计算机科学
COMPUTER SCIENCE

融合情感和常识知识的对话生成模型

程金凤, 蒋宗礼

引用本文

程金凤, 蒋宗礼. 融合情感和常识知识的对话生成模型[J]. 计算机科学, 2025, 52(1): 307-314.

CHENG Jinfeng, JIANG Zongli. Dialogue Generation Model Integrating Emotional and Commonsense Knowledge [J]. Computer Science, 2025, 52(1): 307-314.

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.231100130

融合情感和常识知识的对话生成模型

程金凤　蒋宗礼
北京工业大学信息学部　北京 100124
(996226508@qq.com)

摘　要　随着深度学习技术的发展,开放域对话系统作为人机对话系统的重要分支也得到了快速发展。但目前开放域对话模型生成的回复语句依然存在同理心较差、多样性较低等问题。对此,提出一种融合情感和常识知识的对话生成模型。首先依据情感词典和常识知识图谱获取每个单词对应的常识知识向量,然后将该向量和单词本身的词嵌入向量一同输入编码器中进行编码,接着通过两阶段解码来生成回复语句:第一个解码阶段预测要生成单词的情感强度,并据此获得该单词对应的情感向量,第二阶段解码结合第一阶段编码的结果和已生成单词的词嵌入向量及其对应的常识知识向量作为输入,预测要生成的单词。实验结果表明,该模型生成的回复语句更具同理心和多样性,并且在 PPL,BLEU,ACC 和 DISTINCT 等指标上相比基线模型都有一定提升。

关键词:对话模型;情感词典;常识知识图谱;两阶段解码;情感强度
中图分类号　TP391

Dialogue Generation Model Integrating Emotional and Commonsense Knowledge

CHENG Jinfeng and JIANG Zongli
Faculty of Information Technology,Beijing University of Technology,Beijing 100124,China

Abstract　With the development of deep learning technology,as an important branch of human-machine dialogue system,open domain dialogue system has also developed rapidly. However,there are still problems such as poor empathy and low diversity in response sentences generated by existing dialogue models in open domains. To address these problems,a dialogue generation model integrating emotional and commonsense knowledge is proposed in this paper. Commonsense knowledge vector corresponding to each word is firstly obtained based on the emotion dictionary and commonsense knowledge graph,and the vector is input into the encoder for encoding along with the word embedding vector of the word itself. Then a two-stage decoding process is used to generate response sentence:the first decoding stage is to predict the emotional intensity of the word to be generated and obtain the corresponding emotional vector for that word based on it,the second decoding stage combines the encoding result of the first stage with the word embedding vector of the generated word and its corresponding common sense knowledge vector as input to predict the word to be generated. Experimental results show that the response sentences generated by the proposed model are more empathetic and diverse,and it has a certain improvement in PPL,BLEU,ACC and DISTINCT evaluation compared with the baseline models.

Keywords　Dialogue model,Emotional dictionary,Commonsense knowledge graph,Two-stage decoding,Emotional intensity

1　引言

开放域对话作为对话系统的一个重要分支,有着非常广泛的应用前景,其特点是没有明确的主题和目标,可以表现为闲聊或者回答相关领域的问题等形式[1]。由于人与人之间的对话具有明显的时间特征,可被视为一个序列,因此 Vinyals 等[2]提出了一个序列到序列(Seq2Seq)模型。此模型在编码阶段将对话上下文中的所有单词编码为一个向量,并在解码阶段对该向量进行逐单词解码,以得到回复语句。

但是 Seq2Seq 模型不能很好地利用上下文信息[3],为解决此问题,学者们进行了多方面的改进。2017 年,谷歌团队提出了一种基于自注意力机制的 Transformer 模型[4],该模型可同时提取所有单词的信息,再将这些信息经过加权平均后融入到当前位置的单词信息中,效果较之前的研究有较大提升,并且由于其可以进行并行运算,模型训练的时间也相对较短,因此成为新的主流研究模型。

尽管上述对话模型取得了一定进步,但目前的对话模型依然存在很多困难和挑战,主要表现为生成的回复语句多样性较低,倾向于生成如“我听到这些很难过”“我很高兴你做到了”等通用回复,主要原因在于此类回复语句在数据集中出现的频次较高,因而经过训练后,生成这种回复的概率相应

到稿日期:2023-11-20　返修日期:2024-05-23
通信作者:蒋宗礼(jiangzl@bjut.edu.cn)

较大。另外，目前的对话模型着重于在对话上下文和回复语句之间建立语义关系，模型生成的回复语句缺乏情感，很难与人们产生情感共鸣[5]。通过对开放域对话的进一步分析发现，在人们的日常对话过程中，情感交流与信息交流占据同样重要的地位，情感交流有助于增强人们对模型的认可度与好感，提升模型的共情能力对于提升模型生成回复语句的质量起着十分关键的作用。

共情指通过理解他人的感受，作出合理回复的能力[6]。这要求对话模型在对话过程中主动感知说话者的情感，并在回复语句中对感知到的情感进行合理的回应。在对共情的理解和表达中，情感是一个重要因素，但不是唯一因素。常识知识同样重要，人们往往依赖常识等外部知识来理解对话的语义信息。图 1 展示的对话示例分析了在对话语句中所包含的情感和常识知识信息。

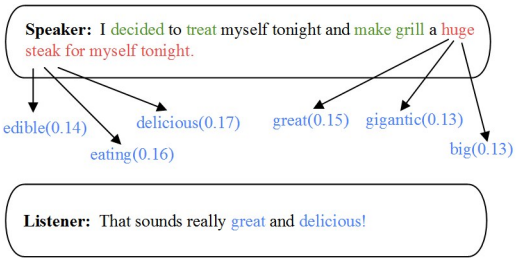

图 1 对话示例(电子版为彩图)

Fig. 1 Dialogue example

图 1 中红色单词为对话语句中具有常识知识的单词，蓝色单词为红色单词所关联的常识知识单词，在蓝色单词旁边扩号中的数值为红色单词和蓝色单词的情感相似度值，绿色单词为情感单词。从图中可以看出，对话历史中的单词和回复语句中的单词通过常识知识关联在一起。

综合上述分析，本文提出了一种融合情感和常识知识的对话生成模型(Emotional Commonsense Dialogue Model，ECDM)，以 Transformer 框架为基础，并在编解码的过程中融合情感和常识知识等信息。本文的主要工作如下：

1)依据常识知识图谱 ConceptNet[7] 和情感词典 NRC_VAD[8] 构建情感知识表，并在此基础上计算得到每个单词对应的常识知识向量。

2)以 Transformer 模型为基础，在编码过程中融入常识知识向量，使模型充分融合常识知识中的隐藏信息，从而更准确地识别上下文的情感类别。为更加充分利用情感信息，将解码分为两个阶段进行：第一阶段解码用于预测单词的情感强度，进而得到单词的情感向量；第二阶段解码将此情感向量结合常识知识向量一同输入解码器中，最终生成情感丰富且与上下文联系更加紧密的回复语句。

3)基于 EmpatheticDialogues[9] 数据集进行实验测试，结果表明，本文所提出的对话模型生成的回复语句通顺连贯且更具同理心，并且相比基线模型，其在 PPL，BLEU，ACC 和 DISTINCT 等指标上都有一定提升。

2 相关工作

2.1 对话生成

随着深度学习技术的快速发展，其相关方法已被广泛应用于对话生成技术研究中。传统的 Seq2Seq 模型[2] 为对话生成技术的发展提供了一个很好的思路，但其因对上下文信息的依赖有限，生成的回复语句普遍存在通用回复等问题。HRED 模型[10] 通过额外增加一个上下文编码器，对句子级别的信息进行编码，提升了对上下文信息的关注程度，因而在一定程度上缓解了 Seq2Seq 模型的问题。VHRED 模型[11] 通过增加一个隐变量，并利用该隐变量使模型更好地追踪上下文信息，从而生成更加流畅的回复语句。Li 等[12] 提出在对话模型中使用强化学习的方法，以 Seq2Seq 模型为基础框架，设计 3 种不同的奖励函数来评估生成的回复语句，在一定程度上减少了 Seq2Seq 模型的通用回复问题，使模型的多样性得到提高。2017 年谷歌提出了一种基于多头自注意力机制的 Transformer 模型[4]，该模型可以同时关注当前单词和其他单词，并通过分配不同的权重使上下文信息得到有效提取，取得了较好的效果。

2.2 共情对话生成

将情感引入到对话模型中，能够有效地提升用户的对话体验[13-14]。Zhou 等[15] 通过指定情感类型来生成符合该情感的回复语句。Lin 等[16] 提出了 MOEL 模型，为每种情感设计一个解码器，并按概率融合多个解码器的输出结果，从而得到混合多种情感的回复语句。Majumder 等[17] 提出了 MIME 模型，将情感分为积极情感和消极情感两种，通过重采样技术提高积极情感并降低消极情感，使得模型可以生成更积极的回复。Li 等[18] 认为常识有助于理解对话中隐藏的情感信息，提出的 KEMP 模型先构建一个情感上下文图，然后将此图融入到改进的编解码器中。Sabour 等[19] 提出的 CEM 模型使用常识知识进行推理训练，使模型具有一定的推理能力，并从情感和认知两个方面来提高模型的共情能力。Wang 等[20] 提出的 SEEK 模型通过分析对话上下文中每个语句的情感变化，提出了一种基于情感和知识的交互方法来提高对话生成的质量。Liu 等[21] 提出的 P2 BOT 模型通过感知对话双方的角色特征，给出了符合人物个性的回复。Che 等[22] 提出的 EmpHi 模型将意图融入共情对话中，通过预测潜在的共情意图，生成与之相对应的回复。

上述模型虽取得了较好的效果，但都没有考虑解码器已生成单词的常识知识信息，而这些信息对于回复语句的生成同样重要。基于这一考虑，本文通过在 Transformer 的编码和解码过程中同时融合单词对应的常识知识信息，并通过两阶段解码生成回复语句，进一步提升对话生成的质量。

3 ECDM 模型

本文提出的 ECDM 模型整体结构如图 2 所示，该模型包括 4 部分：1)情感知识表，由情感词典和常识知识图谱构成；2)编码器，该编码器基于情感知识表，实现对融合情感和常识知识信息之后的上下文进行编码；3)解码器 1，用于预测回复语句单词情感强度；4)解码器 2，用于预测回复语句单词。

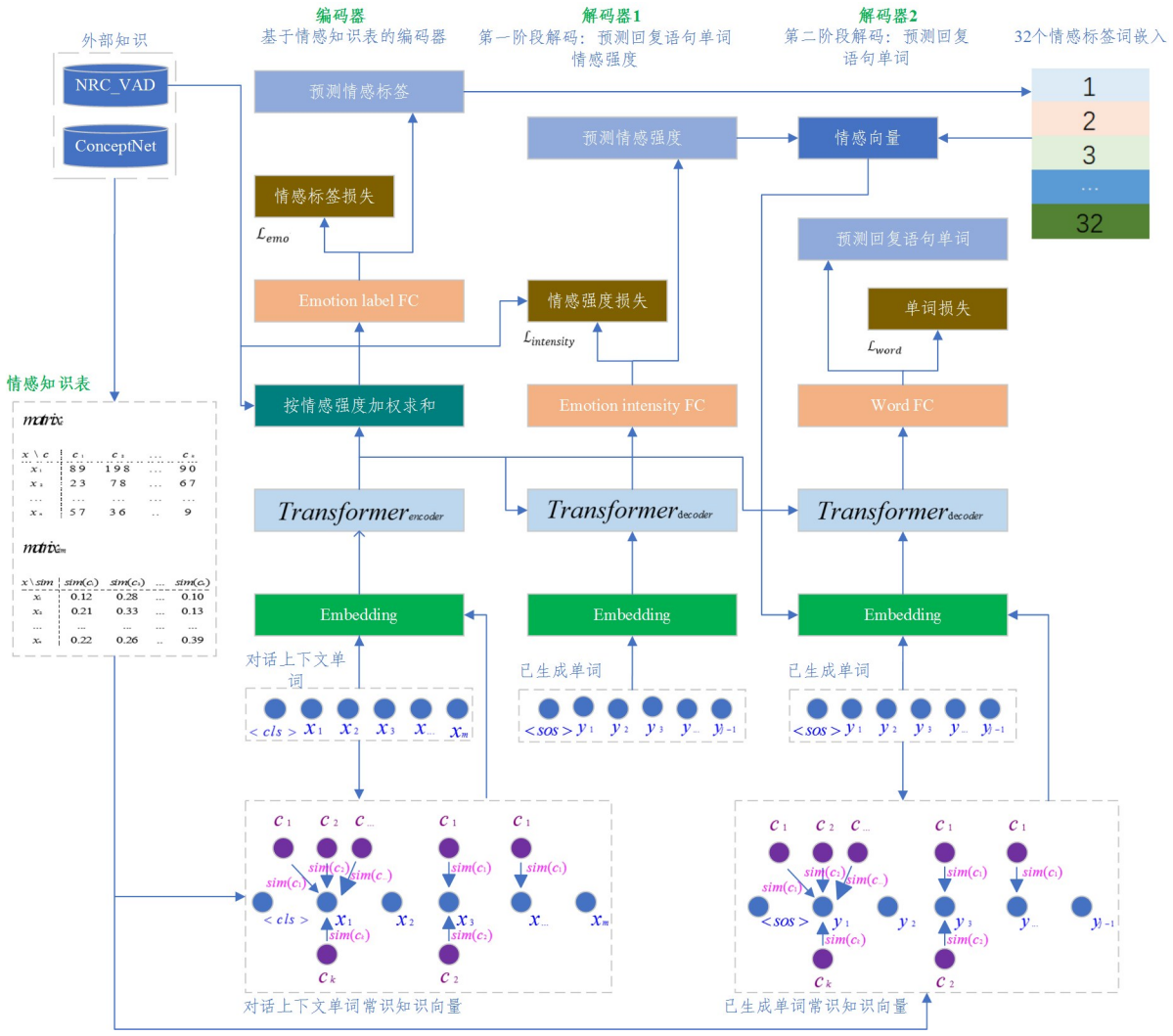

图 2 ECDM 模型整体结构

Fig. 2 Overall architecture of ECDM

ECDM 模型需完成如下任务：将对话上下文 H 中所有语句的单词按顺序串联成一个由 m 个单词组成的序列，标记为 $H=[x_1,x_2,\cdots,x_i,\cdots,x_m]$，通过查询常识知识图谱得到单词对应的常识知识，标记为 K，通过查询情感词典，得到单词对应的三维情感向量值(Valence Arousal Dominance，VAD)，标记为 V，将对话上下文的情感标签标记为 e^*。在给出上述信息的基础上，生成语句连贯且具有一定同理心的回复语句 Y，其概率表示为：

$$P(Y|H,K,V,e^*;\theta)=\prod_{j=1}^{n}P(y_j|y^{<j},H,K,V,e^*;\theta) \quad (1)$$

其中，θ 为模型中的可训练参数，n 为回复语句 Y 中单词的个数，$y^{<j}$ 为回复语句 Y 中已生成的前 $j-1$ 个单词，y_j 为回复语句 Y 中要生成的第 j 个单词。

3.1 情感知识表

本文通过融合常识知识图谱 ConceptNet 和 NRC_VAD 情感词典来构建情感知识表。ConceptNet 使用自然语言描述人类所熟知的一般性常识知识，核心概念由三元组构成，即〈头实体，关系标签，尾实体〉，如 flower 单词对应的其中一个常识三元组为：〈flower，RelatedTo，beautiful〉。NRC_VAD 是一个三维向量集合，即从 3 个不同的维度[Valence(态度)，Arousal(唤起)，Dominance(支配)]描述一个单词，如 flower 单词对应的 VAD 值为：[0.802，0.189，0.254]，这 3 个维度的含义如表 1 所列。

表 1 NRC_VAD 维度说明

Table 1 Interpretations of NRC_VAD dimension

Dimensions	Values	Interpretations
Valence	[0-1]	Negative-Positive
Arousal	[0-1]	Calm-Excited
Dominance	[0-1]	Submissive-Dominant

基于上述两种外部知识，本文以 EmpatheticDialogues 对话数据集中所有单词构成的词汇表为核心，构建情感知识表 $\mathcal{T}$。若词汇表中的单词为 ConceptNet 三元组中的实体，则三元组中对应的另一个实体称为该单词的概念单词。情感知识表 $\mathcal{T}$ 的构建过程为：对词汇表中的每个单词 x，首先查询 ConceptNet 得到该单词所对应的三元组集合，对这些三元组按照关系标签进行过滤，优先选择包含 Synonym，SimilarTo，HasContext，HasSubevent，UsedFor，MotivatedByGoal，HasProperty，CapableOf，Causes，Desires，HasPrerequisite，ReceivesActions 关系标签的三元组，若不存在包含上述关系标签的三元组，则接着选择 RelatedTo 关系标签对应的三元组；若经过

关系标签过滤后，匹配到 l 个三元组，则提取这些三元组中对应的概念单词 $C=[c_1,c_2,\cdots,c_j,\cdots,c_l]$；然后计算这些概念单词与当前单词 x 的情感相似度 Sim。

$$Sim(x,c_j)=\text{softmax}(\eta(c_j)+1-Dist(x,c_j)) \tag{2}$$

其中，$\eta(c_j)$为基于 NRC_VAD 的三维向量计算得到的单词的情感强度。

$$\eta(c_j)=\min-\max\left(\left\|V_a(c_j)-\frac{1}{2},\frac{A_r(c_j)}{2}\right\|_2\right) \tag{3}$$

其中，$\min-\max(\cdot)$为最小值最大值标准化处理函数，$V_a(c_j)$和$A_r(c_j)$分别为 NRC_VAD 三维向量中 Valence 和 Arousal这两个维度对应的数值，$\|\cdot\|_2$ 表示使用 L2 范式正则化。若三元组中概念单词不在 NRC_VAD 中，则 $\eta(c_j)$的值默认为 0.5。

在式(2)中，$Dist(x,c_j)$为词汇表中的单词 x 和其概念单词 c_j 的情感距离。

$$Dist(x,c_j)=abs(V_a(x)-V_a(c_j))/2+abs(A_r(x)-A_r(c_j))/2 \tag{4}$$

其中，$abs(\cdot)$为取绝对值函数；$V_a(\cdot)$和 $A_r(\cdot)$分别为单词在 NRC_VAD 三维向量中对应的 Valence 和 Arousal 维度值，如果词汇表中的单词 x 或概念单词 c_j 不在 NRC_VAD 中，则 $V_a(\cdot)$和 $A_r(\cdot)$的值分别为 0.5 和 0。

将词汇表中的单词 x 对应的 l 个情感相似度 $Sim(x,c_j)$ 按从大到小排序后，选择前 k 个概念单词，以及这些单词对应的情感相似度值，共同构成情感知识表 $\mathcal{T}=[\boldsymbol{matrix}_C,\boldsymbol{matrix}_{sim}]$。该情感知识表由两个二维矩阵构成，$\boldsymbol{matrix}_C$ 描述词汇表中每个单词 x 所关联的概念单词 c_j 在词汇表中的索引值，$\boldsymbol{matrix}_{sim}$ 描述单词 x 和概念单词 c_j 的情感相似度。因此，这两个矩阵的形状是相同的，均为 $\mathbb{R}^{v\times k}$，其中 v 为词汇表中单词的个数，k 为选择的概念单词的个数。

3.2 基于情感知识表的编码器

在编码阶段，首先获取对话上下文 H 中每个单词 x_i 的向量表示 $\boldsymbol{v}(x_i)$，包括词嵌入向量、位置嵌入向量、状态嵌入向量以及常识知识向量 4 部分。

$$\boldsymbol{v}(x_i)=\boldsymbol{E}_w(x_i)+\boldsymbol{E}_p(x_i)+\boldsymbol{E}_w(S(x_i))+\dot{\boldsymbol{v}}(x_i) \tag{5}$$

其中，$\boldsymbol{E}_w\in\mathbb{R}^d$ 为词嵌入层；$\boldsymbol{E}_p\in\mathbb{R}^d$ 为位置嵌入层；d 表示嵌入层的维度；$S(x_i)$表示当前单词所处的状态，若当前单词所在语句为用户所说，则其值为〈USER〉，若当前单词所在语句为系统所说，则其值为〈SYS〉。$\dot{\boldsymbol{v}}(x_i)$为单词的常识知识向量，计算方法如式(6)所示：

$$\dot{\boldsymbol{v}}(x_i)=\sum_{j=1}^{k}\boldsymbol{E}_w(c_j)\odot s_j \tag{6}$$

其中，$\odot$为点乘运算；c_j 为查询情感知识表$\mathcal{T}$中的 $\boldsymbol{matrix}_C$ 矩阵后，得到的单词所关联的概念单词列表$[c_1,c_2,\cdots,c_j,\cdots,c_k]$中的元素；$s_j$ 为查询情感知识表$\mathcal{T}$中的 $\boldsymbol{matrix}_{sim}$ 矩阵后，得到的概念单词对应的情感相似度值列表$[s_1,s_2,\cdots,s_j,\cdots,s_k]$中的元素。概念单词列表中的单词 c_j 经过词嵌入层后，得到概念单词的词嵌入向量 $\boldsymbol{E}_w(c_j)$，将所有概念单词的词嵌入向量与其对应的情感相似度相乘后再进行相加，得到单词 x_i 对应的常识知识向量。

使用 $\boldsymbol{V}_H$ 表示对话上下文 H 中所有单词向量的集合，即 $\boldsymbol{V}_H=\{\boldsymbol{v}(x_i)\}_{i=1,\cdots,m}$，并将其送入 Transformer 的编码器中进行编码。

$$\widetilde{\boldsymbol{V}}_H=Transformer_{\text{Encoder}}(\boldsymbol{V}_H) \tag{7}$$

经过 Transformer 的编码器对单词的向量进行编码后，每个单词的向量都发生了变化，此时单词的向量表明了单词在当前对话上下文中的含义。

3.3 预测对话上下文的情感标签

在完成对上下文单词向量的编码后，接着预测对话上下文的情感标签 $\widetilde{e}$，进而可以使用预测情感标签去指导回复语句的生成。首先通过计算单词的情感强度，得到单词在预测情感标签时的权重，并使用 softmax 函数对得到的权重进行归一化处理。

$$A_H=\text{softmax}(\{\eta(x_i)\}_{i=1,\cdots,m}) \tag{8}$$

其中，$\eta(x_i)$为式(3)中计算单词情感强度的方法。

本文认为，对话上下文中单词的情感强度对于预测情感标签起着重要作用，情感强度越强的单词，对情感标签的影响越大。对于得到的每个单词权重，将其和单词的向量表示进行相乘，再将所有单词的向量进行相加操作，即可得到进行预测情感标签的向量 $\boldsymbol{v}(\widetilde{e})$。

$$\boldsymbol{v}(\widetilde{e})=\sum_{i=1}^{m}a(x_i)\odot\widetilde{\boldsymbol{v}}(x_i) \tag{9}$$

其中，$a(x_i)$表示在 A_H 集合中单词 x_i 对应的权重，$\widetilde{\boldsymbol{v}}(x_i)$为 $\widetilde{\boldsymbol{V}}_H$ 集合中单词 x_i 对应的向量。

$\boldsymbol{v}(\widetilde{e})$经过一个全连接输出层和一个 softmax 激活函数之后，得到具体的情感标签预测分布。

$$p(\widetilde{e})=\text{softmax}(\boldsymbol{w}_e\boldsymbol{v}(\widetilde{e})) \tag{10}$$

其中，$\boldsymbol{w}_e\in\mathbb{R}^{q\times d}$为情感全连接输出层，$q$ 为情感标签个数，在 EmpatheticDialogues 数据集中，q 为 32。

在训练时，本文使用负对数似然函数作为情感标签预测损失函数，以对模型参数进行调整。

$$\mathcal{L}_{\text{emo}}=-\log p(\widetilde{e}=e^*) \tag{11}$$

3.4 基于情感知识表的解码器

在解码过程中，实际上是依据已生成的前 $j-1$ 单词及头部添加的〈SOS〉标签单词，即使用 $y^{<j}=\{y_i\}_{i=0,\cdots,j-1}$，$y_0$ 为〈SOS〉，来预测第 j 个单词。本文的解码过程可以分为两个部分：第一部分预测回复语句中单词的情感强度，第二部分预测回复语句中的单词。

3.4.1 预测回复语句中单词的情感强度

对于第一阶段编码，首先获取回复语句中已生成单词 $y^{<j}$中的每个单词 y_i 的向量表示 $\boldsymbol{v}(y_i)$，其由两部分构成，即词嵌入向量和位置嵌入向量。

$$\boldsymbol{v}(y_i)=\boldsymbol{E}_w(y_i)+\boldsymbol{E}_p(y_i) \tag{12}$$

其中，$\boldsymbol{E}_w\in\mathbb{R}^d$ 为词嵌入层，$\boldsymbol{E}_p\in\mathbb{R}^d$ 为位置嵌入层，d 表示嵌入层的维度。

使用 $\boldsymbol{V}_{y^{<j}}$ 表示 $y^{<j}$中所有已生成单词向量的集合，即 $\boldsymbol{V}_{y^{<j}}=\{\boldsymbol{v}(y_i)\}_{i=0,\cdots,j-1}$，并结合前面 Transformer 编码器的输出结果 $\widetilde{\boldsymbol{V}}_H$，一起送入 Transformer 的解码器中进行解码，从而得到被解码后的单词向量表示。

$$\widetilde{\boldsymbol{V}}_{y^{<j}}=Transformer_{\text{Decoder}}(\boldsymbol{V}_{y^{<j}},\widetilde{\boldsymbol{V}}_H) \tag{13}$$

解码后得到的向量集合再经过一个全连接输出层和一个sigmoid激活函数，即得到情感强度预测值集合。

$$\hat{\eta}_{y^{<j}}=\text{sigmoid}(\boldsymbol{w}_{\eta}\widetilde{\boldsymbol{V}}_{y^{<j}}) \tag{14}$$

其中，$\boldsymbol{w}_{\eta}\in\mathbb{R}^{1\times d}$表示情感强度全连接输出层。

要生成的第 j 个单词的情感强度预测值 $\hat{\eta}(y_j)$ 为 $\hat{\eta}_{y^{<j}}$ 集合中最后一个位置对应的情感强度值。

在训练过程中，使用均方差函数作为单词情感强度的预测损失函数，以对模型参数进行调整。

$$\mathcal{L}_{\text{intensity}}=\frac{1}{n}\sum_{i=0}^{n-1}(\hat{\eta}(y_i)-\eta(y_{i+1}))^2 \tag{15}$$

其中，$\hat{\eta}(y_i)$为 $\hat{\eta}_{y^{<j}}$ 集合中第 i 个位置(即单词 y_i)对应的预测情感强度值，$\eta(y_{i+1})$为数据集给出的标准回复语句中第 $i+1$ 个位置对应单词的情感强度值，其计算方法如式(3)所示，n 为标准回复语句中单词的个数。

3.4.2 预测回复语句中的单词

对于第二阶段编码，首先获取回复语句中已生成单词 $y^{<j}$ 中的每个单词 y_i 的向量表示 $v'(y_i)$，其由四部分构成：词嵌入向量、位置嵌入向量、常识知识向量，以及情感向量。

$$\boldsymbol{v}'(y_i)=\boldsymbol{E}_w(y_i)+\boldsymbol{E}_p(y_i)+\dot{\boldsymbol{v}}(y_i)+\breve{\boldsymbol{v}}(y_i) \tag{16}$$

其中，$\boldsymbol{E}_w\in\mathbb{R}^d$ 为词嵌入层，$\boldsymbol{E}_p\in\mathbb{R}^d$ 为位置嵌入层；$\dot{\boldsymbol{v}}(y_i)$为单词的常识知识向量，其计算式如式(6)所示；$\breve{\boldsymbol{v}}(y_i)$为单词的情感向量，其计算式如式(17)所示。

$$\breve{\boldsymbol{v}}(y_i)=\hat{\eta}(y_i)\odot\boldsymbol{E}_e(\tilde{e}) \tag{17}$$

其中，$\hat{\eta}(y_i)$为单词 y_i 对应的预测情感强度值，$\boldsymbol{E}_e(\tilde{e})$表示情感词嵌入向量，$\boldsymbol{E}_e\in\mathbb{R}^d$ 为情感词嵌入，d 表示嵌入层的维度，$\tilde{e}$ 表示在第3.3章节中预测的对话上下文情感标签。情感强度越大的单词，其获得的情感标签向量的值也应该越大，因此将情感词嵌入向量 $\boldsymbol{E}_e(\tilde{e})$和预测单词的情感强度值 $\hat{\eta}(y_i)$做点乘，即可得到单词的情感向量。

使用 $\boldsymbol{V}'_{y^{<j}}$ 表示 $y^{<j}$ 中所有单词向量的集合，即 $\boldsymbol{V}'_{y^{<j}}=\{\boldsymbol{v}'(y_i)\}_{i=0,\cdots,j-1}$。将其与前面 Transformer 编码器的输出结果 $\widetilde{\boldsymbol{V}}_H$ 一起送入另一个 Transformer 解码器中进行解码，从而得到被解码后的单词向量表示。

$$\widetilde{\boldsymbol{V}}'_{y^{<j}}=Transformer_{\text{Decoder}}(\boldsymbol{V}'_{y^{<j}},\widetilde{\boldsymbol{V}}_H) \tag{18}$$

解码后得到的单词向量经过一个全连接输出层和一个softmax激活函数，得到单词的概率分布。

$$p(y_j)=\text{softmax}(\boldsymbol{w}_v\widetilde{\boldsymbol{v}}'(y_{j-1})) \tag{19}$$

其中，$\boldsymbol{w}_v\in\mathbb{R}^{v\times d}$ 为单词全连接输出层，v 为词汇表中总的词汇个数，$\widetilde{\boldsymbol{v}}'(y_{j-1})$为 $\widetilde{\boldsymbol{V}}'_{y^{<j}}$ 集合中最后一个单词对应的向量。

在训练过程中，使用 LabelSmoothing 函数作为单词预测损失函数，以对模型的参数进行调整。

$$\mathcal{L}_{\text{word}}=LabelSmoothing(p(y_j=y_j^*)) \tag{20}$$

其中，y_j^* 为数据集中给出的标准回复语句中的单词。

为加快训练速度，在计算单词 y_i 对应的情感向量时，即针对式(17)，首先以较大的概率选择数据集中给出的标准回复中单词的情感强度值 $\eta(y_i)$；随着训练的进行，逐步提高使用预测得到的情感强度值 $\hat{\eta}(y_i)$的概率，当训练超过18000个批次后，以60%的概率选择 $\hat{\eta}(y_i)$，以40%的概率选择 $\eta(y_i)$。

最终，使用多任务学习模型，实现情感标签预测误差、情感强度预测误差及单词预测误差的同时学习。模型的总损失函数为：

$$\mathcal{L}=\gamma_1\mathcal{L}_{\text{emo}}+\gamma_2\mathcal{L}_{\text{intensity}}+\gamma_3\mathcal{L}_{\text{word}} \tag{21}$$

其中，γ_1，γ_2，γ_3 为超参数，在训练过程中，其值分别为0.65，0.1，0.25。

4 实验与结果分析

4.1 数据集

本文在实验过程中使用 EmpatheticDialogues 数据集，该数据集包含约25000个开放域对话及32种情感标签。每个对话均包含两个角色，一个是说话者，一个是倾听者。说话者首先说出符合特定情感标签的语句，倾听者通过理解说话者的语句，推测其要表达的情感，并据此给出符合该情感的回复语句。

4.2 基线模型

为验证本文方法的有效性，选取了以下基线模型作对比。

Transformer[4]：标准 Transformer 模型。该模型只使用标准的 Transformer 的编码器层和解码器层进行训练，仅使用对话上下文来预测输出语句。

Transformer-multi[9]：在标准 Transformer 模型的基础上，增加对情感标签预测的训练，使得模型可以感知到当前对话的情感，从而生成合适的回复语句。

MoEL[14]：融合多个情感解码器的对话模型，即通过为每种情感标签分别独立设置与其相对应的解码器，从不同的情感角度预生成回复语句，并将这些预生成回复语句进行统一汇总，最后生成真正的回复语句。

KEMP[16]：融合外部知识的对话模型，通过外部知识来丰富对话上下文，并据此构建一个情感知识上下文图，然后使用改进的编码器和解码器从该图中预测情感标签信息，并指导回复语句的生成。

4.3 实验参数设定

对 EmpatheticDialogues 数据集中的所有对话语句进行分词、转小写等处理之后，得到一个包含22335个单词的词汇表，使用预训练的300维 Glove 向量来初始化词汇表对应的词嵌入向量。

本文模型以 PyTorch 框架为基础，对模型训练时使用一张 RTX3060 显卡，基础 Transformer 模型的配置参数和 CEM 模型[17]中的配置保持一致，其他超参数配置如表2所列。在实际生成回复语句时，回复语句最多包含30个单词，整个训练过程花费约30 min，总共执行36000个训练批次。

表2 模型参数设置

Table 2 Model parameter settings

参数	取值
pointer_gen	False
batch_size	16
Smoothing	0.1
Dropout	0
Concept_num	8

4.4 评估指标说明

本文使用自动评估和人工评价相结合的方式对模型效果进行验证。

4.4.1 自动评估指标

ACC:情感标签预测准确度,即依据对话上下文预测的情感标签与数据集中给出的情感标签的匹配程度,其值越大,说明模型识别情感标签的准确性越好。

Perplexity:困惑度,用于检测模型生成的回复语句相对于数据集中给定的回复语句的平均生成质量,其值越小,表明模型生成的回复语句质量越好。

BLEU:也是一种评价当前模型生成回复语句与数据集中给定的回复语句相似度的指标,其值越大,说明模型生成的回复语句与数据集中给定的回复语句的相似度越高,模型性能也就越好。

Dist-1 和 Dist-2:评估模型生成回复语句的多样性,主要从单个单词和两个单词组成的词组在对应的整个词组中所占的比例来计算,其值越大,说明模型生成的回复语句的单词越丰富,模型的多样性也越好。

4.4.2 人工评价指标

自动评估指标虽然可以验证模型的效果,但并不能完全模拟人们的主观感受。为弥补这种缺陷,本文从语义相关、流畅度和情感匹配 3 个方面对模型生成的回复语句进行人工评价,评分采用 10 分制(1~10)。

语义相关:用于评估回复语句与对话上下文的关联程度,关联程度越高则得分越高。

流畅度:用于从可读性角度评估回复语句,可读性越好则得分越高。

情感匹配:用于评估回复语句中的情感和对话上下文中的情感是否匹配,匹配程度越高,则得分越高。

4.5 实验结果及分析

4.5.1 自动评估

从表 3 的结果可以看出,本文提出的 ECDM 模型在各项指标上均显著优于基线模型,这体现了 ECDM 模型的优越性。具体地,与基线模型的最佳实验结果相比,本文模型的 PPL 指标降低了 38%,BLEU 指标提升了 13%,ACC 指标提升了 3%,Dist-1 和 Dist-2 指标分别提升了 10%,38%。总的来看,相对于基线模型,本文提出 ECDM 模型能更准确地识别对话上下文的情感,生成的回复语句在相似度和多样性等方面的性能提升比较明显。

表 3 ECDM 模型与基线模型自动评估结果对比

Table 3 Automatic evaluation results comparison between ECDM and baseline models

methods	*PPL*	*BLEU*/%	*ACC*/%	*Dist*-1/%	*Dist*-2/%
ECDM	**22.97**	**3.12**	**38.86**	**0.55**	**2.66**
Transformer	37.64	2.67	—	0.41	1.72
Transformer-multi	37.47	2.31	31.13	0.50	1.89
MOEL	37.18	2.71	29.61	0.39	1.92
KEMP	38.25	2.75	37.64	0.36	1.46

4.5.2 人工评价

从测试集中随机抽取 50 组对话样本,使用本文提出的模型和基线模型分别生成回复语句,并邀请 3 名受过良好高等教育的人员作为评审员依据人工评价指标对上述回复语句进行人工打分。为确保公平性,生成回复语句的模型名称对评审员完全匿名。人工评价结果如表 4 所列。与基线模型相比,本文提出的 ECDM 模型在各个指标中均取得了最高分,这表明本文提出的模型在对话上下文语义相关和情感匹配等方面都有一定提升,可以生成高质量的回复语句,有效提升人机对话体验效果。

表 4 ECDM 模型与基线模型人工评价结果对比

Table 4 Human evaluation results comparison between ECDM and baseline models

模型名称	语义相关得分	流畅度得分	情感匹配得分
ECDM	**5.7**	**6.3**	**5.7**
Transformer	4.1	5.1	4.8
Transformer-multi	4.0	5.4	4.3
MOEL	5.2	6.0	5.5
KEMP	4.8	5.9	5.3

4.5.3 消融实验

为更好地分析 ECDM 中各个部分的作用,本文做了以下消融实验:1)w/o ec 在编码阶段将单词的常识知识向量部分去掉,即针对式(5),将 $\dot{\boldsymbol{v}}(x_i)$ 部分去掉;2)w/o dc 在第二阶段解码时将已生成单词的常识知识向量部分去掉,即针对式(16),将 $\dot{\boldsymbol{v}}(y_i)$ 部分去掉;3)w/o de 在第二阶段解码时将情感向量部分去掉,即针对式(16),将 $\check{\boldsymbol{v}}(y_i)$ 部分去掉,相当于取消第一阶段解码,将两阶段解码改为仅使用 Transformer 解码器进行解码。

从表 5 可以看出,ECDM 模型的结果整体比较有优势,消融实验的模型(w/o)结果部分较好,但与 ECDM 模型的实验结果相比,并没较明显的优势。

从 PPL 指标结果数据来看,在第二阶段解码过程中是否使用情感向量对结果的影响较大,如果去掉情感向量,则 PPL 结果会有较大上升(w/o de 的实验结果)。该结果表明,准确地识别对话上下文的情感标签以及预测当前要生成单词的情感强度对于提升模型生成回复语句的质量十分关键。

表 5 消融实验

Table 5 Ablation experiment

模型名称	*PPL*	*BLEU*/%	*ACC*/%	*Dist*-1/%	*Dist*-2/%
ECDM	22.97	**3.12**	38.86	**0.55**	**2.66**
w/o ec	23.32	2.75	37.85	0.55	2.61
w/o dc	**22.90**	2.46	**39.90**	0.53	2.45
w/o de	36.09	2.67	38.25	0.48	2.15

从 BLEU 指标结果数据来看,在第二阶段解码过程中是否使用已生成单词的常识知识向量对结果的影响较大,若不使用,则 BLEU 会有明显下降(w/o dc 的实验结果)。该结果表明,已生成单词的常识知识对于后面要生成单词的影响较大。在模型生成回复语句时,不仅要考虑编码阶段单词的常识知识,还应同时考虑在解码阶段时已生成单词的常识知识。

从 ACC 的结果数据来看,在编码阶段是否使用单词的常识知识向量对结果的影响较大,如果不使用,则 ACC 会有明显下降(w/o ec 的实验结果)。分析其原因是在预测情感标签时,当前模型仅依赖于编码器的输出结果。因此,在编码阶段融合外部知识,会使模型在进行训练时更偏向于提高情感

标签的预测准确性。

从 Dist-1 和 Dist-2 指标来看,在编码和解码的过程中使用单词的常识知识向量,以及在解码过程中使用情感向量对其都有影响,缺少其中任何一个因素,这些指标都会降低。其中,在解码阶段使用情感向量这个因素影响较大,分析其原因是这两个指标的主要作用为评价生成语句的多样性,无论哪个阶段都会影响到生成语句的质量,而解码阶段对生成语句的影响更大。

4.5.4 PPL 指标分析

从 4.5.1 节自动评估的实验结果可以看出,本文模型的 PPL 指标相对于基线模型明显降低,结合 4.5.3 节中 w/o de 消融实验的结果进行分析可以看出,本文模型提出的在第一阶段解码对要生成单词的情感强度进行预测是导致此结果的重要原因。

由于在第一阶段解码中需要对要生成单词的情感强度进行预测,因此模型新增了一个解码器。为验证增加解码器对 PPL 值的影响,本文在 w/o de 消融实验的基础上进行实验,即仅使用 Transformer 解码器预测单词本身,通过增加解码器中的解码层数,验证不同解码层数对 PPL 指标的影响,实验结果如表 6 所列。

从表 6 可以看出,只增加解码层数 PPL 指标明并不会出现显降低,而本文模型提出的两阶段解码相当于是从两个不同方面对要生成的单词进行预测,一个是直接预测要生成的单词本身,另一个是通过预测情感强度来间接预测要生成的单词。通常来说,回复语句都是包含情感的,本文模型相当于从单词本身和单词情感两方面对要生成的单词进行预测,因而效果比单一对单词本身进行预测更好。

表 6 解码层数对 PPL 值影响

Table 6 Effect of the number of decoder layer on PPL value

解码层数	PPL 值
2	35.27
3	34.46
4	34.37
5	34.14
6	34.69

4.6 案例分析

ECDM 模型和基线模型生成的回复语句如表 7 所列,其中,标准回复为对话数据集给出的回复语句。可以看出,ECDM 模型生成的回复语句内容丰富,且更具有同理心。

从第一个例子来看,本文模型生成的回复语句表达了和说话者相同的情绪,在情感上更容易和说话者产生共鸣,而 Transformer 模型生成的回复语句在语义上和对话上下文不一致,其他 3 种模型生成的回复语句虽然也表达了相同的情感,但回复语句的内容偏向于通用回复。

表 7 ECDM 与基线模型生成的回复语句对比

Table 7 Comparison of response statements generated by ECDM and baseline models

情感标签	terrified
对话上下文	Speaker:one night,i was home alone,when suddenly,i heard a loud bang ! Listener:oh goodness,what was it ? Speaker:it was a ghost ! i ran into the kitchen to investigate,and i saw what looked like a white dress flapping out the window.
标准回复	wow,that must have been a terrible experience.
Transformer	i am glad you were able to get a little scary !
Transformer-multi	oh no ! i hate when that happens.
MOEL	oh no ! i hope youdid not hurt yourself !
KEMP	i would be too ! i am glad you are okay now !
ECDM	oh wow,that is terrifying ! i would have been so scared of the same thing.
情感标签	hopeful
对话上下文	Speaker:i am hopeful that my friend will be able to make it to my party
标准回复	That's good,is it very important that this friend be there ?
Transformer	that is great ! i am sure you will get better !
Transformer-multi	that is good. i hope you have a good time.
MOEL	what kind of friend you are in ?
KEMP	that is good to hear. i hope you guys can find it soon.
ECDM	that is great. what are you going to do with your friend ?

从第二个例子来看,ECDM 模型生成的回复语句在情感上与说话者保持一致,同时对话题进行了进一步扩展,有助于引导对话的继续进行。而 MOEL 模型生成的语句是使用反问的方式,虽然可以引导进一步对话,但在情感上很难引起说话者的共鸣。其他 3 种模型生成的回复在情感上和说话者相近,但生成的回复语句仍然偏向于通用回复。

结束语 本文提出的融合情感和常识知识的对话生成模型 ECDM 通过融合情感词典和常识知识图谱等信息,使得模型更容易识别对话上下文中的情感标签,生成的回复语句流畅且更具同理心。实验结果表明,ECDM 模型在各个指标上都优于基线模型。另外,实验还验证了模型中的各个功能模块在整个模型中的作用。实验发现,如果可以提升第一次解码过程中对单词情感强度准确性的预测,那么整个模型的性能会有更进一步的提升,这也是后续提升模型性能的一个方向。

参 考 文 献

[1] CHEN X,ZHOU Q. A survey of research onopen domain dialogue systems[J]. Journal of Chinese Information Processing,2021,35(11):1-12.

[2] VINYALS O,LE Q. A neural conversational model[J]. arXiv:1506.05869,2015.

[3] SHAO L, GOUWS S, BRITZ D, et al. Generating high-quality and informative conversation responses with sequence-to-sequence models[C]// Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. 2017: 2210-2219.

[4] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]// Advances in Neural Information Processing Systems. 2017: 5998-6008.

[5] ZHANG S X, LI J, ZHU G L, et al. Dialogue generation model based on improved encoder-decoder and emotion dictionary[J]. Computer Engineering and Design, 2023, 44(2): 570-575.

[6] KESKIN C S. From what isn't Empathy to Empathic Learning Process[J]. Procedia-Social and Behavioral Sciences, 2014, 116: 4932-4938.

[7] SPEER R, CHIN J, HAVASI C. ConceptNet 5.5: An open multilingual graph of general knowledge[J]. arXiv: 1612.03975, 2017.

[8] MOHAMMAD S. Obtaining reliable human ratings of valence, arousal, and dominance for 20000 English words[C]// Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (volume 1: Long papers). 2018: 174-184.

[9] RASHKIN H, SMITH E M, LI M, et al. Towards empathetic open-domain conversation models: A new benchmark and dataset[J]. arXiv. 1811.00207, 2018.

[10] SERBAN I V, SORDONI A, BENGIO Y, et al. Building end-to-end dialogue systems using generative hierarchical neural network models[C]// Proceedings of the AAAI Conference on Artificial Intelligence. 2016.

[11] SERBAN I V, KLINGER T, TESAURO G. Multiresolution recurrent neural networks: An application to dialogue response generation[C]// Proceedings of the AAAI Conference on Artificial Intelligence. 2017.

[12] LI J, MONROE W, RITTER A, et al. Deep reinforcement learning for dialogue generation[J]. arXiv: 1606.01541, 2016.

[13] LIU L M, CHEN Y Z, ZHEN J T. Multi-turn Dialogue Model for Domain Data Augmentation and Multi-granularity Semantic Understanding[J]. Journal of Chinese Computer Systems. 2024, 45(7): 1585-1591.

[14] CHENG T T, YAO C L, YU X Q, et al. Empathetic Dialogue Generation by Incorporating Commonsense Knowledge Based on Multi-Head Attention Mechanism[J]. Computer Engineering, 2024, 50(6): 94-101.

[15] ZHOU H, HUANG M, ZHANG T, et al. Emotional chatting machine: Emotional conversation generation with internal and external memory[C]// Proceedings of the AAAI Conference on Artificial Intelligence. 2018.

[16] LIN Z, MADOTTO A, SHINI J, et al. MoEL: Mixture of empathetic listeners[C]// Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing. 2019: 121-132.

[17] MAJUMDER N, HONG P, PENG S, et al. MIME: MIMicking emotions for empathetic response generation[C]// Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing. 2020: 8968-8979.

[18] LI Q, LI P, REN Z, et al. Knowledge bridging for empathetic dialogue generation[C]// Proceedings of the AAAI Conference on Artificial Intelligence. British Columbia, Canada, 2022: 10993-11001.

[19] SABOUR S, ZHENG C, HUANG M. Cem: Commonsense-aware empathetic response generation[C]// Proceedings of the AAAI Conference on Artificial Intelligence. 2022: 11229-11237.

[20] WANG L, LI J, LIN Z, et al. Empathetic dialogue generation via sensitive emotion recognition and sensible knowledge selection. [C]// Findings of the Association for Computational Linguistics. 2022: 4634-4645.

[21] LIU Q, CHEN Y, CHEN B, et al. You impress me: Dialogue generation via mutual persona perception[C]// Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. 2020: 1417-1427.

[22] CHE M Y, LI S, YANG Y. EmpHi: Generating empathetic responses with human-like intents[C]// Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2022: 1063-1074.

CHENG Jinfeng, born in 1990, postgraduate. Her main research interests include natural language processing and so on.

JIANG Zongli, born in 1956, professor, Ph.D, Ph.D supervisor. His main research interests include network information processing and so on.

(责任编辑:何杨)