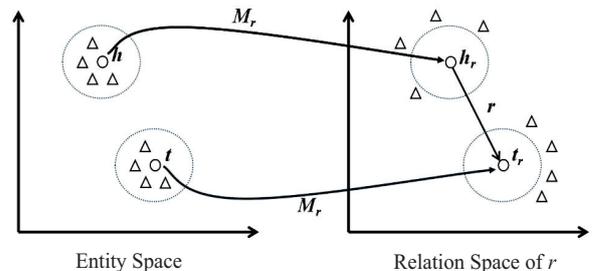


图 4 TransR 模型

与 TransE、TransH 和 TransR 相比, TransD(见图 5)计算复杂度小, 模型优化效果明显。基于 TransD 后验生成模型与基于注意力机制的后验生成模型相比,

其主要区别在于知识编码器不同以及编码器数量有所不同。

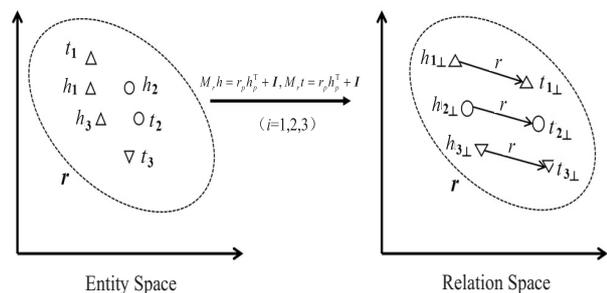


图 5 TransD 模型

使用 TransD 模型将所有的知识图谱向量化,即针对每一个知识三元组  $K_i = \{h_i, r_i, t_i\}$ ,将头实体向量  $h_i$ 、关系向量  $r_i$  和尾实体向量  $t_i$  进行相连,最后经过全连接层生成  $k_i$ ,其计算公式如公式 23 所示。

$$k_i = \text{FUN}([h_i, r_i, t_i]) \quad (23)$$

## 4 数据集及评价指标

### 4.1 数据集介绍

DuConv 数据集于 2019 年由 Wu 等人<sup>[33]</sup>在 ACL2019<sup>[34]</sup>首次公开,主要涵盖内容为基于知识图谱的多轮对话数据集。Ubuntu 数据集在 2015 年由 Lowe 等人首次公开,主要涵盖内容为一个大规模开放域下对话生成的数据集,其广泛用于对话生成任务中<sup>[35-36]</sup>。Cui 等人<sup>[37]</sup>于 ACL2020 发表了多轮对话与推理的数据集 MuTual,针对性强地用于评价模型在多轮对话过程中的逻辑与推理能力。

### 4.2 评价标准

针对基于知识图谱的多轮对话的实验结果,研究者主要采用以下两个客观指标进行评价。

(1) 平均对话轮数。对话轮数指的是从语句输入到最终对话结束一共持续的对话轮数。当对话中出现类似于“嗯嗯”“好的”等这些人为实现定义的回复或者对话一直是重复的无效的,系统则默认对话过程已经结束。

(2) 多样性。采用统计方法模拟在对话过程中产生出来的各种互相不重叠关系的一元文法(unigram)和二元文法(bigram)之间所约占比的百分比,以进一步反映对话结果产生的复杂性。一元文法与二元文法都是语言模型理论中重要的概念,这种指标通常可以用来判断表示出其最终的输出结果的语言模型丰富程度的程度。由于多轮对话回答问题的特殊性,其回答结果也不存在绝对唯一性,故也可以同时出现一个问题或出现多个回答。因此,传统的 BLUE 等传统评价方式也不适用于针对多轮值对话结果进行的评价。

## 5 结束语

对话系统是自然语言处理以及人工智能领域研究的一个重要领域,并且得到了广泛的商用。基于知识图谱的多轮对话技术是结合实体和实体间的关系,将知识三元组和原始对话上下文数据融合在一起,更好地实现多轮对话。研究人员多次运用并证明了将知识图谱融入到多轮对话中对多轮对话技术的发展是有帮助的并且效果是非常显著的。然而目前的研究进展存在诸多挑战,例如针对开放域的信息抽取,主要包括实体抽取、关系抽取以及属性抽取这三个问题。其中,多种语言文本信息、开放域下非结构化纯文本信息抽取等问题是当前面临的重要挑战之一。知识图谱的重要性不仅是一项技术,更加推动了自然语言处理和深度学习的发展。文章希望更多的研究者能够参与并且投入到这份研究工作中。

### 参考文献:

- [1] 刘伟,瞿小童. 2019 年人工智能研发热点回眸[J]. 科技导报,2020,38(1):151-157.
- [2] 陈庆惠,郭福燕. 一种基于物联网的智能家居系统设计[J]. 电子技术与软件工程,2018(7):53-54.
- [3] LOBACCARO G, CARLUCCI S, LÖFSTRÖM E. A review of systems and technologies for smart homes and smart grids[J]. Energies,2016,9(5):348.
- [4] MIKOLOV T, CHEN K, CORRADO G, et al. Efficient estimation of word representations in vector space[J]. arXiv:1301.3781,2013.
- [5] LE Q, MIKOLOV T. Distributed representations of sentences and documents[C]//International conference on machine learning. New York:ACM,2014:1188-1196.
- [6] WIEMER-HASTINGS P. How latent is latent semantic analysis?[C]//Proceedings of the 16th international joint conference on artificial intelligence. Varna:Morgan Kaufmann,1999:932-941.
- [7] CARRERA-TREJO J V, SIDOROV G, MIRANDA-JIMÉNEZ S, et al. Latent Dirichlet allocation complement in the vector space model for multi-label text classification[J]. International Journal of Combinatorial Optimization Problems and Informatics,2015,6(1):7-19.
- [8] 徐增林,盛泳潘,贺丽荣,等. 知识图谱技术综述[J]. 电子科技大学学报,2016,45(4):589-606.
- [9] KENDALL A, MARTIROSYAN H, DASGUPTA S, et al. End-to-end learning of geometry and context for deep stereo regression[C]//Proceedings of the IEEE international conference on computer vision. Piscataway:IEEE,2017:66-75.
- [10] PETERS M E, NEUMANN M, LOGAN IV R L, et al. Knowledge enhanced contextual word representations[J]. arXiv:1909.04164,2019.
- [11] YANG Z, DAI Z, YANG Y, et al. Xlnet:generalized autoregressive pretraining for language understanding[C]//Pro-

- ceedings of the 33rd annual conference on neural information processing systems. Vancouver; NIPS, 2019; 32–42.
- [12] LIU Z, LIN W, SHI Y, et al. A robustly optimized BERT pre-training approach with post-training [C]//China national conference on Chinese computational linguistics. [s. l.]: Springer, 2021; 471–484.
- [13] LAN Z, CHEN M, GOODMAN S, et al. Albert: a litebert for self-supervised learning of language representations [J]. arXiv; 1909.11942, 2019.
- [14] ZHANG Z, WU Y, ZHAO H, et al. Semantics-aware BERT for language understanding [C]//Proceedings of the 34th AAAI conference on artificial intelligence. Menlo Park; AAAI, 2020; 9628–9635.
- [15] ZHANG Z, WU Y, ZHOU J, et al. SG-Net: syntax-guided machine reading comprehension [C]//Proceedings of the 34th AAAI conference on artificial intelligence. Menlo Park; AAAI, 2020; 9636–9643.
- [16] ZHANG Z, LI J, ZHU P, et al. Modeling multi-turn conversation with deep utterance aggregation [J]. arXiv; 1806.09102, 2018.
- [17] LOWE R, POW N, SERBAN I V, et al. Training end-to-end dialogue systems with the ubuntu dialogue corpus [J]. Dialogue & Discourse, 2017, 8(1); 31–65.
- [18] LOWE R, POW N, SERBAN I V, et al. The ubuntu dialogue corpus: a large dataset for research in unstructured multi-turn dialogue systems [J]. arXiv; 1506.08909, 2015.
- [19] WANG S, JIANG J. Learning natural language inference with LSTM [J]. arXiv; 1512.08849, 2015.
- [20] ZHANG S, ZHAO H, WU Y, et al. Dual co-matching network for multi-choice reading comprehension [J]. arXiv; 1901.09381, 2019.
- [21] ZHOU X, DONG D, WU H, et al. Multi-view response selection for human-computer conversation [C]//Proceedings of the 2016 conference on empirical methods in natural language processing. Austin; The Association for Computational Linguistics (ACL), 2016; 372–381.
- [22] ZHOU X, LI L, DONG D, et al. Multi-turn response selection for chatbots with deep attention matching network [C]//Proceedings of the 56th annual meeting of the association for computational linguistics. Melbourne; The Association for Computational Linguistics (ACL), 2018; 1118–1127.
- [23] TAO C, WU W, XU C, et al. One time of interaction may not be enough: go deep with an interaction-over-interaction network for response selection in dialogues [C]//Proceedings of the 57th annual meeting of the association for computational linguistics. Florence; The Association for Computational Linguistics (ACL), 2019; 1–11.
- [24] TAO C, WU W, XU C, et al. Multi-representation fusion network for multi-turn response selection in retrieval-based chatbots [C]//Proceedings of the twelfth ACM international conference on web search and data mining. Virtual Event; ACM, 2019; 267–275.
- [25] HE S, LIU C, LIU K, et al. Generating natural answers by incorporating copying and retrieving mechanisms in sequence-to-sequence learning [C]//Proceedings of the 55th annual meeting of the association for computational linguistics. Vancouver; The Association for Computational Linguistics (ACL), 2017; 199–208.
- [26] BAHDANAU D, CHO K, BENGIO Y. Neural machine translation by jointly learning to align and translate [J]. arXiv; 1409.0473, 2014.
- [27] LUONG M T, PHAM H, MANNING C D. Effective approaches to attention-based neural machine translation [J]. arXiv; 1508.04025, 2015.
- [28] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [C]//Proceedings of the 33rd annual conference on neural information processing systems. Long Beach; NIPS, 2017; 5998–6008.
- [29] BORDES A, USUNIER N, GARCIA-DURAN A, et al. Translating embeddings for modeling multi-relational data [C]//Proceedings of the 27th annual conference on neural information processing systems. Lake Tahoe; NIPS, 2013; 2787–2795.
- [30] WANG Z, ZHANG J, FENG J, et al. Knowledge graph embedding by translating on hyperplanes [C]//Proceedings of the 28th AAAI conference on artificial intelligence. Québec City; AAAI, 2014; 1112–1119.
- [31] LIN Y, LIU Z, SUN M, et al. Learning entity and relation embeddings for knowledge graph completion [C]//Proceedings of the 29th AAAI conference on artificial intelligence. Austin; AAAI, 2015; 2181–2187.
- [32] JI G, HE S, XU L, et al. Knowledge graph embedding via dynamic mapping matrix [C]//Proceedings of the 53rd annual meeting of the association for computational linguistics and the 7th international joint conference on natural language processing. Beijing; ACL, 2015; 687–696.
- [33] LIU K. A survey on neural relation extraction [J]. Science China Technological Sciences, 2020, 63(10); 1971–1989.
- [34] LI L, LI C, JI D. Deep context modeling for multi-turn response selection in dialogue systems [J]. Information Processing & Management, 2021, 58(1); 102415.
- [35] SERBAN I, KLINGER T, TESAURO G, et al. Multiresolution recurrent neural networks: an application to dialogue response generation [C]//Proceedings of the 31st AAAI conference on artificial intelligence. Menlo Park; AAAI, 2017; 3288–3294.
- [36] SERBAN I, SORDONI A, LOWE R, et al. A hierarchical latent variable encoder-decoder model for generating dialogues [C]//Proceedings of the 31st AAAI conference on artificial intelligence. Menlo Park; AAAI, 2017; 3295–3301.
- [37] CUI L, WU Y, LIU S, et al. Mutual: a dataset for multi-turn dialogue reasoning [J]. arXiv; 2004.04494, 2020.