

基于 Bi-LSTM 的多层面隐喻识别方法

朱嘉莹, 王荣波*, 黄孝喜, 谌志群

(杭州电子科技大学 认知与智能计算研究所, 浙江 杭州 310018)

摘要: 以双向长短期记忆网络(Bi-LSTM)为核心,结合多层卷积神经网络以及单向长短期记忆网络构建了多层面隐喻识别模型.基于多特征协同作用的思想,利用依存关系特征、语义特征、词性特征等多特征融合输入方法,丰富了模型的学习信息.为降低信息干扰,利用基于统计学的规范化文本输入方法提升模型识别效果.在英文语料词层面和句层面实验中,各个特征均表现出明显的正向作用.裁剪和填充处理及多特征协调作用在英文语料词层面研究中使 F_1 值分别提升 2.5% 和 5.1%,在句层面研究中 F_1 值分别提升 3.1% 和 1.9%.在中文语料句层面实验中,最优效果的 F_1 值可达 88.8%.

关键词: 自然语言理解;隐喻识别;CNN;Bi-LSTM;依存关系

中图分类号: TP391

文献标识码: A

doi:10.7511/dllgxb202002013

0 引言

隐喻作为一种语言现象,在日常生活中被广泛应用.这种包含了复杂想象力的语言现象,可以帮助协调人们对世界的认知体验和概念系统^[1].这种协调方式体现了某种跨领域的映射关系.黄孝喜^[2]具体地提出了隐喻的本质其实就是以—个较为熟悉或简单的概念来理解另—个较为复杂或困难的概念.如“幸福像花儿—样”,由于“花儿”有甜蜜美好的属性,为了形象表现“幸福”同有这—属性,作为—个抽象名词,“幸福”需要借助—个存在相似特征的具象名词来理解,因此才会出现上文所示的语言表达方式.在“幸福像花儿—样”中,“花儿”甜蜜美好的认知是人们获得的经验认知,“幸福”是人们获得的另—个概念;在这种跨领域映射关系中,“花儿”作为源域,“幸福”作为目标域.

隐喻识别对于更好地挖掘语义和情感信息有重要的作用,在机器翻译、问答系统、情感分析等领域有广泛的应用^[3].因此,隐喻识别的进步关系到自然语言处理多个领域的提升.但隐喻识别仍

然面临着许多不可避免的挑战.首先,隐喻具有较强的主观性,需要特定领域的知识作为支撑.因此,要求研究者不仅具备使用计算机能力还需要熟悉语言学知识,并且研究语料的标注要求多位学者共同工作以保证其客观性,这也成为隐喻研究语料库稀少的因素之—.其次,隐喻不具备特定的词汇搭配形态或者信号^[4].隐喻的使用可以是简单的也可以是复杂的^[5].再次,隐喻常常需要联系上下文来克服其不确定性.例如英文语句“He is playing with fire”有两种意义,一种是字面意义,他在玩火;另—种是隐喻意义,他在冒险.如果脱离上下文,则无法判定这句话的真实意义.

多年来,研究者们针对隐喻识别提出了各种解决方法.根据不同的研究视角可以分为句法方面研究和语义方面研究. Sullivan^[6]提出了句法结构可被应用于识别隐喻中的源域和目标域,认为动词会激活源域,动词的论元则会激活目标域. Stowe 等^[7]分析了句法结构与隐喻现象之间的关系,而不仅仅拘泥于检测源域和目标域的思想,试图通过对基于语料库的论元结构及其隐喻特征的分析来确定有效句法结构,以此实现隐喻识别.

收稿日期: 2019-09-22; 修回日期: 2020-01-03.

基金项目: 国家自然科学基金青年基金资助项目(61202281);教育部人文社科规划青年基金资助项目(12YJCZH201);国家社会科学基金资助项目(18ZDA290);教育部人文社科项目规划基金资助项目(18YJA740016).

作者简介: 朱嘉莹(1995-),女,硕士生, E-mail: zhujy@hdu.edu.cn;王荣波*(1978-),男,副教授, E-mail: wangrongbo@hdu.edu.cn.

Lederer^[8]提出了不一样的理念,他在研究中表明只关注源域的观点,认为源域和目标域在词法上的分离不利于隐喻的识别,且一个源域可以对应多个目标域(目标域是可变的),因此选择源域作为研究对象.他以经济语言为例,基于统计学方法对伴随源域日常出现的搭配词做出评估,通过一定的指标判定其是否为源域出现的触发词,随后将符合要求的词录入数据集用于挖掘更多的隐喻数据.但是这种方法非常依赖人工分析,是一种半自动识别方式,且只能识别常规隐喻,对新颖隐喻束手无策.

Bulat 等^[9]选择从语义角度研究隐喻识别,认为隐喻计算任务的关键在于寻找语义表现的最优抽象级别.先对 McRae 等^[10]建立的属性-规范数据集(property-norm dataset)中的规范性属性训练学习,再利用基于 Mikolov 等^[11]的 skip-gram 模型上下文预测和统计方法建立语言表示方法,最终建立交叉-模态映射模型,实现自动选择合适的属性来比较两个词.然后采用 SVM 来分类预测该词的表述是否为隐喻.这种基于语义属性的研究为未来研究工作提供了深入发展的机会. Zayed 等^[4]使用分布语义识别动-名词结构中的动词隐喻,主要通过动-名词词向量相似度计算来判定是否属于隐喻. Bizzoni 等^[12]认为隐喻识别依赖上下文信息,因此关注到了词序列,采用 Bi-LSTM 获取上下文语义关系.另外,他们还添加了每个词的抽象度作为额外特征输入模型,很大程度地改善了结果.苏畅等^[13]提出了基于动态分类的隐喻识别方法,动态地从概念的客观属性、认知属性、语义类别、整体部件等 4 个维度选择合适的维度比较相似度.

近年,机器学习和深度学习的盛行为隐喻识别自动化提供了更广阔的发展空间. Rai 等^[14]使用了监督学习方法,通过提取句法、概念、情感和上下文 4 个特征,一并将这些特征向量化与词向量拼接起来,选用 CRF 作为分类器输出类别. Mosolova 等^[15]在此基础上将 4 个特征替换成词性、抽象度等 10 个特征,并尝试了 3 种梯度下降方法分别训练模型.随后有研究者尝试用神经网络代替机器学习来研究问题. Pramanick 等^[16]提出了基于 Bi-LSTM 和 CRF 模型的隐喻识别,并提出了在数据预处理阶段将词还原为词根的方法.相比于 Bizzoni 等^[12]的研究,他们更关注语义顺序和每个词的抽象与具象对比度等文本特征,

认为不同程度的对比度对识别结果有着权重不同的影响. Wu 等^[17]认为词的局部语义及长距离语义为隐喻识别提供了有效信息支持,可采用不同窗口大小的 CNN 获取不同范围内语义信息,再结合 Bi-LSTM 综合上下文信息进行隐喻识别.

如上所述,大多数研究都是词和短语层面的隐喻识别,句层面的研究工作相对较少,且关注的特征相对单一.针对隐喻识别依赖上下文信息的特点,本文提出构造 CNN-Bi-LSTM 模型,并将其应用于词层面和句层面的隐喻识别研究.该模型充分利用不同范围内的语义、句法、词性等多特征协同作用进行隐喻识别.

1 研究方法

本文的研究思想主要参考了 Wu 等^[17]工作.在此基础上,注意到隐喻识别不能仅仅依靠单一语义信息判定,且语义信息分布相对分散,捕捉任务上有一定难度.因此提出借助句法信息,对特定部分提取依存关系特征,以此获得论元及其位置信息的方法.本文对 VUAMC(VU Amsterdam metaphor corpus)隐喻数据中各词性的占比 λ 情况进行逐一统计.统计结果如图 1 所示.

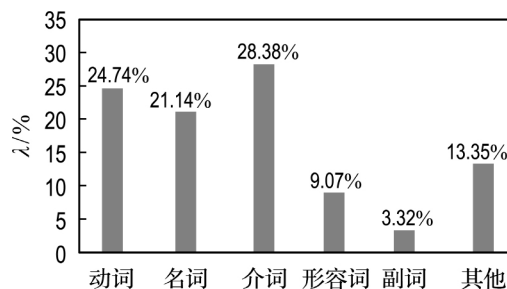


图 1 VUAMC 数据集中各词性隐喻占比情况

Fig. 1 The proportion of metaphors in different parts of speech in VUAMC dataset

由图 1 可知,介词隐喻、动词隐喻和名词隐喻是 3 种最为普遍的隐喻现象.深入分析语料文本又可发现其他隐喻常常伴随于动词隐喻出现.因此,重点关注了动词结构的隐喻,认为隐喻常出现在动词及其周围,将动词作为突破口来挖掘多种类型的隐喻.多层面隐喻识别模型如图 2 所示.

模型的第 1 部分用作句子的处理,首先输入词序列,并通过 NLTK 中的 WordNetLemmatizer 还原成词根,以此获取准确的词向量,并减少表外词的出现.而后,对词序列进行长度规范化以降低

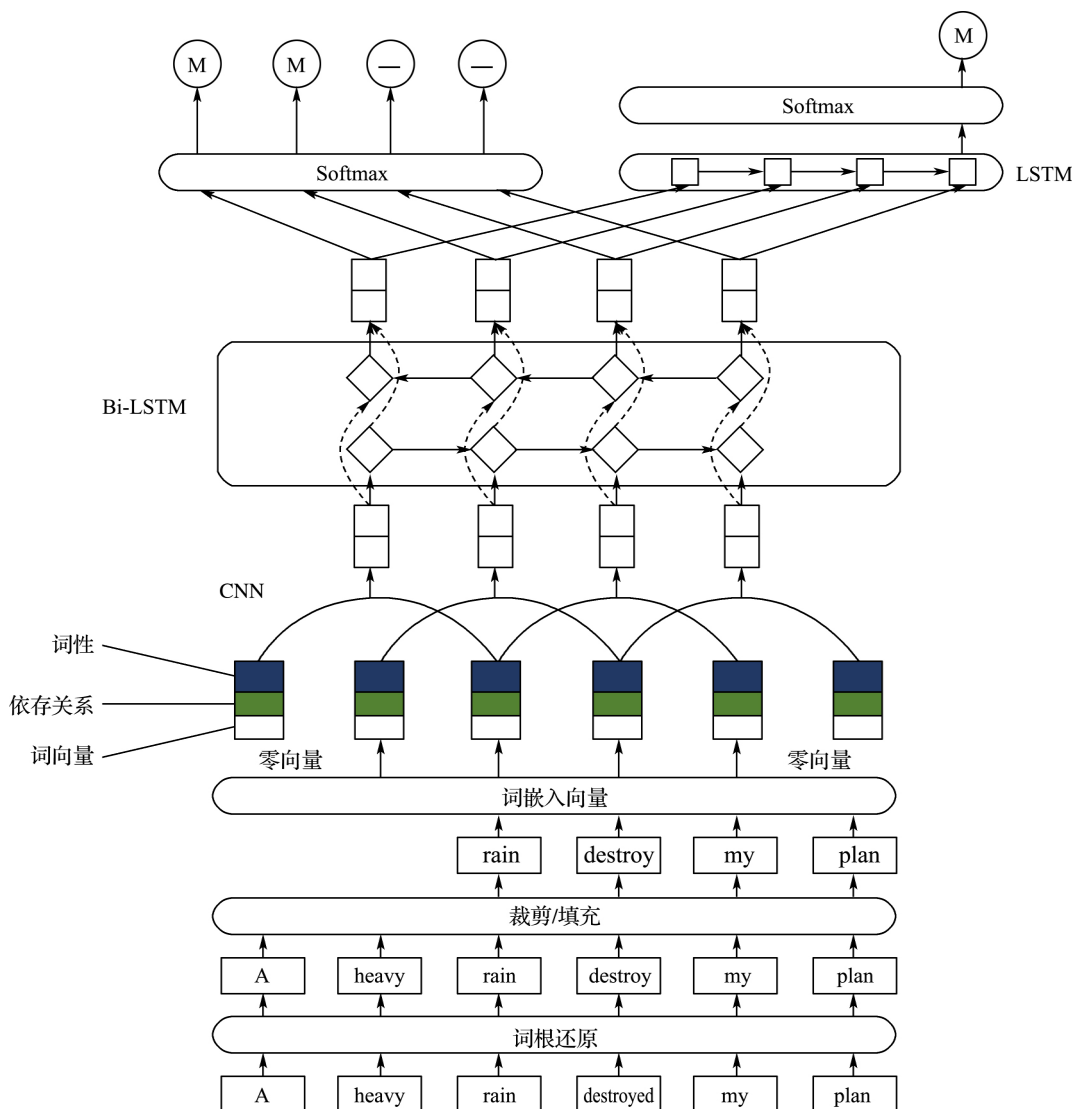


图2 多层面隐喻识别模型

Fig. 2 A multi-level model of metaphor recognition

模型训练的信息干扰,且长序列输入不利于模型的训练,这一处理还从一定程度上缩短了模型训练时间.本文之所以选择保留部分完整序列是因为句子的连贯性会为模型提供细微的有效信息.

模型的第2部分是词向量层.词向量中包含大量的语义信息.考虑到多方面的文本特征信息会产生协同作用,并行输入会为模型提供最大化的特征信息效果,将其编码为独热向量,并将词向量与它们依次串联.

模型的第3部分是卷积神经网络(CNN).卷积神经网络可以通过改变窗口大小,来控制获取不同范围文本的特征信息.根据隐喻信息涉及范围的不确定性,选择滑动窗口大小为2、3、4、5的神经网络来依次获取二元组、三元组、四元组、五

元组等多规模结构内的局部信息,并串联这些不同窗口大小的输出结果向量.

模型的第4部分是Bi-LSTM层,用于从串联起来的卷积神经网络特征信息中获取上下文的全局信息.

模型的第5部分是一个可选的LSTM层,当且仅当模型被用于隐喻句识别时该层被选.该层用于保留由Bi-LSTM层输出的长距离文本信息,并将它们集成于序列末尾.LSTM层中每个单元的遗忘门、输入门、输出门和记忆细胞^[18],为获取长距离文本重要信息的关键.LSTM层的遗忘门用于决策当前词信息的舍弃情况,算法如下:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (1)$$

其中 W_f 是遗忘门的权重矩阵, b_f 是遗忘门的偏

置矩阵, h_{t-1} 是前一个词经过 LSTM 单元运算变换后的输出, x_t 是当前词的输入, σ 表示 sigmoid 函数. LSTM 输入门用于决策当前词信息的保留情况, 算法如下:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (2)$$

其中 W_i 是输入门的权重矩阵, b_i 是输入门的偏置矩阵. LSTM 用于表示当前单元的单元状态, 算法如下:

$$\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (3)$$

其中 W_c 是控制当前单元的权重矩阵, b_c 是控制当前单元的偏置矩阵. LSTM 用于更新当前单元的单元状态, 算法如下:

$$C_t = f_t C_{t-1} + i_t \tilde{C}_t \quad (4)$$

其中 C_{t-1} 为前一单元的单元状态. LSTM 输出门算法如下:

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (5)$$

其中 W_o 是输出门的权重矩阵, b_o 是输出门的偏置矩阵. LSTM 当前单元的输出如下:

$$h_t = o_t \tanh(C_t) \quad (6)$$

经过 N 个 LSTM 单元的运算, 最后一个 LSTM 单元聚集了所有被选中需要保留的文本信息, 包括语义信息和依存关系信息.

最后一部分是判别器, 用于输出判别结果, “M”和“—”分别表示隐喻和非隐喻. 在本文中, 句层面和词层面的实验均采用 Softmax 判别器.

2 实 验

2.1 英文词层面的实验

2.1.1 数据集 实验中使用 VUAMC, 该数据集是 BNC-Baby 的一部分, 包含 4 个领域的文本: 学术、新闻、科幻小说、对话. 每种文本大约包含 50 000 个单词. 一共有 16 202 个句子, 其中 12 122 个为训练集数据, 4 080 个为测试集数据. 数据集的建立者为文本中每个单词都标定了隐喻标签. 之所以选择这个数据集是因为这个数据集中的文本来源于各个领域, 具有比较大的参考价值. 另外, 这个数据集包含多类型的隐喻, 有利于更全面地考虑问题和解决问题.

2.1.2 数据处理 对单词进行还原处理, 目的是为了获取更准确的词向量, 减小表外词的出现概率, 但是这并不能避免它们的出现. 因此, 在实验中, 将会用零向量来代替表外词的词向量.

之后, 对句子长度进行统计, 分布结果如图 3 所示. 纵轴表示句子的数量 n , 横轴表示词的个数 m .

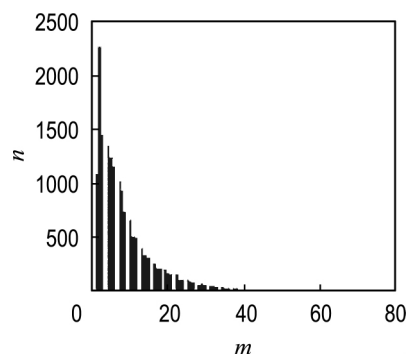


图 3 英文句子长度分布图

Fig. 3 English sentences length distribution diagram

本文采用以下算法来确定合适的句子长度:

$$l = \bar{l} + 2 \sqrt{\frac{1}{n-1} \sum_{i=1}^n (l_i - \bar{l})^2} \quad (7)$$

其中 \bar{l} 表示句子的平均长度, n 表示句子的数量, l_i 表示第 i 个句子的长度.

本文采用这样的方法来规范化模型输入长度. 经过算法计算得到的句子长度为 23, 覆盖了 94.07% 的句子. 随后对数据进行裁剪和填充处理, 处理的数据包括特征序列和标签序列. 当处理长度超过 23 的句子时, 只保留末尾的 23 个序列. 处理长度不足 23 的句子时, 则在整个序列之前填充零向量来补全空缺.

2.1.3 词向量 本文使用的是利用 word2vec 模型在谷歌新闻语料上训练生成的 300 维词向量, 包含 3×10^6 词和短语的词向量. 由于词量庞大, 先对训练集和测试集中出现的单词做了归纳, 生成一张词查找表用于后续的词向量获取.

2.1.4 特征提取 本文通过 StanfordCoreNLP 软件包来解析句子的依存关系, 通过句子中的动词来获得与之相关的论元及论元信息. 为此, 在实验中将动词的主语和宾语结构提取出来, 用二维独热向量编码表示它们在句子中的成分. 通过 Stanford parser 获取词性信息, 并用 29 维独热向量逐一编码.

2.2 英文句层面的实验

在数据预处理阶段, 将隐喻标签序列处理为句子对应的单个标签并生成句层面的数据集. 表 1 为该数据集中隐喻句的占比情况.

表1 VUAMC 数据集中隐喻句占比

Tab.1 The proportion of metaphorical sentences in VUAMC dataset

| 数据集 | 占比/% |
|-----|-------|
| 训练集 | 59.92 |
| 测试集 | 58.16 |

2.3 中文句层面的实验

2.3.1 数据集 本文采用2018年大连理工大学发布的用于中文动词隐喻识别的数据集。该数据集一共有4394条数据,其中90%的数据(3954条)作为训练集,10%的数据(440条)作为测试集。

2.3.2 数据处理 对句子的分词处理,使用了由哈尔滨工业大学开发的软件包LTP^[19]。LTP主要运用在线学习预测训练实例的标注,并且有比较全面的文本处理模型,包括分词模型、词性标注模型、命名实体识别模型、依存关系分析模型、语义角色标注模型等。在中文自然语言处理领域中被广泛应用。

中文句子长度分布图如图4所示。采用2.1.2中同样的方法获取句子长度为25。

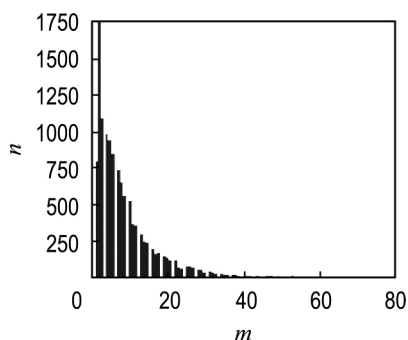


图4 中文句子长度分布图

Fig.4 Chinese sentences length distribution diagram

2.3.3 词向量 采用在维基百科语料(<https://dumps.wikimedia.org/>)上经过 word2vec 模型训练的词向量,一共含有815095个词向量,每个词向量400维。

2.3.4 特征提取 本文同样使用LTP获取词语的词性并解析句子依存关系,将词性和依存关系分别用29维和4维的独热向量编码。

3 实验结果

为验证各方法的正向作用效果,本文对每个

方法依次实验。表2是与baseline(基线)的对比结果,本文选取的baseline是Wu等^[17]的实验结果。实验结果表明,经过长度规范化处理的模型取得了最优的结果,各个指标均优于baseline,准确率、召回率、 F_1 值分别提升5.5%、4.7%、5.1%。同时优于Wu等^[17]研究中最好的结果(使用基于CNN-LSTM与Softmax的集成模型),对应表2中的“best in Wu”。本文结果与之相比在准确率、 F_1 值上分别提升11.2%、2.1%。因此,验证了经过裁剪和填充的数据减少了冗长信息的干扰,可有效提升识别效果。此外,多特征的协同作用为模型提供大量有效的论元信息,使模型更准确地识别隐喻位置,实验结果对应表2中“+relation+pos”, F_1 值提升2.5%。英文句层面的baseline与词层面相同。表中relation表示依赖关系特征, pos表示词性特征,CP表示输入数据经过裁剪和填充处理,none表示输入数据未经过任何处理且未增加任何特征的实验。

表2 英文词层面实验对比结果

Tab.2 The comparison results of English word level experiments

| 方法 | 准确率 | 召回率 | F_1 值 |
|---------------|--------------|--------------|--------------|
| baseline | 0.665 | 0.583 | 0.621 |
| +relation+pos | 0.655 | 0.619 | 0.646 |
| +CP | 0.720 | 0.630 | 0.672 |
| best in Wu | 0.608 | 0.700 | 0.651 |

实验结果最好的模型是做了裁剪和填充处理并加入多种特征的方法。由表3可见,对比baseline的 F_1 值,该方法有4.0%的明显提升。可见该方法对于不同层面的隐喻识别任务均表现出不同程度的提升效果。由此验证了模型具有较强的泛化能力。

表3 英文句层面实验对比结果

Tab.3 The comparison results of English sentence level experiments

| 方法 | 准确率 | 召回率 | F_1 值 |
|------------------|-------|-------|--------------|
| baseline | 0.777 | 0.692 | 0.732 |
| +relation+pos | 0.737 | 0.791 | 0.763 |
| +CP | 0.753 | 0.750 | 0.751 |
| +CP+relation+pos | 0.738 | 0.788 | 0.772 |

对于中文句层面的研究,对比的是一项2018年CCL发布的评测任务结果(<http://ir.dlut.edu.cn/news/detail/526>)。目前尚未公开参赛者

的研究方法,只公开了结果.本文选取前 5 名结果,见表 4.

表 4 参赛队伍评测结果

Tab. 4 Evaluation results of participating teams

| 队名 | F_1 值 |
|----------|---------|
| hqu | 0.833 |
| faun | 0.831 |
| YNU-HPCC | 0.831 |
| MITLAB | 0.827 |
| prism | 0.821 |

由于本文模型在捕捉上下文信息方面存在高效性,在不获取外部特征信息的情况下具有较高的 F_1 值,达 86.0%. 对数据进行裁剪和填充处理并加入词性、依存关系等多种文本特征之后,实验结果在 F_1 值上有了进一步提升,增幅达 2.8%. 对比评测任务公布的最好结果,本文方法 F_1 值提高了 5.5%. 实验结果如表 5 所示.

表 5 中文句层面实验对比结果

Tab. 5 The comparison results of Chinese sentence level experiments

| 方法 | 准确率 | 召回率 | F_1 值 |
|------------------|-------|-------|--------------|
| none | 0.889 | 0.833 | 0.860 |
| +CP | 0.867 | 0.871 | 0.869 |
| +relation+pos | 0.867 | 0.889 | 0.878 |
| +CP+relation+pos | 0.881 | 0.896 | 0.888 |

4 结 语

本文基于 Bi-LSTM 提出的模型,在词层面和句层面的隐喻都得到了较好的实验效果,体现了减少文本信息干扰和文本多样化信息的协同作用对隐喻识别的优化作用.此外,规范化模型输入长度的方法和多特征协同作用的方法在不同任务中表现的作用强度有所不同,例如在英文词层面的研究中长度规范化的作用较依存关系特征更突出,在英文句层面中则相反,而中文句层面的情况与英文句层面一致.这与方法的具体作用方面不同,裁剪和填充处理主要作用于文本序列这个宏观方面,而多特征协同作用主要作用于单个输入这个微观方面,且与任务本身的特性相互呼应.

本文提出的模型依赖于大规模的训练数据,对小规模数据的研究工作造成了限制,在以后的工作中会采用诸如 Transformer 模型来解决这个弊端.

参考文献:

- [1] LAKOFF G, JOHNSON M. *Metaphors We Live by* [M]. Chicago: The University of Chicago Press, 1980.
- [2] 黄孝喜. 隐喻机器理解的若干关键问题研究 [D]. 杭州: 浙江大学, 2009.
HUANG Xiaoxi. Research on some key issues of metaphor machine understanding [D]. Hangzhou: Zhejiang University, 2009. (in Chinese)
- [3] TSVETKOV Y, BOYTSOV L, GERSHMAN A, *et al.* Metaphor detection with cross-lingual model transfer [C] // **52nd Annual Meeting of the Association for Computational Linguistics ACL 2014 - Proceeding of the Conference**. Baltimore: Association for Computational Linguistics, 2014: 248-258.
- [4] ZAYED O, MCCRAE J P, BUITELAAR P. Phrase-level metaphor identification using distributed representations of word meaning [C] // **Proceedings of the Workshop on Figurative Language Processing**. New Orleans: Association for Computational Linguistics, 2018: 81-90.
- [5] MYKOWIECKA A, MARCINIAK M, WAWER A. Literal, metaphorical or both? detecting metaphoricity in isolated adjective-noun phrases [C] // **Proceedings of the Workshop on Figurative Language Processing**. New Orleans: Association for Computational Linguistics, 2018: 27-33.
- [6] SULLIVAN K. *Frames and Constructions in Metaphoric Language* [M]. Amsterdam: John Benjamins Publishing Company, 2013.
- [7] STOWE K, PALMER M. Leveraging syntactic constructions for metaphor identification [C] // **Proceedings of the Workshop on Figurative Language Processing**. New Orleans: Association for Computational Linguistics, 2018: 17-26.
- [8] LEDERER J. Finding metaphorical triggers through source (not target) domain lexicalization patterns [C] // **Proceedings of the Fourth Workshop on Metaphor in NLP**. San Diego: Association for Computational Linguistics, 2016: 1-9.
- [9] BULAT L, CLARK S, SHUTOVA E. Modelling metaphor with attribute-based semantics [C] // **Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics**. Valencia: Association for Computational Linguistics, 2017, **2**: 523-528.
- [10] MCRAE K, CREE G S, SEIDENBERG M S,

- et al.* Semantic feature production norms for a large set of living and nonliving things [J]. **Behavior Research Methods**, 2005, 37(4): 547-559.
- [11] MIKOLOV T, CHEN K, CORRADO G, *et al.* Efficient estimation of word representations in vector space [C] // **1st International Conference on Learning Representations, ICLR 2013 - Workshop Track Proceedings**. Scottsdale: International Conference on Learning Representations, ICLR, 2013.
- [12] BIZZONI Y, GHANIMIFARD M. Bigrams and BiLSTMs two neural networks for sequential metaphor detection [C] // **Proceedings of the Workshop on Figurative Language Processing**. New Orleans: Association for Computational Linguistics, 2018: 91-101.
- [13] 苏 畅, 付 泽, 郑发魁, 等. 基于动态分类的隐喻识别方法 [J]. 软件学报, 2019, 30(11): 3340-3354.
- SU Chang, FU Ze, ZHENG Fakui, *et al.* Method of metaphor recognition based on dynamic categorization [J]. **Journal of Software**, 2019, 30(11): 3340-3354. (in Chinese)
- [14] RAI S, CHAKRAVERTY S, TAYAL D K. Supervised metaphor detection using conditional random fields [C] // **Proceedings of the Fourth Workshop on Metaphor in NLP**. San Diego: Association for Computational Linguistics, 2016: 18-27.
- [15] MOSOLOVA A, BONDARENKO I, FOMIN V. Conditional random fields for metaphor detection [C] // **Proceedings of the Workshop on Figurative Language Processing**. New Orleans: Association for Computational Linguistics, 2018: 121-123.
- [16] PRAMANICK M, GUPTA A, MITRA P. An LSTM-CRF based approach to token-level metaphor detection [C] // **Proceedings of the Workshop on Figurative Language Processing**. New Orleans: Association for Computational Linguistics, 2018: 67-75.
- [17] WU C H, WU F Z, CHEN Y B, *et al.* Neural metaphor detecting with CNN-LSTM model [C] // **Proceedings of the Workshop on Figurative Language Processing**. New Orleans: Association for Computational Linguistics, 2018: 110-114.
- [18] YUAN H, WANG J, ZANG X J. YNU-HPCC at SemEval - 2018 task 11: using an attention-based CNN-LSTM for machine comprehension using commonsense knowledge [C] // **Proceedings of The 12th International Workshop on Semantic Evaluation**. New Orleans: Association for Computational Linguistics, 2018: 1058-1062.
- [19] CHE W X, LI Z H, LIU T. LTP: A Chinese language technology platform [C] // **Coling 2010 - 23rd International Conference on Computational Linguistics, Proceedings of the Conference**. Beijing: Tsinghua University Press, 2010: 13-16.

Multi-level metaphor detection method based on Bi-LSTM

ZHU Jiaying, WANG Rongbo*, HUANG Xiaoxi, CHEN Zhiqun

(Institute of Cognitive and Intelligent Computing, Hangzhou Dianzi University, Hangzhou 310018, China)

Abstract: Taking Bi-directional long short-term memory network (Bi-LSTM) as the core, combining with multi convolutional neural network layers and unidirectional long short-term memory network, a multi-level metaphor recognition model is built. Based on the idea of multi-feature synergism, the learning information of the model is enriched by using the methods of inputting multi-feature such as dependency feature, semantic feature and part-of-speech feature in parallel. In order to reduce information interference, a standardized text input method based on statistics is used to improve the recognition effect of the model. In the experiments on word level and sentence level of English corpus, each feature has obvious positive effect. In the word level research of English corpus, the F_1 -scores of cropping and filling treatment, multi-feature synergism increase by 2.5% and 5.1% respectively, while in the sentence level research, the F_1 -scores increase by 3.1% and 1.9% respectively. In the sentence level experiment of Chinese corpus, the F_1 -score of the optimal effect can reach 88.8%.

Key words: natural language understanding; metaphor detection; CNN; Bi-LSTM; dependence relationship