

面向蒙古语情感对话生成的 构词形态编码单元*

王奎波¹, 马志强^{1,2}, 王洪彬¹, 王子洲¹, 宋经纬¹, 银慧¹

(1. 内蒙古工业大学智能科学与技术学院, 呼和浩特 010080; 2. 内蒙古自治区北疆网络空间
安全重点实验室, 呼和浩特 010080)

摘要: 蒙古语具有黏着语特性, 一个词根可派生出多个语气和情感强度不同的词, 现有模型多依赖词和字符表示, 缺乏对蒙古语词根与后缀结构的建模, 导致生成的回复内容单调且与语境不符。本文提出一种面向蒙古语情感对话生成的构词形态编码方法, 在Transformer架构上引入形态编码单元, 包括词根嵌入层、后缀层级位置编码和形态注意力机制, 以显式建模词根-后缀关系, 并通过概率融合机制将形态向量与语义向量结合, 提升模型对派生词和情感表达的理解能力。在自建的蒙古语情感对话数据集上进行实验, 结果表明, 该方法在BLEU、Rouge、Distinct-1和情感分类准确率方面均优于多种基线模型, 其中正向情感准确率达93.2%, 负向情感准确率达90.4%。消融实验和形态注意力机制可视化分析进一步验证了形态信息与融合机制的有效性。该研究为低资源语言的情感对话生成提供了新的建模思路。

关键词: 蒙古语; 情感对话生成; 构词形态编码; Transformer

中图分类号: TP391 **文献标志码:** A

情感对话生成任务是使机器在对话中生成蕴含情感、有针对性的回复。Zhou等^[1]提出情感聊天机(ECM), 首次将情感信息融入对话生成模型中, 让机器生成的回复更具有情感色彩。当前情感对话生成研究聚焦于中文或英文等资源丰富的语言, 而在蒙古语这类低资源语言场景中仍面临诸多挑战。蒙古语具有高度黏着性的构词特点, 一个情感词根往往能派生出多个具有不同语气或情感强度的派生词。现有的情感对话生成模型在对蒙古语建模时, 缺乏对其词根与后缀结构的建模机制, 模型无法正确编码这些蒙古语词的语义, 导致生成回复语义缺失。因此, 研究适合蒙古语构词规律的情感对话生成模型, 对于提升低资源语言的人机交互质量具有重要意义。

在先前的研究中, 多位学者已经关注到蒙古语的形态学信息并将其应用到自然语言处理任务中。Liu等^[2]为了提高蒙古语形态分割的准确性, 提出一种神经网络模型, 该模型包含两个编码器和一个解码器, 分别用于提取词内特征和词外特征, 并最终生成分割结果。Ji等^[3]提出一种基于生成对抗网络(GAN)的模型, 通过添加形态噪声来缓解未知的词汇问题, 从而提高了蒙古语-中文神经机器翻译系统的性能。Liu等^[4]发现蒙古语语音合成系统中韵律短语预测的准确率不高, 是因为

* 收稿日期: 2025-10-08; 修回日期: 2026-03-23

基金项目: 国家自然科学基金项目(62166029); 内蒙古自治区科研基础条件及平台(2025KYPT0014); 内蒙古自治区高等学校创新团队发展计划项目(NMGIRT2506); 内蒙古自治区高等学校碳达峰碳中和研究项目(STZX202307); 内蒙古自治区直属高校基本科研业务费项目(JY20220281)

作者简介: 王奎波(2000—), 男, 内蒙古赤峰人, 2023级硕士研究生。E-mail: 1622394709@qq.com

通信作者: 马志强(1972—), 男, 内蒙古呼和浩特人, 教授, 硕士。主要从事情感计算、语音识别和模型安全研究。E-mail: mzq_bim@imut.edu.cn

蒙古语是一种黏着语,其词汇量庞大且形态变化多样,导致模型难以准确理解蒙古语的语义。因此,将形态和音位分解并加入输入文本,使用自注意力神经分类器进行短语断句预测,以提高韵律短语预测的准确率。尽管上述研究在各方面取得了进展,但在目前的蒙古语情感对话生成系统中形态学信息的应用仍显不足,尤其缺乏对词根与后缀结构的深度建模。如图1所示,用户表达了自己难过的心情,并要求模型给用户提供一些建议。“难过”这个词在蒙古语中由词根派生而来,图1(a)中的Transformer编码器采用基于词或字符的表示方式,忽视了蒙古语构词过程中词根与多个层级后缀之间的语义与情感。这种结构性信息的缺失使得模型难以对复杂的派生情感词进行有效建模,导致生成的回复效果不佳。在图1(b)中,当引入蒙古语构词形态结构后,模型捕捉到用户的情感状态,可以从用户输入的句子语义中分析出产生这种状态的原因,从而生成开导用户的回复。

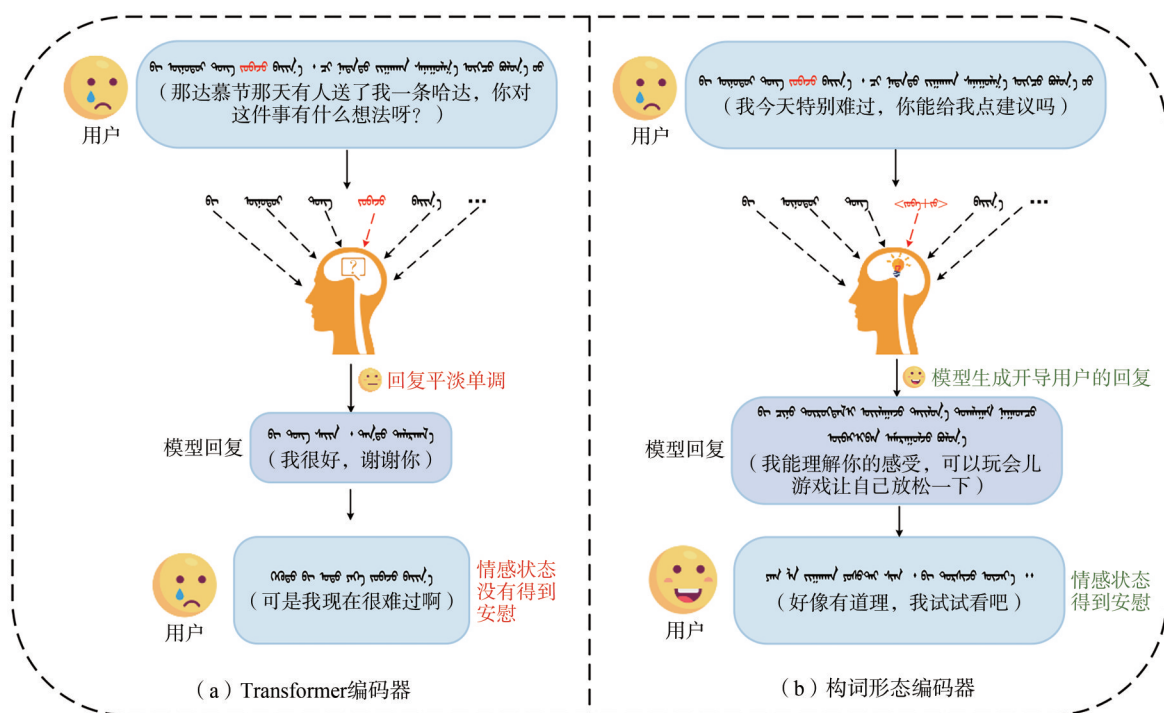


图1 蒙古语人机情感对话示例

Fig. 1 Example of human-computer emotional dialogue in Mongolian

为解决上述问题,提出一种面向蒙古语情感对话生成模型的构词形态编码单元,在Transformer情感对话生成模型的基础上,引入一种形态结构感知的编码模块,通过显式建模蒙古语词根与后缀的组合关系,提取词形结构中的语义与情感特征,并将其与上下文语义特征进行融合,从而使模型生成合理的情感回复。

本文的贡献为以下几点。

① 针对蒙古语复杂的词根-后缀构词规律,设计了一种形态编码器(Morphological encoder),在词根嵌入、后缀层级位置编码和形态注意力机制中建模构词形态特征,从而提升了模型对词形变化和派生词的泛化能力。

② 在Transformer对话生成模型中引入形态向量表示,有效缓解了传统对话生成模型对词根及其派生形式理解不足的问题。

③ 实验结果表明,形态编码器提升了模型的生成质量,也在情感准确性上显著优于对比模型,为低资源语言的对话建模提供了借鉴。

1 相关工作

1.1 传统编码方法

对话生成的回复质量取决于编码器对语义信息的建模能力,早期的研究大多依赖基于循环神经网络(RNN)的端到端结构。Zhou等^[1]采用Seq2Seq^[5]在解码器中融入情绪类别,以控制模型生成指定情绪类别的回复,其虽然能够有效捕捉序列中的上下文依赖,但由于梯度消失和并行计算效率低,在长文本建模和复杂对话场景中的表现不佳。Serban等^[6]提出层级递归编码(HRED),用RNN对句子内部的词序列进行编码,得到句子表示,再用另一层RNN对对话中的句子表示进行建模,捕捉上下文对话历史,进一步增强了模型对对话轮次结构的建模能力。Shen等^[7]提出条件变分框架(SPHRED),使用两个独立的HRED对用户输入与机器回复分开编码,然后使用变分自编码器融合情感和语义,为后续多样化和上下文一致的对话生成提供了支持。情感对话生成的研究强调编码器对情感特征的捕捉。Li等^[8]提出多尺度交互式共情对话生成模型(EmpDG),利用情绪与认知因素增强对话上下文的编码,使得生成结果更具情感色彩。Song等^[9]提出特定情感对话生成模型(EmoDS),进一步结合情感识别与生成,使编码器能够同时建模语义与情感双重特征。Transformer架构^[10]的提出极大地推动了对话生成的发展。Transformer通过多头自注意力机制实现全局依赖建模,解决了RNN的长距离依赖难题,并具备良好的并行训练性能,因此成为主流编码器框架。DialogPT^[11]、BART^[12]和ST5^[13]等基于大规模语料训练的Transformer编码器,能够学习到丰富的语义与语境表示,在开放域对话任务中表现突出。这类模型的优势在于通用性与可迁移性,但在低资源语言场景下仍面临词汇稀疏与知识覆盖不足的问题。

1.2 构词形态编码方法

形态结构现已在语言模型中使用。Qiu等^[14]为了解决英文中上下文信息不足而导致的稀有词和未知词难以生成高质量的词表示这一问题,利用增强word2vec对整个单词及对应的形态进行嵌入。Cotterell等^[15]针对形态丰富的德语中传统词嵌入模型无法有效捕捉词形相似性的问题,提出了形态学词嵌入模型(Morph-LBL),通过引入形态学标注数据、多任务学习和手工特征等方法,有效解决了词形相似性问题。Zheng等^[16]针对预训练语言模型(PLM)在处理中文自然语言时忽略汉字形态特征的问题,使用中文字符形态学知识增强预训练语言模型(SGBERT),通过对比学习和字形感知窗口的方式提取汉字形态中的语义知识,并将其与PLM表示融合,从而增强PLM对自然语言的理解能力。文献[17-19]表明,将形态学知识纳入标记化可以对模型性能产生积极影响,特别是在自然语言推理任务中。

1.3 跨语言情感对话生成方法

近年来,大型预训练语言模型的跨语言情感分析与对话生成研究发展迅速,通过对多语言预训练模型进行微调、参数调优等,大型预训练语言模型在跨语言情感对话生成效果上取得了显著提升。Kadiyala^[20]提出了一种基于大语言模型和集成方法的跨语言情感检测系统,通过结合多种大语言模型,使用微调和零样本配置,有效提高了对话中情感检测的准确性。Cheng等^[21]提出代理工作流程(AWF)方法,利用多个大语言模型代理之间的协作来提高情感检测性能,包括多迭代代理工作流程(MIAWF)和多二元分类器代理工作流程(MB-CAWF),将不同模型和AWF方法的预测结果进行整合,通过集成学习来提高模型整体的情感分析性能。Thakkar等^[22]通过将现有的单模态情感数据集转换为多模态格式,并使用包含文本和图像嵌入特征的多模态模型进行训练,实现了对多语言内容的情感分析。

2 问题定义

情感对话生成旨在根据用户输入语句生成具有情感色彩的回复,对于输入对话句子序列 $X = \{x_0, x_1, \dots, x_i, x_n\}$,其中 x_i 表示句子 X 中的第 i 个词,模型需要在情感目标 e 下生成输出回复序列 $Y =$

$\{y_0, y_1, \dots, y_j, y_m\}$, 使得 Y 在语义内容上与 X 保持一致, 在情感表达上符合情感目标 e 。然而, 蒙古语中语义内容和情感词大多蕴含在复杂的 $\langle \text{Root}(\text{词根}) + \text{Suffix}(\text{后缀}) \rangle$ 构词结构中, 一个基本的词根 r_i 通过添加一系列的后缀 $\{s_i^{(1)}, s_i^{(2)}, \dots, s_i^{(n)}\}$, 可生成具有不同语气、强度或倾向的派生词, 而现有情感对话生成模型大多基于字或词的平面嵌入方式, 形式上可表示为

$$H = \text{Encoder}(E_{\text{char/word}}(X)), Y = \text{Decoder}(H, e),$$

其中, $E_{\text{char/word}}$ 为字符级或词级嵌入, H 为上下文表示。对于蒙古语输入中的派生词 W_i , $E_{\text{char/word}}$ 无法捕捉其构词结构, 导致回复 Y 在语义上与 X 不相关, 在情感表达上不符合情感目标 e 。

3 方法

3.1 本文方法的创新与区别

本文所提方法与已有形态增强 Transformer 的主要区别在于: ①引入层级后缀位置编码, 显式建模后缀语义顺序; ②设计形态注意力机制, 动态学习词根与各级后缀的情感贡献, 而非简单的拼接或加权; ③提出概率引导融合模块, 联合语义概率与形态概率确定模态权重, 保证了在情感对话生成任务中的可解释性与鲁棒性。这些模块共同构成了一个面向黏着语的泛化形态结构建模框架。为解决蒙古语情感对话生成模型中缺乏词根与后缀结构建模而导致生成回复与用户输入内容不匹配的问题, 在标准 Transformer 架构上构建了基于蒙古语构词形态结构的情感对话生成模型, 如图 2 所示。模型以蒙古语对话文本为输入, 经形态编码器提取词形态表示, 经 Transformer 编码器获得语义表示, 二者再通过概率引导融合模块动态结合, 最终由 Transformer 解码器生成与用户语义一致且情感匹配的回复。

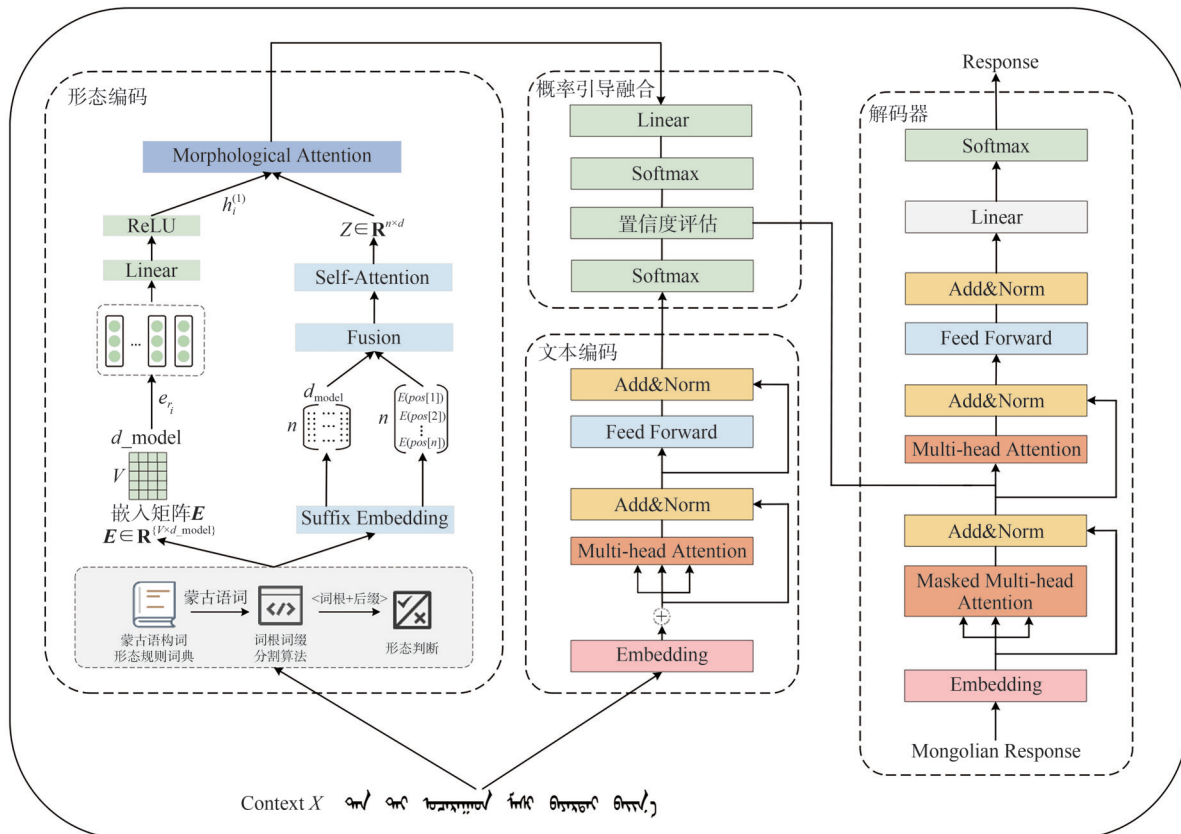


图2 融入构词形态编码单元的蒙古语情感对话生成模型架构

Fig. 2 Architecture of the Mongolian emotional dialogue generation model with morphological encoding unit

3.2 蒙古语构词形态切分

为了实现对蒙古语黏着结构的有效建模,构建了一套基于规则的蒙古语构词形态切分系统。首先,依据蒙古语构词规则与构形规则,建立包含词根词典、构词词缀词典、构形词缀词典和合法派生组合规则库在内的蒙古语构词形态规则资源。然后,对输入的蒙古语句子进行逐词处理,对于每个词,首先匹配候选词根,并按规则库检索所有可接续的后缀序列,利用自顶向下的匹配算法逐级识别一级后缀、二级后缀等,构建候选切分路径,并根据最优匹配原则确定最终的词根与后缀切分结果。最终,系统对得到的词根-后缀进行形态判断,判定是否引入语法意义或语义意义变化,为后续形态编码器中的词根嵌入、后缀层级表示和形态注意力机制提供基础输入。

3.3 形态编码单元

形态编码单元结构如图2所示,输入一条蒙古语对话文本序列 $X = \{x_0, x_1, \dots, x_i, x_n\}$, 其中 x_i 表示句子 X 中的第 i 个词,单词 x_i 经过形态学工具进行词形切分,得到词形结构。切分公式表示为

$$x_i = \{r_i, s_i^{(1)}, s_i^{(2)}, \dots, s_i^{(j)}, \dots, s_i^{(n)}\},$$

其中, r_i 为词根, $s_i^{(j)}$ 为第 j 个后缀。形态编码器针对每个词的〈词根+后缀〉结构进行层次化建模,主要包括3个关键模块:词根嵌入层、后缀位置编码层和形态注意力机制。

3.3.1 词根嵌入层

在蒙古语中,词根是构词的核心语义单位,对其进行高质量的表示是后续形态建模的基础。设计一个词根嵌入矩阵 $E_{\text{root}} = \mathbf{R}^{V_r \times d}$, 其中 V_r 表示词根词表的大小, d 为嵌入维度。每个输入词经过词形切分后提取出词根 r_i , 查表获得其嵌入向量 e_{r_i} 。为了增强嵌入表示的表达能力,将词根嵌入向量输入前馈网络,表示为

$$h_i^{(1)} = \text{LayerNorm}(e_{r_i} + \text{FFN}(e_{r_i})),$$

其中,FFN包括两个线性变换和一个ReLU激活,LayerNorm用于保持模型稳定训练。

最终将词根嵌入表示 $h_i^{(1)}$ 输入形态注意力模块,用于联合建模词根与后缀的层次结构和语义交互关系。

3.3.2 后缀位置编码层

为了有效建模蒙古语中后缀的层级关系与顺序,设计了具有语义分层意识的后缀位置编码层。不同于传统Transformer模型中仅对token的位置进行线性编码,我们在形态编码器中引入了后缀的语义层级结构,并通过位置编码函数赋予不同类型后缀以区分位置表示,进一步提升模型对构词结构中层次信息的建模能力。

单词 x_i 经过形态切分后得到后缀序列 $\{s_i^{(1)}, s_i^{(2)}, \dots, s_i^{(j)}, \dots, s_i^{(n)}\}$, 其中 n 表示该单词的后缀数量。为区分“词根”“一级后缀”“二级后缀”等不同层级,为每个后缀 $s_i^{(j)}$ 分配一个层级位置标签 l_{ij} 且 $l_{ij} \in \{0, 1, 2, \dots\}$, 结合其在词中出现的顺序 $\text{pos}(i)$ 定义后缀 $s_i^{(j)}$ 的最终位置编码向量为

$$\text{PE}(s_i) = E_{l_{ij}} + \text{PE}_{\sin/\cos}(\text{pos}(i)),$$

其中, $E_{l_{ij}} \in \mathbf{R}^d$ 是与后缀层级 l_{ij} 对应的层级嵌入向量,用于区分不同构词层级; $\text{PE}_{\sin/\cos}(\text{pos}(i))$ 是Transformer中正弦、余弦位置编码函数,用于保留词中顺序信息,定义为

$$\begin{aligned} \text{PE}_{\sin/\cos}(\text{pos}(i), 2k) &= \sin\left(\frac{\text{pos}(i)}{10\,000^{2k/d}}\right), \\ \text{PE}_{\sin/\cos}(\text{pos}(i), 2k+1) &= \cos\left(\frac{\text{pos}(i)}{10\,000^{2k/d}}\right). \end{aligned}$$

最终,后缀 $s_i^{(j)}$ 的位置编码通过层级嵌入与顺序编码的加法计算得到,其意义在于同时表示出后缀的层级类型和相对顺序:

$$\text{HPE}(s_i^j) = E_{l_{ij}} + \text{PE}_{\sin/\cos}(\text{pos}(i)).$$

3.3.3 形态注意力机制

形态注意力机制可以动态建模后缀对词根的情感影响,学习哪些后缀具有情感倾向,其计算公式为

$$a_j = \text{Softmax}(w^T \tanh(W_1 h_i^{(1)} + W_2 h_i^{(j)})),$$

$$h_i^{\text{morph}} = \sum_{j=1}^n a_j h_i^{(j)},$$

其中, $h_i^{(j)}$ 为第 j 个后缀的编码表示, a_j 为注意力权重, h_i^{morph} 为词的最终形态表示。最终形态编码器输出为

$$H^{\text{morph}} = [h_1^{\text{morph}}, h_2^{\text{morph}}, \dots, h_L^{\text{morph}}]。$$

3.4 概率引导融合

在文本向量与形态向量两种特征融合过程中,需要对贡献较大的模态特征赋予更大的权重。与门控机制依赖于某个特定向量的信息来决定其他向量的方法不同,本文设计一种概率引导融合模型,通过文本编码器和形态编码器预测出的语义概率与形态概率来动态地为文本向量和形态向量分配权重。

为了估计每个模态在当前 token 上的可靠性,分别在两种特征上引入轻量级分类器:

$$p_i^{(t)} = \text{Softmax}(W_t h_i^{(t)} + b_t),$$

$$p_i^{(m)} = \text{Softmax}(W_m h_i^{(m)} + b_m),$$

其中, $p_i^{(t)}$ 与 $p_i^{(m)}$ 为 token i 的文本模态与形态模态的预测概率分布。然后对模态置信度进行归一化,得到融合权重,过程如下:

$$\alpha_i = \frac{\exp(f(p_i^{(t)}))}{\exp(f(p_i^{(t)})) + \exp(f(p_i^{(m)}))}, \beta_i = 1 - \alpha_i,$$

其中, $f(\cdot)$ 表示从概率分布中提取与情感相关的置信度。最终每个 token 的融合表示为

$$h_i^{(f)} = \alpha_i h_i^{(t)} + \beta_i h_i^{(m)},$$

融合后的表示为 $H^{(f)} = \{h_1^{(f)}, \dots, h_n^{(f)}\}$, 将其送入 Transformer 解码器生成情感对话回复。

4 实验

4.1 数据集

鉴于目前缺乏公开可用的蒙古语情感对话数据集,本文构建了一个蒙古语情感对话数据集。该数据集主要基于中文情感对话语料库和蒙古语文本语料库,通过先进的机器翻译模型将中文情感对话语料翻译为蒙古语版本,然后由蒙古语母语人员对翻译文本的准确性和自然度进行审核。在审核完成后,进行数据预处理,包括文本标准化和去重等步骤。最终,构建了一个包含 10 万个对话对的蒙古语情感对话数据集,其中正向、中性、负向情感样本的占比分别为 33%、42%、25%。为保证模型评估的稳健性与结果的可靠性,将数据集按 8:1:1 划分为训练集、验证集与测试集。

4.2 实验参数设置

实验在 NVIDIA GeForce RTX 3060 上完成,深度学习框架采用 PyTorch 1.12.1。对于 Transformer 模型,词向量维度和隐藏层维度均设为 256,编码器和解码器均包含 6 层,每层使用 8 个注意力头,前馈网络维度为 1024。训练过程中采用 Adam 优化器对模型参数进行更新,其中一阶矩估计系数设置为 0.9,二阶矩估计系数设置为 0.98,数值稳定项设置为 1×10^{-9} ,学习率采用 Transformer 模型中的 warm-up 调度策略,将预热步数设置为 4000,以提高模型训练过程的稳定性。batch size 设为 64,最大序列长度为 128。为了避免过拟合,dropout 设置为 0.1, label smoothing 设置为 0.1,并在训练过程中将梯度裁剪设为 1.0。在所提出的形态编码器部分,词表基于字符级构建,共 268 个蒙古文字符,词根和后缀的编码维度均设为 128,形态注意力机制采用 4 个注意力头,最后通过概率融合方式将形态向量与 Transformer 的语义向量结合。

4.3 对比模型

为了验证所提方法的有效性,设置4类对比模型,分别为①基础对话生成模型(Basic):基础Seq2Seq模型(S2S),作为生成式对话系统的基本结构;S2S+Att模型,在S2S基础上加入Attention机制以增强对上下文的建模能力;当前主流Transformer对话生成模型,编码-解码完全基于多头注意力机制,无需RNN,适合长文本和对话场景。②结构增强类模型(Structural):经典的层次递归编码器-解码器模型(HRED),设计了词级别和句级别两种层次结构辅助建模对话;Meta AI开源的BlenderBot^[23],它是迄今为止较大的开放域对话模型之一,通过引入长期记忆和检索组件,使得在对话中生成更丰富和更一致的回复;SPHRED使用两个独立的HRED对用户输入与机器回复分开编码,然后使用变分自编码器融合情感和语义,为后续多样化和上下文一致的对话生成提供支持。③预训练模型(Pre-train):使用HuggingFace开源的mBART-50多语言预训练生成模型对蒙古语进行微调,该模型具备跨语言迁移能力,但未显式建模蒙古语的形态结构;DialoGPT是基于GPT-2架构的对话生成模型,从大规模对话数据中自监督训练而成;ST5是一个统一的预训练语言模型框架,将所有NLP任务都转化为“文本到文本”的形式,通过输入特定的prompt生成对应的回复,从而在对话生成中展现出强大的迁移和生成能力。④情感对话生成模型(Sentiment):EmpDG利用粗粒度对话级别和细粒度令牌级别的情感,后者有助于更好地捕捉用户情感的细微差别。还通过引入一个交互式对抗性学习框架,利用用户反馈来识别生成的回复是否会在对话中激起情感感知;ECM通过引入情感向量来控制对话情感;EmoDS可以为帖子生成具有连贯结构的、有意义的响应,并在统一的框架内明确或隐含地表达所需的情感。

4.4 评价指标

使用BLEU^[24]、Distinct-1^[25]、PPL(Perplexity)^[26]、Rouge^[27]作为对话生成任务的评价指标,用于衡量生成响应与目标响应之间的相似性。BLEU- N 表示目标响应和生成响应之间的 n -gram重叠数,较高的BLEU- N 分数表示更多的重叠和更高的相似性。Distinct-1用来衡量模型生成文本的多样性,表示生成结果中不重复单词占有所有单词的比例,其值越大,说明模型生成的回复越丰富,重复度越低。PPL是语言模型的困惑度,表示模型对数据的预测能力,其值越低越好。Rouge用于衡量生成文本与参考文本之间的重合度,包括词级匹配和短语级匹配等,其值越高越好。为评价模型在情感控制方面的有效性,使用预训练的蒙古语情感分类器对生成回复的情感类别进行判别,并与输入的目标情感标签进行比对,统计正向和负向情感的分类准确率。

4.5 基线对比实验

表1展示了所提模型与多种基线模型在蒙古语情感对话生成任务上的综合评价结果,包括BLEU、Rouge、Distinct-1、PPL以及正向和负向情感准确率。从指标整体来看,所提模型在生成文本的BLEU和情感准确率上表现优异,远超传统的基础序列到序列模型,展现出更好的文本生成质量和多样性。此外,PPL值为46.35,略高于部分预训练模型,但在情感准确率方面明显领先,正向情感准确率达到93.2%,负向情感准确率达到90.4%,均显著优于其他模型。在结构化模型和预训练模型中,预训练模型DialoGPT和ST5表现较好。针对情感对话生成模型,尽管ECM、EmpDG和EmoDS在BLEU和Rouge等生成质量指标上略低于所提模型,但它们在情感分类准确率上展现出稳定且相对平衡的表现,三者的正向情感准确率分别达到88.3%、89.1%和89.6%,负向情感准确率分别为84.7%、85.2%和85.9%,表明这些模型能够在一定程度上理解输入的情感倾向并生成符合预期情感的回复。

4.6 消融实验

表2消融实验进一步验证了形态编码模块和融合机制的作用。表2结果显示,去除形态编码(w/o Morph)后,模型BLEU和情感准确率均有较大幅度下降,表明形态信息对提升模型泛化和情感理解能力至关重要;去除融合模块(w/o Fusion)后,模型虽然在Rouge和PPL上表现尚可,但情感准确率有所下降,表明融合机制在模型综合性能中具有重要贡献。

表 1 蒙古语构词形态编码方法对比实验结果

Table 1 Comparison experimental results of Mongolian word formation morphological coding methods							
类别	模型	BLEU	Rouge	Distinct-1	PPL	正向情感准确率/%	负向情感准确率/%
Basic	Transformer	0.3495	16.21	0.57	30.75	87.5	83.7
	S2S	0.2450	14.01	0.67	56.89	85.9	82.4
	S2S+Attention	0.2870	14.92	0.70	50.12	—	—
Structural	HRED	0.3120	15.30	0.68	48.00	—	—
	SPHRED	0.3200	16.00	0.63	44.20	—	—
	BlenderBot	0.4110	21.54	0.69	30.09	—	—
Pre-train	DialoGPT	0.4425	22.10	0.75	28.24	—	—
	mBART-50	0.3958	20.24	0.70	32.18	—	—
	ST5	0.4350	21.90	0.72	27.34	—	—
Sentiment	ECM	0.4500	22.00	0.71	26.50	88.3	84.7
	EmpDG	0.4620	23.15	0.71	25.10	89.1	85.2
	EmoDS	0.4716	19.46	0.65	24.17	89.6	85.9
	本文模型	0.4873	22.30	0.72	46.35	93.2	90.4

注:黑体表示该模型在该指标上的最优结果。

表 2 消融实验结果

Table 2 Ablation experimental results						
模型	BLEU	Rouge	Distinct-1	PPL	正向情感准确率/%	负向情感准确率/%
本文模型	0.4873	22.30	0.72	46.35	93.2	90.4
去除形态编码	0.3754	17.91	0.68	50.50	89.8	86.1
去除融合模块	0.4180	21.20	0.68	48.49	90.5	87.3

注:黑体表示该模型在该指标上的最优结果。

4.7 形态注意力权重可视化

为深入理解模型如何利用构词信息进行情感建模,图3展示了形态注意力模块中不同派生后缀的权重分布可视化结果。其中,对于由词根“*ᠪᠠᠭ*”(典礼,喜事)派生出的单词“*ᠪᠠᠭᠠᠨᠠᠭ*”(高兴的动词形式)和“*ᠪᠠᠭᠠᠨᠠᠭᠤ*”(高兴的名词形式),模型给予了较高的注意力权重,分别为0.63和0.59,而对于动词化后缀和名词化后缀则分配较低的注意力权重,分别为0.37和0.41,表明模型倾向于捕捉那些对情感强度和句法角色影响较大的词根或后缀。这一结果说明形态注意力机制确实有效地捕捉了词根或后缀在情感表达中的差异性作用。

4.8 语义与情感一致性

为了进一步验证模型在生成回复时是否能够保持语义和情感表达的一致性,设计了语义与情感一致性实验。首先,利用语义相似度算法对生成回复与参考回复之间的语义一致性进行评分。同时,基于预训练情感分类器对生成回复和参考回复进行情感极性评估,通过对比两者标签获得情感一致性得分。在不同模型下分别计算语义一致性与情感一致性的均值,并绘制语义情感一致性散点图,以展示两者的动态关系(图4)。虽然ECM、EmoDS和EmpDG模型可以保持较高的情感一致性,但它们在语义一致性上的表现普遍较差,往往生成语境相关性不足的回复。Transformer和Seq2Seq模型的语义一致性优于情感一致性,容易出现语义连贯但缺乏情感色彩的情况。本文模型在两个维

- International Conference on Neural Information Processing Systems. Montreal, Canada: MIT Press, 2014: 3104-3112.
- [6] 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. Phoenix, Arizona USA: AAAI, 2016: 3776-3783.
- [7] Shen X Y, Su H, Li Y R, et al. A conditional variational framework for dialog generation[C]//Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics. Vancouver, Canada: Association for Computational Linguistics, 2017: 504-509.
- [8] Li Q T, Chen H S, Ren Z C, et al. EmpDG: Multi-resolution interactive empathetic dialogue generation[C]//Proceedings of the 28th International Conference on Computational Linguistics. Barcelona, Spain: International Committee on Computational Linguistics, 2020: 4454-4466.
- [9] Song Z Q, Zheng X Q, Liu L, et al. Generating responses with a specific emotion in dialog[C]//Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence, Italy: Association for Computational Linguistics, 2019: 3685-3695.
- [10] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[C]//31st Conference on Neural Information Processing Systems. Long Beach, CA, USA: NIPS. 2017: 1-11.
- [11] Zhang Y Z, Sun S Q, Galley M, et al. DialoGPT: Large-scale generative pre-training for conversational response generation[C/OL]//Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations. 2020: 270-278. <https://aclanthology.org/2020.acl-demos.30/>.
- [12] Lewos M, Liu Y H, Goyal N, et al. BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension[C/OL]//Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. 2020: 7871-7880. <https://aclanthology.org/2020.acl-main.703/>.
- [13] Ni J M, Abrego G H, Constant N, et al. Sentence-T5(ST5): Scalable sentence encoders from pre-trained text-to-text models[C]//Findings of the Association for Computational Linguistics: ACL 2022. Dublin, Ireland: Association for Computational Linguistics, 2022: 1864-1874.
- [14] Qiu S Y, Cui Q, Bian J, et al. Co-learning of word representations and morpheme representations[C]//Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers. Dublin, Ireland: COLING, 2014: 141-150.
- [15] Cotterell R, Schutze H. Morphological word-embeddings[C]//Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Denver, Colorado: Association for Computational Linguistics, 2015: 1287-1292.
- [16] Zheng Z Z, Wu X M, Liu X Z. Enhancing pre-trained language models with Chinese character morphological knowledge[J]. Information Processing & Management, 2025, 62(1): 103945.
- [17] Tan S, Joty S, Varshney L R, et al. Mind your inflections! Improving NLP for non-standard Englishes with base-inflection encoding[C/OL]//Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing. 2020: 5647-5663. <https://aclanthology.org/2020.emnlp-main.455/>.
- [18] Hou J, Katinskaia A, Vu A D, et al. Effects of sub-word segmentation on performance of transformer language models[C]//Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing. Singapore: Association for Computational Linguistics, 2023: 7413-7425.
- [19] Jabbar H. MorphPiece: Moving away from statistical language representation[EB/OL]. (2023-07-14) [2025-09-25]. <https://doi.org/10.48550/arXiv.2307.07262>.
- [20] Kadiyala R M R. Cross-lingual emotion detection through large language models[C]//Proceedings of the 14th Workshop on Computational Approaches to Subjectivity. Bangkok, Thailand: Association for Computational Linguistics, 2024: 464-469.
- [21] Cheng L, Shao Q H, Zhao C, et al. TEII: Think, explain, interact and iterate with large language models to solve cross-lingual emotion detection[C]//Proceedings of the 14th Workshop on Computational Approaches to Subjectivity, Sentiment, & Social Media Analysis. Bangkok, Thailand: Association for Computational Linguistics, 2024: 495-504.
- [22] Thakkar G, Hakimov S, Tadic M. M2SA: Multimodal and multilingual model for sentiment analysis of tweets[C]//Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and

- Evaluation. Torino, Italia: ELRA and ICCL, 2024: 10833-10845.
- [23] Shuster K, Xu J, Komeili M, et al. BlenderBot 3: A deployed conversational agent that continually learns to responsibly engage[EB/OL]. (2022-08-04)[2025-09-25]. <https://arxiv.org/abs/2208.03188>.
- [24] Papineni K, Roukos S, Ward T, et al. BLEU: A method for automatic evaluation of machine translation[C]//Proceedings of the 40th Annual Meeting on Association for Computational Linguistics. Philadelphia, Pennsylvania: Association for Computational Linguistics, 2002: 311-318.
- [25] Li J W, Galley M, Brockett C, et al. A diversity-promoting objective function for neural conversation models[C]//Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. San Diego, California: Association for Computational Linguistics, 2016: 110-119.
- [26] Vinyals O, Fortunato M, Jaitly N. Pointer networks[C]//Proceedings of the 28th International Conference on Neural Information Processing Systems. Montreal, Canada: MIT Press, 2015: 2692-2700.
- [27] Lin C Y. ROUGE: A package for automatic evaluation of summaries[C]//Text Summarization Branches Out. Barcelona, Spain: Association for Computational Linguistics, 2004: 74-81.

(责任编辑 黄宝琦)

Morphological Encoding Unit for Mongolian Emotional Dialogue Generation

WANG Kuibo¹, MA Zhiqiang^{1,2}, WANG Hongbin¹, WANG Zizhou¹,
SONG Jingwei¹, YIN Hui¹

(1. College of Intelligent Science and Technology, Inner Mongolia University of
Technology, Hohhot 010080, China;

2. Inner Mongolia Autonomous Region North Border Cybersecurity Key Laboratory,
Hohhot 010080, China)

Abstract: Mongolian, an agglutinative language, allows a single root to derive numerous words with varying nuances and emotional intensities. However, existing models primarily rely on word- and character-level representations, lacking explicit modeling of the root-suffix structure inherent in Mongolian. The deficiency often leads to generated responses that are monotonous and contextually inappropriate. To address this issue, a morphological encoding method for Mongolian emotional dialogue generation is proposed. Built upon the Transformer architecture, a morphological encoding unit is introduced, which comprises a root embedding layer, a hierarchical positional encoding for suffixes, and a morphological attention mechanism to explicitly model the root-suffix relationship. Furthermore, a probabilistic fusion mechanism is employed to combine morphological vectors with semantic vectors, enhancing the model's comprehension of derived words and emotional expressions. Experiments conducted on a self-constructed Mongolian emotional dialogue dataset show that the method outperforms various baseline models across multiple metrics, including BLEU, Rouge, Distinct-1, and emotion classification accuracy. Notably, it achieves an accuracy of 93.2% for positive emotions and 90.4% for negative emotions. Ablation studies and visualization analysis of the morphological attention mechanism further validate the effectiveness of the incorporated morphological information and the fusion mechanism. The research provides a novel modeling approach for emotional dialogue generation in low-resource languages.

Key words: Mongolian; emotional dialogue generation; morphological encoding; Transformer