

基于图卷积网络的知识感知对话生成^①

曾碧卿, 丁明浩, 宋逸云

(华南师范大学 软件学院, 佛山 528225)

通信作者: 曾碧卿, E-mail: zengbiqing@scnu.edu.cn



摘要: 引入结构化知识的对话系统因为能够生成流畅度更高、多样性更丰富的对话回复而受到广泛关注, 但是以往的研究只注重于结构化知识中的实体, 却忽略了实体之间的关系以及知识的完整性. 本文提出了一种基于图卷积网络的知识感知对话生成模型 (KCG). 该模型通过知识编码器分别捕获实体与关系的语义信息并利用图卷积网络增强实体表征; 再利用知识选择模块获得与对话上下文相关的实体与关系的知识选择概率分布; 最后将知识选择概率分布与词表概率分布融合, 解码器以此选择知识或词表字词. 本文在中文公开数据集 DuConv 上进行实验, 结果表明, KCG 在自动评估指标上优于目前的基线模型, 能生成更加流畅并且内容更加丰富的回复.

关键词: 对话生成; 多轮对话; 图卷积网络; 结构化知识; 知识选择

引用格式: 曾碧卿, 丁明浩, 宋逸云. 基于图卷积网络的知识感知对话生成. 计算机系统应用, 2023, 32(7): 47-56. <http://www.c-s-a.org.cn/1003-3254/9192.html>

Knowledge-aware Conversation Generation Based on Graph Convolutional Network

ZENG Bi-Qing, DING Ming-Hao, SONG Yi-Yun

(School of Software, South China Normal University, Foshan 528225, China)

Abstract: The dialogue system that introduces structured knowledge has attracted widespread attention as it can generate more fluent and diverse dialogue replies. However, previous studies only focus on entities in structured knowledge, ignoring the relation between entities and the integrity of knowledge. In this study, a knowledge-aware conversation generation (KCG) model based on the graph convolutional network is proposed. The semantic information of the entity and relation is captured by the knowledge encoder and the representation of the entity is enhanced by the graph convolutional network. Then, the knowledge selection module is applied to obtain the knowledge selection probability distribution of the entities and relations related to the dialogue context. Finally, the knowledge selection probability distribution is fused with the vocabulary probability distribution so that the decoder can select the knowledge or words. In this study, the experiments are conducted on DuConv, a Chinese public data set. The results show that KCG is superior to the current baseline model in terms of automatic evaluation metrics and can generate more fluent and informative replies.

Key words: conversation generation; multi-turn conversations; graph convolutional network (GCN); structured knowledge; knowledge selection

人机对话系统一直以来都是自然语言处理领域的研究热点之一, 随着对话系统技术的发展, 任务型对话

系统已经趋于成熟, 这类对话系统可以根据用户意图完成特定领域的任务, 在日常生活乃至工业界已经得

① 基金项目: 国家自然科学基金 (12001208); 广东省基础与应用基础研究基金 (2021A1515011171); 广东省普通高校人工智能重点领域专项 (2019KZDZX1033); 广州市基础研究计划、基础与应用基础研究项目 (202102080282)

收稿时间: 2022-12-06; 修改时间: 2023-01-19; 采用时间: 2023-03-14; csa 在线出版时间: 2023-05-22

CNKI 网络首发时间: 2023-05-24

到了广泛的应用.然而开放领域对话系统没有特定的领域和目标,并且对话话题的开放性更高,涉及的领域范围更广,因此相较于任务型对话系统将会面临更多挑战.

开放领域对话系统根据实现方式的不同可以分为检索式对话系统与生成式对话系统.检索式对话系统主要从数据库中查找与用户语句匹配的回复,十分依赖数据库的规模与质量.生成式对话系统主要基于深度学习模型,可以生成数据库中没有出现过的语句,灵活性强并且不依赖庞大的检索数据库,但是可控性较差.随着深度学习技术的发展,目前常采用的基于端到端模型的生成式方法取得了较好的效果^[1,2].

由于深度学习模型目标函数的特性以及参差不齐的数据集质量依然存在一些问题,生成式对话系统容易生成“是的”“我不知道”等通用回复甚至是不符合上下文语义的回复.为了解决上述问题,有研究者从深度学习模型的目标函数着手, Li 等人^[3]提出了一种基于最大互信息的目标函数,使回复生成依赖于给定回复生成对话历史的概率,限制通用回复从而提高回复的多样性. Cao 等人^[4]提出了一种基于潜变量的单轮对话生成模型,该模型在解码器中引入变分自编码器,使对话上下文与潜变量同时影响模型回复生成,提高了回复多样性.这些方法取得了一定效果,但是对话生成模型产生低质量回复的另一个重要原因在于模型仅从数据集中难以学习复杂的背景知识^[5],在面临一个新场景的时候容易生成语义不连贯的回复^[6],因此如何引入以及利用外部知识就显得尤为重要.

外部知识可以分为非结构化知识(如文本)和结构化知识(如三元组).非结构化知识主要由自然语句组成,而结构化知识多由名词词组、动词短语组成,信息更加简洁精炼.之前关于引入结构化知识的研究^[7-9]仅考虑了结构化知识中的实体,而忽略了实体之间的关系,因此在解码生成时无法有效地利用关系本身的信息.然而在对话过程中,实体之间的关系不仅是选择知识的依据同时还包含着重要信息.如图1所示,其中,下划线标注的与加粗的分别是关系、实体中的词.“极地大冒险2”和“男孩之歌”都是来自“爱尔兰”的电影,“主演”与“导演”分别是“埃里克·卡尔森”与“艾斯林·沃尔什”,并且“主演”与“导演”也作为电影的相关属性出现在对话上下文中.

基于以上因素,本文提出了一种基于图卷积网络(graph convolutional networks, GCN)^[10]的知识感知对

话生成模型(knowledge-aware conversation generation based on graph convolutional network, KCG).在该模型中,首先对知识三元组进行解析得到实体集合与关系集合;使用基于双向门控循环单元(bi-directional gated recurrent unit, Bi-GRU)^[11]的编码器对对话上下文、目标回复以及实体、关系进行编码,获得对话上下文特征、目标回复特征以及实体、关系知识特征;根据实体之间的关系使用图卷积网络对实体表征进行聚合增强;在知识选择模块中将对话上下文作为先验信息,目标回复作为后验信息,计算实体和关系两种知识来源的先验、后验知识选择概率分布,并且利用KL散度(Kullback-Leibler divergence)减小先验、后验知识选择概率分布的差异以提高知识选择的准确率;在解码器中,根据对话上下文特征计算词表生成概率.最后通过指针生成网络^[12]对实体选择概率、关系选择概率、词表生成概率进行融合获得回复生成概率.本文在中文公开数据集 DuConv^[13]上将KCG与多个基线模型进行对比实验.实验结果表明,KCG生成的回复在自动评估指标以及人工评估指标上均有一定提升.

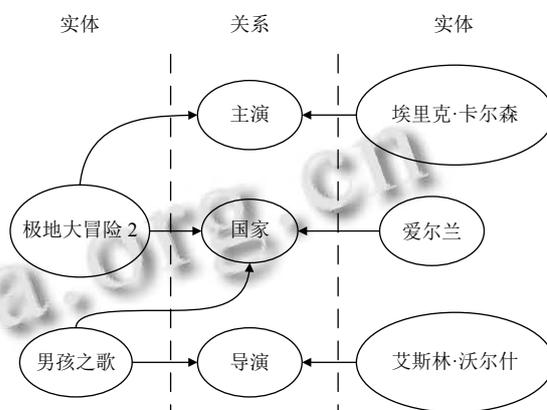


图1 引入结构化知识的对话示例

本文提出的模型KCG利用图卷积网络将结构化知识引入到对话系统中并通过知识选择模块利用知识,主要工作如下.

(1) 设计了一种基于图卷积网络的实体聚合器.根据知识三元组中的关系使用图卷积网络提升实体特征

的表征能力。

(2) 设计了一种多种知识来源情景下的知识选择模块。利用目标回复中的后验信息为多种知识来源的知识选择提供依据,提高模型对不同来源知识的感知能力。

(3) 提出了一种基于图卷积网络的知识感知对话生成模型。该模型可以生成内容丰富、符合对话上下文语境的回复响应。实验结果表明, KCG 生成的回复优于对比的基线模型。

1 相关工作

目前引入结构化知识的对话回复生成工作大致可以分为两个方面。

1.1 结构化知识

对话回复生成研究中的结构化知识主要表现形式为知识三元组^[7,8,14,15],一种方法是将知识三元组视为序列文本对其进行编码。而另一种方法则是将知识三元组中的实体视为节点,实体之间的关系视为边,采用基于图神经网络的方法对节点进行编码,以此作为外部知识特征。

Zhou 等人^[16]首次将大型知识图谱引入到对话回复生成中,他们认为以往的研究对知识三元组的利用是分离的,因此设计了一种静态图注意力机制对知识进行编码,以及一种动态图注意力机制在生成阶段对知识进行有选择地复制生成。Zhang 等人^[17]将对话视为一种在知识图谱上的主题转换过程,并且通过图注意力机制将这种转变导向更有意义的方向,生成更有意义的回复。Wu 等人^[18]提出了一种模式融合机制,这种机制可以将对话上下文以及与对话上下文高度相关的知识实体进行融合,从而提高了回复生成的质量。Li 等人^[19]设计了一种基于知识图谱的共情对话生成模型,根据对话上下文构建情感图,利用图注意力网络提取情感特征来增强对话系统的情感感知能力。但是这些研究只考虑了结构化知识中的实体,忽略了实体之间的关系,而关系往往包含了与实体属性相关的信息导致对知识的利用不足。Ji 等人^[8]对多个关系进行编码,将关系特征引入到图卷积网络中,从而实现对不同关系的实体进行聚合,但是在解码时只考虑了实体特征,没有对引入的关系特征进行有效利用。

1.2 知识选择

引入结构化知识的对话回复生成的另一个重要问题就是知识选择,目前常用的方法主要是基于 Lian 等

人^[20]提出的一种知识选择机制,这种机制引入了先验信息与后验信息并通过 KL 散度目标函数减小了模型与人类之间知识选择的差异,使得模型在回复生成时能够选择正确的知识。Lin 等人^[9]设计了一种知识集成复制生成模型,通过在解码的每个时间刻动态更新当前时间刻隐藏状态对知识的注意力,使解码器获得最可靠的知识注意力分布。Bai 等人^[21]设计了一种目标导向的知识复制网络,通过知识发现机制寻找与对话目标相关的知识,通过指针生成网络在解码时选择相关的实体。上述的研究主要关注知识与对话上下文、主题的相关性,除此之外多轮对话场景下知识的转换也是影响知识选择准确率的重要因素。Kim 等人^[22]将知识选择看作一个连续的过程,使用 BERT 预训练模型^[23]对上下文以及目标回复进行编码,使用潜变量对多轮对话中的知识选择进行建模。Zhan 等人^[24]关注对话过程不同回合的知识转换,通过预训练一个基于 Transformer 模型^[25]的回复生成器来选择每回合相关的知识。Zheng 等人^[26]认为对话过程中每一轮的知识选择之间的差异可以为当前轮的知识选择提供线索,使用 Bi-GRU 计算当前轮句子与之前轮句子的差值,并以此指导知识选择。但是,这些研究只考虑利用目标回复中的后验信息减小单一知识来源的知识选择差异,没有考虑后验信息对多个知识来源的影响,而在实际对话过程中往往可能涉及多个不同的知识来源。

2 基于图卷积网络的对话生成模型

本文模型的任务主要如下,给定一个知识驱动对话数据集 $D = \{(U_i, Y_i, K_i)\}_{i=1}^N$,其中 N 为数据集样本数量,对于每条样本都包含对话历史 $U = \{u_1, u_2, \dots, u_n\}$ 以及相应的目标回复 $Y = \{y_1, y_2, \dots, y_m\}$,并且提供对话相关的知识集合 $K = \{k_1, k_2, \dots, k_s\}$ 。其中 n, m 分别为对话历史 U 和目标回复 Y 的字数, s 表示知识集合 K 中的知识条目数量。知识集合 K 中的每条知识 k 的形式为三元组 $(fact^{head}, relation, fact^{tail})$,统计所有知识三元组可以构成关系集合 $\{relation_i\}_{i=1}^{|R|}$ 与实体集合 $\{fact_i\}_{i=1}^{|F|}$,记为 R 与 F ,其中 $|R|$ 为关系集合中关系的数量, $|F|$ 为实体集合中实体的数量。本文提出的 KCG 的学习目标 $P(Y|U, R, F)$ 是在给定对话历史 U 和相关的实体集合以及关系集合的情况下,生成符合上下文语义、内容丰富的回复 Y 。本文提出的模型总体框架如图 2 所示,主要包含以下几个模块:上下文编码器;知识编码器;知识选择模块;解码器。

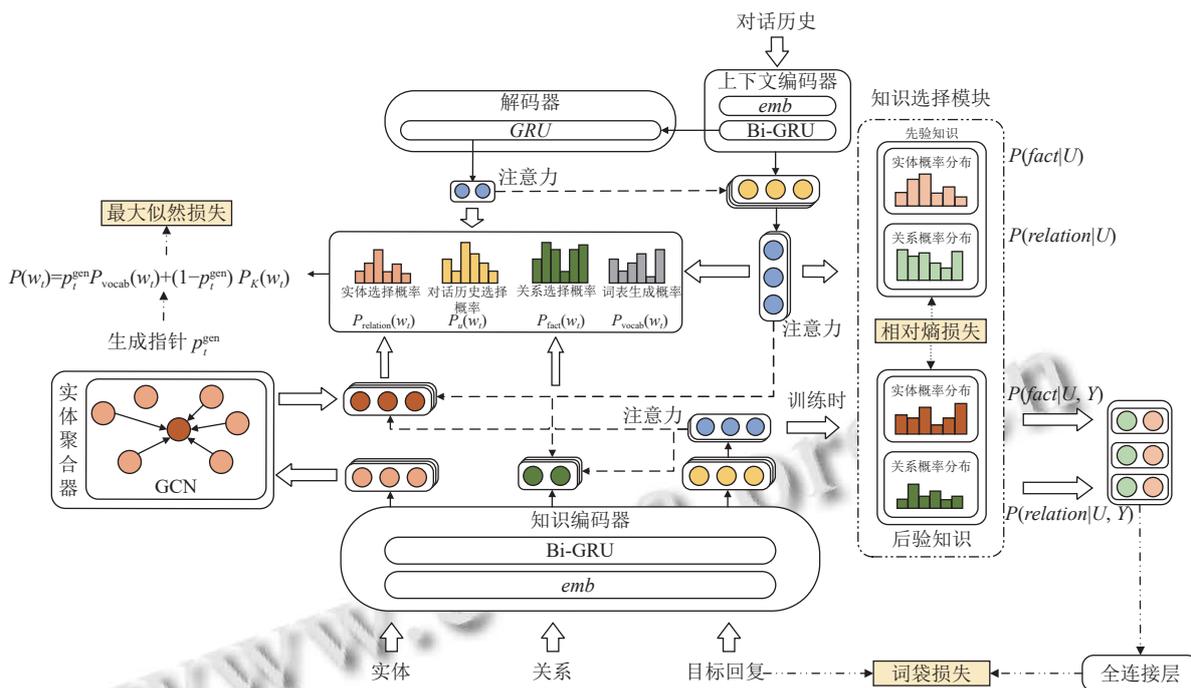


图2 基于图卷积网络的对话生成模型框架

2.1 编码器

本文模型的编码器包括上下文编码器与知识编码器, 两者均为 Bi-GRU, 其中包含一个前向的 GRU_{fw} 与一个后向的 GRU_{bw} , 同时捕捉文本前向与后向的语义信息. 上下文编码器对对话历史 $U = (u_1, u_2, \dots, u_n)$ 编码, 根据上一时刻的前向隐藏状态 h_{t-1}^{fw} 与当前时刻的字词 u_t 的词嵌入向量可以获得当前时刻的前向隐藏状态 h_t^{fw} , 同理可以得到后向隐藏状态 h_t^{bw} , 并将两者进行拼接获得当前时刻的隐藏状态 h_t :

$$\begin{cases} h_t^{fw} = GRU_{fw}(emb(u_t), h_{t-1}^{fw}) \\ h_t^{bw} = GRU_{bw}(emb(u_t), h_{t-1}^{bw}) \\ h_t = [h_t^{fw}; h_t^{bw}] \end{cases} \quad (1)$$

其中, emb 表示词嵌入层, $[h_t^{fw}; h_t^{bw}]$ 表示拼接向量 h_t^{fw} 与向量 h_t^{bw} . 最后计算所有时刻的隐藏状态可以得到对话历史隐藏状态 h^u , 将第 n 时刻的对话历史隐藏状态 h_n^u 作为对话历史上下文状态 C^u . 知识编码器对目标回复 Y 、实体集合、关系集合编码, 根据上述式子可以得到目标回复上下文状态 C^y 与实体上下文状态 C^{fact} 以及关系上下文状态 $C^{relation}$:

$$\begin{cases} h^u = \{h_1, h_2, \dots, h_n\} \\ C^u = h_n^u \end{cases} \quad (2)$$

其中, 上下文编码器与知识编码器除词嵌入层外不共

享权重.

2.2 实体聚合器

本文将实体作为图的节点, 关系作为连接两个节点的边, 把知识集合 K 转化为知识图 $G = (V, E)$ 并且使用 GCN 对每个节点周围的邻居节点进行聚合以增强节点特征, 其中 V 为节点集合, E 为边集合. 对于每个 $v \in V$, 在第 l 层时, 使用线性卷积核 W^l 对其周围邻居节点 $N(v)$ 进行一次线性变换并求和取平均, 最后获得节点特征 h_v^{l+1} . 当 $l = 0$ 时, 使用实体上下文状态 C^{fact} 初始化节点特征. 具体计算过程如下:

$$h_v^{l+1} = \frac{1}{|N(v)|+1} \left(\sum_{u \in N(v)} W^l h_u^l + h_v^l \right) \quad (3)$$

其中, $|N(v)|$ 表示节点 v 周围邻居节点的数量. 将最后一层的节点表征 h_v^L 用于更新实体上下文状态 C^{fact} , 其中 L 为图卷积层的数量:

$$C^{fact} = h_v^L \quad (4)$$

2.3 知识选择模块

本文认为知识三元组中的关系与实体在提供信息方面同样重要, 为了更加充分地利用实体与关系中的知识, 在知识选择模块中将两者同等考虑. 知识选择模块的输入为对话历史上下文状态 C^u 、目标回复上下文

状态 C^y 、关系上下文状态 $\{C_i^{\text{relation}}\}_{i=1}^{|R|}$ 以及实体上下文状态 $\{C_i^{\text{fact}}\}_{i=1}^{|F|}$, 输出为关系概率分布和实体概率分布. 本模块包含两个子模块, 分别为先验知识模块以及后验知识模块.

先验知识模块负责计算对话历史与关系 $relation_i$, 实体 $fact_i$ 之间的注意力权重并根据注意力权重获得关系先验概率分布 $P(relation_i|U)$ 与实体先验概率分布 $P(fact_i|U)$:

$$\begin{cases} P(relation_i|U) = \frac{\exp(C_i^{\text{relation}} \cdot C^u)}{\sum_{s=1}^{|R|} \exp(C_s^{\text{relation}} \cdot C^u)} \\ P(fact_i|U) = \frac{\exp(C_i^{\text{fact}} \cdot C^u)}{\sum_{s=1}^{|F|} \exp(C_s^{\text{fact}} \cdot C^u)} \end{cases} \quad (5)$$

关系先验概率分布与实体先验概率分布的前提条件只与对话历史 U 相关, 如果目标回复中出现了与对话历史不同的信息则模型将难以准确捕捉. 因此为了提高知识选择的准确率, 受到 Lian 等人^[20]启发, 本文将目标回复作为一种后验信息引入到后验知识模块中, 计算关系后验概率分布 $P(relation_i|U, Y)$ 与实体后验概率分布 $P(fact_i|U, Y)$:

$$\begin{cases} P(relation_i|U, Y) = \frac{\exp(C_i^{\text{relation}} \cdot Q_{\text{post}})}{\sum_{s=1}^{|R|} \exp(C_s^{\text{relation}} \cdot Q_{\text{post}})} \\ P(fact_i|U, Y) = \frac{\exp(C_i^{\text{fact}} \cdot Q_{\text{post}})}{\sum_{s=1}^{|F|} \exp(C_s^{\text{fact}} \cdot Q_{\text{post}})} \end{cases} \quad (6)$$

其中, $Q_{\text{post}} = MLP([C^u; C^y])$, MLP 为全连接层. 相较于先验概率分布, 后验概率分布不仅包含了对话历史 U 中的先验信息也包含了目标回复 Y 中的后验信息, 因此能够同时计算与对话历史、目标回复相关知识的注意力并且与目标回复相关的知识注意力权重占比会更大, 能够捕捉到只依赖先验信息难以捕捉的知识.

在训练时, 目标回复对于模型是已知的, 知识选择模块可以根据目标回复中的后验信息计算关系后验概率分布与实体后验概率分布, 并分别作为关系概率分布以及实体概率分布在解码生成时作为选择知识的依据. 但是在推理时, 目标回复对于模型是未知的, 知识选择模块只能计算先验概率分布, 此时模型只能根据先验概率分布选择知识. 因此, 为了减小先验概率分布之间的差异, 本文引入 KL 散度计算两者之间的损失:

$$Loss_{\text{KL}}^{\text{relation}} = \sum_{i=1}^{|R|} P(relation_i|U, Y) \log \left(\frac{P(relation_i|U, Y)}{P(relation_i|U)} \right) \quad (7)$$

$$Loss_{\text{KL}}^{\text{fact}} = \sum_{i=1}^{|F|} P(fact_i|U, Y) \log \left(\frac{P(fact_i|U, Y)}{P(fact_i|U)} \right) \quad (8)$$

$$Loss_{\text{KL}} = Loss_{\text{KL}}^{\text{relation}} + Loss_{\text{KL}}^{\text{fact}} \quad (9)$$

其中, KL 散度损失 $Loss_{\text{KL}}$ 包含关系 KL 散度损失 $Loss_{\text{KL}}^{\text{relation}}$ 以及实体 KL 散度损失 $Loss_{\text{KL}}^{\text{fact}}$, 两者分别估计了关系先、后验概率分布之间的差异和实体先、后验概率分布之间的差异. 为了让先验概率分布尽可能接近后验概率分布, 本文在优化 KL 散度损失时将后验概率分布固定, 不计算相关梯度.

2.4 解码器

解码器部分主要由前向GRU组成, 解码器会根据上一时刻的解码器隐藏状态 h_{t-1} 与解码器上下文状态 c_{t-1} 以及目标回复单词 y_t 的词嵌入向量计算得到当前时刻的解码器隐藏状态 h_t :

$$h_t = GRU([emb(y_t); c_{t-1}, h_{t-1}]) \quad (10)$$

本文将知识三元组解析为关系集合与实体集合分别进行编码, 为了集成不同来源的知识以及提高模型生成回复的多样性与信息量, 引入指针生成网络^[12]从不同来源的知识中复制字词, 并且将对话历史加入到知识中作为指针复制生成的对象以解决未登录词 (out of vocabulary, OOV) 问题.

对于关系集合而言, 本文计算当前时刻解码器隐藏状态 h_t 对关系上下文状态 $\{C_i^{\text{relation}}\}_{i=1}^{|R|}$ 的 Bahdanau 注意力^[27], 获得关系注意力权重 a_t^{relation} 以及关系上下文加权状态 c_t^{relation} :

$$\begin{cases} e'_i = v^T \tanh(W_r C_i^{\text{relation}} + W_t h_t + b) \\ a_t^{\text{relation}} = \text{Softmax}(e') \\ c_t^{\text{relation}} = \sum_{i=1}^{|R|} a_{t,i}^{\text{relation}} C_i^{\text{relation}} \end{cases} \quad (11)$$

其中, W_r 、 W_t 、 b 以及 v 为可训练参数, T 表示转置操作, 激活函数为 Softmax . 上述的关系注意力权重 a_t^{relation} 的计算只考虑了 t 时刻以及 t 时刻之前的解码器状态, 为了充分利用目标回复中的后验信息, 本文通过门控机制将知识选择模块的关系概率分布 $P(relation)$ 与关系注意力权重 a_t^{relation} 融合, 共同计算关系选择概率 $P_{\text{relation}}(w_t)$:

$$\begin{cases} \delta = \text{Sigmoid}(MLP([h_t; c_t^{\text{relation}}])) \\ P_{\text{relation}}(w_t) = \delta P(relation) + (1 - \delta) a_t^{\text{relation}} \end{cases} \quad (12)$$

其中,关系概率分布 $P(\text{relation})$ 在训练时为关系后验概率分布 $P(\text{relation}|U, Y)$,推理时为关系先验概率分布 $P(\text{relation}|U)$,实体概率分布 $P(\text{fact})$ 同样如此。

对于实体集合而言,同样需要计算当前时刻解码器隐藏状态 h_t 对实体上下文状态 $\{C_i^{\text{fact}}\}_{i=1}^{|F|}$ 的Bahdanau注意力,得到实体注意力权重 a_t^{fact} 与实体上下文加权状态 c_t^{fact} 注意力,计算过程简化为如下:

$$a_t^{\text{fact}}, c_t^{\text{fact}} = \text{Attention}(h_t, C^{\text{fact}}) \quad (13)$$

通过上述门控机制将实体概率分布 $P(\text{fact})$ 与实体注意力权重 a_t^{fact} 融合,得到实体选择概率 $P_{\text{fact}}(w_t)$,如下:

$$\begin{cases} \delta = \text{Sigmoid}(\text{MLP}([h_t; c_t^{\text{fact}}])) \\ P_{\text{fact}}(w_t) = \delta P(\text{fact}) + (1 - \delta) a_t^{\text{fact}} \end{cases} \quad (14)$$

对于对话历史,本文使用上述注意力计算公式得到当前时刻解码器隐藏状态 h_t 对对话历史隐藏状态 h^u 的对话历史注意力权重 a_t^u 以及对话历史上下文加权状态 c_t^u ,并将 a_t^u 作为对话历史选择概率 $P_u(w_t)$:

$$\begin{cases} a_t^u, c_t^u = \text{Attention}(h_t, C^u) \\ P_u(w_t) = a_t^u \end{cases} \quad (15)$$

当前时刻解码器上下文状态 c_t 可以根据当前时刻的关系上下文加权状态 c_t^{relation} ,实体上下文加权状态 c_t^{fact} 以及对话历史上下文加权状态 c_t^u 进行更新,从而控制下一时刻的解码过程,更新过程如下:

$$a_t, c_t = \text{Attention}(h_t, [c_t^{\text{relation}}, c_t^{\text{fact}}, c_t^u]) \quad (16)$$

当前时刻的词表生成概率 $P_{\text{vocab}}(w_t)$ 可以通过当前时刻的解码器隐藏状态 h_t 与解码器上下文状态 c_t 获得.词表生成概率分布是一维向量,向量大小为词表字词数,每个元素代表词表中对应字词的生成概率,如下:

$$P_{\text{vocab}}(w_t) = \text{Softmax}(\text{MLP}([h_t; c_t])) \quad (17)$$

a_t 为解码器对不同上下文加权状态的注意力分布,包含3个元素,不同元素视为解码器对不同知识来源的注意力.本文使用 a_t 来平衡关系选择概率 $P_{\text{relation}}(w_t)$ 、实体选择概率 $P_{\text{fact}}(w_t)$ 以及对话历史选择概率 $P_u(w_t)$,得到知识选择概率 $P_K(w_t)$:

$$P_K(w_t) = a_{t,1} P_{\text{relation}}(w_t) + a_{t,2} P_{\text{fact}}(w_t) + a_{t,3} P_u(w_t) \quad (18)$$

为了平衡词表生成概率 $P_{\text{vocab}}(w_t)$ 与知识选择概率 $P_K(w_t)$,本文使用生成指针 p_t^{gen} 来平衡两者的比例,从而最终的生成概率 $P(w_t)$:

$$\begin{cases} p_t^{\text{gen}} = \text{Sigmoid}(W_c c_t + W_h h_t + W_y \text{emb}(y_{t-1}) + b_{\text{gen}}) \\ P(w_t) = p_t^{\text{gen}} P_{\text{vocab}}(w_t) + (1 - p_t^{\text{gen}}) P_K(w_t) \end{cases} \quad (19)$$

其中, W_c 、 W_h 、 W_y 以及 b_{gen} 为可训练参数, p_t^{gen} 为一个实数。

2.5 损失函数

除了KL散度损失外,本文还使用了词袋损失^[28]与最大似然损失.其中词袋损失用于评估知识分布与目标回复之间的相似度,关系概率分布 $P(\text{relation})$ 与关系上下文状态 C^{relation} 相乘求和得到关系知识表征 s^{relation} ,实体概率分布 $P(\text{fact})$ 与实体上下文状态 C^{fact} 相乘求和得到关系知识表征 s^{fact} ,将两者拼接得到知识表征 s :

$$\begin{cases} s^{\text{relation}} = \sum_{i=1}^{|\mathcal{R}|} P(\text{relation}_i) C_i^{\text{relation}} \\ s^{\text{fact}} = \sum_{i=1}^{|\mathcal{F}|} P(\text{fact}_i) C_i^{\text{fact}} \\ s = [s^{\text{relation}}; s^{\text{fact}}] \end{cases} \quad (20)$$

将知识表征 s 输入一个由Softmax函数激活的全连接层得到概率分布 $p(y|s)$,此概率分布可表现为一维向量,向量大小与词表字词数相等,根据下式计算得到词袋损失 $Loss_{\text{BOW}}$:

$$Loss_{\text{BOW}} = -\frac{1}{m} \sum_{t=1}^m \log(p(y_t|s)) \quad (21)$$

最大似然损失 $Loss_{\text{NLL}}$ 用于计算模型生成的回复与真实回复之间的差异,如下:

$$Loss_{\text{NLL}} = -\frac{1}{m} \sum_{t=1}^m \log(p(y_t|y_{<t}, U, R, F)) \quad (22)$$

其中, $y_{<t}$ 表示模型 t 时刻之前生成的字词.模型的最终损失表示如下:

$$Loss = Loss_{\text{KL}} + Loss_{\text{BOW}} + Loss_{\text{NLL}} \quad (23)$$

3 实验

3.1 实验设计

本文使用基于知识图谱的主动聊天中文公开数据集DuConv作为实验数据集,对话过程主要涉及电影相关领域.此数据集共有30000条对话样本,大约120000轮对话,其中包括20000条训练集对话样本,2000条验证集样本以及8000条测试集样本.平均每条对话样本包含4-8轮对话,并且提供三元组形式的背景知识.本文统计了DuConv数据集中背景知识的关系数量共44种,包括导演、评论、售票、主演等相关信息.

本文模型主要基于PyTorch深度学习框架实现的.

编码器与解码器的 *GRU* 层数为 2 层, 每个 *GRU* 层的隐藏维度大小为 400. 实体聚合器中的 *GCN* 隐藏维度大小为 400. 词嵌入维度大小设置为 300. 本文将数据集中的词按照词频从高到低排序, 并选取前 15 000 的词作为词表, 将词表之外的词设置为 UNK 标记. 本文使用 Adam 优化器优化目标函数, 批次大小设置为 24, 学习率设置为 0.000 1, 动态梯度下降因子设置为 0.9 以及 0.999.

本文实验选取 3 个基线模型作为本文模型 KCG 的对比方案: 一个是基于后验知识选择机制的对话模型 PostKS^[20], 该模型引入了回复中的后验信息并且通过一种层次门控融合单元将知识融入解码器中; 另一个是基于预训练知识选择器的对话模型 SKT-KG^[24], 该模型中包含一个预训练的知识导向生成器, 能够给予被选择的知识更高的注意力权重; 最后一个是基于知识集成复制机制的对话模型 KIC^[9], 该模型引入了指针复制网络并在解码器的每个时刻动态更新知识注意力分布.

3.2 实验结果与分析

本文使用人工评价与自动评估两种方式评估回复的质量. 对于人工评价, 本文随机选取 DuConv 测试集中的 100 轮对话, 并在不告知模型名称的情况下要求评分人员对不同模型生成的回复进行评分. 评分的内容主要有两个方面, 包括一致性与信息量, 分数范围为 0-2 分, 分数越高代表回复质量越好, 反之质量越差. 一致性表示生成的回复与对话历史在语义上是否连贯, 信息量表示生成的回复内容是否丰富.

自动评估指标用于评估 4 个模型生成的回复的流畅度与多样性. 流畅度指标使用 BLEU (bi-lingual evaluation understudy), 该指标通常给定一个字词片段长度并计算生成回复中该长度的字词片段出现在真实回复中的占比, 占比越高代表生成回复的流畅度越高, 本文采用的字词片段长度为 1 或者 2, 表示为 BLEU-1、BLEU-2. 多样性指标使用 Distinct-1 与 Distinct-2, 该指标根据字词片段长度计算生成回复中不重复的字词片段在全部字词片段中的占比, 占比越高代表生成回复的多样性越高.

人工评价结果如表 1 所示. 其中, 重点考虑了结构化知识中关系的 KCG 生成的回复在信息量方面与其他 3 个模型相比具有一定的提升. 这是由于 KCG 不仅通过知识三元组中的关系指导实体知识选择, 还将关

系纳入了知识选择范畴, 提高了知识的利用率, 因此模型生成的回复具有更高的信息量. 在一致性方面, KCG 生成的回复相对于 PostKS 具有一定的提升, 说明目标回复中的后验信息不仅对单一知识来源的知识选择有补助作用, 能够弥补先验信息的不足, 提高知识选择的准确率, 对于本文中的关系集合和实体集合等多种来源知识也具有同样的效果, 使得生成回复中的相关知识与真实回复更加接近, 从而提高回复的一致性.

表 1 4 种模型的人工评价结果

模型	一致性	信息量
PostKS	1.59	1.45
KIC	1.65	1.57
SKT-KG	1.63	1.50
KCG	1.71	1.65

自动评估结果如表 2 所示, KCG 在流畅度指标 BLEU-1、BLEU-2 上效果最优, 相对于 PostKS 提升了 17.5%、36.8%, 相对于 KIC 提升了 8.2%、3.4%, 相对于 SKT-KG 提升了 7.9%、3.0%. 在多样性指标 Distinct-1 上, KCG 相对于 PostKS 提升了 77%. 在多样性指标 Distinct-2 上, KCG 相对于 PostKS 与 SKT-KG 分别提升了 74%、16.3%. 上述结果说明将结构化知识构建为图, 并且将实体、关系纳入知识选择范畴的 KCG, 可以生成符合上下文语义、内容丰富的高质量回复. 这是由于本文将知识三元组的实体与关系重新构建为图, 通过图卷积网络将相关联的实体特征聚合, 使得原本独立的实体之间具有了关联性, 相较于基线模型将知识三元组作为文本进行处理的方式, 从实体集合与关系集合中选择知识的范围更小, 不会出现重复冗余的知识, 并且聚合后的实体特征同时关联多个邻居实体, 为知识选择提供了更加充分的依据, 能够选择更加符合上下文语义的知识, 从而生成更加流畅的回复. 而且本文在知识选择时, 考虑到实体与关系的先验、后验概率之间的差异, 在训练时利用回复中的后验信息同时计算并优化两者的交叉熵损失, 以此提高知识选择的准确率.

表 2 4 种模型的自动评估结果

模型	BLEU-1	BLEU-2	Distinct-1	Distinct-2
PostKS	0.347	0.198	0.057	0.155
KIC	0.377	0.262	0.123	0.308
SKT-KG	0.378	0.263	0.105	0.232
KCG	0.408	0.271	0.101	0.270

但是与 KIC 相比, KCG 在多样性指标上并没有提升, 可能是本文知识编码器由双向 *GRU* 网络构成, 该

网络更加适用于长文本, 不适合对实体、关系这类短文本进行编码, 本文之后的工作也会继续探究如何通过其他方法增强编码器的编码能力.

3.3 消融实验

为了验证多来源知识选择以及实体聚合器的有效性, 本文在 DuConv 数据集上对 3 种模型生成的回复进行了实验比较. 首先, 将只有实体知识选择并且没有实体聚合器的模型作为基线模型 (KCG_N). 为了验证实体聚合器的有效性, 在 KCG_N 的基础上加入实体聚合器作为对比模型 (KCG_G). 为了验证多来源知识选择的有效性, 本文同时将关系、实体纳入知识选择范畴, 在 KCG_G 中加入关系知识选择结构作为对比模型, 即 KCG.

从表 3 中可以看出, 随着实体聚合器和关系知识选择的引入, 流畅度与多样性方面的自动化评估指标都有逐步地提高. 与 KCG_N 相比, 加入实体聚合器的 KCG_G 在 BLEU-1、BLEU-2 上提高了 2.3% 与 4.9%, 在 Distinct-1、Distinct-2 上提高了 12.9% 与 6.6%. 在 KCG_G 的基础上将关系纳入知识选择的范畴, 多样性指标得到了进一步的提升, 在 BLEU-1、BLEU-2 上提高了 5.1% 与 6.2%, 在 Distinct-1、Distinct-2 上提高了

5.2% 与 5%. 上述是实验结果证明了本文提出的实体聚合器以及关系知识选择的有效性, 可以有效地提高生成回复的质量.

表 3 消融实验结果

模型	BLEU-1	BLEU-2	Distinct-1	Distinct-2
KCG _N	0.379	0.243	0.085	0.241
KCG _G	0.388	0.255	0.096	0.257
KCG	0.408	0.271	0.101	0.270

3.4 案例分析

图 3 展示了一个由回复样例, 包括对话历史、知识, 以及目标回复和 PostKS、KCG 生成的回复. 其中, 下划线标注的与加粗的分别是关系、实体中的词. 从回复的信息量来看, 与 PostKS 相比 KCG 生成的回复包含更多相关的信息这得益于将关系纳入了知识选择范畴, 因此, 模型可以充分地利用并准确地选择知识. 从回复的流畅度来看, KCG 生成的回复上下文语义更加连贯, 可以流畅地过渡到不同的话题. 在语义层面上, 相较于 PostKS, KCG 生成的回复更加接近目标回复. 图中样例表明, KCG 能够生成内容丰富、语义连贯的高质量回复.

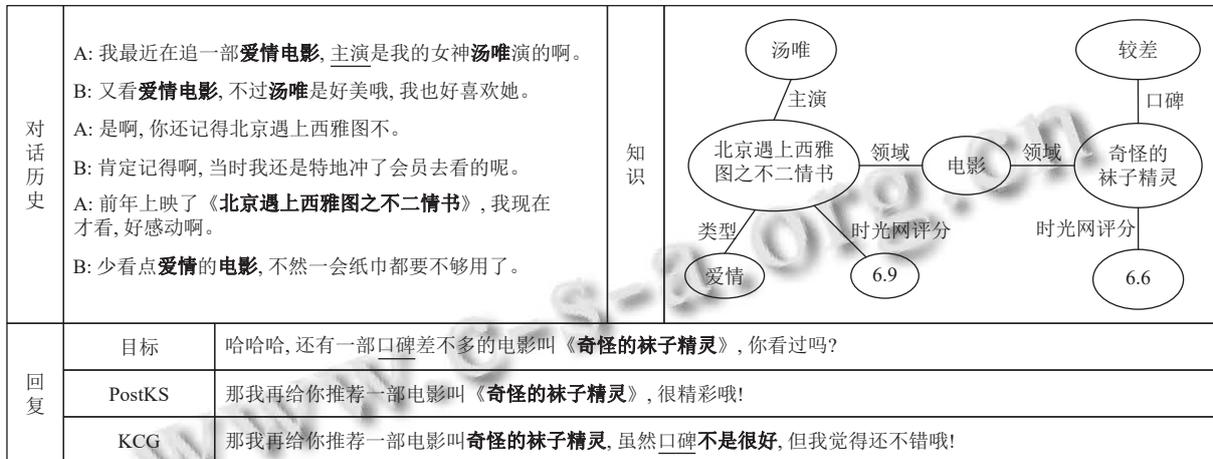


图 3 不同模型生成的回复样例

4 结论

本文提出了一种基于图卷积网络的知识感知对话生成模型 (KCG). 该模型首先在编码器模块中通过图卷积网络利用结构化知识中潜在的图结构增强实体特征的代表能力; 其次, 为了将实体以及实体之间的关系同时作为知识选择的对象, 在知识选择模块中引入目标回复中的后验信息提高两者知识选择的准确率; 最后, 在解码器中通过指针生成网络以及门控机制平衡

不同来源的知识选择分布, 得到最终的生成概率. 人工评价、自动评估以及消融实验表明, KCG 能够生成高质量的回复. 在未来的工作中, 会继续提升引入结构化知识的对话生成质量, 进一步研究如何更加准确地选择知识, 提高回复的多样性.

参考文献

1 Sutskever I, Vinyals O, Le QV. Sequence to sequence

- learning with neural networks. Proceedings of the 27th International Conference on Neural Information Processing Systems. Montreal: MIT Press, 2014. 3104–3112.
- 2 Shang LF, Lu ZD, Li H. Neural responding machine for short-text conversation. Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). Beijing: ACL, 2015. 1577–1586.
 - 3 Li JW, Galley M, Brockett C, *et al.* A diversity-promoting objective function for neural conversation models. Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. San Diego: ACL, 2016. 110–119.
 - 4 Cao K, Clark S. Latent variable dialogue models and their diversity. Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics. Valencia: ACL, 2017. 182–187.
 - 5 陈晨, 朱晴晴, 严睿, 等. 基于深度学习的开放领域对话系统研究综述. 计算机学报, 2019, 42(7): 1439–1466. [doi: 10.11897/SP.J.1016.2019.01439]
 - 6 Zhu WY, Mo KX, Zhang Y, *et al.* Flexible end-to-end dialogue system for knowledge grounded conversation. arXiv:1709.04264, 2017.
 - 7 Liu Y, Wan Y, He LF, *et al.* KG-BART: Knowledge graph-augmented BART for generative commonsense reasoning. Proceedings of the 35th AAAI Conference on Artificial Intelligence. AAAI Press, 2021. 6418–6425.
 - 8 Ji HZ, Ke P, Huang SH, *et al.* Language generation with multi-hop reasoning on commonsense knowledge graph. Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP). ACL, 2020. 725–736.
 - 9 Lin XX, Jian WY, He JS, *et al.* Generating informative conversational response using recurrent knowledge-interaction and knowledge-copy. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. ACL, 2020. 41–52.
 - 10 Kipf TN, Welling M. Semi-supervised classification with graph convolutional networks. Proceedings of the 5th International Conference on Learning Representations. Toulon: OpenReview.net, 2017. 1–14.
 - 11 Cho K, van Merriënboer B, Gulçehre Ç, *et al.* Learning phrase representations using RNN encoder-decoder for statistical machine translation. Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). Doha: ACL, 2014. 1724–1734.
 - 12 See A, Liu PJ, Manning CD. Get to the point: Summarization with pointer-generator networks. Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Vancouver: ACL, 2017. 1073–1083.
 - 13 Wu WQ, Guo Z, Zhou XY, *et al.* Proactive human-machine conversation with explicit conversation goal. Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence: ACL, 2019. 3794–3804.
 - 14 Koncel-Kedziorski R, Bekal D, Luan Y, *et al.* Text generation from knowledge graphs with graph transformers. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers). Minneapolis: ACL, 2019. 2284–2293.
 - 15 Moghe N, Vijayan P, Ravindran B, *et al.* On incorporating structural information to improve dialogue response generation. Proceedings of the 2nd Workshop on Natural Language Processing for Conversational AI. ACL, 2020. 11–24.
 - 16 Zhou H, Young T, Huang ML, *et al.* Commonsense knowledge aware conversation generation with graph attention. Proceedings of the 27th International Joint Conference on Artificial Intelligence. Stockholm: AAAI Press, 2018. 4623–4629.
 - 17 Zhang HY, Liu ZH, Xiong CY, *et al.* Grounded conversation generation as guided traverses in commonsense knowledge graphs. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. ACL, 2020. 2031–2043.
 - 18 Wu SX, Li Y, Zhang DW, *et al.* Diverse and informative dialogue generation with context-specific commonsense knowledge awareness. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. ACL, 2020. 5811–5820.
 - 19 Li QT, Li PJ, Ren ZC, *et al.* Knowledge bridging for empathetic dialogue generation. Proceedings of the 36th AAAI Conference on Artificial Intelligence. AAAI Press, 2022. 10993–11001.
 - 20 Lian RZ, Xie M, Wang F, *et al.* Learning to select knowledge for response generation in dialog systems. Proceedings of the 28th International Joint Conference on Artificial Intelligence. Macao: IJCAI.org, 2019. 5081–5087.

- 21 Bai JQ, Yang Z, Liang XN, *et al.* Learning to copy coherent knowledge for response generation. Proceedings of the 35th AAAI Conference on Artificial Intelligence. AAAI Press, 2021. 12535–12543.
- 22 Kim B, Ahn J, Kim G. Sequential latent knowledge selection for knowledge-grounded dialogue. Proceedings of the 2019 International Conference on Learning Representations. Addis Ababa: OpenReview.net, 2019. 1–14.
- 23 Devlin J, Chang MW, Lee K, *et al.* BERT: Pre-training of deep bidirectional Transformers for language understanding. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Minneapolis: ACL, 2019. 4171–4186.
- 24 Zhan HL, Zhang HN, Chen HS, *et al.* Augmenting knowledge-grounded conversations with sequential knowledge transition. Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. ACL, 2021. 5621–5630.
- 25 Vaswani A, Shazeer N, Parmar N, *et al.* Attention is all you need. Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach: Curran Associates Inc., 2017. 6000–6010.
- 26 Zheng CJ, Cao YB, Jiang DX, *et al.* Difference-aware knowledge selection for knowledge-grounded conversation generation. Findings of the Association for Computational Linguistics: EMNLP 2020. ACL, 2020. 115–125.
- 27 Luong MT, Pham H, Manning CD. Effective approaches to attention-based neural machine translation. Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. Lisbon: ACL, 2015. 1412–1421.
- 28 Zhao TC, Zhao R, Eskenazi M. Learning discourse-level diversity for neural dialog models using conditional variational autoencoders. Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Vancouver: ACL, 2017. 654–664.

(校对责编:牛欣悦)