

采用深度学习的小语种舆情监控方法

宋千里^{1,2}, 赖 华^{1,2*}

1. 昆明理工大学 信息工程与自动化学院, 云南 昆明 650500;
2. 昆明理工大学 云南省人工智能重点实验室, 云南 昆明 650500)

摘要: 在小语种舆情监控领域, 由于小语种的标注语料难以获取, 导致深度学习的训练效果较差。对于民间及媒体发表的新闻内容很难准确抽取其核心观点句, 从而影响了进一步的舆情分析效果。为了将研究问题具体化, 以越南语为例, 提出一种融入共享主题特征的汉越跨语言新闻观点句的抽取方法, 可以借助充足的汉语标注语料解决小语种资源稀缺问题, 并利用双语可比语料间可共享的主题信息来优化抽取效果, 进而提升舆情监控效果。具体方法为, 提取汉越可比新闻的隐含狄利克雷分布 (Latent Dirichlet Allocation, LDA) 主题来构建共享主题特征, 借助共享主题词典和情感词典训练双语词嵌入模型来共享汉越语义空间表征, 将特征融入词向量, 通过将语义信息与主题、情感、位置信息相结合来提升抽取效果。在汉越可比新闻数据集里进行的实验结果表明, 融入共享主题特征能够提升小语种新闻观点句的抽取效果, F_1 值达到 0.721, 对小语种舆情监控起到支撑作用。

关键词: 小语种舆情监控; 跨语言观点句抽取; 汉越可比新闻; 双语词嵌入; 共享主题表征
中图分类号: TP391 **文献标志码:** A **DOI:** 10.3788/IRLA20210298

Monitoring method of public opinion in minor languages using deep learning

Song Qianli^{1,2}, Lai Hua^{1,2*}

1. Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Yunnan 650500, China;
2. Yunnan Key Laboratory of Artificial Intelligence, Kunming University of Science and Technology, Yunnan 650500, China)

Abstract: In the field of public opinion monitoring in minor languages, it is difficult to obtain annotated corpus in minor languages, resulting in poor practice of deep learning. It is difficult to extract the core opinions of the information published by the private and the media for further analysis of public opinion. Taking Vietnamese as an example, a method for extracting Chinese-Vietnamese news opinion sentences that incorporated shared topic features was proposed. The problem of scarcity of small language resources was solved with the help of sufficient Chinese annotation corpus, and the topic information was used to shared between bilingual comparable corpora. It could optimize the extraction effect, and then enhance the public opinion monitoring effect. First, the topics of Chinese-Vietnamese comparable news were extracted separately to construct shared topic features through LDA topic modeling; then, the bilingual word embedding model was trained to achieve shared semantic spatial

收稿日期: 2021-05-08; 修订日期: 2021-05-26

基金项目: 国家自然科学基金 (61972186, 61762056, 61472168); 云南省重大科技专项计划项目 (202002AD080001)

作者简介: 宋千里, 男, 硕士生, 主要从事自然语言处理、小语种跨语言情感方面的研究。

导师(通讯作者)简介: 赖华, 男, 副教授, 硕士生导师, 主要研究方向为复杂过程建模与控制、智能信息处理。

representation of Chinese and Vietnamese; Finally, the features were integrated with word vectors for combining semantic information with topics, emotion and location information to enhance the effect of extracting consequent. The experimental results in the Chinese Vietnamese comparable news dataset show that the integration of shared topic features can improve the extraction of Chinese Vietnamese news opinion sentences, the value of F_1 is 0.721, which supports the monitoring of public opinion in minor languages.

Key words: public opinion monitoring in minor languages; cross-lingual opinion sentence extraction; Chinese-Vietnamese comparable news; bilingual word embedding; shared topic representation

0 引言

在当今国际形势下,为了更好地把握与周边国家的国际关系和地缘政治,以越南为例,需要实时监控越南民间和官方媒体的舆情动向。目前主流的方法是利用爬虫获取大量训练语料,再通过深度学习获得舆情监控结果,而小语种有标注资源稀缺,需要通过其他语言语料进行辅助训练。

在舆情监控领域,观点句抽取是舆情监控的重要支撑,影响着后续对观点的情感分析效果。跨语言观点句抽取任务的核心是通过丰富的源语言标注资源弥补稀缺目标语言标注资源,准确且高效地抽取出篇章中代表观点的句子。单语观点句的抽取问题较普遍,而跨语言领域的观点句抽取问题研究较少,具有一定的研究价值。

观点句抽取任务是指给出一个包含多个句子的文档,识别并抽取文档中表达篇章观点的句子二分类问题,新闻的观点句也通常需要依赖观点句特征进行抽取。如 Pang^[1]利用 unigram 特征训练 SVM 分类模型和朴素贝叶斯分类模型将电影评论分为正面和负面两类。还有一部分方法融入了情感词典和情感特征,以及通过加权强化对重点信息的关注,如 Liu 等^[2]通过抽取主客观特征进行句子的序列标注来获取观点句。Zhao 等^[3]采用了集成学习方法,实现了基于句子的主题、位置、情感、特征词词性的观点句识别。

跨语言观点句抽取是在单语观点句抽取研究的基础上增加了对源语言的利用,主要分为基于双语词典、机器翻译、平行语料和双语词嵌入的方法进行跨语言观点句抽取。核心思想都是将源语言语料迁移到目标语言语义空间,弥补目标语言资源稀缺的问题,提升目标语言观点句抽取效果。

基于双语词嵌入方法属于目前的主流做法,实现了目标语言和源语言的语义空间对齐,核心问题转变为如何解决不同语言的语义表达差异问题^[4]。Zhang 等^[5]利用 tf-idf 词典和 LDA (Latent Dirichlet Allocation) 主题词典构建共享语义空间以实现语义空间共享。Liu 等^[6]采用了结合要素关联和情感关联的汉越双语新闻观点句抽取方法。Lin 等^[7]在双语词嵌入模型基础上,在分类模型中融入了主题、位置和情感特征来实现跨语言观点句识别。现有的特征与跨语言结合的方法大多将融入的特征和观点句抽取作为两个独立的部分,没有充分利用汉越语料间的关联关系,对汉语资源的利用不足,影响了最终的目标语言观点句抽取效果。

考虑到跨语言新闻篇章可能描述同一事件主题内容的特点,用汉越可比新闻作为训练资源进行研究。通过对主题接近的双语可比语料进行分析,发现可比语料描述的主题高度一致,主题词信息有差异,情感信息接近,无法互相转译。主题信息与观点句抽取任务有着紧密的联系,所以获取的高质量的句子主题信息可以提升观点句抽取效果。

综上所述,文中结合跨语言新闻篇章可能描述同一事件主题内容的特点,用汉越可比新闻作为训练资源,提出一种融入共享主题信息,结合深度学习框架和共享语义空间,融合多特征的跨语言新闻篇章观点句抽取方法。

1 汉越可比新闻观点句抽取算法

文中提出融入主题特征的汉越新闻观点句抽取模型,包含共享主题模块,双语词嵌入模块和观点句抽取模块。图 1 为融入共享主题特征的汉越新闻观点句抽取模型流程图。

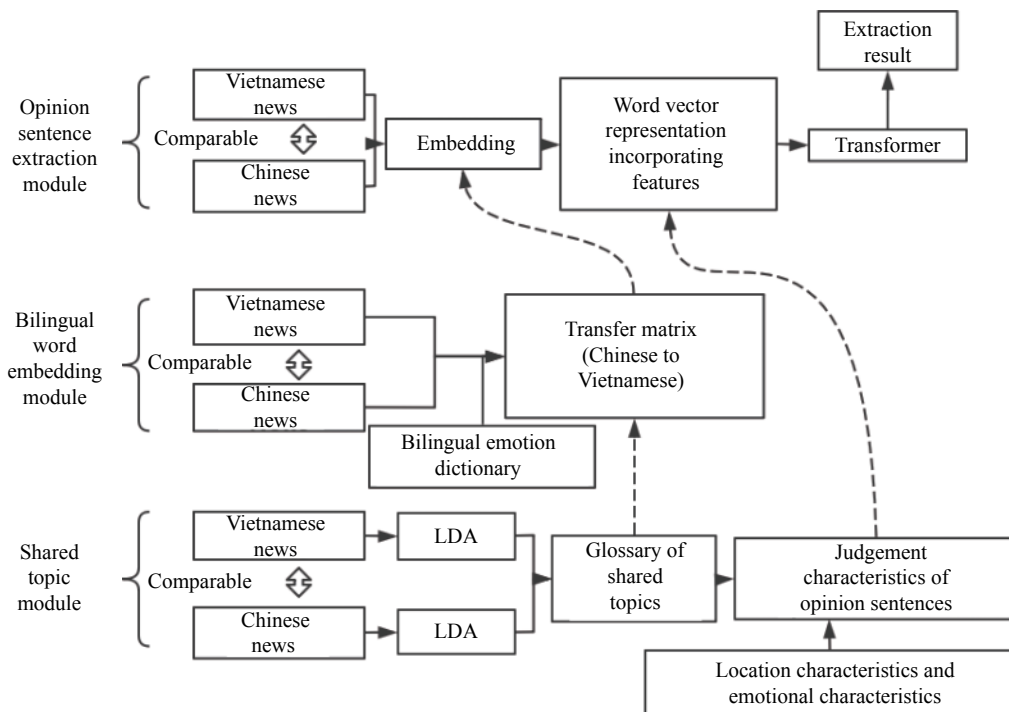


图 1 融入共享主题特征的汉越新闻观点句抽取流程图

Fig.1 Flow chart of Chinese-Vietnamese news opinion sentence extraction incorporating the characteristics of shared topics

1.1 共享主题信息的获取

文中核心思想是利用可比新闻篇章语料间的主题关联关系,通过共享主题特征来优化观点句抽取效果。所提到的共享主题信息即是越南语新闻主题和其可比汉语新闻的主题。因此训练语料资源采用汉越可比语料,采用 LDA 主题模型来获取新闻的主题特征。

隐含狄利克雷分布 (Latent Dirichlet Allocation, LDA)^[8]可以借助无监督学习来获取大规模的文档的主题分布信息,并且以词袋模型的形式进行文本表示,最后获得每篇文档的主题分类及该分类下的主题词语。对于大规模文档的集合,通过主题分布模块获取一篇文档的主题,从该主题对应的单词分布中取出一个单词,遍历所有单词以获取整个文档的 LDA 主题分类类别,其中每个类别有其独有的 LDA 主题词表,每个文档都有其类别分类^[9]。

主题特征信息用双语主题词表 B 表示。对一篇越南语新闻,其主题词表提取前 n 个词 $c_1 \sim c_n$,其可比汉语主题词表提取前 n 个词 $c'_1 \sim c'_n$ 。所以对于一篇越南语新闻可以获得 $2n$ 个词组成的共享主题词表 B 。

$$c = \{c_1, c_2, \dots, c_n\} \quad (1)$$

$$c' = \{c'_1, c'_2, \dots, c'_n\} \quad (2)$$

$$B = \{c, c'\} \quad (3)$$

LDA 不需要做提前标注,只需要对主题类别数量做出定义即可。经过训练可以得到越南语新闻的主题类别,主题词表及其对应可比汉语新闻的主题词表,两个词表经过拼接可以得到共享主题特征矩阵,在抽取任务中作为模型引入的外部知识增强抽取效果。同时得到的共享主题词表作为双语词嵌入模型做监督信号。

1.2 共享语义空间的构建

双语词嵌入模型可以使源语言和目标语言中的近义词在共享语义空间中的距离靠近。目前主流的双语词嵌入模型构建主要通过丰富的双语平行语料进行。汉越平行语料目前没有大规模公开数据,不容易获取。无监督双语词嵌入的效果在大规模语料上效果明显,但是对于越南语这种小规模语料效果很差,文中采用半监督结合外部词典方式优化双语词嵌入的效果。

将 LDA 主题抽取模型获得的汉语主题词表 c 翻译成越南语 d ,越南语主题词表 c' 翻译成汉语 d' ,构成

双语主题词典 D 。对汉越 hownet 情感词表进行翻译得到双语情感词典 E 。双语关键词词典 $F = \{D, E\}$ 。

篇章中的每个词先通过一种基于神经网络的词嵌入模型 (word2vec) 获得单语词向量表征, 越南语语义空间为 Y , 汉语语义空间为 X 。 F 中包含越南语和汉语的翻译对, $F = \{Y_i, X_i\}_{i=1}^n$, 其中 i 为翻译对序号。因为任务目标是解决越南语的观点句抽取问题, 因此考虑将汉语语料转移到越南语的语义空间获得其表示, 需要获得从汉语到越南语的转移矩阵 M 。

$$X_i^Y = X_i \cdot M \quad (4)$$

式中: X_i^Y 为汉语第 i 个词在越南语语义空间的表示。

由于获得了双语重点词词典 D , 在越南语义空间下, 越南语及其翻译结果应该空间分布接近, 即词典中翻译对的距离最小, 其平均平方差 L 为:

$$L = \frac{1}{n} \sum_{i=1}^n (X_i^Y - Y_i^X)^2 \quad (5)$$

由于已获取双语情感词典 E , 对于汉越两种语言笔者可以分别获取其中的正面情感词典 X_{pos} 、 Y_{pos} 和负面情感词典 X_{neg} 、 Y_{neg} 。 Zhang 等^[5] 曾提出令含有不同情感极性的情感词在语义空间中空间距离最大, 使获得的双语词嵌入在情感区分上更加明显, 利于后续对情感倾向的分析。训练时使投影后的不同极性情感词的平均余弦距离 L' 最小。

$$L' = \cos(X_{pos}, X_{neg}) + \cos(Y_{pos}, Y_{neg}) \quad (6)$$

通过使共享空间内源语言和目标语言的分布接近, 且使不同极性的情感词在共享空间内分开, 最终根据公式 (4) 得到从汉语到越南语的转移矩阵 M 。

1.3 融入共享主题的汉越新闻观点句抽取模型

上述工作已获得每篇新闻的双语主题词表 B 及每个共享语义空间的词向量表示。传统的观点句抽取是依赖观点句特征进行分类, 本质是一个二分类任务。文中提出一种基于 Transformer 深度学习模型, 融入主题特征的跨语言观点句抽取模型。所提模型中的观点句特征包括共享主题特征、句子位置特征和句子情感特征。图 2 为抽取模型原理图。

句子的情感特征, 用来判断该句子是否带有情感倾向。有情感倾向的多为观点表达句, Lin 等^[7] 就通过情感词典判断句子里是否含有情感表达词汇。通

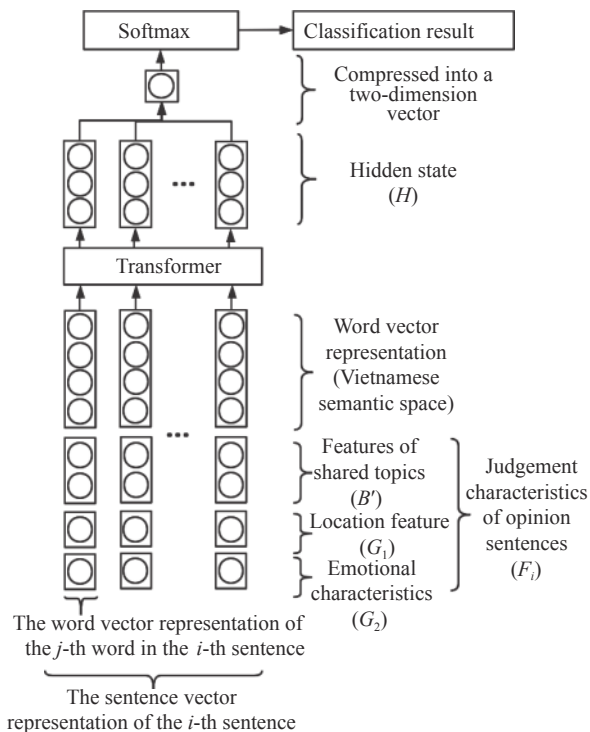


图 2 融入观点句判别特征的观点句抽取模型

Fig.2 Opinion sentence extraction model incorporating the discriminative features of opinion sentences

过计算得出句子所包含的情感表达强烈程度, 定义为情感分数 g_1 :

$$g_1 = \frac{\sum_{k=1}^m emotion(w_{i,k})}{m} \quad (7)$$

式中: m 是该句中的词的个数; $w_{i,k}$ 是第 i 句中的第 k 个词。 $emotion(w_{i,k})$ 是情感词判定函数, $w_{i,k}$ 不是情感词, 则 $emotion(w_{i,k})$ 为 0; 是情感词, 则 $emotion(w_{i,k})$ 为 1。

$$emotion(w_{i,k}) = \begin{cases} 0, w_{i,k} \notin E \\ 1, w_{i,k} \in E \end{cases} \quad (8)$$

句子位置特征, 用来判断句子在文档中是否符合观点句的位置特点。新闻文档的开头或结尾多数为主旨句, 更容易出现作者的观点, 属于总结性的句子。因此使文本中位置靠前或靠后的句子获得较高的权重, 最后可以计算得到句子在文档中的位置分数 g_2 :

$$g_2 = \frac{1}{\min(i, n - i + 1)} \quad (9)$$

式中: n 为该篇文章的句子总数; i 表示该句在该篇章所有句子中所处的位置为第 i 句。

$$C = emb(c) \in R^{1 \times dim} \quad (10)$$

$$C' = emb(c') \in R^{1 \times dim} \quad (11)$$

$$G_1 = emb(g_1) \in R^{1 \times dim} \quad (12)$$

$$G_2 = emb(g_2) \in R^{1 \times dim} \quad (13)$$

式中: $emb()$ 为随机初始化的词嵌入过程, 对 c, c', g_1, g_2 通过词嵌入进行向量化表示, 再将维度统一为 dim_1 , 对统一维度后的向量进行拼接, 拼接后维度为 dim_2 。其中, $dim_1=128, dim_2=512$ 。共享主题特征 $B'=C \oplus C'$ 。

为每个句子计算其观点句判别特征矩阵 F_i , 并拼接到该句子的每个词向量后方, 每个词可获得 dim_3 维的带有观点句判别特征的词向量表示 F'_i 。

$$F_i = (C \oplus C' \oplus G_1 \oplus G_2) R^{1 \times dim} \quad (14)$$

式中: i 为句子序号; $dim_3=1024$ 。

经过 Transformer 编码端获取含有观点句判别特征信息的句子表征, 通过自注意力机制能更好地关注该句的观点句判别信息。自注意力机制是由多头注意力 Multi-Head Attention 实现, 在 Attention 机制的基础上对输入多路线性变换后, 计算 Attention 结果, 再

拼接后, 通过线性变换得到输出。其中的求和与归一化可以促进梯度传播和模型收敛。

将通过抽取模型生成的句子表征即网络中每个神经元输出的隐向量组成的矩阵 H 压缩成二维向量表示。最后通过 softmax 将其转化成条件概率分布, 判别每个句子是否是该篇新闻的观点句。

$$y = \text{softmax}(WH + b) \quad (15)$$

式中: W 和 b 是参数。

最后使用交叉熵作为模型损失进行计算, 使模型损失最小化, 得到训练后的模型。

$$loss = - \sum_i \sum_j y_i^j \log \hat{y}_i^j + \lambda \|\omega\|^2 \quad (16)$$

式中: $\lambda \|\omega\|^2$ 是 L_2 正则化项, 防止模型过拟合; y 是第 i 个句子的目标分布; \hat{y} 是通过模型预测的观点句分布; j 是观点句判定结果。

2 实验

2.1 数据准备

文中借助爬虫获取的大量新闻语料, 从中选择部分含观点的越南语新闻及与其可比的汉语新闻语料。表 1 为语料篇章划分情况。

表 1 汉语和越南语新闻语料的训练集、测试集、验证集分布

Tab.1 Distribution of training set, test set, and verification set of Chinese and Vietnamese news corpus

	Number of Chinese news articles	Number of Vietnamese news articles
Training set	450	450
Test set	25	25
Validation set	25	25

借助 jieba 分词工具对中文新闻语料进行分词, 借助 Vncorenlp 工具^[10]对越南语新闻语料进行分词, 最后利用停用词表去停用词。对语料进行人工观点句标注, 标注出符合篇章观点表达的句子作为汉越新闻观点句抽取的数据集, 每篇可能有多个观点句。

2.2 评价指标

采用准确率 P 、召回率 R 、 F_1 值作为评价指标,

$$P = \frac{a}{a+b} \times 100\% \quad (17)$$

$$R = \frac{a}{a+c} \times 100\% \quad (18)$$

$$F_1 = \frac{2 \times P \times R}{P+R} \times 100\% \quad (19)$$

式中: a 为预测正确的次数; b 为将负例判断为正例次数; c 为将正例判断为负例的次数。

2.3 参数设置

采用 word2vec 训练单语词嵌入, 词嵌入大小设置为 512 维。对汉语和越南语用 skip-gram^[11] 分别训

练单语词嵌入。下采样为 10^{-4} ,负采样为 10。去掉出现 2 次以下的词,上下文滑动窗口大小设置为 5,双语词嵌入学习率为 0.1。

LDA 模型中的主题数设置为 5 类,每个主题获取前 n 个主题词, n 设为 10。观点句判别特征矩阵维度 dim_2 设为 512 维,子特征维度 dim_1 设为 128 维。

采用 Transformer 编码层获取越南语新闻语义特征,训练批次大小设置为一个新闻篇章,训练 200 轮,学习率设置为 0.1,正则化权重为 0.01。对数据进行筛选,句子数量限制在 20 以内,句子词数限制在 40 以内,对每个句子标注其篇章序号,输出结果可以显示每个句子是否为其所在篇章的观点句。在模型训练过程中加入梯度裁剪,最大梯度裁剪为 5,

dropout 设置为 0.3。

2.4 对比实验

为了验证观点句抽取采用不同模型对观点句抽取效果的影响,文中设计了实验一。Lin 等^[12]在跨语言观点句抽取任务中采用 Bi-LSTM(双向循环神经网络)结合多种特征融入方法,取得了不错的效果,因此文中实验在其研究基础上,采用 LSTM(长短期记忆网络)、Bi-LSTM、Transformer 模型进行观点句抽取对比实验。如表 2 所示,LSTM 效果最差,Bi-LSTM 可以更好的关注上下文信息,效果有明显提升。Transformer 模型效果最好,其自注意力机制可以自动对提升抽取效果的重点信息给予重点关注,最终提高了观点句抽取效果。

表 2 不同模型下的观点句抽取效果对比

Tab.2 Comparison of the effect of opinion sentence extraction under different models

Opinion sentence extraction model	P	R	F_1
LSTM + opinion sentence discriminative features	0.623	0.639	0.631
Bi-LSTM + discriminant features of opinion sentences	0.658	0.667	0.662
Transformer+discriminant features of opinion sentences	0.711	0.732	0.721

2.5 消融实验

实验二对比了将各特征融入词向量对观点句抽取效果的影响。在词向量拼接时,将共享主题,位置,情感维度分别变换成 $\text{dim}_2=512$ 维,再与词向量拼接。如表 3 所示,只融入一种判别特征时,提升效果最高为情感特征,然后是共享主题、位置特征。融入三种

特征的效果要好于单特征。同时对比单语主题与共享主题区别,其中采用共享主题特征拼接的抽取效果要比单语主题特征拼接提升 0.071 的 F_1 值。该实验说明了三种特征拼接融入的有效性,也证明了用汉越共享主题取代单语主题的合理性。

表 3 不同特征下的观点句抽取效果对比

Tab.3 Comparison of the effect of opinion sentence extraction under different characteristic

Discriminant features of opinion sentences	P	R	F_1
None	0.638	0.650	0.644
Shared topic	0.686	0.695	0.690
Position	0.664	0.683	0.673
Emotion	0.699	0.707	0.703
Vietnamese theme features + location + emotion	0.676	0.709	0.650
Shared theme topic + location + emotion	0.711	0.732	0.721

2.6 实例分析

通过表 4 中实例的观点句抽取结果可以发现,可比汉越新闻虽然描述同一个主题和内容,但是其中可能含有不同的观点表达、语义表达和主题词,而这些

可以在汉越新闻间共享适用的主题词无疑可以对观点句判别特征起到丰富加强的作用。结论是单独融入单语主题信息可能造成观点句判别错误或漏判,融入共享主题特征可以强化观点句的抽取效果。

表 4 汉越新闻实例的观点句抽取效果对比

Tab.4 Comparison of the effect of opinion sentence extraction in Chinese-Vietnamese news cases

	Chinese news	Vietnamese news
Text	越通社河内—12月21日和22日,越南有关部门、越南驻美国大使馆、越南国家航空公司同当地政府有关部门配合,将在美国滞留的近360名越南公民安全接回国。...今后,将在国外滞留的公民接回国工作将继续根据公民的愿望和国内疫情和隔离能力等情况展开。	Trong hai ngày 21-22/12, các cơ quan chức năng Việt Nam, các cơ quan đại diện Việt Nam tại Hoa Kỳ, hãng Hàng không Quốc gia Việt Nam đã phối hợp với các cơ quan chức năng sở tại đưa gần 360 công dân Việt Nam về nước an toàn....Thời gian tới, việc đưa công dân có hoàn cảnh đặc biệt khó khăn về nước sẽ được sắp xếp theo nguyện vọng của công dân, phù hợp với tình hình dịch bệnh và năng lực cách ly trong nước.
Subject headings	越南, 疫情...	Việt Nam(越南), cách ly (隔离)...
No shared theme feature	Opinion sentence number is 1, 8(8 sentences in total)	Opinion sentence number is 1(9 sentences in total)
Incorporate shared theme features	Opinion sentence number is 1, 8(8 sentences in total)	Opinion sentence number is 1, 9(9 sentences in total)
Artificially annotated opinion sentence	Opinion sentence number is 1, 8(8 sentences in total)	Opinion sentence number is 1, 9(9 sentences in total)

3 结 论

通过该模型的相关实验,可以得出如下结论:对越南语观点句抽取任务,在训练时采用双语词嵌入引入汉语标注语料,在词向量中融入包含共享主题特征、位置特征、情感特征的观点句判别特征,采用Transformer做抽取模型,都会提升越南语观点句抽取效果(实验结果中 F_1 值达到0.721),进而可以提升小语种舆情监控效果。在下一步研究中将继续研究如何利用已获取的观点句抽取结果来提升越南语新闻情感分类的效果,以及在其他小语种中应用该方法的具体参数调整。

参考文献:

[1] Pang Bo, Lee Lillian, Vaithyanathan Shivakumar, et al. Thumbs up? Sentiment classification using machine learning techniques[C]//Proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP-2002), 2002: 79-86.

[2] Liu P Y, Xun J, Fei S D, et al. Subjective sentence recognition based on Hidden Markov Model [J]. *Journal of Chinese Information Processing*, 2016, 30(4): 206-212. (in Chinese)

[3] Zhao H J, Liu H L, Ren J W, et al. News-oriented emotional key sentence extraction and polarity determination [J]. *Journal of Shanxi University (Natural Science Edition)*, 2014, 37(4): 588-594. (in Chinese)

[4] Wang J, Tang S, Hang Y X, et al. Chinese-Vietnamese bilingual multi-document news opinion sentence recognition based on sentence association graph [J]. *Journal of Computer*

Applications, 2020, 40(10): 2845-2849. (in Chinese)

[5] Zhang M M. Cross-language sentiment classification based on shared space [J]. *Information Technology and Informatization*, 2020(5): 202-207. (in Chinese)

[6] Liu S L, Research on extraction and analysis methods of Chinese and Vietnamese bilingual news opinion sentences [D]. Kunming: Kunming University of Science and Technology, 2017. (in Chinese)

[7] Lin S Q, Yu Z T, Guo J J, et al. Chinese-Vietnamese news perspective sentence extraction methods incorporating multiple features [J]. *Journal of Chinese Information Processing*, 2019, 33(11): 101-106. (in Chinese)

[8] Wang Q, Tian M J, Cui R Y, et al. Bilingual topic word embedding for Chinese-Korean cross-lingual text classification [J]. *Journal of Chinese Information Processing*, 2020, 34(12): 39-47. (in Chinese)

[9] Kang C, Zheng S H, Li W L. Short text classification combining LDA topic model and 2D convolution [J]. *Computer Applications and Software*, 2020, 37(11): 127-131, 153. (in Chinese)

[10] Vu T, Nguyen D Q, Nguyen D Q, et al. VnCoreNLP: A vietnamese natural language processing toolkit[C]//Proceedings of NAACL-HLT, 2018: 56-60.

[11] Zhang Jing. Research on viewpoint extraction of chinese comments based on deep learning [D]. Chengdu: Southwest Jiaotong University, 2018. (in Chinese)

[12] Lin S Q, Yu Z T, Guo J J, et al. Chinese-Vietnamese bilingual news sentiment classifications incorporating perspective sentence features [J]. *Journal of Kunming University of Science and Technology (Natural Science Edition)*, 2020, 45(6): 67-73. (in Chinese)