

引入常识和情感词典知识的对话情绪识别

杨开来

院（系）： 计算学部

专 业： 计算机科学与技术

学 号： 1170300709

指导教师： 刘秉权

2021 年 6 月

哈爾濱工業大學

畢業設計（論文）

題 目 引入常識和情感詞典知識的

對話情緒識別

專 業 計算機科學與技術

學 號 1170300709

學 生 楊開來

指 導 教 師 劉秉權

答 辯 日 期 2021 年 6 月

摘 要

近年来，对话情绪识别（ERC）任务已经在 NLP 社区获得了广泛的关注。该任务关注于识别对话中每一个句子的情感，在社交媒体分析，对话系统等方面有重要的应用价值。一些 ERC 工作关注于在对话表示中引入词级别的结构化常识知识来帮助模型进行语义推理。然而，这些模型直接将对话和知识表示进行拼接作为知识增强的表示，但忽略了它们内部的交互和推理过程。此外，部分工作利用多任务学习来引入任务相关的信息，这些信息可以帮助模型进行复杂的情绪推理。但是这些模型采用与情绪不直接相关的副任务，因此这些任务只能为 ERC 任务提供有限的指导信息，并且理解这些指导信息需要较复杂的推理过程。

为了解决以上问题，我们提出一个基于常识知识引入和情感极性强度敏感的多任务学习的模型，简称为 KI-Net，该模型利用常识知识和情感词典知识来丰富语义信息。具体地，为解决对话和知识表示交互不足的问题，我们首次在 ERC 任务中引入一个自匹配模块，该模块是一个自注意力相关的机制。自匹配模块被用于拼接后的知识增强表示，以进一步完成该表示的内部交互以及句子和知识的相关性推理。为了解决多任务学习中副任务提供的信息与情绪不直接相关的问题，考虑到和 ERC 任务的相关性，我们引入一个新的副任务：短语级别的情感极性强度的预测任务（SPIP）。利用该任务，我们通过指导模型预测句子中关键短语的情感极性值的方式，首次利用多任务学习向 ERC 模型中引入情感词典知识。情感词典知识与情绪直接相关，且无需复杂推理过程就可以向 ERC 任务提供有效的指导信息。

实验表明，常识知识引入，自匹配模块以及 SPIP 模块均提升了模型在三个测试数据集上的表现。与非常有竞争力的基线模型相比，我们的 KI-Net 模型在这些数据集上获得了当前的最佳表现。此外，我们通过实验结果分析模型取得优良效果的原因，知识推理过程中存在的问题，以及为了改进这些问题，未来可以探索的工作方向。

关键词：对话情绪识别；常识知识；情感词典；多任务学习；自匹配；

Abstract

In recent years, Emotion Recognition in Conversation (ERC) has gained much attention from the NLP community. This task focuses on recognizing the emotion of each utterance within a conversation, and has many applications such as in social media analysis and dialogue systems. Some ERC works concentrate on leveraging word-level structured commonsense knowledge to help complicated semantic reasoning. However, these models directly concatenate utterance and knowledge representations as knowledge-enriched representations, but neglect their internal interaction and reasoning process. In addition, some works utilize multi-task learning to introduce task-related information, and these information can aid complicated emotion reasoning process. But these tasks adopt emotion-indirect auxiliary tasks, which provide limited guiding information for ERC task. These guiding information requires hard reasoning process. To address the above issues, we propose a Knowledge-Interactive Network with sentiment polarity intensity-aware multi-task learning, namely KI-Net, which leverages both commonsense knowledge and sentiment lexicon to augment semantic information. Specifically, to address the problem of limited interaction between utterances and knowledge, for the first time, we use a self attention-based self-matching module. Self-matching module is used for concatenated knowledge-enriched representations, to further enable their internal interaction and reasoning process of utterance-knowledge relations. To enhance the emotion-indirect auxiliary tasks, considering correlations with the ERC task, a phrase-level Sentiment Polarity Intensity Prediction (SPIP) task is devised as an auxiliary task. This task instructs the model to predict sentiment polarity values of key phrases within utterances, and firstly introduces sentiment lexicon knowledge to ERC using multi-task learning. Sentiment lexicon knowledge is directly-related to emotions, and is able to provide effective guiding information without complicated reasoning process. Experiments show that all commonsense knowledge integration, self-matching and SPIP modules improve the model performance respectively on three datasets. Compared with competitive baseline models, our KI-Net model achieves state-of-the-art results on these datasets. Besides, through the experimental results we analyse the reasons of our superior model performance, existing weaknesses and future work to improve these weaknesses.

Keywords: ERC, commonsense knowledge, sentiment lexicon, multi-task learning, self-matching

目 录

摘要.....	1
第 1 章 绪论	4
1.1 课题背景及研究的目的和意义.....	4
1.2 知识引入在情绪相关任务中的发展概况.....	4
1.3 对话情绪识别任务的发展概况.....	5
1.4 本文的主要研究内容.....	7
第 2 章 引入常识知识和情感极性敏感的多任务学习的模型	9
2.1 任务定义和模型概述.....	9
2.2 上下文和对话关系敏感的编码器.....	9
2.3 常识知识引入及知识增强.....	11
2.3.1 知识抽取与基于图注意力的表示学习.....	12
2.3.2 知识增强表示的自匹配机制.....	13
2.4 引入情感词典知识的多任务学习.....	14
2.4.1 情感词典 SenticNet 简介及知识抽取.....	14
2.4.2 情感极性预测任务及多任务学习设置.....	15
第 3 章 实验设置及基线介绍	17
3.1 基线数据集简介.....	17
3.2 基线模型简介.....	18
3.3 实验设置及评测指标.....	19
第 4 章 实验结果与分析	20
4.1 整体实验结果分析.....	20
4.2 测试结果中的典例和误差分析.....	22
4.3 关键模块的销蚀实验.....	24
4.4 常识知识的关键维度.....	25
结论.....	27
参考文献.....	28
原创性声明.....	35
致谢.....	36

第 1 章 绪论

1.1 课题背景及研究的目的和意义

情绪是人类的一种内心活动。因此，对于优秀的人工智能系统，理解人类情绪并做出正确的反应是非常关键的一环。对话情绪识别研究（Emotion Recognition in Conversations, ERC）作为通过来自社交媒体（如 Facebook, YouTube, Reddit, Twitter）的大量的公开的对话数据中挖掘情绪相关信息的一种新兴的自然语言处理前沿研究任务，正在获得越来越多的关注。ERC 任务的目标是识别对话中每一个句子的情感，这个任务要求机器理解对话中表达情绪的方式^[1]。由于对话系统需要在聊天过程中理解用户的情绪并提供正确反馈，ERC 上的研究可以帮助构造情绪敏感的对话系统^[4,5]，进而提升整体的人机交互体验。因此，ERC 任务拥有广泛的应用领域，例如社交媒体分析^[2,3]，智慧家居和对话机器人等智能系统^[54]。除此之外，该任务还可用于健康分析系统（心理健康分析）^[43]，教育心理学（检测学生学习效果与情绪的关系）^[42]等。因此，这些潜在的需求和应用场景，要求我们开发出更高效和可靠的对话情绪识别算法。

1.2 知识引入在情绪相关任务中的发展概况

在情绪相关任务中，经常会遇到有标注数据不足，在实际应用过程中泛化能力差的情况。因此，很多相关工作考虑引入外部知识帮助模型进行推理。下面我们主要介绍四类在情绪相关任务上引入知识，并对我们的工作很有启发性的典型方法。

引入情感词典知识 Teng 等人^[44]首先找出文本中的情感相关词汇，并计算其对文本整体情感极性的贡献程度，然后结合每个词的贡献值及其情感得分作为局部的情感极性值，最终加上全局的情感极性预测值作为整个文本的情感得分。虽然上述工作在计算情感得分时，考虑了 not、very 等否定词、强化词的得分信息，但是没有显式把这些词对上下文的情感语义表示的影响刻画出来。Qian 等人^[45]考虑到情感词、否定词、强化词在情感语义组合过程中起到的不同作用，对文本建模过程中不同位置词的情感分布加以约束。具体的约束方式是设置针对不同方面（情感词，否定词等）的正则器（regularizer），并将其度量的结果合并到损失函数中进行训练。

引入无标注语料 Tang 等人^[46]观察到一般的词向量对于“good”和“bad”这种上下文相近但极性相反的词，给出的向量表示没有很强的区分性，不利于情感分析任务。而社交媒体中有海量包含表情符的文本，利用这些情感极性明确的表情符可以过滤得到大量弱标注的情感文本。作者使用这些语料，在原有的 C&W 训练方式基础上，在词向量训练中引入情感得分相关的损失，将情感信息融入词向量表示中，使上下文相近但情感不同的词的向量表示有明显的差异。

引入特征提取算法 在很多情况下，知识不是显性存在于大规模的知识图谱中，而是存在于对应的人工标注数据中。目前一般利用学习算法从这些数据中训练用于提取特征的模型。Tian 等人^[47]在 BERT 语言模型基础上，将文本中的评价对象（属性）、情感词及情感词极性等情感元素引入 Masked Language Model 预训练任务。类似地，Ke 等人^[48]在预训练语言模型（PLMs）中引入词级别的情感、词性知识。他们先为每个词预测词性信息，然后依据词性信息从 SentiWordNet 中推断其情感极性。

引入常识知识 为了在情感分析以及跨领域情感分析中有更好表现，很多方法尝试引入常识知识来增强文本表示。Dragoni 等人^[49]利用语义和情感知识库 SenticNe 和 General Inquirer 包含的词语及情感表示的相关知识，结合句法分析进行跨领域的情感分析任务。Ghosal 等人^[50]在模型中引入 ConceptNet 包含的常识知识，找出每个文档中独有的名词、形容词、副词集合，从中抽取出一个文档相关的子图，进而提供一个由 ConceptNet 提取而来的特征表示，与文档本身的表示一起做最后的情感分类。类似的方法也被使用于领域情感分类。Hang 等人^[51]在跨目标的 stance detection 任务中通过 SenticNet 和 EmoLex 构建语义-情绪图，并使用图卷积网络（GCN）^[55]对其进行编码，并将这些知识通过记忆网络融入到句子的语义表示中。

1.3 对话情绪识别任务的发展概况

相比于其它情绪相关任务，对话情绪识别在最近几年才获得 NLP 社区的关注^[17, 18]，这是因为更多高质量的对话数据集的公开，如 IEMOCAP^[19]、MELD^[20]和 DailyDialog^[21]，可供 ERC 的相关研究。与传统情感分析任务不同，ERC 任务要求建模对话参与方的交互关系。考虑到这个要求，很多工作采用基于循环神经网络（RNN）的模型对上下文中的句子进行序列建模。例如 CMN^[22]和 ICON^[23]，采用多个 RNN 来建模每一个对话参与方的状态，并采用多跳的记忆网络来建模历史对话。基于这些工作，Majumder 等人^[24]提出一个基于注意力 RNN 的模型 DialogueRNN，在引入新的 RNN 建模对话参与方的状态和情绪变动的同时，在情

绪分类过程中引入注意力机制使得模型关注当前状态的各个不同方面。

作为一个在很多 NLP 任务中表现优秀的表示学习模型，Transformer^[25]也已经在很多工作中被用来建模输入序列，并且在很多基线数据集中表现超过了基于 RNN 的模型。BERT 等 PLMs 的成功已经证明 Transformer 拥有更强的知识推理能力，因此它在 ERC 任务中常被用于编码外部知识增强的表示。例如 Zhong 等人^[26]采用一个层级的 Transformer 结构来编码知识增强的对话向量，这也是首个在 ERC 任务中引入 Transformer 的工作。类似地，Zhang 等人^[27]利用一个增量式的 Transformer 编码多轮对话。

为了更好的建模上下文信息和句子依赖，一些结构例如记忆网络^[28, 29]和基于图的模型也被用于学习表示。图神经网络在建模对话参与方的依赖性方面具有先天优势，因为它们的边可以保存丰富的关系信息。Ghosal 等人^[30]预定义多个句子级别的关系为有向边，并在每个对话上构造一个图，将句子级别的表示作为节点。然后一个 GCN 被用于图编码。为建模说话者相关依赖，Zhang 等人^[59]结合说话者节点到图中，并使用新的边来连接其与句子节点。Sheng 等人^[56]使用异质图（Heterogeneous graph）建模主题相关的情绪交互，添加从句子中抽取的情绪相关的短语作为节点到图中。此外，Liang 等人^[57]向图中添加如面部表情特征等多模态信息到异质图中，以丰富情绪相关知识。考虑到图结构中位置信息的缺失，Ishiwatari 等人^[31]计算每个节点的相对位置编码来为图提供位置信息。

受当前可用的高质量数据集的大小的限制，一些工作尝试通过引入任务相关的知识来帮助模型的表现。一些工作提出任务相关的预训练任务来提升模型的泛化能力。Jiao 等人^[33]设计了一个句子级的对话补充（conversation completion）任务，在给定上下文的情况下从一个随机采样的句子集合中预测一个被遮盖的句子。传统的预训练任务，例如 Masked Language Model 和 Generalized Autoregressive Pre-training 也在对话的预训练中取得了成功^[34]。此外，Hazariika 等人^[58]使用一个 Hierarchical Recurrent Encoder-Decoder (HRED)来进行生成式的预训练，然后该 HRED 编码器被用于 ERC 任务。

为形成知识增强的表示以帮助进一步的推理，一些工作从大型知识图谱中显式地抽取常识知识并将其与词表示进行拼接。例如 Zhong 等人^[26]利用一个动态的上下文敏感的图注意力机制来编码知识图谱中的结构化知识。Zhang 等人^[27]则利用一个层级的图注意力机制来编码这些知识。此外，一些工作通过多任务学习隐式地引入知识。例如 Zhang 等人^[27]通过引入判断情绪是否为中性的副任务来缓解数据

集中的标签不平衡问题。Wang 等人^[32]通过一个神经主题模型，将预测对话的整体主题，顾客意图和服务者意图作为副任务，在模型的情绪推理过程中引入主题信息。Li 等人^[15]通过一个简单的句子-说话者关系判断的副任务，使模型能够关注说话者的对话风格和情绪变化。

1.4 本文的主要研究内容

与传统的句子中的情绪识别不同，对话中的上下文建模对于 ERC 任务是非常关键的。早期基于循环神经网络（RNN）的 ERC 工作采用记忆网络（memory networks）^[22]来存储对话历史信息。近期，随着 PLMs 的发展，例如 BERT^[6]和 XLNet^[7]，许多下游任务已经获得了很大的提升，例如对话系统^[8,9]和机器阅读理解^[12,39]。然而，研究仍然表明 PLMs 可能缺失某些维度的知识^[11]，它们会限制其在相关下游任务中的表现。因此，近期的大部分基于 PLMs 的 ERC 工作采用层级结构，通过 PLMs 来获得词级别的或句子级别的表示，然后采用其它专门设计的模块来补充知识。它们中的一些工作显式地向模型中结合结构化的常识知识^[12]，并形成知识增强的表示。对于较为抽象或非结构化的知识，一些其他模型引入多任务学习^[13, 14, 15]来隐式地弥补模型缺失的知识维度。

然而，引入常识知识时，当前所有的方法均没有考虑对话和知识表示的内部交互，而直接进行拼接。这种忽视导致对于知识增强表示的内部情绪推理的缺失，进而严重影响模型表现。这种严重性我们将会在后期的实验中证明。除此之外，大部分多任务学习方法的副任务都是情绪不相关的，例如主题推理^[32]和对话-说话者关系验证^[15]，这些方法并没有提供额外的直接情绪相关的信息。还有研究表明，不同知识源之间的冗余是很少的^[11]，因此对于知识驱动的任务，不同维度的知识补充可以更充分地帮助推理过程。在图 1 中，我们展示一个常识知识和情感词典知识均对情绪推理过程有帮助的例子，其中圆柱体表示常识知识，矩形表示情感词典知识。考虑第一个句子中的关键词“divorce”，将“an_affair”作为可能的原因，“stop_being_married”作为行动和“divide_a_family”作为结果，模型可以对于“divorce”形成一个完整的增强语义链条。情感词典对“divorce”给予一个非常消极的情感极性值“-0.83”，可以直接指导模型确定消极的情绪。明显地，在结果“Frustrated”的情绪推理过程中，这两种知识源均对模型有同等重要的帮助，并且少有重叠。

为了解决以上的挑战，我们提出一个基于知识交互和情感极性强度敏感的多任务学习的模型（KI-Net）。我们首先采用一个上下文和相关性敏感的编码器，这个

编码器使 PLMs 关注于上下文建模中的关键因素。为了进一步提升词级别的表示，我们利用来自大型常识知识图谱和情感词典的知识。受到 Yang 等人^[12]的启发，我们使用一个自匹配（Self-matching）机制来结合常识知识表示到词级别的表示中，这种机制允许词级别表示和知识表示产生充分的内部交互关系。此外，我们还引入一个短语级别的情感极性强度的预测任务（SPIP）作为副任务，利用来自情感词典的情感极性值，向模型提供关键词的情感极性信息，直接为目标句子的情感探测提供指导。

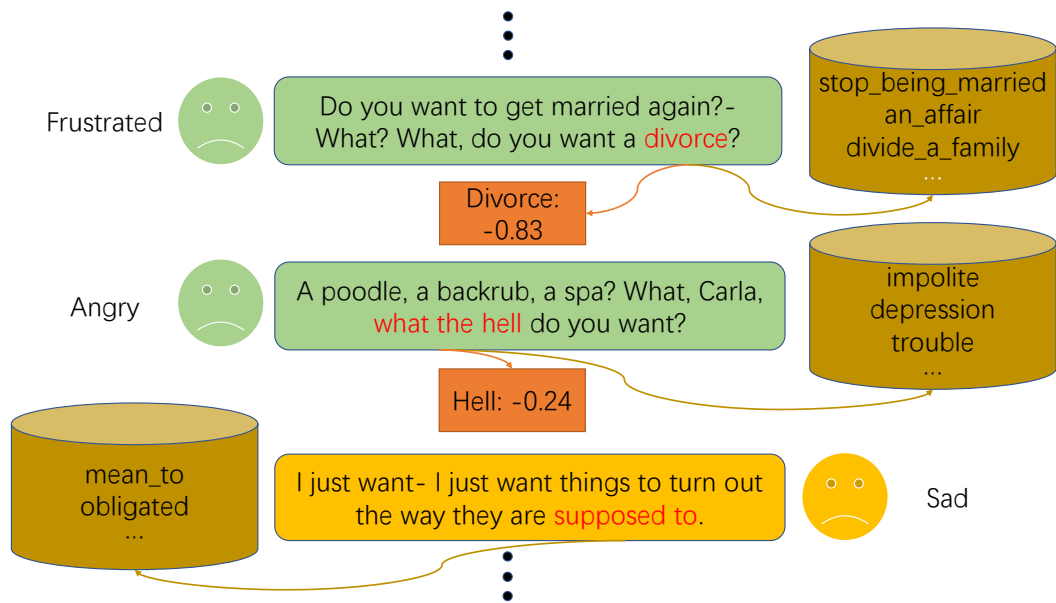


图 1：一个常识知识和情感词典知识都可以帮助 ERC 任务的例子

总结来说，这篇论文做出了以下贡献：

- 我们利用多种来源的知识尝试弥补 PLMs 中的部分缺失的知识维度。随后的销蚀实验表明我们引入的知识的确对模型表现有积极影响。
- 我们第一次在 ERC 任务上探索和讨论句子和知识表示之间显式交互的必要性，为以后的知识结合相关工作提供指导性意见。
- 我们在 ERC 任务上采用一个新的副任务，称为短语级别的情感极性强度的预测(SPIP)任务。这个任务与情绪直接相关。实验表明 SPIP 任务可以为 ERC 任务提供关键的情绪信息。

第 2 章 引入常识知识和情感极性敏感的多任务学习的模型

本节中我们正式定义对话情绪识别任务，并对我们的基于常识知识引入和情感极性强度敏感的多任务学习的模型的各部分进行简要概述。此外，我们详细介绍该模型的重要基础部分：上下文和对话关系敏感的编码器，常识知识引入及知识增强，引入情感词典知识的多任务学习。

2.1 任务定义和模型概述

我们以如下的方式正式定义对话情绪识别(ERC)任务：给定 $\{\{X^i\}, Y^i\}$ ，其中 $i = 1, \dots, N, j = 1, \dots, N^i$ ，作为对话数据集中一个对话的{对话，标签}对的集合。对话 X 包含 N 个句子，其中每一个句子 X^i 包含 N^i 个 token，称为 $X^i = \{X_1^i, X_2^i, \dots, X_{N^i}^i\}$ 。每一个 X^i 由 $p(X^i)$ 说出，其中 P 是说话者的集合。离散值 $Y^i \in S$ 被用来表示情绪标签，其中 S 表示情绪标签的集合，并且 $|S| = h_c$ 。这个任务的目标是给定上文和 X 与 P 之间的映射的情况下预测句子 X^i 的情绪标签 Y^i 。

为补充 PLMs 缺失的知识维度，我们设计一个层次化的模型，其关键的想法是引入不同来源的多维度的相关知识来增强 PLMs，并使对话和知识表示充分交互。基于以上的想法，我们如图 2 一样构建 KI-Net 模型。该模型包含四个主要的部分：我们首先设计一个改进注意力机制的编码器，用于计算上下文和对话关系敏感表示（2.2 节）。然后一个知识结合模块被用于常识知识的抽取，并形成基于图注意力机制的表示（3.2 节）。一个自匹配模块被用于基于自注意力机制的对话-知识之间的交互（3.3 节）。我们通过设计一个 SPIP 任务，向模型中引入情绪相关的情感词典知识，并利用一个多任务学习的设置来结合 ERC 和 SPIP 任务（4.2 节）。具体地，在图 2 中右侧是 KI-Net 模型的整体结构。Rep.代表向量表示。在（a）中我们展示一个从 ConceptNet 中使用关键词“happy”抽取的子图作为例子。在（b）中我们展示 SenticNet 为关键词“bless”提供的情感信息的例子，其中包含情感极性（灰色），极性值（粉色），情感标签（绿色）和语义（蓝色）。

2.2 上下文和对话关系敏感的编码器

对于 ERC 任务，语义知识和依赖性建模对情绪探测都是非常关键的。丰富的语义知识可以由 PLMs 引入，但对话中多方的依赖关系却无法通过传统的自注意力机制建模。因此，在 XLNet^[7]的基础上，我们设计一个上下文和对话关系敏感的编码器（CDA Encoder），通过改进传统的自注意力机制，使编码器可以同时关注以上两个要素。

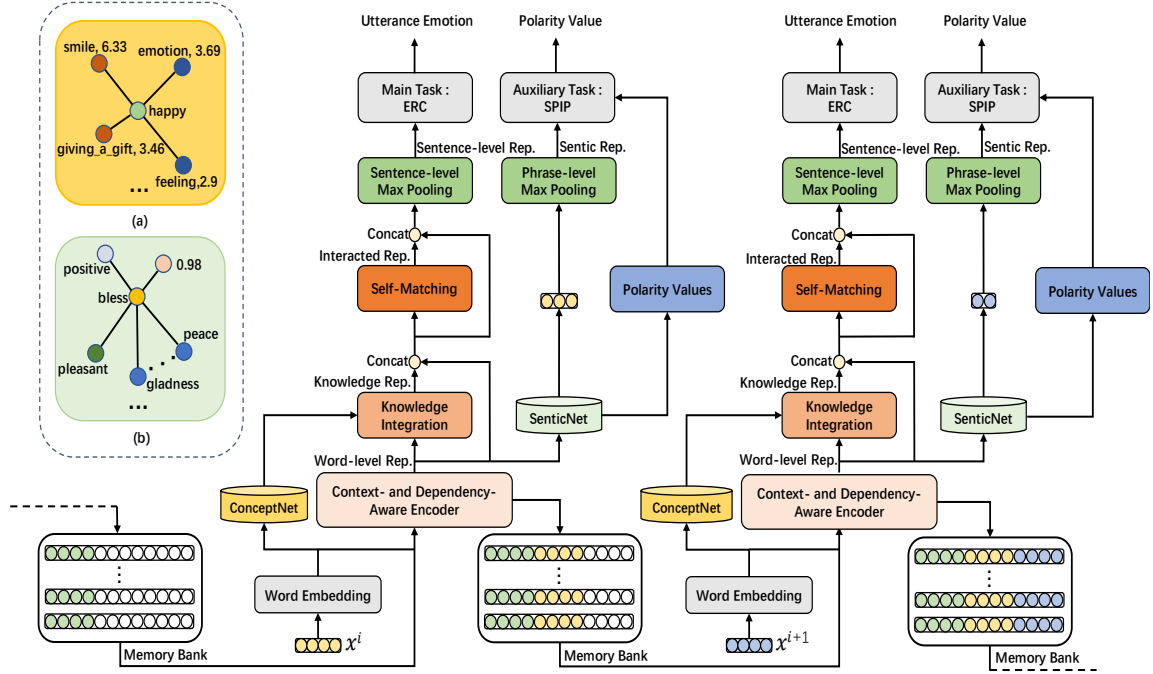


图 2: KI-Net 模型整体结构

对于时间步 i ，我们在当前句子 X^i 之前添加符号 “[CLS]” 来表示句子的开始：

$x^i = \{[CLS], X_1^i, X_2^i, \dots, X_{N_i}^i\}$ 。此后 x^i 通过一个编码层：

$$h_0^i = \text{Embedding}(x^i) \quad (1)$$

其中 $h_0^i \in \mathbb{R}^{N^i \times D_h}$ ，并且 D_h 表示 XLNet-base 的输入维度。编码层的输出 $h_0^i = \{H_0^{i0}, H_0^{i1}, H_0^{i2}, \dots, H_0^{iN_i}\}$ 被当作输入 PLMs 第一层的输入状态。同样地， h_0^i 被当作知识结合模块的 concept 编码层，我们将在 3.3 节进行相关讨论。

传统的 PLMs 没有考虑多方关系，例如说话者内部和说话者之间的关系。为了解决这个问题，我们的编码器引入上下文和对话关系敏感的自注意力机制来代替传统的自注意力机制，形成改进的 Transformer 结构。我们定义以下四种自注意力机制使得模型关注于当前对话的不同方面：

全局自注意力机制 全局自注意力机制将所有的历史上下文和目标句子均纳入考虑窗口中。它与传统自注意力机制类似，目标句子关注整个上下文。这种设置允许我们的模型关注较遥远，但仍然有用的上文句子^[24]。

局部自注意力机制 局部自注意力机制只考虑历史对话中离目标句子最近的 ω 个句子，其中 ω 是一个超参数。使用该注意力机制的目的是，直觉上说话者的情绪受到最近的上文影响的可能性最大。在局部自注意力机制中，我们掩盖上文中超出目标句子 ω 个句子之前的上文。

说话者自注意力机制 说话者自注意力机制只考虑由目标句子的说话者说出的上文中的句子。该机制的目的是通过寻找当前说话者的历史句子中的情绪线索来建模说话者内部的情绪依赖关系^[30]。在说话者自注意力机制中，我们掩盖上文中除目标句子的说话者外的其他参与方的对话历史。

聆听者自注意力机制 聆听者自注意力机制只考虑上文中除目标句子的说话者外的其他说话者的对话历史。该机制的目的是建模对话者间的依赖关系^[30]，因为当前说话者的情绪可能受到其他说话者对话的影响。在聆听者自注意力机制中，我们掩盖目标句子说话者的对话历史句子。

我们通过拼接来结合这四种自注意力机制，并进一步形成上下文和对话关系敏感的自注意力机制的最终输出。得到输入状态 h_0^i 之后，CDA Encoder 采用 L 层改进 Transformer 进行词级别的编码，其中 L 是一个超参数。跟随 DialogXL^[16] 的设置，上一个时间步 $i - 1$ 的上文记忆模块 m^{i-1} 被结合到其中，来完成自注意力机制和 utterance recurrence 的计算。为了方便，我们利用以下的公式表示这个过程：

$$h_L^i = \text{Encoder}(h_0^i, m^{i-1}, L) \quad (2)$$

其中 $h_L^i \in \mathbb{R}^{N^i \times D_h}$ ，并且 $m^{i-1} \in \mathbb{R}^{L \times D_m \times D_h}$ ， D_m 是一个预定义的最大记忆长度，Encoder 表示编码过程。

2.3 常识知识引入及知识增强

我们已经说明，在 ERC 任务的情绪推理过程中，词级别的常识知识可以帮助模型形成增强的语义链条，使得模型更好地完成细粒度情感推理。因此，在本节中我们介绍在 KI-Net 中引入的常识知识源 ConceptNet，引入常识知识的主要方法以及使对话和知识表示更充分交互的自匹配机制。

2.3.1 知识抽取与基于图注意力的表示学习

在这一节中我们简要介绍常识知识来源。我们使用 ConceptNet^[40]作为我们的模型中的知识库。ConceptNet 是一个大型多语言的常识知识图谱，被设计用于使计算机理解人类语言的关系和意义。该知识图谱的知识来源除众包的标注数据外，还结合了 DBPedia¹收集的维基百科相关知识，多语言词典 Wiktionary²和来自 Open Multilingual WordNet³中的词典知识等。整个知识图谱可表示为一个有向图，其中每一个节点表示一个短语级别的概念，每一个边表示一个关系。每一个四元组 $\langle \text{concept1}, \text{relation}, \text{concept2}, \text{weight} \rangle$ 表示一个断言，其中 concept1 和 concept2 表示该断言连接的两个概念节点，relation 表示两个节点间的关系，其中 $\text{relation} \in S$ ， S 表示一个预定义的关系集合。weight 表示一个关于该断言的自信度打分。由于我们的工作均在英语数据集上进行，我们只使用该图谱的英语部分，其中包括 590 万条断言，310 万个概念，且关系集合 S 中包含 38 个关系。

接下来我们首先介绍 ConceptNet 中知识抽取的过程。一般地，对于 token t ，我们抽取一个子图 g_t ，其中包括 t 在 ConceptNet 中的直接邻居。对于每一个 g_t ，我们丢弃是停用词或者不在 CDA Encoder 的词典 V 中的概念，并且除去自信度打分小于 1 的断言，以避免在知识表示中引入噪声。根据以上描述，我们可以将 g_t 形式化为 $g_t = \{(c_1, \omega_1), (c_2, \omega_2), \dots, (c_k, \omega_k)\}$ ，其中 c_i 表示 t 的第 i 个直接邻居概念， ω_k 表示该概念相关的自信度打分。 g_t 的一个例子被展示在图 2 (a) 中。对于相同概念中的词语和相同词语中的子词，我们分配相同的连接概念和自信度打分^[12]。

我们采用一个图注意力机制来形成知识表示。和上文中的知识抽取类似，对于每一个非停用词 token $X_j^i \in x^i$ ，我们有一个概念的图 g_j^i 。对于 X_j^i 和 $c_p \in g_j^i$ ，我们通过公式 (1) 中提到的编码层来获得它们的 token 的编码 h_0^{ij} 和 h_0^{cp} 。我们用如下的方式计算知识表示 k_j^i ：

$$t_p = h_0^{ij} \cdot h_0^{cp} \quad (3)$$

¹ <https://dbpedia.org/>

² https://en.wiktionary.org/wiki/Wiktionary:Main_Page

³ <http://compling.hss.ntu.edu.sg/omw/>

$$\alpha_p = \frac{\exp(t_p * w_p)}{\sum_{p=1}^{N_c^{ij}} \exp(t_p * w_p)} \quad (4)$$

$$k_j^i = \sum_{p=1}^{N_c^{ij}} \alpha_p * h_0^{c_p} \quad (5)$$

其中 $h_0^{ij}, h_0^{c_p}, k_j^i \in \mathbb{R}^{D_h}$, \cdot 表示向量的点积操作, 并且 N_c^{ij} 表示 g_j^i 中的概念的数量。如果 $N_c^{ij} = 0$, 我们将 k_j^i 设置为所有节点表示的平均值^[26]。

2.3.2 知识增强表示的自匹配机制

我们提出一个自匹配机制来完成模型中句子和知识的内部交互。自匹配通过一个基于自注意力的机制来获得知识增强的表示, 使得对话和知识进行进一步的交互。我们已经获得了上下文和对话关系敏感表示 h_L^i (3.2 节) 和句子 x^i 的知识表示 k^i (3.3 节), 其中 k_j^i 是 k^i 的第 j 个输入。利用这两个表示, 我们形成句子 x^i 的知识增强的表示 u^i 。具体地, 对于每个 token x_j^i , 我们用如下方式获得 u_j^i :

$$u_j^i = [h_L^{ij}; k_j^i] \quad (6)$$

其中 $[\cdot]$ 表示拼接操作, $h_L^{ij} \in \mathbb{R}^{D_h}$ 并且 $u_j^i \in \mathbb{R}^{2D_h}$ 。 u_j^i 同时关注上下文信息和对话关系, 并且是知识增强的。自注意力机制可以使得表示之间实现内部交互, 因此我们利用一个基于自注意力机制的自匹配模块。对于同一个句子中的两个 token X_j^i 和 X_m^i , 我们通过一个 trilinear function [35] 计算它们的相似度矩阵:

$$\hat{r}_m^j = W^T [u_j^i; u_m^i; u_j^i \odot u_m^i] \quad (7)$$

其中 $W \in \mathbb{R}^{6D_h}$ 是模型参数, 并且 \odot 表示 element-wise multiplication。类似地, 我们获得相似度矩阵 \hat{R} , 其中 \hat{r}_j^i 是该矩阵的第 ij 个输入。然后我们使用如下的公式获得自注意力矩阵 Q :

$$q_m^j = \frac{\exp(\hat{r}_m^j)}{\sum_m \exp(\hat{r}_m^j)} \quad (8)$$

其中 q_m^j 是 Q 的第 jm 个输入。除此之外, 间接的交互也可以帮助模型学习知识增

强表示的更深层的语义关系。例如关于“Cola”的增强知识在直觉上可以连接“Coca”和“Pepsi”。直觉上，使用间接交互后，所有的 token 对可以通过该句子中其他所有的 token 进行交互。为了实现间接交互，我们使用一个 Q 的 Self-multiplication:

$$\hat{Q} = QQ^T \quad (9)$$

我们获得 Q 和 \hat{Q} 后,我们为每一个 token X_j^i 计算两个经注意力计算之后的向量:

$$v_j^i = \sum_m q_m^j u_m^i \quad (10)$$

$$\hat{v}_j^i = \sum_m \hat{q}_m^j u_m^i \quad (11)$$

其中 $v_j^i, \hat{v}_j^i \in \mathbb{R}^{2D_h}$, q_m^j 是 Q 的第 jm 个输入, 并且 \hat{q}_m^j 是 \hat{Q} 的第 jm 个输入。我们使用不同的方式来拼接这两个经注意力之后的向量, 来帮助模型更充分地对它们进行理解:

$$c_j^i = [u_j^i; v_j^i; u_j^i - v_j^i; u_j^i \odot v_j^i; \hat{v}_j^i; u_j^i - \hat{v}_j^i] \quad (12)$$

其中 $c_j^i \in \mathbb{R}^{12D_h}$, 并且 c_j^i 表示自匹配输出矩阵 C 的第 ij 个输入。 C 结合了对话 token 间的语义和知识交互, 对常识知识推理过程有很大益处。

2.4 引入情感词典知识的多任务学习

情感词典知识无疑非常有用, 因为它提供的知识与 ERC 任务是非常相关的。因此, 我们提出一个短语级别的情感极性强度的预测任务 (SPIP) 来隐式地从情感词典中向模型引入知识, 并利用一个多任务学习的设置来结合 SPIP 任务到主任务 ERC 中。

2.4.1 情感词典 SenticNet 简介及知识抽取

我们采用一个大型情感词典 SenticNet^[41]作为知识来源。SenticNet 是一个概念级的情感知识库, 其主要目标是使机器更容易地理解自然语言中的概念和情绪信息, 使这些信息更加可解释。SenticNet 包含二十万条短语和其相关的情感信息。对于 SenticNet 中的每一个情感短语 s^i , 均包含一个元组 <极性标签, 极性值, 情绪标签, 语义>。极性标签表示该短语的极性积极和消极。极性值是一个在 -1 和 +1

之间的浮点数，用于表示 s^i 的积极性指数（其中-1是极消极，+1是极积极）。情绪标签表示 s^i 的更加细粒度的情感类别。具体地，对于短语 s^i ，它的情绪标签 $m^i \subset M$ ，其中 M 是一个预定义的情绪描述短语的集合。语义 e^i 定义一个与 s^i 语义相关的概念集合。这些元组的一个例子在图2(b)中展示。情绪标签和语义都能够直接指导情绪推理，因此我们将它们结合到在3.3节中抽取的常识知识库。具体地，对于 $s^i \in V$ ，我们构建带权值的情绪标签集合 $\hat{m}^i = \{(m_1^i, w_0), (m_2^i, w_0), \dots, (m_{N_m^i}^i, w_0)\}$ ，其中 $w_0 = 2.0$ ， N_m^i 是 m^i 中情绪标签的个数。类似地，我们构建带权值的语义集合 $\hat{e}^i = \{(e_1^i, \hat{w}_0), (e_2^i, \hat{w}_0), \dots, (e_{N_e^i}^i, \hat{w}_0)\}$ ，其中 $\hat{w}_0 = 1.0$ ， N_e^i 是 e^i 中语义的个数。得到 \hat{m}^i 和 \hat{e}^i 后，我们采用如下的方式构建一个增强的知识图谱 $\hat{g}_{s^i} = g_{s^i} \cup \hat{m}^i \cup \hat{e}^i$ ，其中 \cup 表示集合的并操作。

得到增强的知识图谱 \hat{g} 后，我们对公式(3)作微小的修改。对于 X_j^i 和 $\hat{c}_p \in \hat{g}_j^i$ ，我们通过公式(1)描述的编码层来获得它们的 token 级别的编码 h_0^{ij} 和 $h_0^{\hat{c}_p}$ 。我们修改公式(3)为如下情况：

$$t_p = h_0^{ij} \cdot h_0^{\hat{c}_p} \quad (13)$$

其中 $h_0^{\hat{c}_p} \in \mathbb{R}^{D_h}$ 。然后 t_p 被用于公式(4)的计算，并且3.3节剩余部分不做修改。

2.4.2 情感极性预测任务及多任务学习设置

我们采用情感词典中的 SPIP 任务作为副任务。具体地，该模型为一个句子中所有的 SenticNet 短语预测情感极性值。对于 x^i ，我们抽取一个集合 $E^i = \{e_k^i | e_k^i \in x^i \text{ 中的 } n\text{-gram}\}$ ， $n = 1, 2, \dots, N_g$ ，其中 N_g 是一个超参数。对于 $e_k^i \in \text{SenticNet}$ ，其中 e_j^i 表示 E^i 的第 k 个短语，我们记录它们在句子中的开始和结束位置 $\langle P_0^{ik}, P_1^{ik} \rangle$ 以及相应的极性值 O_k^i 。因此，对于每一个句子 x^i 我们有 $\{\langle P_0^{ik}, P_1^{ik} \rangle, O_k^i\}$ ， $k = 1, \dots, \hat{N}_i$ ，其中 \hat{N}_i 表示该句子 x^i 中 SenticNet 短语的个数。

对于每一个句子 x^i ，我们通过公式2获得它的词级别表示 h_L^i 。对第 k 个 SenticNet 短语 n_k^i ，我们使用一个短语级别的最大池化层来获得它的表示 r_k^i ：

$$\hat{h}_k^i = h_L^i[P_0^{ik}:P_1^{ik}] \quad (14)$$

$$r_k^i = \text{MaxPooling}(\hat{h}_k^i W_0 + b_0) \quad (15)$$

其中 $\hat{h}_k^i \in \mathbb{R}^{|P_1^{ik} - P_0^{ik}| \times D_h}$, $r_k^i \in \mathbb{R}^{D_h}$, $W_0 \in \mathbb{R}^{D_h \times h}$, $b_0 \in \mathbb{R}^h$ 是模型参数, h 表示一个预定义的隐藏层维度, $[\cdot]$ 表示矩阵的切片操作。 *MaxPooling* 表示最大池化操作。然后我们通过以下公式计算最终的预测分数:

$$\hat{O}_k^i = \tanh(r_k^i W_1 + b_1) \quad (16)$$

其中 $W_1 \in \mathbb{R}^{h \times 1}$, $b_1 \in \mathbb{R}^1$ 都是模型参数。作为训练目标, 我们为任务计算标准均方 (MSE) 误差:

$$\text{loss}_a = \frac{1}{N * \hat{N}_i} \sum_{i=1}^N \sum_{k=1}^{\hat{N}_i} (\hat{O}_k^i - O_k^i)^2 \quad (17)$$

对于句子 x^i , 我们已经通过自匹配模块 (3.4 节) 获得了它的词级别表示 c^i , 其中 c^i 表示 C 的第 i 个输入。我们通过最大池化层计算它的句子级别的表示:

$$\hat{c}^i = \text{MaxPooling}(c^i W_2 + b_2) \quad (18)$$

其中 $c^i \in \mathbb{R}^{N^i \times 12D_h}$, $W_2 \in \mathbb{R}^{12D_h \times h}$ 和 $b_2 \in \mathbb{R}^h$ 是模型参数。我们通过如下公式计算最终的概率:

$$\hat{Y}^i = \text{Softmax}(\hat{c}^i W_3 + b_3) \quad (19)$$

其中 $\hat{c}^i \in \mathbb{R}^h$, $W_3 \in \mathbb{R}^{h \times h_c}$ 和 $b_3 \in \mathbb{R}^{h_c}$ 是模型参数。 h_c 表示情绪标签的维度。 *Softmax* 表示 Softmax 操作。我们使用标准交叉熵 (cross entropy) 损失函数计算 ERC 任务的 loss:

$$\text{loss}_m = -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^{h_c} Y_j^i \log \hat{Y}_j^i + (1 - Y_j^i) \log(1 - \hat{Y}_j^i) \quad (20)$$

通过以上的过程, 我们已经获得了主任务 ERC 的 loss_m 和副任务 SPIP 的 loss_a 。我们使用如下的式子计算任务总体的 loss:

$$\text{loss} = \frac{\text{loss}_m + \epsilon \text{loss}_a}{1 + \epsilon} \quad (21)$$

其中 $\epsilon \in [0, 1]$ 是 loss_a 预定义的权重系数。

第 3 章 实验设置及基线介绍

在本节我们介绍主要的实验设置和基线，包括基线数据集和模型的简介，主要实验参数设置，主要实验硬件条件，以及各个数据集上的评测指标等。

3.1 基线数据集简介

我们在如下三个对话情绪识别数据集上验证我们的模型。它们相关的统计数据我们展示在表 1 中。

表 1: 各数据集的相关统计数据

<i>Dataset</i>	<i>Domain</i>	<i>Conv. (Train/Val/Test)</i>	<i>Utter. (Train/Val/Test)</i>
IEMOCAP	Emotional Dialogues	100/20/31	4810/1000/1623
MELD	TV Show Scripts	1038/114/280	9989/1109/2610
DailyDialog	Daily Conversations	11118/1000/1000	87170/8069/7740

IEMOCAP^[19]: 一个多模态的对话数据集，它包含大约 12 小时的视听数据，包括视频、语音、面部运动捕捉、文本转录。视频录制过程每个对话均包含两个说话方，参与者在其中执行即兴表演或脚本场景，着重引出情感表达。情绪标签包含类别 *neutral*, *happiness*, *sadness*, *anger*, *frustrated* 和 *excited*。该语料库是现有数据库的重要补充，有助于研究和建模多模态和表达性的人类交流。该数据集中非中性情绪占比 77%，但缺点是规模较小

MELD^[20]: MELD 是一个由 Emotion Lines 数据集增强得来的多模态的数据集，从电视剧老友记的台词中收集得来。该数据集的标注包含标签 *neutral*, *happiness*, *surprise*, *sadness*, *anger*, *disgust* 和 *fear*。该数据集中非中性情绪占比 53%，优点是数据集质量较高并且有多模态信息，缺点是数据集中的对话涉及到的剧情背景太多，情绪识别难度很大。

DailyDialog^[21]: 人类撰写的高质量多轮对话数据集，对话反映不同主题的日常生活，但无固定说话人。该数据集只包含文本模态，且噪声较小。数据集标注了 7 类情绪: *Neutral*, *Happiness*, *Surprise*, *Sadness*, *Anger*, *Disgust*, *Fear*。除了 7 类情绪标注，还有 10 类主题标注以及 4 类对话行为标注。该数据集中非中性情绪占比

16.8%，数据规模大，但中性情绪占比过高。

3.2 基线模型简介

我们使用我们的模型与如下的模型相比较：

BERT_BASE^[36]：由 BERT-BASE 的预训练后的参数初始化，该模型被精调后用于 ERC 任务。当前对话中的上文句子和目标句子被直接拼接作为输入。BERT 形成词级别的表示后，使用目标句子的开始符"[CLS]"的表示作为整个句子的表示被用于情绪分类任务。

DialogueRNN^[24]：DialogueRNN 使用两个 GRU 来建模说话者状态和对话的整体上文。第三个 GRU 被用于追踪情绪状态。在每一个句子的情绪分类过程中使用一个注意力机制关注不同方面的状态。该模型被认为可以在多方参与的对话中建模说话者间关系，并关注较遥远的重要上文信息。

DialogueGCN^[30]：较早地在 ERC 任务上使用基于图的结构来建模对话中句子关系的工作。该模型使用句子级的表示作为节点，并预定义多种句子间的关系，将其建模为边。这些关系考虑句子间的相对位置和说话人关系。为了学习图的表示，一个 GCN 被用于特征提取。最后，ERC 任务被建模为图上结点的分类任务。

KET^[26]：第一个使用上下文敏感的情感注意力机制来结合结构化常识知识到句子表示中的模型。该结构使得模型在选取 ConceptNet 中的概念时同时考虑上下文和概念的情感词典信息。该模型还首次利用一个层级的 Transformer 来建模句子级的对话信息。

AGHMN^[37]：该模型使用一个层级记忆网络来建模和存储上下文表示。在每一个时间步，该上下文记忆网络经由一个基于注意力机制的 GRU 模块结合进行情绪探测，使得模型可以在每个时间步选择性地关注上文的不同方面。

HiTrans^[15]：基于层级 Transformer 结构，该模型使用多任务学习，设计一个句子-说话者关系预测任务作为副任务，来帮助模型关注说话者信息。这个副任务可以使模型在情绪推理过程中更加关注每一个说话者的情绪变化和对话习惯。

KATIML^[13]：基于一个增量式的 Transformer，该模型使用一个双重级别的图注意力机制来引入常识知识，使得在知识提取过程中同时考虑知识图谱的节点（概念）和边（关系种类）的信息。此外，模型使用多任务学习设置，通过引入一个判断标签是否为中性的副任务来解决标签不均衡问题。

RGAT^[31]: 基于图结构的设计, 该模型利用预训练后的 BERT 编码得到句子级表示, 其中 BERT 可以为模型提供更多的语义信息。此外, 该模型在图结构中为每一个句子节点添加一个相对位置编码, 来向图结构中引入序列信息。

DialogXL^[16]: 当前在 ERC 任务上的最佳模型。DialogXL 使用对话敏感的自注意力机制来向 XLNet 中引入对说话人内部和说话人之间的信息的关注。在每个时间步上 DialogXL 接受一个句子作为输入, 并且使用一个增强的记忆模块来引入上下文信息。

3.3 实验设置及评测指标

我们使用由 Hugging face⁴提供的 XLNet Tokenizer 来预处理所有的数据集, ConceptNet 和 SenticNet, 使得这些信息与 CDA Encoder 保持相同的词表。对于 ConceptNet 和 SenticNet, 不在词表中的概念和信息直接被丢弃。对于超参数设置, $D_h = 768$, $h = 300$, $L = 12$, $N_g = 4$, h_c 和 D_m 取决于数据集。我们使用预训练的 XLNet-base 的参数来初始化 CDA Encoder,并在训练过程中使用 AdamW Optimizer^[38]。所有的实验结果均仅使用文本模态获得。

关于模型评测指标, 对于数据集 IEMOCAP 和 MELD, 我们使用加权 F1 值, 对于 DailyDialog, 我们使用 micro-F1 值。我们报告的实验结果均来自在测试集上的 5 次随机运行。我们所有的实验均在一个拥有 768GB 内存的 Xeon(R) Silver 4110 CPU 和一个拥有 11GB 显存的 Nvidia GeForce GTX 1080 Ti GPU 上完成。

⁴ Hugging Face 官方网站: <https://huggingface.co/>

第 4 章 实验结果与分析

4.1 整体实验结果分析

我们的模型和基线模型的结果列在表 2 和表 3 中。在表 2 中我们将每一个情绪上最好的两个结果加粗，而在表 3 中我们用粗线标出各个数据集上最佳的结果。由“*”标记的数据基于我们的复现结果。根据该结果，在基线模型和最佳模型中，DialogXL, RGAT 和 DialogueGCN 在测试集上的表现优于其它模型。三个模型均使用精心设计的模块建模对话中的依赖关系，例如对话敏感的自注意力机制和基于图的结构。它们的成功从实验上进一步证明定制的对话建模模块对于 ERC 任务是有积极影响的。我们还注意到例如 KET 等模型在基于 Transformer 的结构上表现优于 BERT，主要的可能原因是因为它们词级别的表示中显示引入常识知识。除此之外，HiTrans 和 KAITML 利用多任务学习结合任务相关的信息，也获得了较大提升。这证明显示和隐式引入外部知识均可以指导模型的情绪推理过程。

表 2: KI-Net 模型和基线模型在 IEMOCAP 测试集上各个标签上的加权平均 F1 值

<i>Model</i>	<i>Happy</i>	<i>Sad</i>	<i>Neutral</i>	<i>Angry</i>	<i>Excited</i>	<i>Frustrated</i>	<i>Avg.</i>
BERT_BASE	37.09	59.53	51.73	54.33	54.26	55.83	53.31
DialogueRNN	33.18	78.8	59.21	65.28	71.86	58.91	62.75
DialogueGCN	42.75	84.54	63.54	64.19	63.08	66.99	64.18
KET	--	--	--	--	--	--	59.56
AGHMN	52.1	73.3	58.4	61.9	69.7	62.3	63.5
HiTrans	--	--	--	--	--	--	64.50
KAITML	--	--	--	--	--	--	61.43
RGAT	51.62	77.32	65.42	63.01	67.95	61.23	65.22
DialogXL	44.06	77.10	64.67	61.59	69.73	66.98	65.88*
KI-Net (Ours)	49.45	73.38	65.63	65.13	71.15	68.38	66.98

我们的 KI-Net 的试验结果表明，该模型在三个数据集上均取得了新的最佳的结果。KI-Net 在 DailyDialog 数据集上取得了 57.30% 的结果，相比 DialogXL 取得

了约 2.5%左右的提升。可能的原因是我们同时采用采用常识知识和情感词典，从两个不同维度来弥补 PLMs 缺失的知识维度。在 MELD 数据集上，KI-Net 取得了 63.24%的结果，相比同样利用常识知识的 KET 有约 5%左右的提升。我们相信原因之一是我们的自匹配模块使得知识增强表示之间形成内部交互，在 MELD 数据集对话短小和需要大量常识推理的情况下，相比于 KET 等直接的拼接操作，使模型更充分地理解常识知识和对话的关系。对于 IEMOCAP 数据集，我们的模型取得 66.98%的结果，相比同样采用多任务学习的 KAITML 等模型取得约 5.5%的提升。我们相信我们的优势之一是 KI-Net 的 SPIP 任务向 ERC 任务中直接引入情绪相关的知识（关键概念的情感分数），这些信息可以直接指导模型确定句子情绪。但是 KAITML 等模型并未在多任务学习中采用直接引入情绪信息的副任务，因此模型需要更复杂的推理才能理解这些知识和情绪推理之间的关系。

表 3: KI-Net 模型在 MELD 和 DailyDialog 测试集上与基线模型的表现比较

<i>Model</i>	<i>MELD</i>	<i>DailyDialog</i>
BERT_BASE	56.21	53.12
DialogueRNN	57.03	50.65
DialogueGCN	58.10	--
KET	58.18	53.37
AGHMN	58.1	--
HiTrans	61.94	--
KAITML	58.97	54.71
RGAT	60.91	54.31
DialogXL	62.41	54.93
KI-Net (Ours)	63.24	57.30

我们在表 2 中展示 KI-Net 和各基线模型在 IEMOCAP 数据集上的具体情绪的测试结果。我们将每一个情绪上最好的两个结果进行加粗。根据测试结果，KI-Net 在大部分情绪上获得了 Top-2 的结果。为了获得总体均衡的和最佳的整体表现，在部分情绪上有可以容忍的下降。具体地，在情绪 Frustrated 和 Neutral 上，我们的

模型取得了最好的结果 65.63%和 68.38%。我们相信在 Neutral 上取得了最佳结果的原因是情感词典为情感极性的确定提供了直接的指导，使得模型可以较为容易地确定句子情感是否为中性。除此之外，常识知识为形成在情绪推理中很关键的增强语义链条提供重要信息，有利于细粒度与更加复杂的情绪的判断，例如 Frustrated 和 Excited。然而，在本应比其它情绪所需的推理过程更简单的情绪 Happy 和 Sad 上，我们的模型仅取得了 49.45%和 73.38%的结果。我们分析可能的原因是知识引入导致模型在这些情绪上的过度推理，导致情绪被误分为更加细粒度的同类情感。



图 3：两个来自 IEMOCAP 数据集测试结果中的典型例子

4.2 测试结果中的典例和误差分析

我们提供两个具体例子来研究 SPIP 任务和 ConceptNet 中的常识知识对模型表现的影响，相关的例子在图 3 中，其中 Golden, CDA.和 W/O SPIP 表示正确的情绪标签，CDA Encoder 的预测，和 KI-Net 除去 SPIP 任务的预测。和 W/O SPIP, KI-Net 相连的部分表示关键 token 的 8 个 attention 值最高的常识概念和情感极性强度的预测结果。这些典例来自于 IEMOCAP 数据集的实际测试过程。在 case #1 中，引入“miss”，“husband”和“wedded”等常识概念后。模型可以更加深入地理解“married”的语义。与此同时，SPIP 分类器给短语“getting_married”分配了相对很强的积极性，因此考虑到短语前的“not”，模型可以更加关注到句子中强烈的消

极性。CDA Encoder 的预测结果进一步表明在这个例子中，上下文和语义信息并不是关键的部分。

在 case #2 中，模型没有通过常识知识如“earned”等获得直接的情绪相关的信息。因此，在这个例子中知识引入模块并没有发挥较大的作用，导致 CDA Encoder 和 W/O SPIP 均进行了错误的分类“Frustrated”。然而，SPIP 分类器为词语“cheap”分配了较强的消极性，模型得以为该句子打上“Angry”的标签，这也是一个消极的情感，但是明显比“Frustrated”拥有更强的消极性。例子分析表明，这两种知识源均可在不同的情况下为情绪推理提供关键线索，因此两种知识的引入都是必要的。

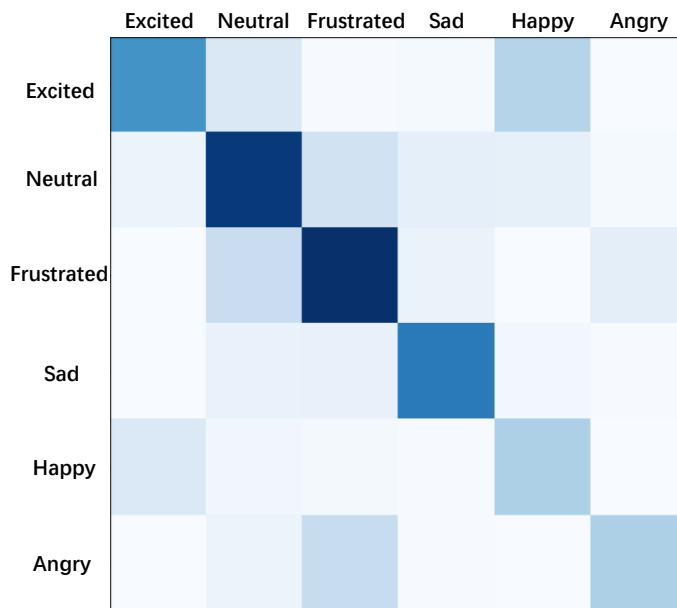


图 4: KI-Net 模型在 IEMOCAP 测试集上的混淆矩阵

此外，我们在图 4 中展示 KI-Net 在 IEMOCAP 数据集上的测试结果的困惑矩阵，并将其用于误差分析。根据该结果，尽管 KI-Net 在“Neutral”情感上已经取得了最佳的结果，很大一部分误判仍来自“Neutral”和其他情感之间。这表明除了 SPIP 的情感词典提供的情感信息外，部分中性情感还需要其它的信息进行确定。此外，其余的误判主要都来自同一极性的情感，比如 Happy 和 Excited 之间的误判，并且很少有跨情感极性的误判（例如 Excited 和 Angry）。这表明模型已经具备了很强区分积极和消极情感的能力。然而，这同时也表明我们的模型仍然缺乏区别相似情感的能力，尽管我们相比于基线模型已经取得了进步。可能的原因是情感词典会给表达相近情感的词语分配相近的分数。使模型具有分辨相近情绪的能力是

使模型获得进一步提升的必要步骤。

我们已经讨论了 KI-Net 模型在测试中仍然缺乏区别相似情绪的能力的问题。正如上文中所分析的,我们在 SPIP 任务中仅使用了每个概念的一维的情感极性值,而这些值缺乏更加细粒度的情感信息。为了解决这个问题,我们注意到在 SenticNet 中,每个短语还被赋予一个四维的语义值。每个语义值包含一个向量<introspection, temper, attitude, sensitivity>, 其中每个维度代表该相关方面的强度。我们相信在 SPIP 任务相关的未来工作中,通过多任务学习的方式向模型中引入这些语义值可以使其更加关注情感中细粒度的差别,并帮助模型区分相似的情感。

除 SenticNet 之外,我们还考虑引入其他包含细粒度情感信息的情感词典知识源。其中一个 SentiWordNet^[52]。该情感词典考虑到每一个概念在不同的上下文中拥有不同的词性,因此会表达不同程度的情感强度。SentiWordNet 为特定短语根据其不同的词性和语境分配不同的情感极性值。此外,我们已经注意到一些在 PLMs 中加入 SentiWordNet 中的知识的工作^[48],并在情感分析任务中获得了较好的结果。

我们还注意到近期在情感相关任务中通过不同方式结合情感词典知识的工作。例如, Margatina 等人^[53]直接构建情感词典特征向量,并提出了三种在对话表示中显示结合情感词典的方法。KET^[26]也在上下文敏感的情感注意力机制中结合情感词典知识来选择常识概念。这些工作为引入情感词典知识提供了不同于多任务学习的新的思路。

4.3 关键模块的销蚀实验

我们为 KI-Net 中的关键模块设计销蚀实验。我们通过移除一些关键组成部分自匹配模块,常识知识引入模块和 SPIP 任务来分析它们对模型表现的影响。KI-Net 在三个测试数据集上的结果如表 4 所示。其中“-Self-matching”表示去除自匹配模块但保留知识引入模块,令句子表示部分和知识表示部分直接拼接。“-Knowledge integration”表示常识知识引入模块被全部移除(包括自匹配)。对于“-SPIP”,我们丢弃 SPIP 任务和多任务设置。“CDA Encoder”则代表 CDA Encoder 直接预测的结果。

我们可以观察到每一个部分被移除时我们的模型在三个测试数据集上的表现均有所下降。对于 IEMOCAP 和 MELD 这两个数据集,当自匹配模块被移除时,测试结果都有相当大的下降,甚至比不包含任何知识引入模块的表现更差。这个

结果表明自匹配对于模型通过充分的词级别交互理解常识知识非常关键，没有这种交互，模型便无法充分理解引入的常识知识和词表示的关系，引入的知识甚至会成为推理过程中的噪音。

表 4: KI-Net 在三个测试数据集上的销蚀实验结果

<i>Method</i>	<i>IEMOCAP</i>	<i>MELD</i>	<i>DailyDialog</i>
KI-Net	66.98	63.24	57.3
-SPIP	66.39(-0.59)	62.89(-0.35)	57.07(-0.23)
-Self-matching	64.89(-2.09)	62.38(-0.86)	55.35(-1.95)
-Knowledge integration	66.58(-0.40)	62.72(-0.52)	56.52(-0.78)
CDA Encoder	65.94(-1.04)	62.41(-0.83)	54.82(-2.48)

我们对其中更加典型的 IEMOCAP 和 MELD 数据集上的结果做进一步的分析。对于 IEMOCAP 数据集，移除 SPIP 副任务后模型表现下降更多。可能的原因是 IEMOCAP 数据集的每个对话包含更多轮数的句子（每段对话包含平均 70 个左右的句子），提供了非常丰富的上下文信息，需要模型通过更复杂的上下文推理完成情绪的确定。在这种情况下，模型会更多得益于为情感探测提供直接情绪相关信息的 SPIP 任务。此外，该数据集中 70% 的句子中都拥有至少一个 SenticNet 中的短语，且平均每个句子中包含 2.2 个短语。因此，SPIP 也能为 IEMOCAP 数据集提供更多的情感相关信息。

对于 MELD 数据集，根据测试结果，ConceptNet 中引入的常识知识对该任务有更多有益的影响。可能的原因是 MELD 数据集中的每一个对话拥有更少轮次的句子（平均每个对话包含 7 个左右的句子），因此常识知识引入可以更加丰富的语义信息。而对于 SPIP 任务，MELD 数据集中仅有 65% 的句子包含至少一个 SenticNet 短语并且平均每个句子仅包含 1.9 个短语。这表明相比于 IEMOCAP 数据集，SenticNet 只能向 MELD 数据集提供更少的情绪相关信息。

4.4 常识知识的关键维度

尽管 ConceptNet 已经被证明在补充 KI-Net 的缺失知识维度方面是成功的，它的知识补充过程却不一定是最有效率的。根据 Ilievski 等人^[11]的研究，预定义的知识

识维度“quality”和“temporal”在基于 PLMs 的常识知识驱动任务（例如常识知识问答）的测试结果中，对模型表现帮助最大。我们在表 5 中展示在四个常用的常识知识源上不同知识维度的分布的部分数据，其中每个知识维度包含相对数量最多的数据被加粗标出。

表 5：四个常用知识源上一些预定义的知识维度的相对数量分布

<i>Dimension</i>	<i>ConceptNet</i>	<i>ATOMIC</i>	<i>WebChild</i>	<i>Wikidata-CS</i>
desire/goal	20	244	--	--
quality	9	143	6510	1
comparative	--	--	813	--
temporal	71	346	2135	3
similarity	255	--	343	1
lexical	704	--	--	0.5
taxonomic	244	--	783	73
part-whole	19	--	5752	8
spatial	28	--	660	0.5
relation-other	1969	--	291	6

根据表中的结果，ConceptNet 在关键知识维度“quality”和“temporal”上相对于 ATOMIC 和 WebChild 包含较少的知识，因此如 ATOMIC 和 WebChild 等知识源在理论上可以更加准确的弥补 PLMs 中缺失的知识维度。在未来工作中，除 ConceptNet 外，我们还将考虑引入这些在关键维度上包含更多知识的常识知识源。

结 论

在本文中我们提出了一个基于常识知识引入和情感极性强度敏感的多任务学习的模型，简称为 **KI-Net**，用于对话情绪识别任务（**ERC**），即识别对话中每个句子包含的情感。首先，我们提出一个常识知识引入模块来结合 **ConceptNet** 中词级别的结构化常识概念知识，并在这个过程中首次利用一个自匹配模块来帮助句子和知识表示进行充分交互，以解决以往知识引入相关工作直接拼接得到知识增强表示，而忽略了其内部推理交互过程的问题。此外，为了解决当前多任务学习相关的 **ERC** 工作没有引入情绪直接相关信息的问题，我们在 **ERC** 任务中通过设计多任务学习的副任务 **SPIP**，引入来自情感词典 **SenticNet** 的知识。**SPIP** 任务通过指导模型预测关键短语在 **SenticNet** 中的情感极性值的方式，为 **ERC** 任务提供直接情绪相关的信息，这种信息无需复杂的推理即可为情绪推理过程提供直接指导。最后，我们通过严密的实验分析 **KI-Net** 各模块的有效性。整体实验结果表明 **KI-Net** 在三个基线数据上的表现优于其它基于知识引入，预训练模型和多任务学习的 **ERC** 工作，且各具体情绪上的结果表明 **KI-Net** 模型可以在各个情绪上获得整体平衡的优秀表现，但在部分情绪上可能存在过度推理的问题。测试结果中的典例分析和关键模块的销蚀实验还证明，常识知识引入和 **SPIP** 任务均对模型的情绪推理从不同方面有积极影响，且自匹配模块对于模型更充分地理解对话和知识的关系非常重要。销蚀实验结果分析还表明，在不同的推理复杂度，背景知识量等情况下，常识知识和情感词典知识对模型整体效果的贡献可以互补。

误差分析表明，我们的模型在区分相似情感方面能力仍然欠缺。可能的原因是 **SPIP** 任务为相似的情感分配相近的情感极性值。因此，接下来我们的研究重点关注通过引入更加细粒度的情感相关知识来提升模型区分相似情绪的能力。此外，新的调研表明，相比于 **ConceptNet**，存在更多在预训练语言模型缺失的关键维度上知识更丰富的常识知识源。因此我们将进一步探索对 **ERC** 任务更加关键的知识维度，并尝试利用在这些知识维度上更加丰富的知识源来补充缺失的知识维度。

参考文献

- [1] Soujanya Poria, Navonil Majumder, Rada Mihalcea, and Eduard H. Hovy. 2019b. Emotion recognition in conversation: Research challenges, datasets, and recent advances. CoRR, abs/1905.02947.
- [2] Runnan Li, Zhiyong Wu, Jia Jia, Yaohua Bu, Sheng Zhao, and Helen Meng. 2019. Towards discriminative representation learning for speech emotion recognition. In Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19, pages 5060–5066. International Joint Conferences on Artificial Intelligence Organization.
- [3] A. Chatterjee, Umang Gupta, Manoj Kumar Chinnakotla, R. Srikanth, Michel Galley, and P. Agrawal. 2019. Understanding emotions in text using deep learning and big data. Comput. Hum. Behav., 93:309–317.
- [4] Sayan Ghosh, Mathieu Chollet, Eugene Laksana, Louis-Philippe Morency, and Stefan Scherer. 2017. Affect-LM: A neural language model for customizable affective text generation. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 634–642, Vancouver, Canada. Association for Computational Linguistics.
- [5] Hao Zhou, Minlie Huang, Tianyang Zhang, Xiaoyan Zhu, and Bing Liu. 2017. Emotional chatting machine: Emotional conversation generation with internal and external memory. CoRR, abs/1704.01074.
- [6] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.
- [7] Zhilin Yang, Zihang Dai, Yiming Yang, Jaime Carbonell, Russ R Salakhutdinov, and Quoc V Le. 2019b. Xlnet: Generalized autoregressive pretraining for language understanding. In Advances in Neural Information Processing Systems, volume 32. Curran Associates, Inc.
- [8] Matthew Henderson, Iñigo Casanueva, Nikola Mrkšić, Pei-Hao Su, Tsung-Hsien

Wen, and Ivan Vulic. 2020. ConveRT: Efficient and accurate conversational representations from transformers. In Findings of the Association for Computational Linguistics: EMNLP 2020, pages 2161–2174, Online. Association for Computational Linguistics

[9] Siqu Bao, Huang He, Fan Wang, Hua Wu, and Haifeng Wang. 2020. PLATO: Pre-trained dialogue generation model with discrete latent variable. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pages 85–96, Online. Association for Computational Linguistics.

[10] Jeremy Howard and Sebastian Ruder. 2018. Universal language model fine-tuning for text classification. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 328–339, Melbourne, Australia. Association for Computational Linguistics.

[11] Filip Ilievski, Alessandro Oltramari, Kaixin Ma, Bin Zhang, Deborah L. McGuinness, and Pedro Szekely. 2021. Dimensions of commonsense knowledge

[12] An Yang, Quan Wang, Jing Liu, Kai Liu, Yajuan Lyu, Hua Wu, Qiaoqiao She, and Sujian Li. 2019a. Enhancing pre-trained language representations with rich knowledge for machine reading comprehension. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pages 2346–2357, Florence, Italy. Association for Computational Linguistics.

[13] Duzhen Zhang, Xiuyi Chen, Shuang Xu, and Bo Xu. 2020. Knowledge aware emotion recognition in textual conversations via multi-task incremental transformer. In Proceedings of the 28th International Conference on Computational Linguistics, pages 4429–4440, Barcelona, Spain (Online). International Committee on Computational Linguistics.

[14] Jiancheng Wang, Jingjing Wang, Changlong Sun, Shoushan Li, Xiaozhong Liu, Luo Si, Min Zhang, and Guodong Zhou. 2020. Sentiment classification in customer service dialogue with topic-aware multi-task learning. Proceedings of the AAAI Conference on Artificial Intelligence, 34(05):9177–9184.

[15] Jingye Li, Donghong Ji, Fei Li, Meishan Zhang, and Yijiang Liu. 2020. HiTrans: A transformer-based context- and speaker-sensitive model for emotion detection in conversations. In Proceedings of the 28th International Conference on Computational Linguistics, pages 4190–4200, Barcelona, Spain (Online). International Committee on

Computational Linguistics.

[16] Weizhou Shen, Junqing Chen, Xiaojun Quan, and Zhixian Xie. 2020. Dialogxl: All-in-one xlnet for multi-party conversation emotion recognition

[17] N. Colneric and J. Demšar. 2020. Emotion recognition on twitter: Comparative study and training a unison model. *IEEE Transactions on Affective Computing*, 11(3):433–446.

[18] Laurence Devillers and Laurence Vidrascu. 2006. Real-life emotions detection with lexical and paralinguistic cues on human-human call center dialogs.

[19] Carlos Busso, Murtaza Bulut, Chi-Chun Lee, Abe Kazemzadeh, Emily Mower Provost, Samuel Kim, Jeannette Chang, Sungbok Lee, and Shrikanth Narayanan. 2008. Iemocap: Interactive emotional dyadic motion capture database. *Language Resources and Evaluation*, 42:335–359.

[20] Soujanya Poria, Devamanyu Hazarika, Navonil Majumder, Gautam Naik, Erik Cambria, and Rada Mihalcea. 2019a. Meld: A multimodal multi-party dataset for emotion recognition in conversations.

[21] Yanran Li, Hui Su, Xiaoyu Shen, Wenjie Li, Ziqiang Cao, and Shuzi Niu. 2017. DailyDialog: A manually labelled multi-turn dialogue dataset. In *Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 986–995, Taipei, Taiwan. Asian Federation of Natural Language Processing.

[22] Devamanyu Hazarika, Soujanya Poria, Amir Zadeh, Erik Cambria, Louis-Philippe Morency, and Roger Zimmermann. 2018b. Conversational memory network for emotion recognition in dyadic dialogue videos. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 2122–2132, New Orleans, Louisiana. Association for Computational Linguistics.

[23] Devamanyu Hazarika, Soujanya Poria, Rada Mihalcea, Erik Cambria, and Roger Zimmermann. 2018a. Icon: Interactive conversational memory network for multimodal emotion detection. pages 2594–2604.

[24] Navonil Majumder, Soujanya Poria, Devamanyu Hazarika, Rada Mihalcea, Alexander Gelbukh, and Erik Cambria. 2019. Dialoguerrn: An attentive rnn for emotion

detection in conversations. *Proceedings of the AAAI Conference on Artificial Intelligence*, 33(01):6818–6825.

[25] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need.

[26] Peixiang Zhong, Di Wang, and Chunyan Miao. 2019. Knowledge-enriched transformer for emotion detection in textual conversations. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 165–176, Hong Kong, China. Association for Computational Linguistics.

[27] Duzhen Zhang, Xiuyi Chen, Shuang Xu, and Bo Xu. 2020. Knowledge aware emotion recognition in textual conversations via multi-task incremental transformer. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 4429–4440, Barcelona, Spain (Online). International Committee on Computational Linguistics.

[28] Wenxiang Jiao, Michael R. Lyu, and Irwin King. 2019b. Real-time emotion recognition via attention gated hierarchical memory network.

[29] S. Xing, S. Mai, and H. Hu. 2020. Adapted dynamic memory network for emotion recognition in conversation. *IEEE Transactions on Affective Computing*, pages 1–1.

[30] Deepanway Ghosal, Navonil Majumder, Soujanya Poria, Niyati Chhaya, and Alexander Gelbukh. 2019. DialogueGCN: A graph convolutional neural network for emotion recognition in conversation. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 154–164, Hong Kong, China. Association for Computational Linguistics.

[31] Taichi Ishiwatari, Yuki Yasuda, Taro Miyazaki, and Jun Goto. 2020. Relation-aware graph attention networks with relational position encodings for emotion recognition in conversations. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 7360–7370, Online. Association for Computational Linguistics.

[32] Jiancheng Wang, Jingjing Wang, Changlong Sun, Shoushan Li, Xiaozhong Liu, Luo Si, Min Zhang, and Guodong Zhou. 2020. Sentiment classification in customer service

dialogue with topic-aware multi-task learning. Proceedings of the AAAI Conference on Artificial Intelligence, 34(05):9177–9184.

[33]Jiao, Wenxiang, Michael R. Lyu and I. King. PT-CoDE: Pre-trained Context-Dependent Encoder for Utterance-level Emotion Recognition. ArXiv abs/1910.08916 (2019): n. pag.

[34]Chapuis, Emile, Pierre Colombo, Matteo Manica, Matthieu Labeau and C. Clavel. Hierarchical Pre-training for Sequence Labelling in Spoken Dialog. ArXiv abs/2009.11152 (2020): n. pag.

[35]Seo, Minjoon, Aniruddha Kembhavi, Ali Farhadi and Hannaneh Hajishirzi. Bidirectional Attention Flow for Machine Comprehension. ArXiv abs/1611.01603 (2017): n. pag.

[36]Devlin, J., Ming-Wei Chang, Kenton Lee and Kristina Toutanova. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. NAACL-HLT (2019).

[37]Jiao, Wenxiang, M. R. Lyu and I. King. “Real-Time Emotion Recognition via Attention Gated Hierarchical Memory Network.” ArXiv abs/1911.09075 (2020): n. pag.

[38]Ilya Loshchilov and Frank Hutter. 2017. Fixing weight decay regularization in adam. ArXiv, abs/1711.05101.

[39] Vered Shwartz, Peter West, Ronan Le Bras, Chandra Bhagavatula, and Yejin Choi. 2020. Unsupervised commonsense question answering with self talk. ArXiv, abs/2004.0548

[40] Robyn Speer, Joshua Chin, and Catherine Havasi. 2017. Conceptnet 5.5: An open multilingual graph of general knowledge. In Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, AAAI’17, page 4444–4451. AAAI Press.

[41] Erik Cambria, Yang Li, Frank Z. Xing, Soujanya Poria, and Kenneth Kwok. 2020. Senticnet 6: Ensemble application of symbolic and subsymbolic ai for sentiment analysis. In Proceedings of the 29th ACM International Conference on Information & Knowledge Management, CIKM ’20, page 105–114, New York, NY, USA. Association for Computing Machinery.

[42] Florian Brachten, Felix Brünker, Nicholas R. J. Frick, Björn Ross, and Stefan

Stieglitz. 2020. On the ability of virtual agents to decrease cognitive load: an experimental study. *Information Systems and e-Business Management*, pages 1–21

[43] John Sweller. 1988. Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2):257–285.

[44] Teng, Zhiyang, Duy-Tin Vo and Yue Zhang. “Context-Sensitive Lexicon Features for Neural Sentiment Analysis.” *EMNLP* (2016).

[45] Qian, Qiao, Minlie Huang, Jinhao Lei and Xiaoyan Zhu. Linguistically Regularized LSTM for Sentiment Classification. *ACL* (2017).

[46] Tang, Duyu, Furu Wei, Nan Yang, M. Zhou, T. Liu and Bing Qin. Learning Sentiment-Specific Word Embedding for Twitter Sentiment Classification. *ACL* (2014).

[47] Tian, Hao, Can Gao, X. Xiao, Hao Liu, Bolei He, Hua Wu, H. Wang and Feng Wu. SKEP: Sentiment Knowledge Enhanced Pre-training for Sentiment Analysis. *ACL* (2020).

[48] Ke, Pei, Haozhe Ji, Siyang Liu, X. Zhu and Minlie Huang. SentiLARE: Sentiment-Aware Language Representation Learning with Linguistic Knowledge. *EMNLP* (2020).

[49] Dragoni, M. and Giulio Petrucci. A fuzzy-based strategy for multi-domain sentiment analysis. *Int. J. Approx. Reason.* 93 (2018): 59-73.

[50] Ghosal, Deepanway, Devamanyu Hazarika, Navonil Majumder, A. Roy, Soujanya Poria and Rada Mihalcea. KinGDOM: Knowledge-Guided DOMain adaptation for sentiment analysis. *ACL* (2020).

[51] Zhang, B., M. Yang, Xutao Li, Yunming Ye, Xiaofei Xu and Kuai Dai. Enhancing Cross-target Stance Detection with Transferable Semantic-Emotion Knowledge. *ACL* (2020).

[52] Stefano Baccianella, Andrea Esuli, and Fabrizio Sebastiani. 2010. SentiWordNet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC’10)*, Valletta, Malta. European Language Resources Association (ELRA).

- [53] Katerina Margatina, Christos Baziotis, and Alexandros Potamianos. 2019. Attention-based conditioning for external knowledge integration. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pages 3944–3951, Florence, Italy. Association for Computational Linguistics.
- [54] Tom Young, Erik Cambria, Iti Chaturvedi, Hao Zhou, Subham Biswas, and Minlie Huang. 2018. Augmenting end-to-end dialogue systems with commonsense knowledge. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 32.
- [55] Thomas Kipf and M. Welling. 2017. Semi-supervised classification with graph convolutional networks. ArXiv, abs/1609.02907.
- [56] Dongming Sheng, Dong Wang, Ying Shen, Haitao Zheng, and Haozhuang Liu. 2020. Summarize before aggregate: A global-to-local heterogeneous graph inference network for conversational emotion recognition. In Proceedings of the 28th International Conference on Computational Linguistics, pages 4153–4163, Barcelona, Spain (Online). International Committee on Computational Linguistics.
- [57] Yunlong Liang, Fandong Meng, Y. Zhang, Jin'an Xu, Yufeng Chen, and J. Zhou. 2020. Infusing multi-source knowledge with heterogeneous graph neural network for emotional conversation generation. ArXiv, abs/2012.04882.
- [58] Devamanyu Hazarika, Soujanya Poria, Roger Zimmermann, and Rada Mihalcea. 2021. Conversational transfer learning for emotion recognition. *Information Fusion*, 65:1–12
- [59] Dong Zhang, Liangqing Wu, Changlong Sun, Shoushan Li, Qiaoming Zhu, and Guodong Zhou. 2019. Modeling both context- and speaker-sensitive dependence for emotion detection in multi-speaker conversations. In Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19, pages 5415–5421. International Joint Conferences on Artificial Intelligence Organization.

哈尔滨工业大学本科毕业设计（论文）原创性声明

本人郑重声明：在哈尔滨工业大学攻读学士学位期间，所提交的毕业设计（论文）《引入常识和情感词典知识的对话情绪识别》，是本人在导师指导下独立进行研究工作所取得的成果。对本文的研究工作做出重要贡献的个人和集体，均已在文中以明确方式注明，其它未注明部分不包含他人已发表或撰写过的研究成果，不存在购买、由他人代写、剽窃和伪造数据等作假行为。

本人愿为此声明承担法律责任。

作者签名：

日期： 年 月 日

致 谢

衷心感谢导师刘秉权教授自开题以来在各方面对本人的精心指导，他的言传身教将使我终生受益。感谢孙承杰教授在实验严谨性和论文写作规范性方面对我提出的宝贵意见。此外要感谢解云鹤师兄在实验和论文写作上对我的建议，以及智能技术与自然语言处理（ITNLP）实验室其他老师和同窗们的热情帮助和支持！感谢 ITNLP 实验室为我提供强大的算力支持，以及在计算服务器出现故障时为我提供的宝贵技术支持，使我能够快速地实践我的新想法，这对于我顺利完成毕业论文是至关重要的。

最后，我真诚地感谢并将本文献给我的母亲，父亲和其它关心我的亲人。在我四年本科生涯中，当我遭遇困难或研究进展不顺利时，他们都毫无保留地支持我，是我走出低谷并取得今天的成果的最大精神支柱。