

多层面语义结构增强的对话情感诱因片段抽取

秦鸣飞, 付国宏

引用本文

秦鸣飞, 付国宏. 多层面语义结构增强的对话情感诱因片段抽取[J]. 计算机科学, 2023, 50(12): 236-245.

QIN Mingfei, FU Guohong. Multi-level Semantic Structure Enhanced Emotional Cause Span Extraction in Conversations [J]. Computer Science, 2023, 50(12): 236-245.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[SemFA: 基于语义特征与关联注意力的大规模多标签文本分类模型](#)

SemFA: Extreme Multi-label Text Classification Model Based on Semantic Features and Association Attention

计算机科学, 2023, 50(12): 270-278. <https://doi.org/10.11896/jsjcx.230300239>

[基于可信细粒度对齐的多模态方面级情感分析](#)

Aspect-based Multimodal Sentiment Analysis Based on Trusted Fine-grained Alignment

计算机科学, 2023, 50(12): 246-254. <https://doi.org/10.11896/jsjcx.221100038>

[基于CodeBERT的设计模式语言模型](#)

CodeBERT-based Language Model for Design Patterns

计算机科学, 2023, 50(12): 75-81. <https://doi.org/10.11896/jsjcx.230100115>

[基于时空注意力机制的多元时间序列异常检测](#)

Spatial-Temporal Attention Mechanism Based Anomaly Detection for Multivariate Times Series

计算机科学, 2023, 50(11A): 230300022-8. <https://doi.org/10.11896/jsjcx.230300022>

[融合无监督SimCSE的短文本聚类研究](#)

Study on Short Text Clustering with Unsupervised SimCSE

计算机科学, 2023, 50(11): 71-76. <https://doi.org/10.11896/jsjcx.220900214>

多层面语义结构增强的对话情感诱因片段抽取

秦鸣飞¹ 付国宏^{1,2}

1 苏州大学计算机科学与技术学院 江苏 苏州 215006

2 苏州大学人工智能研究院 江苏 苏州 215006

(mingfei_qin@outlook.com)

摘要 对话情感诱因片段抽取旨在从对话历史中抽取出诱发目标情感表达的原因片段,在情感对话系统中起到枢纽的作用。然而,已有方法抽取出的诱因片段仍存在话轮定位错误、边界识别错误等亟待解决的问题。为此,提出了一种多层面语义结构增强的对话情感诱因片段抽取方法。该方法基于篇章层面的指代结构,增强对诱因片段所处话轮的定位;基于句子层面的句法结构,增强对诱因片段边界的识别。首先,依据预处理后的语义结构及对话内容特征表示,使用图注意力网络分别在词符级别与话轮级别构图、建模对话,并通过双仿射机制促进两种级别构图的交互与融合,从而获得结构增强的语义综合表示;然后,使用线性层抽取诱因片段。在两个公开数据集上进行实验,结果表明,与基准模型相比,该模型的 $F1$ 值和 EM_{pos} 值最高分别提升了 2.42% 和 2.26%;同时,在 $F1_{pos}$ 和 EM_{pos} 指标上的性能均优于其他基线模型,且该模型也能有效兼容话轮级别的对话情感诱因蕴含。

关键词: 自然语言处理; 结构增强; 对话情感; 诱因片段抽取; 图注意力网络

中图法分类号 TP391

Multi-level Semantic Structure Enhanced Emotional Cause Span Extraction in Conversations

QIN Mingfei¹ and FU Guohong^{1,2}

1 School of Computer Science and Technology, Soochow University, Suzhou, Jiangsu 215006, China

2 Institute of Artificial Intelligence, Soochow University, Suzhou, Jiangsu 215006, China

Abstract Emotional cause span extraction in conversations aims to extract causal spans that induce target emotion expression from conversational history, which plays a pivotal role in emotional conversation systems. However, causal spans extracted by existing methods still have problems to be solved urgently, such as utterance position errors and boundary recognition errors. To this end, this paper proposes a multi-level semantic structure enhanced emotional cause span extraction method in conversations. The discourse-level coreferential structure is used to enhance the positioning of utterances where causal spans are located. The sentence-level syntactic structure is used to enhance the recognition of causal span boundaries. Firstly, according to preprocessed semantic structures and conversational content feature representations, the graph attention network is utilized to construct comprehensive graphs and model conversations at token level and utterance level, respectively. Meanwhile, the biaffine mechanism is utilized to promote interactions and integrations between two-level graphs, and structure-enhanced semantic comprehensive representations are obtained. Then, the linear layer is applied to extract causal spans. Experimental results on the two public datasets show that compared with the benchmark model, the $F1$ value and EM_{pos} value are improved by 2.42% and 2.26%, respectively. The proposed model also outperforms other baseline models in both $F1_{pos}$ and EM_{pos} metrics, and can also be effectively compatible with utterance-level emotion cause entailment.

Keywords Natural language processing, Structure enhancement, Conversational emotion, Causal span extraction, Graph attention networks

1 引言

对话情感诱因片段抽取 (Emotional Cause Span Extraction in Conversations, ECSEC)^[1] 是情感对话语义理解任务中

的重要分支,在情感对话系统中起到承上启下的枢纽作用,其承接上游任务识别出的情感^[2],指导下游任务生成更有针对性的情感回复^[3],在电商客服、人机交互等领域拥有广阔的应用前景,具备较高的学术研究价值与实际应用价值。

到稿日期:2022-11-23 返修日期:2023-02-19

基金项目:国家自然科学基金(62076173)

This work was supported by the National Natural Science Foundation of China(62076173).

通信作者:付国宏(ghfu@suda.edu.cn)

以图1为例,Poria等^[1]将诱因片段(Causal Span)定义为历史话轮(Utterance)中可精确解释目标情感表达的最大子串(Maximal Sub-string),如话轮1中的“your speech was excellent”,话轮3中的“Sure it was”。然而,已有方法抽取出的诱因片段仍存在两类亟待解决的问题:1)诱因片段所处话轮定位错误,如将诱因片段定位在话轮2中;2)诱因片段边界识别错误,如将话轮3中的诱因片段抽取为“it was”。

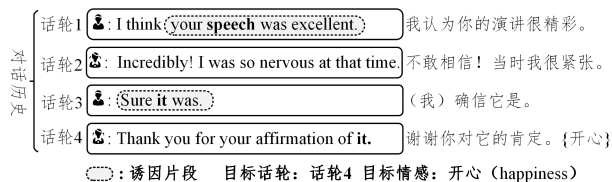


图1 ECSEC 样例

Fig. 1 Sample of ECSEC

图1中,话轮3与话轮4中的指称“it”均指代话轮1中的实体“speech”,而样例的诱因片段正是处在话轮1与话轮3中。进一步分析两个公开的ECSEC数据集^[1]发现:存在指代消解现象的对话平均比例达72.75%,而诱因话轮及目标话轮中包含实体或指称的平均比例达64.16%,这说明指代消解的过程与情感回溯至诱因的过程可能存在较大比例的交集。与此同时,Liu等^[4]证明了指代结构有助于总结对话的内容,Han等^[5]证明了指代结构有助于跟踪对话的状态。因此得出如下推断结果:在ECSEC任务中,指代结构蕴含着对话内容交互及演进的线索,可能有助于解决诱因片段所处话轮定位错误的问题。

然而,即使诱因话轮定位正确,抽取出的诱因片段仍可能存在模棱两可的边界错误。相比新闻、博客等表述严格的正式文本,对话的语言表达更加口语化,如话轮3中省略了“I am”,这给模型精确识别诱因片段边界带来了干扰。而已有方法大多依赖通用的预训练模型编码语义内容,尽管有效,但没有明确考虑句子层面的依存句法关系,因此句子中词符(token)之间的交互不充分^[6],这可能会降低模型对诱因片段边界的精准感知,而Sadvilkar等^[7]也证明了基于依存句法关系可增强对文本边界的消歧。因此得出如下推断结果:在ECSEC中,句法结构可能有助于解决诱因片段边界识别错误的问题。

以管道(pipeline)式的方式能够更清楚地解析并归类诱因片段抽取过程中存在的核心问题,但管道式的模型存在错误传播^[2]。因此,本文提出了一种端到端(end-to-end)的ECSEC模型,该模型依据上述两种语义结构及预处理后的对话内容特征表示,使用图注意力网络分别在词符级别与话轮级别综合构图、建模对话,并通过双仿射机制促进两种级别构图的交互与融合。然后,基于语义综合表示,使用线性层抽取诱因片段。

本文的贡献点主要包括:

1)为解决ECSEC中的两类问题,提出了一种多层次语义结构增强的方法,验证了其有效性,并明晰了结构与问题的具体相关性。

2)设计了一种多层次语义结构增强的ECSEC模型。该

模型在4项评测指标上均优于基准模型的性能。同时,在正样本的评测指标上均优于其他基线模型的性能,且本文模型也能有效兼容话轮级别的对话情感诱因蕴含任务。

2 相关工作

2.1 情感诱因抽取方法的发展

Lee等^[8]首次定义了情感诱因分析任务,构建了诱因分析语料库,并在后续工作中基于规则检索诱因^[9],而Ghazi等^[10]则基于条件随机场抽取诱因片段。随着深度学习的发展以及情感诱因抽取数据集的不断构建,情感诱因的抽取方法也同步演变出如下5个角度:1)从内容本身交互的角度,基于注意力机制^[11]增强情感与上下文内容的交互^[12];2)从知识增强的角度,引入外部知识^[13]以增强对隐含信息的推理能力,但这比较依赖知识的质量;3)从结构增强的角度,Hu等^[14]基于句法结构使用图卷积网络建模子句(clause)内部及其之间的交互,但依据句子层面的结构建模篇章层面的交互存在较大的局限性;4)从数据分布特点的角度,基于诱因位置的分布特点设计位置敏感的模型^[15],而诱因外在形式上的分布特点一定程度上是多层次语义结构内在交互机制的体现^[16];5)从联合增强的角度^[13,17-18],考虑到情感与诱因的相互影响,Xia等^[17]进一步定义了子句级别的情感-诱因对抽取任务,使用多任务学习的方式同时识别情感与诱因所处的子句并为之配对。

Poria等^[1]首次定义了对话情感诱因识别任务RECCON (Recognizing Emotion Cause in Conversations),该任务共分为两项具有递进关系的子任务:话轮级别情感诱因蕴含以及更细粒度的片段级别情感诱因抽取。同时,基于DD^[19] (DailyDialog)与IE^[20] (IEMOCAP)构建了RECCON-DD与RECCON-IE两个基准数据集。针对话轮级别子任务,Zhang等^[21]从数据分布特点的角度,依据对话者的交互线索增强话轮间的交互;Li等^[22]从知识增强的角度,提出了一种过滤常识的方法,用于增强对中性情感的区分;Bhat等^[23]从联合增强的角度,延续Turcan等^[13]多任务的做法,联合学习情感预测与诱因识别。在下游任务中,Kim等^[3]通过情感诱因引导模型生成情感回复,但Li等^[12]指出以片段为诱因抽取的单位具有噪声少、冗余度低等优点,因此更有利于下游任务生成更精确的情感回复。此外,由于诱因片段是诱因话轮的子集^[1],因此片段级别子任务的模型也能兼容话轮级别子任务的输出,反之则不然,但具备以上优势的片段级别子任务尚未被深入探索且仍存在两类问题亟待解决。为了弥补此研究差距,本文以片段级别的子任务为研究对象,重点探索多层次语义结构增强的对话情感诱因抽取方法。

2.2 结构增强方法的应用

结构增强方法的可行性与有效性在多个自然语言处理任务上得到了验证^[4-7,24-26],但以往方法多聚焦在单层面的结构增强。在篇章层面,针对机器阅读理解任务,Huang等^[24]验证了使用图卷积网络显式地融合指代结构的有效性;在句子层面,针对对话情感识别任务,Shou等^[25]证明了使用注意力机制融合句法结构可有效增强对情感词汇的捕捉。

然而,在ECSEC任务中,诱因话轮定位错误的问题处于

篇章层面,而诱因片段边界识别错误的问题处于句子层面。因此,以往基于单层面结构增强的方法对于解决存在于两种层面上的问题具有较大的局限性^[14]。

2.3 本文模型的创新点

相比外部知识增强、规则设计以及位置感知等方法,本文模型重点考虑的是自然语言本身天然固有的根本属性之一^[16],即多层面的语义结构。因此无需耗费额外的精力、财力来专门构建高质量的外部知识,也不依赖于特定数据集中固定的匹配规则以及特有的诱因位置分布特点,因此本文模型更通用、更易用。

相比条件随机场、联合学习等仅考虑了语义内容交互的方法,本文模型在语义内容交互的基础上,同时充分考虑了被以往工作忽略掉的多层面语义结构信息,进一步增强了话轮内部及其之间的交互,能够针对性地解决或缓解 ECSEC 任务中尚存在的两类亟待解决的问题,同时能够弥补以往仅考虑语义内容交互导致的性能差距。

值得注意的是,不同于以往单层面结构增强的方法无法同时针对性解决处于两种层面的问题^[14],本文提出了一种新颖的多层面语义结构增强的模型方法。该方法并非仅基于单层面的语义结构笼统解决处于多种层面的问题,而是针对特定层面的问题引入对应层面的语义结构,并提出了两种级别的构图与建模方式。同时,特别设计了一种不同级别构图间交互与融合的方法,使得模型既能够感知句子内部的局部语义信息,还能捕捉句子之间远距离的全局语义信息,这也为有效兼容话轮级别的子任务提供了基础。

3 多层面语义结构增强的对话情感诱因片段抽取模型

3.1 任务定义

为了能与基准^[1]形成公平且有效的对比,延续 Poria 等^[1]的任务定义及范式设定,假设对话历史 HU 由 t 个连续的话轮构成,即 $U_1, U_2, \dots, U_i, \dots, U_t$ 。 U_i 为目标话轮, E_i 为目标情感, U_i 为候选话轮 ($i \in [1, t]$), U_i 由 $L(i)$ 个词符 (token) 构成,即 $W_i^1, W_i^2, \dots, W_i^{L(i)}$, CU 为诱因话轮集合, CS 为诱因片段集合 ($CS \subseteq CU$)。则 Poria^[1] 将 ECSEC 任务定义为:给定 $HU = \{U_i\}_{i \in [1, t]}$, U_i 与 E_i , 抽取 CS , 形式化如式(1)所示:

$$CS = f_{ECSEC}(\{U_i\}_{i \in [1, t]}, U_i, E_i) \quad (1)$$

Poria 等^[1]进一步将 ECSEC 任务设定为机器阅读理解范式,由上下文 (Context, C)、问题 (Question, Q)、答案 (Answer, A) 构成。其中, $C = HU$, $A = CS$; Q 由问题模板填充 U_i, E_i, U_i 构成,包含 $L(Q)$ 个词符,问题模板详见 Poria 等^[1]的设定。

3.2 模型架构

如图 2 所示,模型输入总信息为 HU, E_i, U_i, U_i , 输出为 CS , 具体体现在诱因片段的起点 (start) 索引与终点 (end) 索引以及诱因所处话轮在对话历史序列中的位置 (position), 模型主要分为 3 部分。1) 结构与内容预处理: 输入 HU , 使用成熟的结构解析包预处理语义结构, 输出指代消解邻接矩阵与依存句法邻接矩阵; 输入 HU, E_i, U_i, U_i , 通过两种注意力机制预处理语义内容, 输出 HU 对 E_i, U_i, U_i 这 3 方面信息感知

后的对话内容特征表示。2) 结构增强的内容交互与融合: 该部分是本文模型的核心, 设计了一种新颖的基于多层面语义结构增强的模型, 该模型输入多层面的语义结构邻接矩阵与对话内容特征表示, 使用图注意力网络 (Graph Attention Networks, GAT)^[27], 分别在词符级别与话轮级别综合构图并建模对话, 并针对性设计了一种促进不同级别构图间交互与融合的方法, 本模块输出结构增强后的语义综合表示。3) 诱因片段抽取: 输入结构增强后的语义综合表示, 使用线性层抽取诱因片段。

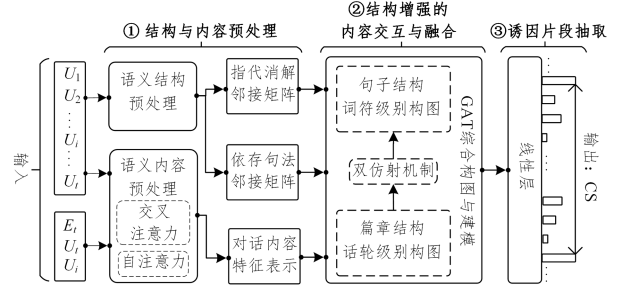


图 2 本文模型的整体架构

Fig. 2 Overall framework of the proposed model

3.3 结构与内容预处理

3.3.1 语义结构预处理

指代消解是将多个指称链接至同一对应实体的过程^[16], 形式上体现为指代簇 (Coreference Cluster, CoreC)。在图 1 的对话样例中, 指代簇为 $[\text{speech}; [\text{speech}, \text{it}, \text{it}]]$ 。如式(2)所示, 在词符级别构建 $L(i)_{\max} \times L(i)_{\max}$ 的指代消解邻接矩阵 CAM^{token} (Coreference Adjacency Matrix)。其中, $L(i)_{\max}$ 为话轮最大序列长度。

$$CAM^{\text{token}}(W_x^j, W_y^k) = \begin{cases} 1, & \text{if } ([W_x^j; W_y^k] \subseteq \text{CoreC}) \text{ or } (x=y, j=k) \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

如式(3)所示, 在话轮级别构建 $t \times t$ 的邻接矩阵 CAM^{utter} :

$$CAM^{\text{utter}}(U_m, U_n) = \begin{cases} 1, & \text{if } (W_x^j \in U_m, W_y^k \in U_n, [W_x^j; W_y^k] \subseteq \text{CoreC}) \\ & \text{or } (m=n) \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

依存句法是句法结构与词汇之间的依存关系^[16], 形式上体现为依存句法树 (Dependency Parsing Tree, DPT)。如式(4)所示, 在词符级别, 构建 $L(i)_{\max} \times L(i)_{\max}$ 的依存句法邻接矩阵 DAM^{token} (Dependency Adjacency Matrix):

$$DAM^{\text{token}}(W_x^j, W_y^k) = \begin{cases} 1, & \text{if } ((W_x^j, W_y^k) \in \text{DPT}) \text{ or } (x=y, j=k) \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

式(2)–式(4)中, W_x^j 是 HU 中第 x 个话轮中第 j 个词符, W_y^k 的含义同理, 当 $x=y$ 且 $j=k$ 时, 为词符级别自环; U_m 与 U_n 是 HU 中第 m 和第 n 个话轮, 当 $m=n$ 时, 为话轮级别自环。

至此, 通过语义结构的预处理获得了指代消解邻接矩阵与依存句法邻接矩阵, 为 3.4 节中的构图提供了语义结构基础。

3.3.2 语义内容预处理

尽管经过预训练模型编码,已经获得初步的语义内容特征表示,但是聚焦到 ECSEC 任务, E_i, U_i, U_i 这 3 方面信息感知后的对话内容特征表示才是后续结构增强的载体与根本基础^[16], 对应 HU 与 Q 的交互。注意力机制能够有效建模语义内容的交互, 已被多项工作论证^[11-12, 27], 因此, 本文模型综合使用自注意力 (SelfAttention) 与交叉注意力 (CrossAttention), 以获取多信息感知的对话内容特征表示。

首先, 通过预训练模型分别对 Q 与 HU 编码, 获得词嵌入表示: $\mathbf{E}_Q = \{e_{w_i}^Q\}_{i \in [1, L(Q)]}$, $\mathbf{E}_C = \{e_{w_j}^U\}_{j \in [1, L(U)]}$, $\mathbf{E}_Q \in \mathbb{R}^{1 \times L(Q) \times d}$, $\mathbf{E}_C \in \mathbb{R}^{l \times L(U) \times d}$, d 是词嵌入特征维度。进一步使用双向门控机制 (Bi-GRU) 分别获取 Q 与 HU 对上下文序列感知后的编码, 即 \mathbf{H}_Q 和 \mathbf{H}_C 。为了建模 Q 与 HU 各自内部的交互并考虑到计算效率, 自注意力机制使用缩放点积注意力作为评分函数。为了建模 Q 与 HU 之间的交叉交互, 且由于 $L(Q) \neq L(HU)$, 交叉注意力机制使用加性注意力作为评分函数, 则 Q 与 HU 各自内部的交互如式 (5)、式 (6) 所示:

$$\mathbf{H}_Q^{\text{self}} = \text{SelfAttention}(\mathbf{H}_Q, \mathbf{H}_Q, \mathbf{H}_Q) \quad (5)$$

$$\mathbf{H}_C^{\text{self}} = \text{SelfAttention}(\mathbf{H}_C, \mathbf{H}_C, \mathbf{H}_C) \quad (6)$$

三方面信息与 HU 之间的交叉交互主要体现在 C 对 Q 的感知, 如式 (7) 所示:

$$\mathbf{H}_C^{\text{cross}} = \text{CrossAttention}(\mathbf{H}_Q^{\text{self}}, \mathbf{H}_C^{\text{self}}, \mathbf{H}_C^{\text{self}}) \quad (7)$$

使用式 (8)、式 (9) 汇聚融合来自两种注意力机制的对话内容特征表示。

$$\xi = \text{sigmoid}(\mathbf{W}_{\text{self}} \mathbf{H}_C^{\text{self}} + \mathbf{W}_{\text{cross}} \mathbf{H}_C^{\text{cross}}) \quad (8)$$

$$\overrightarrow{H}_C = \text{Fuse}(\mathbf{H}_C^{\text{self}}, \mathbf{H}_C^{\text{cross}}) = \xi \odot \mathbf{H}_C^{\text{self}} + (1 - \xi) \odot \mathbf{H}_C^{\text{cross}} \quad (9)$$

其中, \odot 为按元素乘积, \mathbf{W}_{self} 与 $\mathbf{W}_{\text{cross}}$ 为可学习参数。最终多信息感知后的对话内容特征表示为 $\overrightarrow{H}_C \in \mathbb{R}^{l \times L(U) \times d_2}$, d_2 为注意力层特征维度。

至此, 通过语义内容的预处理获得了对话内容特征表示, 为 3.4 节的构图提供了语义内容基础。

3.4 结构增强的内容交互与融合

相比图卷积网络, 基于多头注意力机制的 GAT 能够动态处理图问题, 并高效率计算得到每一个邻居节点对当前节点不同的注意力分值, 从而更精准地识别出最重要的邻居节点^[27]。而以上特性恰好能够较好地满足本文提出的基于多层面语义结构构图并建模对话的需求, 进而较好地契合了两类核心问题的解决方案: 在词符级别能够更精准地识别出最有可能的候选诱因片段的两端边界词符, 在话轮级别能够更精准地定位到最有可能的候选诱因话轮。因此, GAT 为同时解决 ECSEC 中尚存在的两类核心问题提供了更有效的途径。

为了简洁起见, 这里着重介绍 GAT 的核心步骤, 为后续两种构图及其交互与融合奠定理论基础。假设当前节点 i 与其邻居节点 j 的特征表示分别为 $\overrightarrow{Node}_i \in \mathbb{R}^D$, $\overrightarrow{Node}_j \in \mathbb{R}^D$, D 为节点特征维度。如式 (10) 所示, 则节点 j 对节点 i 更新特征的贡献值 ϕ_{ij} 为:

$$\phi_{ij} = \frac{\exp(\text{LeakyReLU}(\overrightarrow{\eta}^T [\mathbf{W}_k \overrightarrow{Node}_i \parallel \mathbf{W}_k \overrightarrow{Node}_j]))}{\sum_{s \in \mathcal{N}_i} \exp(\text{LeakyReLU}(\overrightarrow{\eta}^T [\mathbf{W}_k \overrightarrow{Node}_i \parallel \mathbf{W}_k \overrightarrow{Node}_s]))} \quad (10)$$

依据 $\overrightarrow{\eta}$ 参数化单层前馈神经网络; 通过 \mathbf{W}_k 对每一个节点线性变换, 以获得更高级别的特征表示; \parallel 为串联拼接操作; \mathcal{N}_i 为当前节点 i 的所有邻居节点集合; LeakyReLU 为非线性激活单元。如式 (11) 所示, 在最后一层, 使用平均化聚合多头注意力, 获得节点的最终特征表示 \overrightarrow{Node}_i' :

$$\overrightarrow{Node}_i' = \sigma\left(\frac{1}{K} \sum_{k=1}^K \sum_{j \in \mathcal{N}_i} \phi_{ij}^k \mathbf{W}^k \overrightarrow{Node}_j\right) \quad (11)$$

其中, K 为多头注意力头的个数, σ 为非线性激活函数, \mathbf{W}^k 为多头映射权重。

3.4.1 词符级别构图

如图 3 所示, 针对诱因片段边界识别错误的问题, 在词符级别, 依据 $\text{CAM}^{\text{token}}$, $\text{DAM}^{\text{token}}$ 及 \overrightarrow{H}_C , 使用 GAT 构建无向图并建模对话内容。

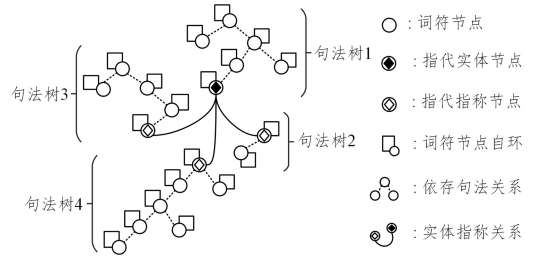


图 3 词符级别构图

Fig. 3 Token-level composition

词符级别节点的设定如下: 以 \overrightarrow{H}_C 中单个词符向量 h_p^o 作为图节点 $\text{Node}_i^{\text{token}}$ 的特征表示, p 为词符所处话轮在 HU 中的位置, o 为词符在话轮中的位置。

词符级别连接边的设定如下: 1) 在话轮内部 (intra-edge), 节点之间的连接边与本话轮的依存句法树连接边一致; 2) 在话轮之间 (inter-edge), 将指代实体节点与指代指称节点连接; 3) 每个词符节点自环 (self_loop^{token})。则词符节点 h_x^k 与词符节点 h_y^k 之间的连接边可形式化为:

$$\text{Edge}^{\text{token}}(h_x^k, h_y^k) = \begin{cases} \text{intra_edge}, & h_x^k \in \text{DAM}^{\text{token}}, h_y^k \in \text{DAM}^{\text{token}} \\ \text{inter_edge}, & [h_x^k, h_y^k] \subseteq \text{CAM}^{\text{token}} \\ \text{self_loop}^{\text{token}}, & x = y, j = k \end{cases}$$

口语化的对话中常出现省略句、长难句甚至病句等^[2], 而在序列化的上下文中, 以往仅考虑了近距离词符间线性关系的方法, 这会导致长距离的词符得不到充分交互, 同时忽略了句中词符间不同的交互频率, 这给模型精准识别诱因片段边界带来了较大的干扰。

而本文模型进一步考虑了更深层次的、非线性的依存句法关系, 并使用图的方式在词符级别构图并建模对话。由于依存句法关系蕴含着当前词符与不同线性距离邻居词符间的交互线索^[16], 因此在本文模型中, 长距离的词符间也能得到精准且有效的交互。同时, 依存句法中当前词符与邻居词符之间是一对多的关系, 这提供了词符之间不同的交互频率, 结合 GAT 在捕捉邻居节点与当前节点重要程度的优势与

特性,以上核心设计均增强了模型对最有可能的候选边界符号的感知,也为诱因片段的准确抽取提供了句子层面的局部语义综合表示。

3.4.2 话轮级别构图

通过词符级别的构图获得了句子层面的交互特征,针对诱因话轮定位错误的问题,还需要获取篇章层面的交互特征。如图4所示,在话轮级别,依据 $\text{CAM}^{\text{utter}}$ 、对话时间线次序以及 \vec{H}_C ,使用GAT构建无向图并建模对话内容。

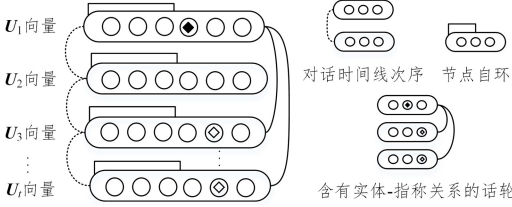


图4 话轮级别构图

Fig. 4 Utterance-level composition

话轮级别节点的设定如下:话轮级别GAT层中的图节点特征表示同样基于 \vec{H}_C ,以 \vec{H}_C 中的话轮向量 \vec{h}_c^r 作为图节点 $\text{Node}^{\text{utter}}$ 的特征表示, r 为当前话轮在 HU 中所处的位置。

话轮级别连接边的设定如下:1)依据对话的时间线次序连接(timeline_edge);2)将指代实体所处话轮与指代指称所处话轮连接(core_edge);3)每个话轮节点自环(self_loop^{utter})。则话轮节点 \vec{h}_c^m 与话轮节点 \vec{h}_c^n 之间的连接边可形式化为:

$$\text{Edge}^{\text{utter}}(\vec{h}_c^m, \vec{h}_c^n) = \begin{cases} \text{core_edge}, & \vec{h}_x^i \in \vec{h}_c^m, \vec{h}_y^j \in \vec{h}_c^n, [\vec{h}_x^i : \vec{h}_y^j] \subseteq \text{CAM}^{\text{utter}} \\ \text{timeline_edge}, & m - n = 1 \\ \text{self_loop}^{\text{utter}}, & m = n \end{cases}$$

为了简洁,口语化的长话轮对话中常出现指代消解现象^[4-5],在对话上文的某话轮中引出实体词,而在后续不同距离的话轮中使用与实体词相对应的指称来表达情感^[1],这给模型精准定位诱因所处话轮带来了较大的干扰。

而本文模型在考虑了对话时间线顺次交互关系的基础上,进一步考虑了更深层次且非线性的指代消解关系,并使用图的方式在话轮级别构图并建模对话。由于指代消解关系蕴含着目标话轮与不同距离话轮间的话题、观点等内容的交互演进线索^[2,16],因此在本文模型中,长远距离的话轮之间也能得到有效且精准的交互。同样地,指代消解中实体与指称之间也是一对多的关系,因此目标话轮内容及其情感能够与对话历史中多个不同距离的候选诱因话轮做交互,如图1中目标话轮4与话轮1、话轮3的交互,进而通过话轮级别的GAT计算出最有可能的候选诱因话轮,同时这也为诱因片段的准确抽取提供了篇章层面的全局语义综合表示。

3.4.3 双仿射机制交互与融合

虽然通过上述两种不同级别的构图分别获得了句子层面的局部语义综合表示与篇章层面的全局语义综合表示,但两种不同级别的构图之间还是相互独立的,因此最终对话内容的语义综合表示所包含的信息并不全面,致使模型依旧无法同时解决处于两种层面的问题。为此,本文基于双仿射机制,针对性地设计了一种不同级别构图间交互与融合的方法。

由于情感诱因片段抽取是在词符级别上的输出,因此两种构图均对 \vec{H}_C 进行更新,但最终更新、融合至词符级别的构图特征表示。如图5所示,假设第 i 层词符级别GAT与话轮级别GAT的输出分别为: $\mathbf{G}_{\text{word}}^i \in \mathbb{R}^{bs \times hd \times sl}$, $\mathbf{G}_{\text{utter}}^i \in \mathbb{R}^{bs \times sl \times hd}$, $i \in [1, L_{\text{num}}]$ 。其中, L_{num} 为GAT总层数, bs 为批次大小, hd 为隐层特征维度, sl 为序列长度。

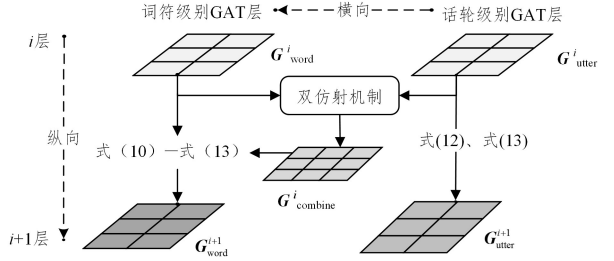


图5 交互与融合机制

Fig. 5 Interaction and integration mechanism

$\mathbf{G}_{\text{utter}}^i$ 到 $\mathbf{G}_{\text{utter}}^{i+1}$ 使用式(10)、式(11)对纵向层之间的节点进行更新。如式(12)所示, $\mathbf{G}_{\text{utter}}^i$ 与 $\mathbf{G}_{\text{word}}^i$ 横向层之间通过双仿射机制得到 $\mathbf{G}_{\text{combine}}^i \in \mathbb{R}^{bs \times sl \times sl}$ 。

$$\mathbf{G}_{\text{combine}}^i = [\mathbf{G}_{\text{utter}}^i; \mathbf{1}] \mathbf{W}_{\text{Biaff}} [\mathbf{G}_{\text{word}}^i; \mathbf{1}]^T \quad (12)$$

其中, $\mathbf{W}_{\text{Biaff}} \in \mathbb{R}^{hd \times hd}$ 为可学习的双仿射矩阵,使用式(8)一式(11)融合 $\mathbf{G}_{\text{combine}}^i$ 与 $\mathbf{G}_{\text{word}}^i$,得到第 $i+1$ 词符级别GAT层的特征表示 $\mathbf{G}_{\text{word}}^{i+1}$ 。以最后一层词符级别GAT的隐层作为最终的语义综合特征表示,即 $\vec{H}_C \in \mathbb{R}^{L \times L(i) \times d_3}$, d_3 为词符特征维度。

在命名实体识别任务中,双仿射机制可有效增强头尾词符特征的交互与融合,解决了实体嵌套的问题^[26]。受此启发,本文提出的不同级别GAT层之间横向与纵向交互、更新及融合的特有设计,使得两种级别的构图形成了一个可互联联动、可同步学习的整体,进而汇聚融合成语义内容与语义结构、局部特征与全局特征兼备的语义综合表示,这为更精准地抽取诱因片段提供了有力的基础,同时也为有效兼容话轮级别的子任务提供了基础。

3.5 诱因片段抽取

基于语义综合表示 \vec{H}_C ,与基准^[1]一致,同样使用线性层抽取诱因片段。如式(13)所示,计算诱因片段的起点索引损失 L_{start} 与终点索引损失 L_{end} ,以及诱因片段所处话轮的位置损失 L_{position} ,则总损失为3部分损失的加权和。

$$\text{Loss} = \alpha L_{\text{start}} + \beta L_{\text{end}} + \gamma L_{\text{position}} \quad (13)$$

一方面, L_{start} 与 L_{end} 对应诱因片段两端的边界识别错误, L_{position} 对应诱因所处话轮的定位错误,但不同训练数据集中两类错误的比重不同;另一方面, L_{start} 与 L_{end} 主要来源于词符级别的构图, L_{position} 主要来源于话轮级别的构图,而词符级别的构图中核心考虑了依存句法信息,话轮级别的构图核心考虑了指代消解信息,然而不同训练数据集中长难句、省略句等干扰项的比重不同,且指代消解的比例以及实体与指称一对多关系的具体复杂度也不同。因此,若平等地对待3项损失并简单地将其堆叠相加,很难灵活且充分地应对上述两方面的问题,致使模型的通用性与易用性大打折扣。

基于以上考虑,本文设置了 α, β, γ 这3项超参数,以精细化调控并适配3项损失的具体比重。由于 L_{start} 与 L_{end} 对应

同一类问题且均对应词符级别的构图,因此设定 $\alpha=\beta$ 。由于边界误差与定位误差共同构成了诱因片段抽取的总误差,因此设定 $\alpha+\beta+\gamma=1$ 。具体实验中,通过在验证集上测试与调整得到 3 项相对较优的权重超参数。抽取过程中的其他设定与基准模型^[1]保持一致,这里不再赘述。

4 实验及分析

4.1 实验数据集

考虑到数据的可用性与可比性^[3,28-30],本文以 Poria 等^[1]构建的两个 ECSEC 任务基准数据集 RECCON-DD 与 RECCON-IE 为实验数据集。为了适配机器阅读理解范式,Poria 等^[1]进一步使用了 3 种不同的策略为对话内容构建正负样本,构建后的数据集对应 Fold1-Fold3。由于 ECSEC 任务的基准模型^[1]是在 Fold1 上取得的最优性能,为了能与其形成直接且有效的对比,本文也在 Fold1 上进行实验,RECCON Fold1 数据集的具体划分信息如表 1 所列。

表 1 RECCON Fold1 数据集划分
Table 1 RECCON Fold1 datasets division

数据集	正/负样本	训练集	验证集	测试集
RECCON-DD	正样本	7 269	347	1 894
	负样本	20 646	838	5 330
RECCON-IE	正样本	—	—	1 080
	负样本	—	—	11 305

值得注意的是,由于 RECCON-IE 只有测试集,因此本文延续了 Poria 等^[1]的做法:在 RECCON-DD 上进行训练,分别在 RECCON-DD 和 RECCON-IE 两测试集上做评测。

4.2 实验参数设置

与基准^[1]最优性能的设置一致,实验中使用 SpanBERT^[31]对文本编码,版本为 spanbert-finetuned-squadv2。对话历史、构造问题与答案的最大序列长度分别为 512,512,200;词符、话轮级别编码与隐层特征维度均为 768;学习率为 1×10^{-5} ;批次大小为 16;迭代轮次为 12;指代结构解析使用 NeuralCoref,版本为 4.0;依存句法分析使用 spaCy,版本为 2.1.0;词符与话轮级别 GAT 层数均为 3;LeakyReLU 函数斜率为 -0.2;超参数 $\xi, \alpha, \beta, \gamma$ 分别为 0.5,0.4,0.4,0.2。

4.3 实验评测指标

评测指标为: $F1_{pos}, F1_{neg}, F1, EM_{pos}$ (Exact Match)。其中,针对话轮定位错误与边界识别错误的问题, EM_{pos} 与 $F1$ 为核心评测指标。 EM_{pos} 代表正样本中预测片段与金标片段完全匹配的百分比,用于衡量模型的精确抽取能力^[1,31]。 $F1$ 代表正负样本中预测片段和金标片段之间的平均重叠程度,用于衡量模型的粗略抽取能力^[1,31]。实验中使用与基准模型^[1]一致的评测脚本计算模型性能,各实验均重复 3 次,实验结果为 3 次性能均值,“±”后为对应的标准差 SD (Standard Deviation),实验结果以四舍五入的方式保留小数点后两位。

4.4 对比实验设置

为了验证本文模型在 ECSEC 任务上的有效性,在 RECCON-DD 与 RECCON-IE 两数据集上,设置了与下列 4 项典型基线模型的对比实验。

1)CRFL。Ghazi 等^[10]通过构建条件随机场学习器抽取

情感诱因片段,该方法在片段抽取的相关任务中普遍使用。

2)BGAT。Li 等^[12]未考虑语义结构,直接以句子为单位构图对话内容,并同样使用了 GAT 建模语义内容的交互。

3)MTLAKM。Turcan 等^[13]使用多任务的方式联合学习情感预测与诱因抽取,并引入了外部知识增强。

4)SBenchmark。Poria 等^[1]设定的 ECSEC 任务基准模型,将对话历史与问题拼接,使用预训练模型对拼接序列编码后接入线性层抽取诱因片段。

4.5 消融实验设置

为了分析本文模型中各结构的具体影响,在两数据集上设置下列消融实验。

1)-DP:表示完全移除 Our model 中的句法结构,对应 3.4.1 小节中的 intra_edge。

2)-CR:表示完全移除 Our model 中的指代结构,对应 3.4.2 小节中的 core_edge 以及 3.4.1 小节中的 inter_edge。

3)-CR&DP:表示同时完全移除 Our model 中的句法结构与指代结构,即-DP 且-CR。

4.6 结构与问题相关性实验设置

1)为了针对性验证句法结构对解决边界识别问题的相关性及其有效性,在两数据集上设置下列实验。

(1)OnlyDP。仅保留本文模型中的句法结构,且在预测抽取阶段直接给模型指定金标诱因话轮。即在诱因话轮定位完全正确的情况下,单独验证句法结构与边界识别问题的相关性。由于负样本的金标诱因话轮内容为空,抽取出的诱因片段必然为空,此时计算 $F1_{neg}$ 与 $F1$ 没有评测意义。因此本实验中针对正样本以 $F1_{pos}$ 与 EM_{pos} 为评测指标。

(2)OnlyDP-DP。完全移除 OnlyDP 中的句法结构,其他设定与 OnlyDP 保持一致。

2)为了针对性验证指代结构对解决话轮定位问题的相关性及其有效性,同时验证本文模型是否也能有效兼容话轮级别情感诱因蕴含,进而证明以片段为诱因提取单位的优势,针对情感诱因蕴含任务,本文在两数据集上设置下列实验。

(1)UBenchmark。Poria 等^[1]设定的话轮级别情感诱因蕴含任务基准模型,与 SBenchmark 类似,使用预训练模型编码拼接序列后依据<CLS>处的特征向量做情感诱因蕴含。

(2)OnlyCR。仅保留本文模型中的指代结构,双仿射机制改为更新、融合至话轮级别构图的特征表示,取话轮级别 GAT 最后一层隐层向量做情感诱因蕴含,同时仅计算话轮位置损失 $L_{position}$,其他设定与 UBenchmark 保持一致。

(3)OnlyCR-CR。完全移除 OnlyCR 中的指代结构,其他设定与 OnlyCR 保持一致。

由于话轮级别情感诱因蕴含任务只需识别出诱因所处话轮在对话历史中的位置,因此本实验中使用与 Poria 等^[1]一致的评测脚本计算模型性能,具体评测指标为: $Pos, F1, Neg, F1, Macro F1$ 。其中, $Macro F1$ 为核心评测指标,用于衡量模型对诱因所处话轮的综合定位能力。

4.7 实验结果与分析

4.7.1 对比实验结果及有效性分析

根据表 2 中的对比实验结果来具体分析本文模型方法的有效性、有效的原因以及优劣势。

表 2 对比实验结果

Table 2 Results of comparison experiments

数据集	模型方法	$F1_{pos}$	$F1_{neg}$	$F1$	EM_{pos}
RECCON-DD	CRFL	46.22	70.66	62.80	23.94
	BGAT	61.58	87.83	76.19	35.52
	MTLAKM	62.72±0.51	88.73±0.35	78.37±0.23	36.76±0.14
	SBenchmark	60.00	86.02	75.71	34.64
	Our model	62.81±0.47	88.26±0.41	78.13±0.18	36.90±0.08
RECCON-IE	CRFL	21.30	70.81	63.34	11.96
	BGAT	38.32	91.42	83.90	23.77
	MTLAKM	38.95±0.62	91.83±0.41	84.19±0.27	24.19±0.35
	SBenchmark	37.80	90.54	82.86	22.41
	Our model	39.07±0.52	91.70±0.32	84.04±0.20	24.38±0.17

此外,如图 6 所示,为了更直观地展现本文模型的效果,在两个数据集上,针对 4 项评测指标,计算了各基线方法与本文模型的性能差距。为了简洁起见,以“*Our Model*”为原点,当横向柱在左边时代表在对应指标上,基线方法的性能低于本文方法的性能,反之则高于本文方法的性能,横向柱的长度代表基线方法与本文方法具体的性能差距。

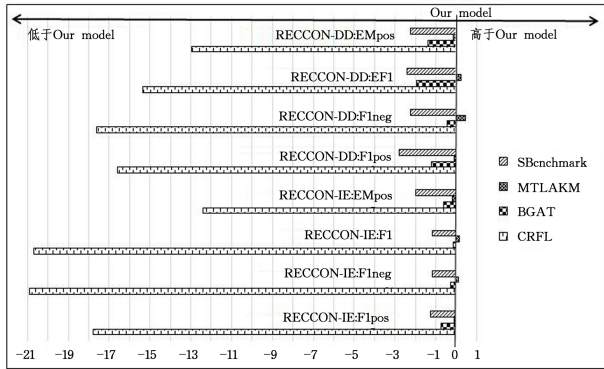


图 6 基线方法与本文模型的性能差距

Fig. 6 Performance gap between each baseline method and our model

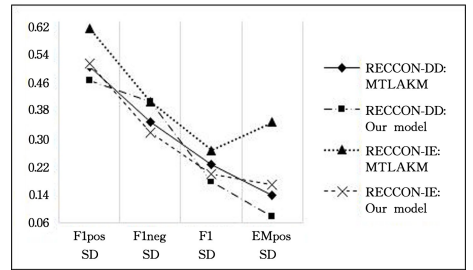
1) 与 SBenchmark 相比,在两个数据集上,本文模型在 4 项评测指标上的性能均超过 ECSEC 任务的基准模型。其中,在 $F1$ 与 EM_{pos} 两核心指标上最高分别提升了 2.42% 和 2.26%,进而从整体上验证了本文模型的有效性。

2) 本文模型及 BGAT 在 4 项指标上的性能均超过未使用 GAT 的 SBenchmark 及 CRFL,这说明了使用 GAT 建模对话的有效性,并验证了 GAT 在识别最有可能的候选边界词符及候选诱因话轮的优势。

3) 与 CRFL 及 BGAT 相比,在两个数据集上,本文模型在 4 项评测指标上的性能均超过两基线。其中,虽然 BGAT 同样使用了 GAT 建模对话,但未能针对处于特定层面的问题依据对应层面的语义结构构图并建模对话,因此不能很好地应对口语化对话中存在的长难句、省略句以及指代消解现象等干扰,因此其在 $F1$ 与 EM_{pos} 两项核心指标上的性能最高分别低于本文模型 1.94% 和 1.38%,结合本节第 2) 点的分析,这也进一步说明了本文提出的多层面语义结构增强的模型方法的有效性及其有效的原因。

4) 与 MTLAKM 相比,一方面,如图 7 所示,在两个数据集上,MTLAKM 在 $F1$ 与 EM_{pos} 两项核心指标上的性能标准差均高于本文方法的性能标准差,这说明 MTLAKM 的性能波动幅度较大,这正是由于基于知识增强的 MTLAKM 的

性能依赖于知识本身的质量以及与对话内容的匹配程度,而本文模型则是基于自然语言本身天然固有的多层面语义结构,不依赖于质量参差不齐的外部知识,因此本文模型的性能表现更加稳定。另一方面,即使本文模型未引入外部的知识做增强,在两个数据集上,本文模型在 $F1_{pos}$ 与 EM_{pos} 两项指标上的性能也均超过 MTLAKM,同时在 $F1$ 与 $F1_{neg}$ 指标上也能与 MTLAKM 保持相近的性能。因此,综合以上两方面的分析,相比 MTLAKM,本文模型的优势为:在保持较高性能的同时性能更稳定,且通用性与易用性更强。

图 7 *Our Model* 与 MTLAKM 的性能标准差比较Fig. 7 Performance standard deviation comparison between *Our Model* and MTLAKM

关于本文模型在 $F1$ 与 $F1_{neg}$ 指标上的性能低于 MTLAKM 性能的原因,可能是基于知识增强的 MTLAKM 具备更广的情景知识检索能力与更深的隐含信息推理能力^[22],因此更容易识别出对话中的隐含信息及负样本中的异常语句。而本文模型仅基于有限的语义内容及语义结构建模对话,缺少了对话者之间默认的情景知识,故在推理隐含信息、识别负样本中的异常语句等方面存在局限性。针对本文模型存在的局限性,我们将在未来的工作中进一步考虑语义结构、语义内容与外部知识的有效结合。

4.7.2 消融实验结果及结构具体影响分析

由表 3 中的消融实验结果分析各语义结构的具体影响。

表 3 消融实验结果

Table 3 Results of ablation experiments

数据集	模型方法	$F1_{pos}$	$F1_{neg}$	$F1$	EM_{pos}
RECCON-DD	Our model	62.81	88.26	78.13	36.90
	-DP	62.21	87.80	77.65	36.06
	-CR	61.50	86.98	76.95	36.27
	-CR&-DP	60.32	86.26	76.00	35.22
RECCON-IE	Our model	39.18	91.70	84.04	24.38
	-DP	38.44	91.11	83.38	23.45
	-CR	38.30	90.78	83.15	23.87
	-CR&-DP	38.06	90.68	83.07	22.92

此外,如图 8 所示,为了更清晰地展现本文模型中各模块的具体效果及其有效性,我们针对 $F1$ 与 EM_{pos} 两项核心指标,比较了本文完整的模型 Our Model 与各模块在两个数据集上的 3 项消融实验(-DP,-CR,-CR&DP)的性能下降幅度。

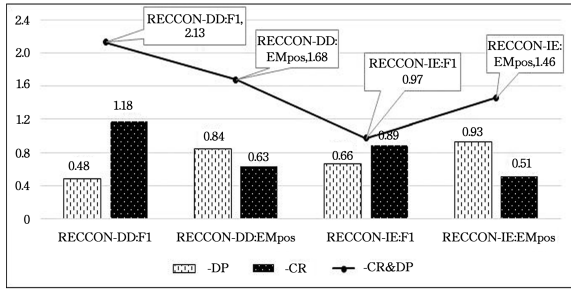


图 8 消融实验性能下降幅度

Fig. 8 Decline amplitude of performance in ablation experiments

1)在两个数据集上,单独移除本文模型中的任何一种结构,均会导致在 4 项指标上的性能下降,这说明了本文模型中的句法结构与指代结构均有助于诱因片段的抽取。

2)当本文模型中两种结构均被移除时,在 4 项指标上的性能下降幅度达到最大且均超过单独移除一种结构的下降幅度,但仍均高于 SBenchmark 的性能。这说明了篇章层面的指代结构与句子层面的句法结构有联合增强诱因片段抽取的效果,进一步验证了本文提出的不同级别构图间交互与融合特别设计的有效性,也反向证明了以往方法基于单层面结构增强的局限性。同时,也验证了 3.3.2 节中综合使用自注意力与交叉注意力机制获取对话内容特征表示的有效性。

3)在两个数据集上,当从模型中单独移除句法结构时, EM_{pos} 下降幅度大于 $F1$ 下降幅度;当从模型中单独移除指代时, $F1$ 下降幅度大于 EM_{pos} 下降幅度。这说明了句子层面的句法结构侧重于对模型精确抽取能力的提升,而篇章层面的指代结构侧重于对模型粗略抽取能力的提升。

4.7.3 结构与问题相关性分析

由表 4、表 5 所列的实验结果可以得出以下结论。

1)句法结构有助于解决边界识别错误的问题。

如表 4 所列,在诱因话轮完全正确且模型 OnlyDP 中仅有句法结构的情况下抽取诱因片段,此时仅存在诱因片段边界识别错误的问题。移除 OnlyDP 中的句法结构后,在两个数据集上, $F1_{pos}$ 值及 EM_{pos} 值均出现了与 -DP 消融实验中相当的性能下降。这说明 -DP 消融实验中的性能下降主要是由于边界识别错误导致的,也解释了句法结构能够侧重提升模型精确抽取能力的原因,进而在句法结构有助于抽取诱因片段的总结论基础上,更细粒度地证明了句法结构有助于解决诱因片段边界识别错误的问题。

表 4 句法结构与边界识别问题相关性的实验结果

Table 4 Experimental results of correlation between syntactic structures and boundary recognition problems

数据集	模型方法	$F1_{pos}$	EM_{pos}
RECCON-DD	OnlyDP	68.37	43.10
	OnlyDP-DP	67.88(↓0.49)	42.34(↓0.76)
RECCON-IE	OnlyDP	43.42	27.99
	OnlyDP-DP	42.81(↓0.61)	27.12(↓0.87)

2)指代结构有助于解决话轮定位错误的问题

如表 5 所列,本文针对话轮级别的情感诱因蕴含任务进行对比实验。将 OnlyCR 与 OnlyCR-CR 进行对比,在模型中仅有指代结构的情况下识别诱因所处的话轮,此时仅存在话轮定位错误的问题。移除模型中的指代结构后,在两数据集上,模型的 3 项指标性能均出现了较大幅度的下降。将 OnlyCR 与 UBenchmark 进行对比,基于指代结构增强的 OnlyCR 在 3 项指标上的性能均超过基准模型 UBenchmark。以上两组对比实验,一方面证明了本文模型也能有效兼容话轮级别的情感诱因蕴含任务,进而证明了以片段为诱因提取单位的优势;另一方面,也解释了指代结构能够侧重提升模型精确粗略能力的原因,进而在指代结构有助于抽取诱因片段的总结论基础上,更细粒度地证明了指代结构有助于解决话轮定位错误的问题。

表 5 指代结构与话轮定位问题相关性的实验结果

Table 5 Experimental results of correlation between coreferential structures and utterance positioning problems

数据集	模型方法	Pos. F1	Neg. F1	Macro F1
RECCON-DD	UBenchmark	66.23	87.89	77.06
	OnlyCR	67.71	88.57	78.14
	OnlyCR-CR	66.59(↓1.12)	87.95(↓0.62)	77.27(↓0.87)
RECCON-IE	UBenchmark	40.83	95.68	68.26
	OnlyCR	41.90	96.12	69.01
	OnlyCR-CR	41.07(↓0.83)	95.75(↓0.37)	68.41(↓0.60)

4.7.4 抽取结果示例及优劣势分析

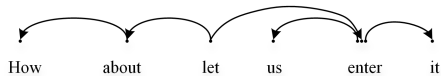
为了更好地理解本文模型的优劣势,本节基于表 6 所列的样例,直观地展示了各模型的诱因片段抽取结果。

表 6 抽取结果

Table 6 Extraction results

样例	模型方法	各模型的抽取结果
对话历史: U_1 : Have you seen that house? U_2 : Yes, looks like a castle. U_3 : You're right. I love it. U_4 : It must be old. U_5 : More than 100 years old. U_6 : How about let us enter it?	CRFL	More than 100 years old; How about let us enter it
目标话轮: U_6 目标情感:happiness 诱因片段: looks like a castle; It must be old; let us enter it	BGAT	It must be old; How about let us enter it
	MTLAKM	looks like a castle; It must be old; How about let us enter it
	SBenchmark	I love it; How about let us enter it
	Our model	It must be old; let us enter it

针对话轮定位错误的问题,相比 CRFL 与 SBenchmark,基于 U_4 中指称“it”与 U_1 中实体“house”的指代关系,本文模型中的话轮级别构图增强了 U_4 节点与 U_1 节点的交互,因此基于指代结构增强的本文模型准确定位到了 U_4 中的诱因;针对边界识别错误的问题,如图 9 所示,基于 U_6 的句法结构,本文模型中的词符级别构图增强了词符“enter”与其一跳邻居词符的交互,而 U_6 中的诱因片段正是由词符“enter”及其一跳邻居词符组成,因此相比 4 项基线,基于句法结构增强的本文模型精确识别出了 U_6 中的诱因片段边界。

图9 U_6 的句法结构Fig. 9 Syntactic structure of U_6

然而,相比知识增强的 MTLAKM,包括本文模型在内的其他模型均未抽取 U_2 中的诱因片段。针对本文模型,一方面是由于 U_2 中省略了主语“it”,致使本文模型使用的结构解析包未能识别出 U_2 中隐含的“it”与 U_1 中“house”的指代关系,此错误又进一步传播至模型的相关构图中;另一方面,如 4.7.1 小节中的第 3) 点分析,本文模型仅依据有限的语义内容及语义结构较难推理出 U_2 中“castle”与 U_1 中“house”的比喻关系,且这种关系的阐述可能会诱发 U_6 中“开心”情感的表达。针对本文模型存在的局限性,我们将在未来的工作中加以改进。

结束语 为了解决 ECSEC 任务中两类亟待解决的问题,本文提出了一种多层面语义结构增强的对话情感诱因片段抽取模型。该模型基于句子层面的句法结构与篇章层面的指代结构,分别在词符级别与话轮级别构图并建模对话,并针对性设计了两不同级别构图间交互与融合的方法。对比实验与消融实验验证了本文方法及模型的有效性,更细粒度的结构与问题相关性实验证明了本文模型中的句法结构有助于解决诱因边界识别错误的问题,而指代结构有助于解决诱因话轮定位错误的问题;同时,证明了本文模型也能有效兼容话轮级别的情感诱因蕴含任务。针对本文模型的局限性,未来将使用多任务的方式联合学习 ECSEC 与语义结构解析,并考虑引入外部知识来增强模型对隐含信息的推理能力。

参考文献

- [1] PORIA S, MAJUMDER N, HAZARIKA D, et al. Recognizing Emotion Cause in Conversations[J]. arXiv:2012.11820, 2021.
- [2] PORIA S, MAJUMDER N, MIHALCEA R, et al. Emotion Recognition in Conversation: Research Challenges, Datasets, and Recent Advances[J]. arXiv:1905.02947, 2019.
- [3] KIM H, KIM B, KIM G. Perspective-taking and Pragmatics for Generating Empathetic Responses Focused on Emotion Causes [J]. arXiv:2109.08828, 2021.
- [4] LIU Z, SHI K, CHEN N F. Coreference-Aware Dialogue Summarization[J]. arXiv:2106.08556, 2021.
- [5] HAN T, HUANG C, PENG W. Coreference Augmentation for Multi-Domain Task-Oriented Dialogue State Tracking [J]. arXiv:2106.08723, 2021.
- [6] ZHANG Y, XIA Q, ZHOU S, et al. Semantic role labeling as dependency parsing: Exploring latent tree structures inside arguments[J]. arXiv:2110.06865, 2021.
- [7] SADVILKAR N, NEUMANN M. PySBD: Pragmatic Sentence Boundary Disambiguation[J]. arXiv:2010.09657, 2020.
- [8] LEE S Y, CHEN Y, LI S, et al. Emotion Cause Events: Corpus Construction and Analysis[C]// Proceedings of the Seventh International Conference on Language Resources and Evaluation, Valletta, Malta: ELRA, 2010.
- [9] LEE S, CHEN Y, HUANG C. A text-driven rule-based system for emotion cause detection [C] // In the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text. California: ACL, 2010: 45-53.
- [10] GHAZI D, INKPEN D, SZPAKOWICZ S. Detecting emotion stimuli in emotion-bearing sentences [C] // In International Conference on Intelligent Text Processing and Computational Linguistics. Cham: Springer, 2015: 152-165.
- [11] VASWANI A, SHAZEER N M, PARMAR N, et al. Attention is All you Need [J]. arXiv:1706.03762v1, 2017.
- [12] LI X, GAO W, FENG S, et al. Span-level emotion cause analysis by bert-based graph attention network [C] // Proceedings of the 30th ACM International Conference on Information & Knowledge Management. Queensland, Australia: ACM Digital Library, 2021: 3221-3226.
- [13] TURCAN E, WANG S, ANUBHAI R, et al. Multi-Task Learning and Adapted Knowledge Models for Emotion-Cause Extraction [C] // Findings of the ACL. Online: Association for Computational Linguistics, 2021: 3975-3989.
- [14] HU G, LU G, ZHAO Y. FSS-GCN: A graph convolutional networks with fusion of semantic and structure for emotion cause analysis [J]. Knowledge-Based Systems, 2021, 212: 106584.
- [15] DING Z, HE H, ZHANG M, et al. From Independent Prediction to Re-ordered Prediction: Integrating Relative Position and Global Label Information to Emotion Cause Identification [J]. arXiv:1906.01230, 2019.
- [16] JURAFSKY D, MARTIN J H. Speech and Language Processing [M]. Stanford University: University of Colorado at Boulder, 2021: 231-548.
- [17] XIA R, DING Z. Emotion-Cause Pair Extraction: A New Task to Emotion Analysis in Texts [J]. arXiv:1906.01267, 2019.
- [18] PU J, BU L, LU Y, et al. Utilizing Heterogeneous Graph Neural Network to Extract Emotion-Cause Pairs Effectively [J]. Computer Science, 2023, 50(1): 205-212.
- [19] LI Y, SU H, SHEN X, et al. DailyDialog: A Manually Labelled Multi-turn Dialogue Dataset [C] // Proceedings of the Eighth International Joint Conference on NLP. Taipei, Taiwan: Asian Federation of NLP, 2017: 986-995.
- [20] BUSSO C, BULUT M, LEE C, et al. IEMOCAP: interactive emotional dyadic motion capture database [J]. Language Resources and Evaluation. 2008, 42(4): 335-359.
- [21] ZHANG D, YANG Z, MENG F, et al. TSAM: A Two-Stream Attention Model for Causal Emotion Entailment [J]. arXiv: 2203.00819, 2022.
- [22] LI J, MENG F, LIN Z, et al. Neutral Utterances are Also Causes: Enhancing Conversational Causal Emotion Entailment with Social Commonsense Knowledge [J]. arXiv: 2205.00759, 2022.
- [23] BHAT A, MODI A. Multi-Task Learning Framework for Extracting Emotion Cause Span and Entailment in Conversations

- [J]. arXiv:2211.03742,2022.
- [24] HUANG B,ZHANG Z,ZHAO H. Tracing Origins:Coreference-aware Machine Reading Comprehension[C]//Proceedings of the 60th Annual Meeting of the ACL. Dublin, Ireland; Association for Computational Linguistics,2022:1281-1292.
- [25] SHOU Y,MENG T,AI W,et al. Conversational emotion recognition studies based on graph convolutional neural networks and a dependent syntactic analysis[J]. Neurocomputing, 2022, 501: 629-639.
- [26] YU J,BOHNET B,POESIO M. Named Entity Recognition as Dependency Parsing[J]. arXiv:2005.07150,2020.
- [27] VELICKOVIC P,CUCURULL G,CASANOVA A,et al. Graph Attention Networks[J]. arXiv:1710.10903,2017.
- [28] WANG F,DING Z,XIA R,et al. Multimodal Emotion-Cause Pair Extraction in Conversations[J]. arXiv:2110.08020,2021.
- [29] LI W,PANDELEA V,GE M,et al. ECPEC:Emotion-Cause Pair Extraction in Conversations[J]. arXiv:2203.13926,2022.
- [30] GHOSAL D,SHEN S,MAJUMDER N, et al. CICERO:A Dataset for Contextualized Commonsense Inference in Dialogue[J].

arXiv:2203.13926,2022.

- [31] JOSHI M,CHEN D,LIU Y, et al. SpanBERT:Improving Pre-training by Representing and Predicting Spans[J]. Transactions of the Association for Computational Linguistics,2020,8:64-77.



QIN Mingfei, born in 1995, postgraduate, is a member of China Computer Federation. His main research interests include natural language processing and emotional conversation systems.



FU Guohong, born in 1968, Ph.D, professor, is a member of China Computer Federation. His main research interests include natural language processing, opinion mining/sentiment analysis and so on.

(责任编辑:喻藜)