



北京大学学报(自然科学版)

Acta Scientiarum Naturalium Universitatis Pekinensis

ISSN 0479-8023, CN 11-2442/N

《北京大学学报(自然科学版)》网络首发论文

题目: 一种融入背景知识的交互文本立场分析方法
作者: 刘常健, 杜嘉晨, 冷佳, 陈荻, 毛瑞彬, 张俊, 徐睿峰
DOI: 10.13209/j.0479-8023.2019.096
收稿日期: 2019-05-20
网络首发日期: 2019-09-29
引用格式: 刘常健, 杜嘉晨, 冷佳, 陈荻, 毛瑞彬, 张俊, 徐睿峰. 一种融入背景知识的交互文本立场分析方法. 北京大学学报(自然科学版).
<https://doi.org/10.13209/j.0479-8023.2019.096>



网络首发: 在编辑部工作流程中, 稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定, 且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式(包括网络呈现版式)排版后的稿件, 可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定; 学术研究成果具有创新性、科学性和先进性, 符合编辑部对刊文的录用要求, 不存在学术不端行为及其他侵权行为; 稿件内容应基本符合国家有关书刊编辑、出版的技术标准, 正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性, 录用定稿一经发布, 不得修改论文题目、作者、机构名称和学术内容, 只可基于编辑规范进行少量文字的修改。

出版确认: 纸质期刊编辑部通过与《中国学术期刊(光盘版)》电子杂志社有限公司签约, 在《中国学术期刊(网络版)》出版传播平台上创办与纸质期刊内容一致的网络版, 以单篇或整期出版形式, 在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊(网络版)》是国家新闻出版广电总局批准的网络连续型出版物(ISSN 2096-4188, CN 11-6037/Z), 所以签约期刊的网络版上网络首发论文视为正式出版。

北京大学学报(自然科学版)

Acta Scientiarum Naturalium Universitatis Pekinensis

doi: 10.13209/j.0479-8023.2019.096

一种融入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 毛瑞彬² 张俊² 徐睿峰^{1,†}

1. 哈尔滨工业大学(深圳)计算机科学与技术学院, 深圳 518055; 2. 深圳证券信息有限公司, 深圳 518028;

† 通信作者, E-mail: xuruifeng@hit.edu.cn

摘要 现有的交互文本立场分析方法主要基于交互文本本身的表示学习, 但往往缺乏对背景知识的利用, 导致其性能受到影响。针对这一问题, 本文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文本, 然后对背景知识文本进行编码并通过深度记忆网络获取相关的背景知识特征, 以此来增强交互文本的表示学习。在三个英文在线辩论数据集上的实验结果显示, 通过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, 结合背景知识的方法可以有效提高交互文本立场分析性能。

关键词 立场分析; 交互文本; 背景知识; 深度记忆网络

An Interactive Stance Classification Method Incorporating Background Knowledge

LIU Changjian¹, DU Jiachen¹, LENG Jia¹, CHEN Di¹, MAO Ruibin², ZHANG Jun², XU Ruifeng^{1†}

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

2. Shenzhen Securites Information Co. Ltd. Shenzhen 518028; † Corresponding author, E-mail: xuruifeng@hit.edu.cn

Abstract Most existing stance classification on interactive text methods are representation learning of interactive text itself., but lack the use of background knowledge, which limits the performance of stance classification. To address this problem, in this paper proposes a stance classification method on interactive text by incorporating background knowledge. This method retrieves relevant background knowledge texts from Wikipedia by using the interactive text as query. The retrieved background knowledge texts are encoded and then utilized to learn the representation of relevant background knowledge through deep memory network for improving the representation learning of interactive text. The experimental results on three English online debate datasets show that the performance of interactive stance classification are be effectively improved by incorporating background knowledge through choosing the appropriate number of background knowledge embedding layers and the connection method of background knowledge embedding layer.

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

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(JSJG20170817140856618)和深圳证券信息联合研究计划资助

收稿日期: 2019-05-20; 修回日期: 2019-09-27

文本立场分析作为绑定话题目标对象的细粒度情感分析，目的是将文本中表达的立场分类为支持/反对/中立，或支持/反对。在此基础上，交互文本立场分析研究的对象是 Q-R 对形式的交互文本，其中 Q 为引述(Quote)，R 为应答(Response)。在该任务中，模型需要联合建模引述和应答以及两者间的交互语境，进而判断对于当前引述 Q，所做出的应答 R 所持的立场。

现有的文本立场分析方法主要包括基于机器学习和基于深度学习的方法。其中基于机器学习的文本立场分析方法以特征构造和筛选为主。此外，结合弱监督方法进行语料扩充，也是一种提高模型性能表现的方法。基于深度学习的文本立场分析方法，以循环神经网络为代表，则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。现有方法在文本立场分析问题上取得了一定进展，但这些方法往往局限于单纯利用交互文本本身的特征，缺乏对背景知识的利用，使得交互文本立场分析的性能提升受到限制。

表 1 展示了一个交互文本立场分析样本实例。

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)	Kennedy: "There is not any doubt we had great influence in Cuba, and I think it is unfortunate that we did not use that influence more vigorously to persuade Castro to hold free, open elections, so that the people of Cuba could have made the choice."(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)	Nixon: "What we must remember too is that the United States has the military power - and Mr. Castro knows this - to throw him out of office tomorrow or the next day or any day that we choose."(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)	President Eisenhower's New Look policy had emphasized the use of nuclear weapons to deter the threat of Soviet aggression.(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。) In his 1960 presidential race, Kennedy strongly criticized Eisenhower's inadequate spending on defense.(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。) Kennedy used the military as a political instrument more often than any other postwar president, ...(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...) Some historians criticize Nixon for not taking greater advantage of Eisenhower's popularity ...(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...)

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论点的反对态度。因此，对于该实例，正确的交互文本立场应为反对。

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，我们可以推断出尼克松实际上是对肯尼迪的论点表示反对态度的。这就启发本文研究结合背景知识的交互文本立场分析方法，通过对背景知识的表示学习和嵌入，使得立场分析方法可以取得更高的性能。

本文方法主要可分为两个部分：1)以交互文本作为查询，从维基百科中检索相关背景知识文本；2)应用多层的深度记忆网络进行背景知识嵌入，用于增强交互文本立场分析。在三个英文辩论文本数据集的实验结果表明，结合背景知识的交互文本立场分析可以取得一定性能提升，显示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^{[1][2][3]}和基于深度学习的方法^{[4][5][6]}两大类。前者主要依赖于特征筛选和构造,以及基于统计机器学习的分类器。Abbott 等^[1]结合词汇特征和依存关系特征,对比了朴素贝叶斯分类器和 JRip χ^2 分类器文本立场分析任务上的效果。Rosenthal 和 Mckeown^[2]通过实验表明对话结构在文本立场分析中能起到重要作用。Menini 等^[3]利用情感特征、语义特征和形态特征,构造支持向量机分类器,解决政治领域的文本立场分析。后者则使用神经网络对文本进行特征表示学习,用于立场分类。Augenstein 等^[4]使用双向长短时记忆网络抽取文本特征用于文本立场分析。Liu 等^[5]使用自注意力机制来构造句子特征表示用于自然语言推断。该方法也可以应用到交互文本立场分析中。Chen 等^[6]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^[7]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。Sukhbaatar 等^[8]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。此外,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^[9]使用深度记忆网络实现面向方面的情感分类,并通过实验发现深度记忆网络能够在该任务多个数据集上取得显著性能提高。

1.3 背景知识嵌入方法

人对文本内容的准确理解通常离不开背景知识。但目前自然语言处理方法往往对背景知识利用不够充分,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^[10]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^[11]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^[12]从知识图谱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^[12]的做法与本文工作有一点相似之处,但不同于本文研究的是交互文本立场分析的任务,利用背景文本知识库,而不是图结构的知识图谱。相比于图结构的知识图谱而言,文本形式的背景知识容易大量获取,并且能够表达更加丰富的语义。

2 总体框架

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 $\hat{y} = f(Q, R) = f(P)$ 。本文研究引入背景知识的交互文本立场分析,记背景知识为 BK , 目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 是与引述-应答对相关的背景知识表示。本文总体框架可以分为交互文本特征抽取模块和背景知识检索及记忆嵌入两大部分,如图 1 所示。

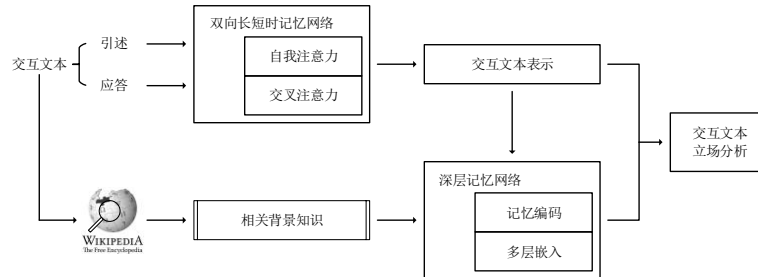


图 1 本文方法总体框架

Fig. 1 Framework of the proposed method

其中，交互文本特征抽取模块对 Q-R 对进行交互建模。在这方面，本文工作采用了 Chen 等^[6]提出的 BiLSTM-hybrid 模型。背景知识检索及记忆嵌入模块是本文工作所研究的重点。它利用交互文本构建查询，从维基百科中检索背景知识文本，然后结合交互文本特征对背景知识文本进行深度嵌入。本文将在第 3 节和第 4 节分别介绍背景知识文本的抽取和背景知识文本的编码嵌入。

3 基于维基百科的背景知识检索

维基百科具有广阔的话题覆盖面，对于交互立场分析任务而言是很好的背景知识来源。为此，本文首先建立一个检索系统，从维基百科中抽取与交互文本话题相关的背景知识文本。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。

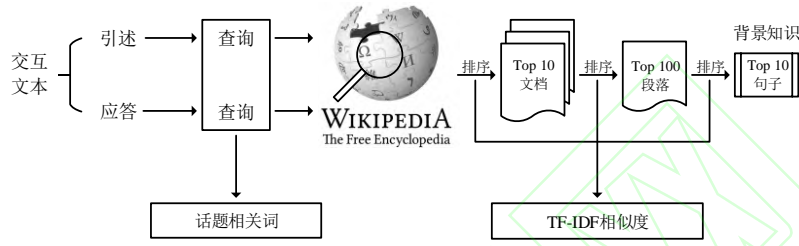


图 2 背景知识检索流程

Fig. 2 Procedure for retrieving background knowledge

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有可能引入噪声。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；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，检索系统借助 Stanford CoreNLP^[13]工具对 Q-R 对形式的交互文本进行去标点符号、去停用词、英文大小写统一等操作，然后提取其中的动词和名词，以及人名、地名和组织名。这些实词的通常是语义表达关键。对于有预置话题目标的数据集，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^[14]的相关工作启发，使用基于 n-grams 统计的方法从句子的动词、名词子序列中抽取短语词组，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。

在得到查询之后，检索系统根据查询文本和目标文本之间的词频率-逆文档率(TF-IDF)词袋表示的点积相似性来检索与上下文相关的背景知识文本。此处，检索系统所采用的 TF-IDF 词袋表示中考虑了 Uni-gram, Bi-gram 等计数，并分别采用常数项 1 和 0.5 进行统计值平滑。随后，检索系统对目标文本进行文档、段落和句子三个级别的筛选过滤后，保留其中 TF-IDF 相似度最高且为正值的不超过 10 个句子作为背景知识文本。

4 背景知识深度嵌入

受 Sukhbaatar 等^[8]和 Tang 等^[9]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。

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。首先，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，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，并记为对应句子的记忆编码 $m_{s_i} = [h_1^i, h_2^i, \dots, h_n^i]$ 。然后，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 $m = [m_{s_1}, m_{s_2}, \dots, m_{s_n}] = [h_1^1, h_2^1, \dots, h_n^n]$ 。

为了引入背景知识，本文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，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，可以通过一个变换层，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。

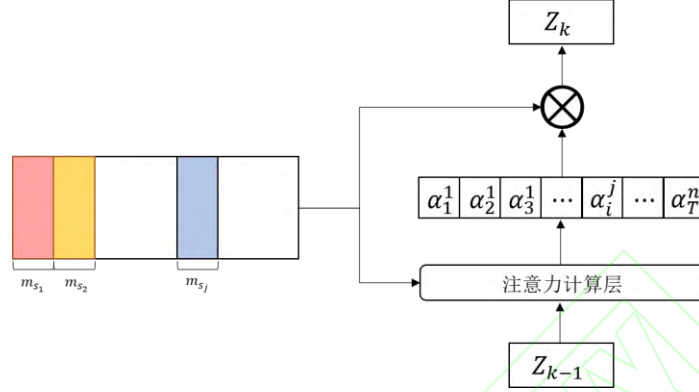


图 3 基于多层深度记忆网络的背景知识多层嵌入

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, 本文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学习更有效的特征表示。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, 这里将 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 数据集(Internet Argument Corpus)^[15], DP 数据集(Debatepedia)^[3]和 ABCD 数据集(Agreement by Create Debaters)^[2]。各个数据集中的交互文本立场

类标分布如表 2 所示。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^{[1][3]}, 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^[2]。

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 5075182 篇文档, 其中包含 9008962 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^[16]进行存储。

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	1113	6157	-	4forums.com
DP	11875	12899	-	debatepedia.org
ABCD	13519	25200	72683	createdebate.com

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为 64 (过长则截断, 过短则用 <PAD> 补全), 其中的词向量维度为 300, 由 GloVe^[17] 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, 过短则用 <PAD> 补全), 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, 根据模型在验证集上的性能选取最优值。

整个模型的训练过程中, 模型优化器 Adam^[18] 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 基线系统

本文工作与以下几个基线系统进行对比:

- **JRip χ^2** : Abbott 等^[1]提出的基于 JRip χ^2 分类器的模型。JRip 是一种基于决策树的规则学习方法, 而 JRip χ^2 在 JRip 的基础上增加了 χ^2 特征选择。特征包括词汇特征和依存关系特征。
- **SVM**: Menini 等^[3]提出的基于向量机分类器的模型。特征包括情感特征、语义特征和形态特征。
- **ME+structure**: Rosenthal 和 Mckeown^[2]提出的基于最大熵分类器的模型。特征包括词汇特征、风格特征、情感特征和对话结构特征。
- **NLI**: Liu 等^[5]提出的自然语言推断模型。该模型使用 BiLSTM 分别对前提 (Premise) 和假设 (Hypothesis) 进行编码, 然后通过自注意力机制构造句子特征。
- **BiLSTM-hybrid**: Chen 等^[6]提出的结合了自注意力机制和交叉注意力机制的交互文本推断模型。

5.3 实验结果分析

本文实验首先评估 IAC、DP 和 ABCD 三个在线评论英文数据集中的立场分析性能。本文方法与上述基线方法所取得的对比结果如表 3 所示。可以发现, 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, 优于绝大多数基线方法。相比主要对比方法 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%。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。该结果表明, 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。

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

表 3 交互文本立场分析实验结果

模型	IAC/准确率	DP/准确率	ABCD/ F1 值
JRip χ^2 ^[1]	68.2%	-	-
SVM ^[3]	-	74.0%	-
ME+structure ^[2]	-	-	77.6%
NLI ^[5]	74.9%	92.4%	76.5%
BiLSTM-hybrid ^[6]	77.0%	93.6%	76.9%
MemNN-share	76.7%	95.3%	77.4%
MemNN-stack	77.1%	95.3%	77.3%

说明: ME+structure 使用了 ABCD 数据集特有的对话结构特征, 故在 ABCD 数据集性能较高。

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三种模型的性能差异, 其结果如图 4 所示。

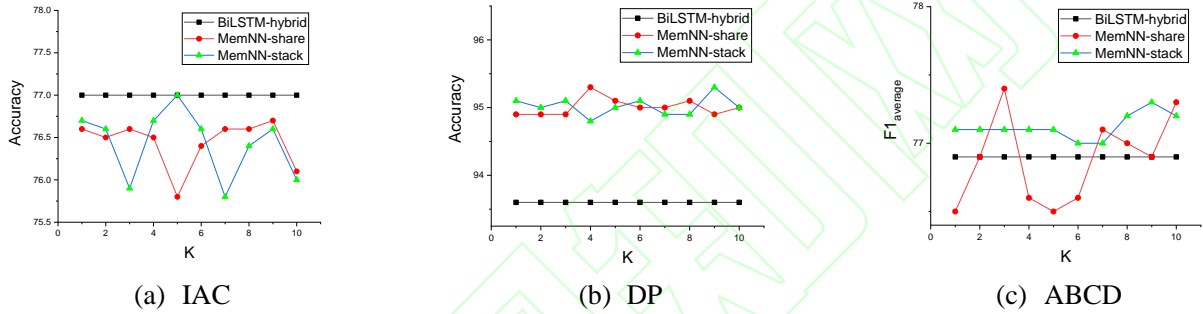


图 4 不同的嵌入层层数对任务性能的影响

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一定的差异。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。在较小规模的 IAC 数据集上, 本文方法会出现轻微性能下降。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简单对话样本, 这些样本的应答直接表示了对引述的立场, 使得背景知识引入的收益下降。同时文本方法所抽取的

背景知识可能包含噪声,对模型性能带来轻微副作用。

6 结语

本文针对现有的文本立场分析方法缺乏背景知识支持的问题,提出了一种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] Abbott R, Walker M, Anand P, et al. How can you say such things!?: recognizing disagreement in informal political argument // Workshop on Languages in Social Media (WLSM), 2012: 2–11
- [2] Rosenthal S, Mckeown K. I couldn't agree more: the role of conversational structure in agreement and disagreement detection in online discussions // Conference of the Special Interest Group on Discourse and Dialogue (SIGDIAL), 2015: 168–177
- [3] Menini S, Tonelli S. Agreement and disagreement: comparison of points of view in the political domain // International Conference on Computational Linguistics (COLING), 2016: 2461–2470
- [4] Augenstein I, Rocktäschel T, Vlachos A, et al. Stance detection with bidirectional conditional encoding // Conference on Empirical Methods in Natural Language Processing (EMNLP), 2016: 876–885
- [5] Liu Yang, Sun Chengjie, Lin Lei, et al. Learning natural language inference using bidirectional LSTM model and inner-attention. Computing Research Repository, 2016: abs/1605.09090
- [6] Chen Di, Du Jiachen, Bing Lidong, et al. Hybrid neural attention for agreement/disagreement inference in online debates // Conference on Empirical Methods in Natural Language Processing (EMNLP), 2018: 665–670
- [7] Weston J, Chopra S, Bordes A. Memory networks. Computing Research Repository, 2014: abs/1410.3916
- [8] Sukhbaatar S, Szlam A, Weston J, et al. End-to-end memory networks // Conference on Neural Information Processing (NIPS), 2015: 2440–2448
- [9] Tang Duyu, Qin Bing, Liu Ting. Aspect level sentiment classification with deep memory network // Conference on Empirical Methods in Natural Language Processing (EMNLP), 2016: 214–224
- [10] Wang Jin, Wang Zhongyuan, Zhang Dawei, et al. Combining knowledge with deep convolutional neural networks for short text classification // International Joint Conference on Artificial Intelligence (IJCAI), 2017: 2915–2921
- [11] Chen Qian, Zhu Xiaodan, Ling Zhenhua, et al. Neural natural language inference models enhanced with external knowledge // Annual Meeting of the Association for Computational Linguistics (ACL), 2018: 2406–2417
- [12] Mihaylov T, Frank A. Knowledgeable reader: enhancing cloze-style reading comprehension with external commonsense knowledge // Annual Meeting of the Association for Computational Linguistics (ACL), 2018: 821–832
- [13] Manning C D, Surdeanu M, Bauer J, et al. The Stanford CoreNLP natural language processing toolkit // Annual Meeting of the Association for Computational Linguistics (ACL), 2014: 55–60
- [14] Turney P D. Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews // Annual Meeting of the Association for Computational Linguistics (ACL), 2002: 417–424

- [15] Walker M A, Anand P, Tree J E F, et al. A corpus for research on deliberation and debate // International Conference on Language Resources & Evaluation (LREC), 2012: 23–25
- [16] Weinberger K Q, Dasgupta A, Langford J, et al. Feature hashing for large scale multitask learning // International Conference on Machine Learning (ICML), 2009: 1113–1120
- [17] Pennington J, Socher R, Manning C D. Glove: global vectors for word representation // Conference on Empirical Methods in Natural Language Processing (EMNLP), 2014: 1532–1543
- [18] Kingma D P, Ba J. Adam: a method for stochastic optimization // Proceedings of International Conference on Learning Representations (ICLR), 2015

