

一种融入非结构化背景知识的交互文本立场分析方法

刘常健¹ 杜嘉晨¹ 冷佳¹ 陈荻¹ 徐睿峰^{1,†}

1. 哈尔滨工业大学（深圳），深圳 518055；

† 通讯作者，E-mail: xuruifeng@hit.edu.cn

摘要 传统交互文本立场分析方法由于缺乏背景知识的支持，仅局限于所建模交互文本的语境，因而使得系统的性能受到限制。本文针对这一问题提出了一种融入背景知识的交互文本立场分析方法，该方法以交互文本作为查询，从维基百科中检索相关的背景知识，对背景知识进行编码，提取与当前交互文本相关的背景知识特征，以增强交互文本的特征表示。实验结果证明，选取适当的背景知识嵌入层数以及背景知识嵌入层连接方式，结合背景知识的交互文本立场分析可以取得更进一步的性能表现，证明了本文方法的有效性。

关键词 立场分析；交互文本；背景知识；深度记忆网络

中图分类号 TN914

An Interactive Stance Classification Method that Incorporate Unstructured Background Knowledge

LIU Changjian¹, DU Jiachen¹, LENG Jia¹, CHEN Di¹, XU Ruifeng^{1,†}

1. Department of Computer Science and Technology, Harbin Institute of Technology (Shenzhen), Shenzhen 518055;

†Corresponding Author, E-mail: xuruifeng@hit.edu.cn

Abstract Traditional interactive stance classification methods reveal a constrained performance, due to the limited context provided by the modeled interactive text and the insufficient support of background knowledge. In this paper, we proposed an interactive stance classification method that incorporate unstructured background knowledge, which retrieves relevant knowledge sentences from Wikipedia with the interactive text as query, encodes them as external memories, and utilizes the encoded memories to enhance the representation of the interactive text with a multi-layer deep memory network. Experiments show that stance classification system can achieve better performance when background knowledge is incorporated with a proper setting of knowledge embedding layers, thus proving the effectiveness of our proposed method.

Key words Stance Classification; Interactive Text; Background Knowledge; Deep Memory Network

文本立场分析作为绑定话题目标对象的细粒度情感分析，可将立场分类为三种类型（支持/反对/中立），或两种类型（支持/反对）。在此基础上，交互文本立场分析所强调的是文本间的交互性，所研究的对象是Q-R对形式的交互文本，其中Q为引述（Quote），R为应答（Response）。该设定使得上下文语境，模型需要联合建模引述和应答以及两者间的交互语境，进而判断应答的立场。目前该项任务相关研究所涉及的主要的数据集有IAC^[1]，DP^[2]和ABCD^[3]三个英文数据集。

现阶段，基于机器学习的文本立场分析相关技术仍以特征构造和筛选为主，并结合弱监督方法进行语

国家自然科学基金(U1636103, 61632011, 61876053)，深圳市基础研究项目(JCYJ20180507183527919, JCYJ20180507183608379)，深圳市技术攻关项目(JSGG20170817140856618)，深圳证券信息联合研究计划资助

收稿日期: 2019-05-22; 修回日期: 2019-07-31; 网络出版时间:

网络出版地址:

料扩充，提升模型的性能表现。另一方面，基于深度学习的文本立场分析相关技术，以循环神经网络为代表，对文本进行序列建模，通过联合话题目标对象，学习与任务更相关的表示。

表 1 中展示了一个来自 DP 数据集的交互文本立场分析样本实例。该实例是肯尼迪与尼克松关于古巴卡斯特罗政权局势的讨论中的节选文段。在该实例中，肯尼迪表达了美国对古巴具有影响力，以及表示了对古巴人民自由选举权的支持，而尼克松则顺水推舟表明美国所具有的强大军事力量足以将古巴卡斯特罗政权赶下台，也揭示了肯尼迪的军事企图。两者都表示了对古巴卡斯特罗政权的否定，但是尼克松从侧面表达了对肯尼迪的反对态度。因此，对于该实例，正确的交互文本立场应为反对。

表 1 交互文本立场分类数据实例
Table 1 Data instance of interactive stance classificaion

引述 (Quote)	肯尼迪：“毫无疑问，我们在古巴有很大的影响力，我认为很不幸，我们没有更积极地利用这种影响力说服卡斯特罗举行自由、公开的选举，这样古巴人民就可以做出选择。”
应答 (Response)	尼克松：“我们必须记住的是，美国拥有军事力量——卡斯特罗先生也知道这一点——明天、第二天或我们选择的任何一天都要把他赶下台。”
背景知识 (Background Knowledge)	艾森豪威尔总统的新政策强调使用核武器来威慑苏联的侵略威胁。 在 1960 年的总统竞选中，肯尼迪强烈批评艾森豪威尔在国防方面的开支不足。 肯尼迪把军队作为政治工具使用的次数比战后任何一位总统都要多… 一些历史学家批评尼克松没有充分利用艾森豪威尔的声望…

但是，以往文本立场分析的相关工作都局限于交互文本的语境，而关键词间的联系可能需要背景知识的支持，且当前一系列的端到端深度学习模型，缺乏背景知识嵌入的方法，使得交互文本立场分析的性能仍然受到限制。比如说，在表 1 实例中，如果缺乏对政治事件、政治行为的背景知识，则无法判断肯尼迪所表达的支持古巴人民选举自由与尼克松所表达的美国拥有用于推翻卡斯特罗政权的军事力量之间的关系，进而导致交互文本立场的误判。相反，如果结合表 1 中的背景知识，我们可以很明确地推导出尼克松实际上是对肯尼迪的立场表示反对态度的。为此，本文研究结合背景知识的交互文本立场分析方法，目的在于解决以上问题。

本文方法主要包含两点贡献：1) 以维基百科作为背景知识来源，以交互文本作为查询，从维基百科中检索相关背景知识；以及 2) 应用多层的深度记忆网络进行背景知识嵌入。实验结果表明，结合背景知识的交互文本立场分析可以取得更进一步的性能表现，证明了背景知识嵌入的有效性。

1 相关工作

1.1 交互文本立场分析

以往文本立场分析相关的研究主要可以划分成基于机器学习的方法^{[2][3][4]}、基于弱监督的方法^{[5][6][7]}，以及基于深度学习的方法^{[8][9]}三大类。相对于以往的文本立场分析，交互文本立场分析更加强调社交媒体的交互性，它所涉及的引述-应答对的交互情景引入了更加复杂的上下文语境信息。通常来说，一般文本立场分析任务上的模型也可以应用到交互文本立场分析中，但存在无法建模复杂的文本交互关系的问题。最近，Chen 等^[10]提出的 BiLSTM-hybrid 从交互文本联合建模的角度，结合自我注意力和交叉注意力两种注意力机制，捕捉与任务更相关的特征表示，取得了当时最好的性能表现。

1.2 深度记忆网络

2014 年，Weston 等^[11]提出记忆网络（Memory Network），其中心思想是构建一个任务相关的长期记忆模块，在任务推理的过程中，该记忆模块可以被读取或改写。具体而言，记忆网络包含一个外部记忆和四个组件：输入组件，泛化组件，输出组件以及响应组件。其中，输入组件负责输入的特征映射，泛化组件根据新的输入更新旧的记忆，输出组件结合当前输入和外部记忆状态进行输出，响应组件基于输出的结果进行响应（分类、回归等）。

Sukhbaatar 等^[12]将深度记忆网络应用到问答系统中，用于从给定的一系列句子中寻找相关线索回答对

应的问题。Sukhbaatar 等^[12]指出，多跳的深度记忆网络可以从外部记忆中提取更抽象的特征，可以进一步提升问答任务的性能。Tang 等^[13]使用深度记忆网络实现面向方面的情感分类，并通过实验证明深度记忆网络能够在该任务多个数据集上取得显著高于以往基于 LSTM 的分类模型的性能。目前，深度记忆网络在问答系统、语言建模等领域已经取得很好的性能表现。

1.3 模型引入背景知识

在自然语言处理中，对于文本内容的准确理解离不开背景知识。对背景知识运用的不充分，导致目前大多数模型都遭遇到了性能瓶颈，因此近年来不断出现一些新的工作，尝试为已有模型引入背景知识的支持。在短文本分类任务上，Wang 等^[14]将短文本表示分为显式表示和隐式表示，将用短文本检索相关的概念的特征序列直接拼接到原文本特征序列末端，作为该短文本的隐式表示。在自然语言推断任务上，Chen 等^[15]在词级别上引入单词同反义关系和上下位关系作为外部知识，加强了对前提-假设文本对的交互建模。在完型填空任务上，Mihaylov 等^[16]从外部知识库中检索与文本内容相关的实体-关系-实体三元组背景知识，并使用记忆网络对背景知识进行读取嵌入。Mihaylov 等^[16]的做法与本文工作有相似之处，但不同在于我们解决的是交互文本立场分析的任务，而且采用的背景知识库是非结构化文本，而不是图结构的知识图谱。相比于图结构的知识图谱而言，非结构化的背景知识容易大量获取，并且能够表达比知识图谱更加丰富的语义信息。此外，我们还研究了不同的嵌入方式对背景知识嵌入的影响。

2 模型总体框架

2.1 问题定义

交互文本立场分析问题是在给定交互文本，即引述-应答对，判别应答对于引述所表达的立场，主要有“支持/反对”和“支持/中立/反对”两种立场划分，但两者区别不大。一般来说，记引述为 Q ，应答为 R ，我们需要通过一个特征抽取器建模两者之间的交互特征，该特征我们记为 P 。以往交互文本立场分析方法给定分类器 f ，直接基于交互文本表示 P 进行立场分类，即有预测标签 $\hat{y} = f(Q, R) = f(P)$ 。本文研究的是引入背景知识的交互文本立场分析，因此记背景知识为 BK ，我们想要做到的是，同时基于背景知识 BK 和交互文本表示 P 对标签进行推断，即有预测标签 $\hat{y} = f(Q, R, BK) = f(P, Z)$ ，其中 Z 表示的是与引述-应答对相关的背景知识表示。

2.2 模型总体框架

本文使用的模型大致可以分类两个模块，一个部分是基于 Chen 等^[10]的模型而构建的交互文本特征抽取模块，另一部分则是本文所提出的背景知识检索及记忆嵌入模块，整体的模型框架见图 1。

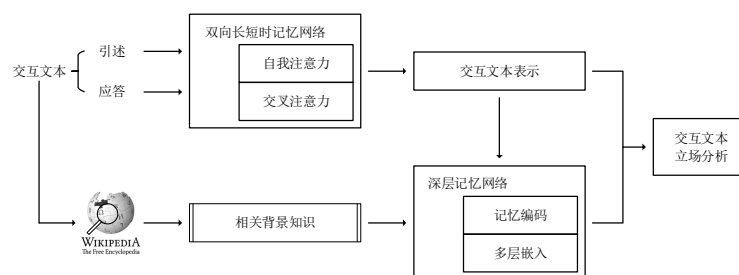


图 1 模型总体框架

Fig.1 Overview of model architecture

3 基于维基百科的背景知识检索

维基百科由于其客观公正性，以及广阔的话题覆盖面，对于本文中交互立场分析任务的三个公开英文在线辩论数据集而言，是很好的背景知识来源。对此，我们构建了一个同时兼顾性能和效率的非机器学习检索系统，从维基百科中抽取与交互文本话题相关的背景知识。背景知识检索过程主要包含文本查询构建和背景知识检索两部分，如图 2 所示。

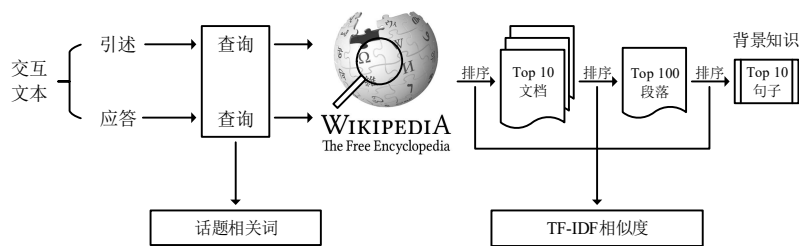


图 2 背景知识检索流程

Fig.2 Pipeline for retrieving background knowledge

3.1 文本查询构建

要从维基百科中检索相关背景知识，首先需要从原始文本构建查询，但文本中的停用词或话题无关词会为背景知识的检索引入噪声。为了尽可能降低噪声对检索的影响，查询的构建需要达到两个目的：1) 限定维基百科检索范围；同时 2) 保持文本话题领域下背景知识的丰富性，使得检索结果与人类的背景知识尽可能相近。

为了对数据进行预处理，我们借助 Stanford CoreNLP^[17]工具对原始交互文本进行处理，步骤如下：

- (1) 对 Q-R 对形式的交互文本分别进行预处理，包括去标点符号、去停用词、英文大小写统一等；
- (2) 对文本进行词性标注，提取其中的动词、名词，这些实词的通常是语义表达关键；
- (3) 对文本进行命名实体识别，提取其中涉及的人名、地名、以及组织名；

(4) 对于有预置话题目标的数据集（如 IAC），可以把相应的话题目标直接加入查询作为背景知识的领域限定。受 Turney^[18]的相关工作启发，可以在上述步骤（2）的基础上，基于 n-grams 统计的方法抽取短语词组，通过计算短语词组与话题目标的互信息，筛选出话题目标相关词。

3.2 背景知识检索

检索系统在开放领域问答、机器阅读理解等任务中已有成熟的应用。类似地，本节的基本思想是将查询和维基百科文档以词频率-逆文档率（TF-IDF）词袋表示的形式映射至同一向量空间，计算两者之间的相似性。其中，对于两类 TF-IDF 词袋表示，我们都考虑了 1-gram, 2-gram 等计数，并分别采用常数项 1 和 0.5 进行统计值平滑。最后，我们用向量点积衡量 TF-IDF 词袋表示的相似性。

在介绍本节的背景知识检索策略前，可以设想人类获取相关背景知识的过程——首先使用关键词在维基百科中进行检索，概览其中几篇相关的文档，然后选择文档中的关键部分（如段落、句子等）进行详细阅读，获取相关背景知识。受益于这一启发，基于 TF-IDF 的文本检索流程具体分三步：

- (1) 计算由当前交互文本构建的查询与维基百科所有文档之间的 TF-IDF 相似度，选取其中的相似度最高的不超过 10 篇文档作为候选文档；
- (2) 对各个候选文档进行段落划分，计算查询与所有段落之间的 TF-IDF 相似度，选取其中相似度最高且为正值的不超过 100 个文本段落作为候选段落；
- (3) 对各个候选段落进行句子划分，再次计算查询与所有句子之间的 TF-IDF 相似度，选取其中相似度最高且为正值的不超过 10 个句子作为背景知识。

4 背景知识深度嵌入

受 Sukhbaatar 等^[12]和 Tang 等^[13]相关工作的启发，在交互文本立场分析任务中，本文提出一种基于多层的深度记忆网络的背景知识嵌入模型，利用双向长短时记忆网络（BiLSTM）将背景知识编码为外部记忆，并且结合注意力机制提取与当前交互文本相关的背景知识特征，嵌入至交互文本表示。

4.1 背景知识记忆编码

检索所得的背景知识需要转化为相应的特征表示，才能够作为深度记忆网络中可读写的外部记忆，此处我们利用 BiLSTM 对背景知识中的文本集合进行记忆编码。每个交互文本对应的背景知识 BK 包含多个句子，将句子中的词映射至词向量空间后，我们独立地对各个句子，使用 BiLSTM 进行记忆编码得到输出的隐状态序列，并记为该句子的记忆编码 $m_{s_j} = [h_1^j, h_2^j, \dots, h_T^j]$ 。最后，我们将不同句子的隐状态序列按

句子长度的维度进行拼接，得到 BK 相应的外部记忆矩阵 $m = [m_{s_1}, m_{s_2}, \dots, m_{s_n}] = [h_1^1, h_2^1, \dots, h_T^n]$ 。

4.2 基于深度记忆网络的背景知识嵌入

为了引入背景知识，我们采用注意力机制从背景知识的记忆编码中抽取与当前交互文本最相关的特征。在交互文本立场分析模型中，我们通过拼接引述和应答的文本特征，得到交互文本表示 P 。考虑到交互文本和背景知识之间的文本特征存在差异，在交互文本表示 P 进行背景知识嵌入之前，我们利用一个变换层，将交互文本的表示映射到背景知识的空间中，得到 Z_0 ，如公式(1)所示。

$$Z_0 = \tanh(W_{trans}P + b_{trans}) \quad (1)$$

其中可学习的参数包括权重 W_{trans} 以及偏置 b_{trans} ， Z_0 表示文本特征经过变换映射后的结果，同时也是背景知识嵌入层的初始输入。

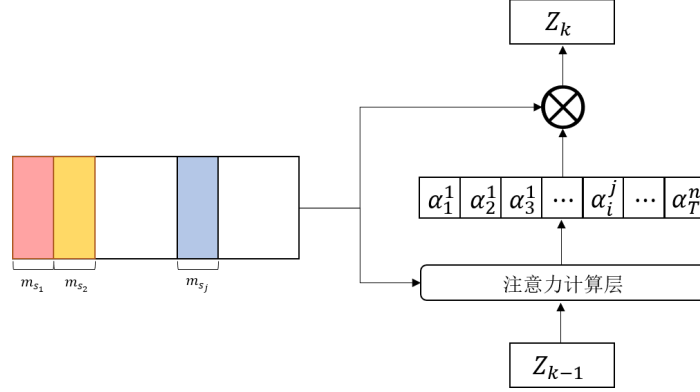


图 3 基于多层深度记忆网络的背景知识多层嵌入

Fig.3 Knowledge embedding with multi-layer deep memory network

实际上，单一嵌入层中的注意力计算可能无法处理交互文本和背景知识之间的复杂关系。因此，我们采用多层的深度记忆网络，通过堆叠背景知识嵌入层，来迭代细化背景知识嵌入表示，为交互文本立场分析任务提供更有效的特征。如图 3 所示， Z_0 输入背景知识嵌入层后，通过多层迭代对背景知识进行嵌入，其中每一步迭代都是对外部记忆矩阵的一次读取。比如说，在第 $k-1$ 步迭代中，通过注意力计算层计算 Z_{k-1} 与外部记忆矩阵之间的注意力权重，并依据该权重对记忆矩阵加权求和，输出当前交互文本的背景知识嵌入表示 Z_k 。

$$\alpha_i^j = \frac{\exp[\tanh(\phi(Z_{k-1})^T h_i^j)]}{\sum_{j,i} \exp[\tanh(\phi(Z_{k-1})^T h_i^j)]} \quad (2)$$

$$\phi(Z_{k-1}) = \tanh(W_Z Z_{k-1} + b_Z) \quad (3)$$

$$Z_k = \sum_{j,i} \alpha_i^j h_i^j \quad (4)$$

这里， $\{W_{Z_k}, b_{Z_k}\}$ 是第 k 层的可学习参数，不同层次之间的参数可以通过以下两种方式进行连接：

(1) 共享式 (MemNN-share)：不同背景知识嵌入层之间的 $\{W_{Z_k}, b_{Z_k}\}$ 相等，即 $W_{Z_1} = W_{Z_2} = \dots = W_{Z_K}$ ， $b_{Z_1} = b_{Z_2} = \dots = b_{Z_K}$ 。这种连接方式减少了深度记忆网络的参数数目并且降低了训练难度，训练速度较快。

(2) 堆叠式 (MemNN-stack)：不同背景知识嵌入层之间的 $\{W_{Z_k}, b_{Z_k}\}$ 不相等。这种连接方式能够提取更复杂的特征表示，但训练难度较大。

假设模型进行 K 层背景知识嵌入，则最后得到的交互文本背景知识嵌入表示为 $Z = Z_K$ 。为了结合背景知识推断交互文本立场，我们将 Z 与交互文本表示 P 拼接之后再通过线性分类器进行标签预测。

$$y = \text{softmax}(W_l(P \oplus Z) + b_l) \quad (5)$$

5 实验与结果分析

5.1 数据集及实验设置

本文的实验在三个公开的英文在线辩论数据集上进行，分别是 IAC^[1]，DP^[2]和 ABCD^[3]。各个数据集交互文本立场类标分布如表 2 所示。其中，IAC 与 DP 数据集均为二分类（支持/反对），常用评价指标为准确率；而 ABCD 数据集则为三分类（支持/中立/反对），常用评价指标为宏平均 F1 值。为与以往的基线方法进行对比实验，本文实验在 IAC 和 DP 数据集上应用准确率（Accuracy）作为交互文本立场分类的评价指标，在 ABCD 数据集上则应用宏平均 F1 值（Macro-average F1）作为交互文本立场分类的评价指标。

表 2 在线辩论数据集交互文本立场类标分布
Table 2 Stance labels distribution of online debate datasets

数据集	# 支持	# 反对	# 中立	网站
IAC	1,113	6,157	-	4forums.com
DP	11,875	12,899	-	debatepedia.org
ABCD	13,519	25,200	72,683	createdebate.com

考虑到三个英文数据集的发布时间，我们选取 2016-12-21 的英文维基百科离线副本¹作为背景知识的来源。经过预处理后，保留了 5,075,182 篇文档，其中包含 9,008,962 个不同的单词或符号。在检索系统中，为了兼顾速度和性能，将维基百科中的文档数据转换为 bi-gram 的组织形式，通过 *murmur3* 哈希算法^[19]进行存储。

本研究实验涉及模型参数的具体设置可以分为两个部分：

（1）交互文本建模部分，交互文本引述和应答的文本长度均设定为 64（过长则截断，过短则补全<PAD>字符），其中的词向量维度为 300，由 GloVe^[20]预训练词向量初始化；BiLSTM 的隐状态特征维度均为 128。

（2）深度记忆网络部分，每个交互文本的背景知识包含至多 10 个相关句子，每个句子的文本长度同样设定为 64（过长则截断，过短则补全<PAD>字符），其中的词向量维度为 300，由 GloVe 预训练词向量初始化；负责背景知识记忆编码的 BiLSTM 的隐状态特征维度为 128。背景知识嵌入层的层数 K 以及连接方式（share/stack）是性能对照实验中的控制变量。

整个模型的训练过程中，模型优化器 Adam^[21]的学习率为 $1e-3$ ，一阶和二阶动量衰减系数分别为 0.9 和 0.999，数值稳定量为 $1e-8$ ，以及权重衰减系数为 $1e-5$ 。所有模型的批训练数据大小为 32，除 ABCD 数据集以外（已被划分出固定的训练集和测试集），其他两个数据集均使用五折交叉验证的方式获得相应模型的交互文本立场分析性能。

5.2 实验结果分析

表 3 交互文本立场分析实验结果
Table 3 Experiment results on interactive text stance classification

IAC		DP		ABCD	
模型	准确率	模型	准确率	模型	宏平均 F1 值
NB ^[4]	60.3%	SVM ^[2]	74.0%	ME ^[3]	50.8%
JRip χ^2 ^[4]	68.2%	-	-	ME+structure ^[3]	77.6%
NLI ^[9]	74.9%	NLI ^[9]	92.4%	NLI ^[9]	76.5%
BiLSTM-hybrid ^[10]	77.0%	BiLSTM-hybrid ^[10]	93.6%	BiLSTM-hybrid ^[10]	76.9%
MemNN-share	76.7%	MemNN-share	95.3%	MemNN-share	77.4%
MemNN-stack	77.0%	MemNN-stack	95.3%	MemNN-stack	77.3%

说明：对话结构（Conversational Structure）是 ABCD 数据集特有的特征，故 ME+structure 性能较高。

¹ <https://dumps.wikimedia.org/enwiki/latest>

在本研究工作所采用的深度记忆网络中，通过进行多层迭代可以建模交互文本和背景知识之间的复杂关系，为交互文本立场分析任务提供更有用的特征。但是背景知识嵌入层层数 K 本身作为超参，最优取值未知。因此，本文实验针对两种链接方式 (share/stack)，首先验证在 K 的不同取值下模型在三个数据集上的各自性能表现，以获得对应的最优的嵌入层层数 K 取值。我们取在最优的嵌入层层数 K 取值下模型的性能作为最终性能，并与以往基线方法进行对比，对比结果如表 3 所示。实验结果表明，选取适当的背景知识嵌入层数以及背景知识嵌入层连接方式，结合背景知识的交互文本立场分析可以取得更进一步的性能表现，这证明了背景知识的嵌入的方法是有效的。

综合分析实验结果发现，在 DP 数据集上，MemNN-share 和 MemNN-stack 均可取得最好的交互文本立场分析性能，相比基线 BiLSTM-hybrid 的精确率提升了 1.7%；而在 ABCD 数据集上，MemNN-share 比 BiLSTM-hybrid 的宏平均 F1 值提升了 0.5%。但综合考虑数据规模、计算效率和性能表现，相比 stack 链接方式，share 链接方式能更有效地提取背景知识的特征。

此外，为了研究控制变量（嵌入层层数 K 、嵌入层连接方式）对模型在各个数据集上的性能表现的影响，我们对比了不同控制变量取值下共 20 种模型设置的性能差异，如图 4 所示。

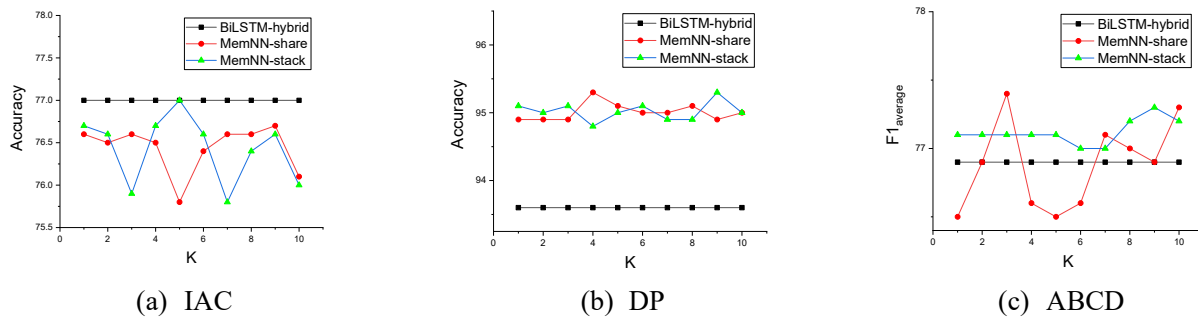


图 4 不同的嵌入层层数对任务性能的影响

Fig.4 Influence of different number of layers to the task performance

在 IAC 数据集上，如图 4(a)所示，MemNN-share 和 MemNN-stack 分别在嵌入层层数为 9 层和 5 层达到最佳性能，各自相应的准确率为 76.7%和 77.0%。当层数为 1 层时，MemNN-share 和 MemNN-stack 可以被认为是两个等价的模型，性能也较为接近。随着深度记忆网络背景知识嵌入层的层数增加，两者的性能均上下波动，其中 MemNN-stack 的嵌入层层数在 3 到 5 层之间时，性能呈现上升趋势；而 MemNN-share 的嵌入层层数在 5 层时性能最低，但其后到 9 层为止性能不断上升。但是两种改进模型的性能相比于 BiLSTM-hybrid 几乎没有变化。

在 DP 数据集上，如图 4(b)所示，MemNN-share 和 MemNN-stack 在不同的嵌入层层数时均相比 BiLSTM-hybrid 有明显的性能提升。总体而言，随着深度记忆网络背景知识嵌入层层数的增加，两种模型的性能均有上升，并在 95.0%上下小幅波动，其中 MemNN-share 在第 4 层时达到最高点 95.3%，MemNN-stack 则在第 9 层时达到最高点 95.3%，均相比基线方法性能提升了 1.7%。实际应用中，从参数空间大小的角度考虑，9 层 stack 嵌入层的参数量是 4 层 share 嵌入层的 9 倍，但后者的性能与前者相当，表明参数共享的有效性。

在 ABCD 数据集上，如图 4(c)所示，MemNN-share 的嵌入层层数在 1 到 3 层之间变化时，相应的性能提升明显，在层数为 3 层时就取得最优的 77.4%，但后续的变化趋势显示，随着嵌入层层数的增加，过高的层数对于 MemNN-share 性能提升帮助不大。另一方面，对于 MemNN-stack，嵌入层层数在 1 到 7 之间的变化对该模型的影响很小，但相应的性能均高于基线方法，并从 7 层开始逐步提升，在 9 层达到最优性能 77.3%。

对比本文方法在三个数据集上的性能表现，我们可以发现其中存在着较大的差异。在较小规模的 IAC 数据集上，本文方法不但没有带来性能提升，反而会出现反效果。经过分析发现，IAC 数据集中大量存在着不需要背景知识支持的样本，这些样本的应答直接表示了对引述的立场，而本文方法所抽取的背景知识由于包含噪声，因此噪声对模型造成了误导作用。另外，在较大规模的 DP 和 ABCD 数据集的实验结果显示，适当增加深度记忆网络背景知识嵌入层的层数，有助于模型学习更有效地背景知识特征表示从而提升分类性能，而过高的层数可能会导致模型的性能下降。其中，本文方法在 DP 数据集上的效果尤为明显，而在 ABCD 数据集上则相对较弱，该现象的出现与数据集本身的构成有着密切关联。

6 结语

本文针对传统的文本立场分析方法缺乏背景知识支持的问题,研究了结合背景知识的交互文本立场分析方法,该方法首先以交互文本作为查询,从维基百科中检索相关的背景知识,利用双向长短时记忆网络将背景知识编码为外部记忆,然后使用多层的深度记忆网络提取与当前交互文本相关的背景知识特征,以增强交互文本的特征表示。本文针对模型中的控制变量,背景知识嵌入层层数和嵌入层链接方式,设计了对照实验,研究了不同控制变量对模型性能的影响。实验结果表明,在适当的背景知识嵌入层数以及背景知识嵌入层连接方式设定下,结合背景知识的交互文本立场分析可以取得优于基线模型的性能表现,证明了本文方法的有效性。但是,本文方法所抽取的背景知识暂时还无法完全排除噪声,该问题的解决可以作为该工作进一步的研究方向。

参考文献

- [1] Walker M A, Anand P, Tree J E F, et al. A Corpus for Research on Deliberation and Debate[C]//Proceedings of the 8th International Conference on Language Resources & Evaluation (LREC), 2012: 23-25.
- [2] Menini S, Tonelli S. Agreement and Disagreement: Comparison of Points of View in the Political Domain[C]//Proceedings of the International Conference on Computational Linguistics (COLING), 2016: 2461-2470.
- [3] Rosenthal S, McKeown K. I Couldn't Agree More: The Role of Conversational Structure in Agreement and Disagreement Detection in Online Discussions[C]//Proceedings of the 2015 Conference of the Special Interest Group on Discourse and Dialogue (SIGDIAL), 2015: 168-177.
- [4] Abbott R, Walker M, Anand P, et al. How can you say such things?!?: Recognizing Disagreement in Informal Political Argument[C]//Proceedings of the Workshop on Languages in Social Media (WLSM), 2012: 2-11.
- [5] Johnson K, Goldwasser D. "All I know about politics is what I read in Twitter": Weakly Supervised Models for Extracting Politicians' Stances from Twitter[C]//Proceedings of the International Conference on Computational Linguistics (COLING), 2016: 2966-2977.
- [6] Ebrahimi J, Dou D, Lowd D. Weakly Supervised Tweet Stance Classification by Relational Bootstrapping[C]//Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2016: 1012-1017.
- [7] Dias M, Becker K. A Heuristics-Based, Weakly-Supervised Approach for Classification of Stance in Tweets[C]//Proceedings of the International Conference on Web Intelligence (WI), 2016: 73-80.
- [8] Augenstein I, Rocktäschel T, Vlachos A, et al. Stance Detection with Bidirectional Conditional Encoding[C]//Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2016: 876-885.
- [9] Liu Y, Sun C, Lin L, et al. Learning Natural Language Inference Using Bidirectional LSTM Model and Inner-Attention[J]. Computing Research Repository, 2016: abs/1605.09090.
- [10] Chen D, Du J, Bing L, et al. Hybrid Neural Attention for Agreement/Disagreement Inference in Online Debates. Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2018: 665-670.
- [11] Weston J, Chopra S, Bordes A. Memory Networks[J]. Computing Research Repository, 2014, abs/1410.3916.
- [12] Sukhbaatar S, Szlam A, Weston J, et al. End-to-End Memory Networks[C]//Proceedings of the Conference and Workshop on Neural Information Processing (NIPS), 2015: 2440-2448.
- [13] Tang D, Qin B, Liu T. Aspect Level Sentiment Classification with Deep Memory Network[C]//Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2016: 214-224.
- [14] Wang J, Wang Z, Zhang D, et al. Combining Knowledge with Deep Convolutional Neural Networks for Short Text Classification[C]//Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence (IJCAI). 2017: 2915-2921.
- [15] Chen Q, Zhu X, Ling Z, et al. Neural Natural Language Inference Models Enhanced with External Knowledge[C]//Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (ACL), 2018: 2406-2417.
- [16] Mihaylov T, Frank A. Knowledgeable Reader: Enhancing Cloze-Style Reading Comprehension with External Commonsense

Knowledge[C]//Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (ACL), 2018: 821-832.

- [17] Manning C D, Surdeanu M, Bauer J, et al. The Stanford CoreNLP Natural Language Processing Toolkit[C]//Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (ACL), 2014: 55-60.
- [18] Turney P D. Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews[C]//Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL), 2002: 417-424.
- [19] Weinberger K Q, Dasgupta A, Langford J, et al. Feature Hashing for Large Scale Multitask Learning[C]//Proceedings of the International Conference on Machine Learning (ICML), 2009: 1113-1120.
- [20] Pennington J, Socher R, Manning C D. Glove: Global Vectors for Word Representation[C]//Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2014. 1532-1543.
- [21] Kingma D P, Ba J. Adam: A Method for Stochastic Optimization[J]. Computer Science, 2014.