

融合双通道音节特征的藏文La格例句自动分类模型

班玛宝^{1,2,3} 才让加^{1,2,3,4,5,†} 张瑞^{1,2,3} 色差甲^{1,2,3} 卓玛扎西^{1,2,3}

1. 青海师范大学计算机学院, 西宁 810016; 2. 藏语智能信息处理及应用国家重点实验室, 西宁 810008;
3. 青海省藏文信息处理工程技术研究中心, 西宁 810008; 4. 青海省藏文信息处理与机器翻译
重点实验室, 西宁 810008; 5. 藏文信息处理教育部重点实验室, 西宁 810008;

† 通信作者, E-mail: zwxzx@163.com

摘要 基于藏文La格(ལ་རྩོམ།)例句的自动分类在藏语自然语言处理领域的重要性, 根据藏文La格的用法和添加规则, 在对藏文La格例句进行分类并定义分类概念的基础上, 提出一种融合双通道音节特征的藏文La格例句自动分类模型。该模型首先使用word2vec和Glove构建双通道藏文音节嵌入, 分别在每路卷积中融合双通道音节特征, 丰富输入特征的表达和提高卷积层的空间表征能力; 然后在每一路卷积均使用结合层级注意力机制的Bi-LSTM学习时序特征后, 拼接多路特征, 提高上下文时序特征的学习能力; 最后通过全链接层和Softmax层实现藏文La格例句自动分类。实验结果表明, 该模型在测试集上的藏文La格例句分类准确率达到90.26%。

关键词 自然语言处理; 双通道音节特征; 藏文La格例句; 自动分类

An Automatic Classification Model of Tibetan La Case Example Sentences with Fusion Dual-channel Syllable Features

BAN Mabao^{1,2,3}, CAI Rangjia^{1,2,3,4,5,†}, ZHANG Rui^{1,2,3}, SE Chajia^{1,2,3}, ZHUO Mazhaxi^{1,2,3}

1. College of Computer Science and Technology, Qinghai Normal University, Xining 810016; 2. The State Key Laboratory of Tibetan Intelligent Information Processing and Application, Xining 810008; 3. Tibetan Information Processing Engineering Technology and Research Center of Qinghai Province, Xining 810008; 4. Tibetan Information Processing and Machine Translation Key Laboratory of Qinghai Province, Xining 810008; 5. Key Laboratory of Tibetan Information Processing, Ministry of Education, Xining 810008; † Corresponding author, E-mail: zwxzx@163.com

Abstract Based on the importance of automatic classification of Tibetan La case (ལ་རྩོམ།) example sentences in Tibetan natural language processing, according to the usage and adding rules of Tibetan La case, this paper classifies Tibetan La case example sentences and defines the classification concept, and proposes an automatic classification model of Tibetan La case example sentences with fusion dual-channel syllable features. The proposed model first uses word2vec and Glove to construct a dual-channel Tibetan syllable embedding, and combines the dual-channel syllable features in each convolution respectively to enrich the expression of input features and improve the spatial representation ability of the convolutional layer. Then in each convolution, the Bi-LSTM combined with the hierarchical attention mechanism is used to learn the timing features, and the multi-channel features are spliced to improve the learning ability of the context timing features. Finally, the automatic classification of Tibetan La case example sentences is realized through the full link layer and the Softmax layer. Experiments show that proposed model has an accuracy of 90.26% in the classification of Tibetan La case example sentences on the test set.

Key words NLP; dual-channel syllabic features; Tibetan La case example sentences; automatic classification

格虚词 ཏུ 表述了某一事物 ཐུ་ཡིག (汉文) 译成另一种事物 ཐོད་ཡིག (藏文) 后在变化结果上具有“同一性”。

时格句: 表示实施某一动作行为时间的句子 (བྱ་བ་གང་ཞིག་བྱེད་པའི་དུས་སྟོན་པའི་ཚེགས།^[13])。如“ཁོང་གིས་ལོ་གསུམ་ལ་བསྐྱར་སྦྱོང་བྱས། (他复习了三年)”, 用 La 格虚词 ལ 将实施动作的时间 ལོ་གསུམ (三年) 和实施的动作 བསྐྱར་སྦྱོང (复习) 连接后, 准确地表达了句意。

1.3 藏文 La 格例句的分类概念

藏文 La 格例句分类的主要任务是, 通过藏文 La 格例句分类模型, 给定一个待分类的藏文 La 格例句集中每个句子的类别标签。该任务可以用数学符号直观地表达, 假定 $LD = \{(x_i, y_j) | 1 \leq i \leq n, 0 \leq j \leq 4\}$ 为训练样本集, 则利用设计的模型学习到一个映射 $f: X \rightarrow Y$, 即藏文 La 格例句集到例句类别标签集的映射, 其中 $x_i \in X$ 表示第 i 个藏文 La 格例句, $y_i \in Y$ 表示例句 x_i 所对应的类别标签。该映射关系见图 1。

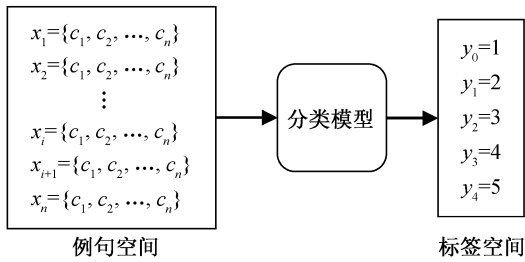


图 1 藏文 La 格及其类别的映射关系

Fig. 1 Tibetan La Case example sentences category relation

图 1 中, 例句空间 X 中包含 n 个藏文 La 格例句, 通过藏文 La 格例句分类模型, 可以得到这些例句对应的标签空间 Y , 标签空间 Y 中包含 m 个类别标签。测试过程中, 通过分类模型, 每一个藏文 La 格例句可获得对应的标签, 获得标签的过程就是藏文 La 格例句的自动分类过程。

2 藏文 La 格例句自动分类模型

融合双通道音节特征的藏文 La 格例句自动分类模型如图 2 所示。模型的输入为双通道音节嵌入, 分别为基于 Word2vec^[14] 和 GloVe^[15] 的音节向量, 可以增加文本表示的空间维度和特征多样性, 丰富特征的表达, 且不受特定分类任务的影响, 进而利于提升模型的泛化能力。与单通道音节嵌入相比, 随着模型训练过程的进行, 在反向传播时对双通道嵌入层的权值进行动态调整, 使之从原本与藏文 La 格例句分类任务关系不大的音节向量变成与该任务相关的语义表征向量, 提高了模型的收敛速度和句子特征表示的准确性。模型采用多路卷积提取空间特征, 每一路使用不同大小的卷积核提取不同感受视野的局部空间特征。为了避免多路卷积特征融合过程对特征时序造成破坏, 本文在每一路均使用结合层级注意力机制的双向长短期记忆网络学习时序特征, 并对各路特征进行拼接, 获得藏文 La 格例句

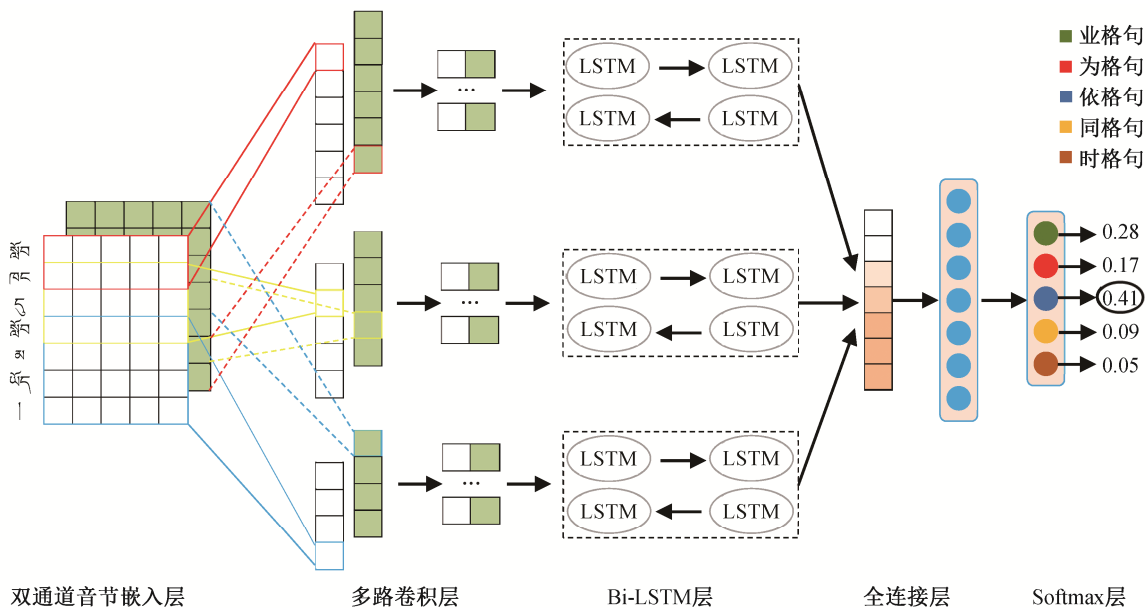


图 2 融合双通道音节特征的藏文 La 格例句自动分类模型

Fig. 2 An automatic classification model of Tibetan La case sentences with two channel syllable features

的最终表示。最后,通过全链接层和 Softmax 层进行藏文 La 格例句的自动分类。

2.1 融合双通道音节特征的卷积层

如图 3 所示,卷积层的输入为双通道音节嵌入矩阵,通过与特定大小的卷积核进行卷积操作,得到两张不同的特征图,之后进行特征融合。令嵌入矩阵最多可包含 n 个音节,音节数大于 n 的句子将被截断,不足 n 的则用 0 填充。假设 $\mathbf{x}_i \in \mathbb{R}^d$ 为当前句子中第 i 个音节的 d 维向量,则句子的音节嵌入矩阵 $\mathbf{X}_{1:n}$ 用式(1)表示。

$$\mathbf{X}_{1:n} = \mathbf{x}_1 \otimes \mathbf{x}_2 \otimes \cdots \otimes \mathbf{x}_n, \quad (1)$$

其中, \otimes 表示拼接运算。卷积操作需在音节嵌入矩阵 $\mathbf{X}_{1:n}$ 上按时间轴自上而下地滑动卷积核 $\mathbf{W}_c \in \mathbb{R}^{hk}$ 来完成。 h 为卷积核的感受视野大小, k 为音节嵌入的维度。特征音节 c_i 是卷积核在音节 $\mathbf{X}_{i:i+h-1}$ 上滑动生成:

$$c_i = f(\mathbf{W}_c \cdot \mathbf{X}_{i:i+h-1} + b_c), \quad (2)$$

其中, f 是非线性激活函数, $b_c \in \mathbb{R}$ 是偏置项。卷积核 \mathbf{W}_c 在句子 $\{\mathbf{X}_{1:h}, \mathbf{X}_{2:h}, \dots, \mathbf{X}_{n-h+1:n}\}$ 上进行卷积生成的特征图 \mathbf{C} 如下:

$$\mathbf{C} = [c_1, c_2, \dots, c_{n-h+1}], \quad (3)$$

由于嵌入层为双通道,故通过卷积后可得到两张不同的特征图 \mathbf{C}_1 和 \mathbf{C}_2 。使用大小为 1×1 , 深度为 2 的卷积核 \mathbf{W}_f 对 \mathbf{C}_1 和 \mathbf{C}_2 进行通道特征融合后的特征图 \mathbf{M} 为

$$\mathbf{M} = f(\mathbf{W}_f \cdot [\mathbf{C}_1, \mathbf{C}_2] + b_f), \quad (4)$$

其中, f 代表非线性函数, b_f 代表偏置项。得到融合双通道音节特征 \mathbf{M} 后,由于卷积神经网络通常会使用 N 个相同大小的卷积核进行卷积运算,所以卷积后可以生成 N 个特征图组成的特征矩阵 \mathbf{M}_o :

$$\mathbf{M}_o = [\mathbf{M}_1, \mathbf{M}_2, \dots, \mathbf{M}_N], \quad (5)$$

由于卷积后的特征矩阵 \mathbf{M}_o 的行维度较大,所以需要进行降维处理。为了不丢失特征的时序性,本文没有采用池化降维法,改为使用步幅为 k 的卷积核 \mathbf{W}_p 对特征矩阵 \mathbf{M}_o 进行卷积降维,降维后生成的特征矩阵 \mathbf{M}_k :

$$\mathbf{M}_k = f(\mathbf{M}_p \cdot \mathbf{M}_o + b_p), \quad (6)$$

其中, f 为线性函数, b_p 为偏置项。 \mathbf{M}_k 保留了时序特

征,可以按行顺序输入到 LSTM,完成时序特征的学习。

2.2 结合注意力机制的 Bi-LSTM 层

与主题分类和文本分类等任务相比,藏文 La 格例句分类任务没有特定属于某一类的特征字或词,使得藏文 La 格例句的分类难度加大,更依赖于上下文语义信息和语义关联度,对特征的时序性要求更高。在输入 LSTM 前,如果先融合多路卷积,则无法保证特征的有序性,造成特征质量下降,影响 LSTM 对句子时序及语义信息的学习过程。所以,本文模型在每一路卷积上均使用 Bi-LSTM 学习时序特征,再拼接多路时序特征来表示最终的藏文 La 格例句,从而避免各路特征在进入 Bi-LSTM 前因融合而导致的特征质量下降问题。为了使模型能够根据上下文信息提高时序特征的学习能力和充分利用每个时刻的输出特征,本文使用结合层级注意力机制的 Bi-LSTM 来提高 LSTM 输出的特征质量。结合层级注意力机制的 Bi-LSTM 层如图 4 所示。

LSTM 按时序接受 \mathbf{x}_t 。LSTM 的最终输出可以由 h_t 表示,其计算过程如下:

$$\mathbf{X} = [h_{t-1}, \mathbf{x}_t], \quad (7)$$

$$i_t = \sigma(\mathbf{W}_i \cdot \mathbf{X} + b_i), \quad (8)$$

$$f_t = \sigma(\mathbf{W}_f \cdot \mathbf{X} + b_f), \quad (9)$$

$$o_t = \sigma(\mathbf{W}_o \cdot \mathbf{X} + b_o), \quad (10)$$

$$C_t = o_t \odot \tanh(\mathbf{W}_c \cdot \mathbf{X} + b_c) + f_t \odot C_{t-1}, \quad (11)$$

$$h_t = o_t \odot \tanh(C_t), \quad (12)$$

其中, \mathbf{x}_t 表示第 $i \in [0, t]$ 时刻的输入向量, C_t 表示 LSTM 的单元状态。 i_t , f_t 和 o_t 分别表示输入门、遗忘门和输出门的特征值, σ 表示 Sigmoid 激活函数, \mathbf{W}_i , \mathbf{W}_f , \mathbf{W}_o , \mathbf{W}_c , b_i , b_f , b_o 和 b_c 分别为各个门需要学习的参数, \mathbf{X} 表示 $t-1$ 时刻隐藏层的值以及 t 时刻的输入。

由于本文使用 Bi-LSTM 学习时序特征,所以 Bi-LSTM 的输出为正向和反向 LSTM 输出的拼接,用 h_t 表示:

$$h_t = [\vec{h}_t \oplus \overleftarrow{h}_t]. \quad (13)$$

由于 LSTM 只能学习到最后一刻的输出向量,不能充分地利用每一时刻的输出,所以本文使用层级注意力机制对各个时刻的输出进行加权融合。假

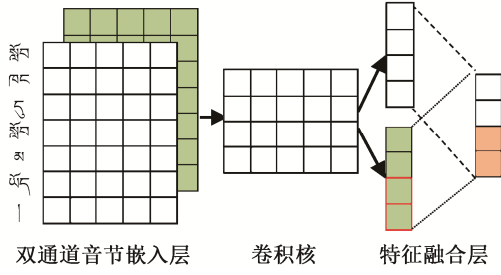


图3 卷积层的设计原理

Fig. 3 Design principle of the convolution layer

设 h_i 表示第 i 时刻 Bi-LSTM 层的输出向量, e_i 表示 h_i 对整个 La 格例句语义表示的重要程度, a_i 表示 h_i 对整个 La 格例句语义表示贡献的权重, 则 Bi-LSTM 层的注意力权重为

$$e_i = \mathbf{u}^T \cdot \tanh(\mathbf{W}_a \cdot \mathbf{h}_i + b_a), \quad (14)$$

$$a_i = \frac{\exp(e_i)}{\sum_{j=0}^t \exp(e_j)}, \quad (15)$$

其中, \mathbf{u}^T , \mathbf{W}_a 和 b_a 都是网络需要学习的参数, \tanh 为非线性激活函数。

至此, 我们已得到 Bi-LSTM 每一时刻的注意力权重, 对其进行加权求和, 便可得到整个 Bi-LSTM 层输出的最终特征向量 \mathbf{Q} :

$$\mathbf{Q} = \sum_{i=0}^t a_i \cdot \mathbf{h}_i. \quad (16)$$

本文 3 路卷积分别经过 Bi-LSTM 之后, 拼接形成的最终藏文 La 格例句向量为

$$\mathbf{Q}_d = \mathbf{Q}_1 \oplus \mathbf{Q}_2 \oplus \mathbf{Q}_3, \quad (17)$$

\mathbf{Q}_i 表示第 i 路卷积特征经过 Bi-LSTM 学习得到的藏文 La 格例句向量。

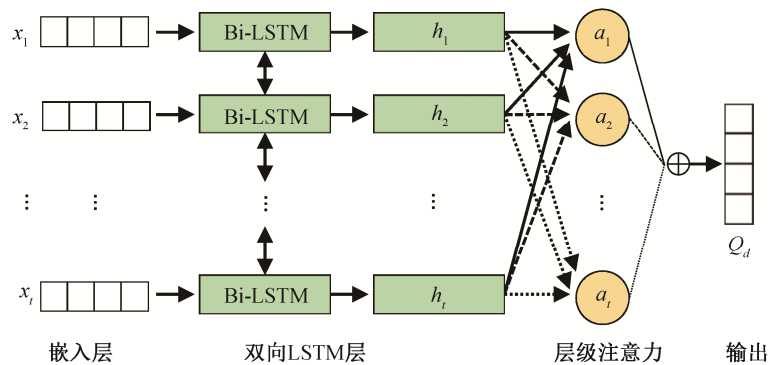


图4 结合层级注意力机制的 Bi-LSTM 层

Fig. 4 Bi-LSTM layer combined with Hierarchical attention mechanism

2.3 藏文 La 格例句分类层

得到藏文 La 格例句的最终表示 \mathbf{Q}_d 后, 经过全连接层和 Softmax 层, 输出最终的预测类别。用 d ($d \in d_c$) 表示 \mathbf{Q}_d 经全连接层输出的向量, 则藏文 La 格例句预测为 c 类的概率 P_c 的计算方法如下:

$$d = \text{softmax}(\mathbf{W}_c \cdot \mathbf{Q}_d + b_c), \quad (18)$$

$$p_c = \frac{\exp(d_c)}{\sum_{m=1}^n \exp(d_m)}, \quad n = 5, \quad (19)$$

其中, $c \in [1, 5]$, 表示藏文 La 格例句的某种类型, n 表示类别数目, d_c 表示 d 中属于类别 c 的分值, \mathbf{W}_c 和 b_c 为全连接层需要学习的参数。

3 实验

3.1 实验数据

本文通过网络爬虫和人工录入等采集方式, 得到 95.4 MB 的藏文文本语料。首先以“ π +空格”、“ π +空格”、“ η +空格”及动词、形容词、终结词等句尾词加单垂符或双垂符为切分点, 对藏文文本语料进行分句(此分句不符合严格意义上现代文法中藏文句子的概念), 得到 33.6551 万条藏文句子。然后通过匹配藏文 La 格虚词及其添接规则的方式, 抽取 15 万条仅含一个藏文 La 格虚词的句子, 以保证语料质量为目的, 通过人工去除所有不符合现代藏文文法句子概念的句子和含 La 格虚词但不属于 La 格的句子, 得到 2 万条藏文 La 格例句。其次, 以人工标注的方式, 用 1~5 的数字对每条藏文 La 格例句进行类别标注。最后, 为了确保标注质量, 通过人工逐条复核类别标签的方式, 完成藏文 La 格例句分类数据集的构建。

实验时,首先对藏文 La 格例句分类数据集进行预处理。在预处理过程中,为了保证句子结构的完整性,分别用 N, G, E 和 C 替换原文中的数字、希腊字母、英文和汉文,并剔除其他不影响句子结构的非藏文字符。然后,将数据按 8:1:1 的比例分成训练集、验证集和测试集。各类藏文 La 格例句的分布见表 2。

3.2 实验参数及评价指标

神经网络模型需要根据数据集的大小、语言的特性以及不同任务进行调参。实验时,所有模型的参数都已进行多次调整,实验结果为各模型最优参数组合下的最佳结果。本文模型的具体参数如表 3 所示。

本文选用分类任务中常用的评价指标精度(P)、召回率(R)、F1 值和准确率(ACC)对模型的性能进行评价。TP 表示预测为正的样本,FP 表示预测为正的负样本,FN 表示预测为负的正样本,TN 表示预测为负的负样本,则精度(P)、召回率(R)、F1 值和准确率(ACC)的计算公式如下:

$$P = \frac{TP}{TP+FP}, \quad (20)$$

$$R = \frac{TP}{TP+FN}, \quad (21)$$

表 2 藏文 La 格例句分布

Table 2 Tibetan La case example sentences distribution

类型	总句子数	句子数	占比/%
业格句	2 万	6964	34.82
为格句		2684	13.42
依格句		3104	15.52
同格句		3595	17.98
时格句		3653	18.26

表 3 模型参数

Table 3 Model parameters

参数名	参数
最长音节数	20
批处理大小	16
音节嵌入维度	128
卷积核数量	200
卷积核尺寸	(2, 3, 4)
损失函数	categorical_crossentropy
辍学率	0.25
学习率	0.0001
优化函数	Adam
迭代次数	40

$$F1 = \frac{2 \times P \times R}{P + R}, \quad (22)$$

$$ACC = \frac{TP+TN}{TP+FP+TN+FN}. \quad (23)$$

3.3 基线方法

由于目前尚未查阅到有关藏文 La 格例句自动分类的文献,加上没有公开的藏文 La 格例句分类数据集,所以本文选用神经网络中常用于分类任务的 7 个经典模型作为基线模型来验证本文模型的有效性。

1) FastText^[16]: 一种快速文本分类算法。

2) LSTM: 使用基于单向长短记忆网络文本分类方法^[17]。

3) Bi-LSTM: 使用基于双向长短记忆网络文本分类方法^[17]。

4) Bi-LSTM+Att^[18]: 一种基于双向长短记忆网络和注意力机制的关系分类方法。

5) CNN^[2]: 首次应用于句子分类的经典卷积神经网络模型。

6) RCNN^[19]: 使用双向循环神经网络加最大池化的方法进行文本分类。

7) C-LSTM^[17]: 使用单通道的多路卷积加长短记忆网络的方式进行文本分类。

SF-C+LSTM+Att 和 SF-C+Bi-LSTM+Att 为本文模型, SF 表示融合双通道音节特征, C 表示去掉池化层的卷积神经网络(CNN), Att 表示层级注意力机制(hierarchical attention mechanism)。

3.4 实验结果与分析

3.4.1 分类性能对比

为了验证本文模型的有效性,分别在几种常用的经典分类神经网络模型上对比藏文 La 格例句的分类效果,实验结果见表 4。

从表 4 可以看出,与各种基线模型相比,本文模型实验效果均有所提升,比单一模型 FastText、LSTM, Bi-LSTM, Bi-LSTM+Att 和 CNN 的分类准确率分别提高 2.16, 3.10, 2.76, 2.51 和 1.76 个百分点,比混合模型 RCNN 和 C-LSTM 的分类准确率分别提高 4.76 和 0.95 个百分点,表明本文模型的性能更好。

本文模型性能更好的原因主要有三方面: 1) 本文融合了双通道音节特征,丰富了嵌入向量的语义

表 4 各种分类模型的对比实验结果(%)
Table 4 Comparison of experimental results of various classification model (%)

实验	模型	P	R	F1	ACC
	FastText	88.67	87.51	88.09	88.10
	LSTM	87.73	86.56	87.14	87.16
	Bi-LSTM	88.07	86.54	87.45	87.50
基线	Bi-LSTM+Att	88.26	86.93	87.59	87.75
	CNN	89.18	87.68	88.42	88.50
	RCNN	86.08	84.73	85.40	85.50
	C-LSTM	88.83	87.76	88.29	88.34
本文	SF-C+LSTM+Att	89.73	88.72	89.22	89.31
	SF-C+Bi-LSTM+Att	90.58	89.77	90.17	90.26

说明: 粗体数字表示性能最佳。

信息; 2) 本文采用多路卷积模式, 提取不同感受视野的文本特征, 丰富了文本表示; 3) 本文在每路卷积后直接拼接结合层级注意力机制的 Bi-LSTM, 更合理地学习文本的时序特征以及每一时刻输出表, 提高了最终文本表示的质量。另外, 通过实验发现, 影响本文模型性能的主要原因是部分藏文 La 格例句无法根据其语法结构和浅层的语义信息来进行分类。例如: “ཤིང་ལ་ལོ་མ་ཡོད།、བ་ལང་ལ་ར་ཡོད། 和 ཅུ་བ་ཁ་མལ་ན་ཡོད།” 属于依格句, 而 “ཤིང་ལ་ཅན་དན་ཡོད།、བ་ལང་ལ་དགའ་ཟུང་ཡོད། 和 ནད་བ་ཁ་མལ་ན་ཡོད།” 属于同格; “ཉ་ལ་ཚུམ་མེད།” 属于为格句, 而 “ཉ་ལ་དུག་མེད།” 属于业格句。可见, 这些藏文 La 格例句需要根据具体的语境及语用目的深入分析其语义才能正确分类。

3.4.2 双通道音节特征的有效性验证

为了验证融合双通道音节特征的有效性, 分别考查融合双通道的音节级 Word2vec 和 GloVe 的模型性能以及使用单通道的音节级 Word2vec 或 GloVe 的模型性能, 实验结果如图 5(a)所示。可以看出, 融合双通道音节特征的模型比使用单通道的音节级 Word2vec 和 GloVe 的分类准确率分别提高 1.1 和 1.25 个百分点, 验证了融合双通道音节特征的有效性。

3.4.3 多粒度卷积模式的有效性验证

为了验证卷积模式对模型性能的影响, 考查不同卷积模式对本文模型性能的影响。模式 1 为单粒度卷积方式, 模式 2 为多粒度卷积方式, 均选择效果最佳的卷积核大小, 卷积核大小分别为 3 和 (2, 3, 4), 实验结果如图 5(b)所示。可以看出, 融合双通

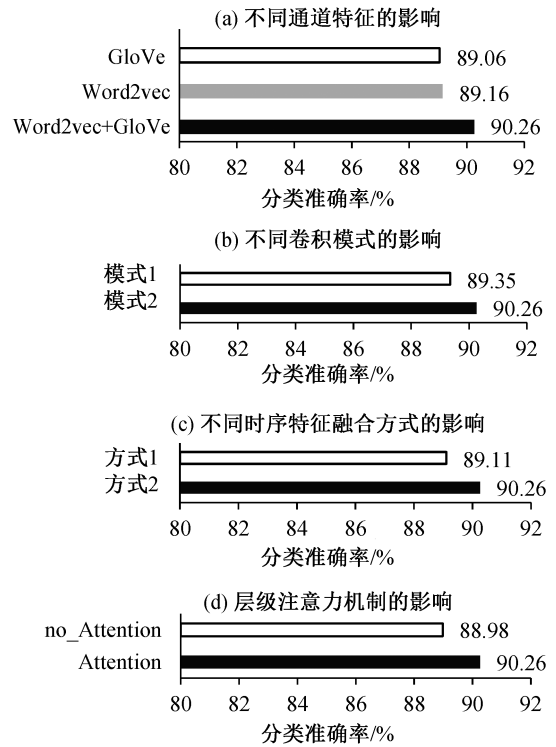


图 5 不同特征学习方式对模型性能的影响
Fig. 5 Effects of different feature learning methods on model performance

道音节特征的多粒度卷积模式的分类准确率比单粒度卷积模式高 0.91 个百分点, 验证了多粒度卷积模式的有效性。

3.4.4 时序特征融合方式的影响

为了验证本文时序特征融合方式对模型性能的影响, 探究多路卷积与 Bi-LSTM 结合方式对最终模型性能的影响。方式 1 为先进行多路卷积特征的融合, 再通过 Bi-LSTM 学习时序特征的实验结果; 方式 2 为每一路卷积后直接用 Bi-LSTM 学习时序特征的实验结果, 实验结果如图 5(c)所示。可以看出, 时序特征学习方式 2 的分类准确率比方式 1 高 1.15 个百分点, 表明每一路卷积后直接用 Bi-LSTM 学习时序特征的分类效果更佳。

3.4.5 层级注意力机制对模型性能的影响

为了验证层级注意力机制对本文模型性能的影响, 分别考查加层级注意力机制时以及不加层级注意力机制时模型的性能, 实验结果如图 5(d)所示。可以看出, 模型加层级注意力机制时的分类准确率比不加层级注意力机制时高 1.28 个百分点, 表明本文模型加层级注意力机制时分类效果更好。

4 结束语

本文提出一种融合双通道音节特征的藏文 La 格例句自动分类模型, 使用 word2vec 和 GloVe 构建双通道藏文音节嵌入, 分别在每路卷积中对双通道音节特征进行融合, 提高了卷积层的空间表征能力。为了更好地学习句子的时序特征, 模型通过在每路卷积之后直接使用结合注意力机制的双向 LSTM 学习时序特征, 得到最终的句子表征, 避免了过早融合卷积特征而破坏句子的时序特征。实验结果显示, 模型的藏文 La 格例句分类准确率能够达到 90.26%, 分类性能优于几种基线模型。在未来的工作中, 我们将通过扩充数据, 研究融合双通道词特征的藏文 La 格例句分类方法, 并尝试融入词性特征, 进一步优化模型的性能。

参考文献

- [1] 吉太加. 藏语语法研究. 西宁: 青海民族出版社, 2016
- [2] Kim Y. Convolutional neural networks for sentence classification // Proceedings of Empirical Methods on Natural Language Processing. Doha, 2014: 1746–1751
- [3] Vieira J, Moura R S. An analysis of convolutional neural networks for sentence classification // Computer Conference. Cordoba, 2017: 1–5
- [4] 高云龙, 左万利, 王英, 等. 基于稀疏自学习卷积神经网络的句子分类模型. 计算机研究与发展, 2018, 55(1): 179–187
- [5] 汪冉, 金忠. 基于极性转移和 LSTM 的树结构网络与句子分类. 计算机应用研究, 2019, 36(1): 64–67
- [6] Fentaw H W, Kim T H. Design and investigation of capsule networks for sentence classification. Applied Sciences, 2019, 9(11): 2200
- [7] Zhao Z, Wu Y. Attention-based convolutional neural networks for sentence classification // Interspeech. San Francisco, 2016: 705–709
- [8] Yang L, Ji L, Huang R, et al. Multi-grained-attention gated convolutional neural networks for sentence classification. Intelligent Data Analysis, 2019, 23(5): 1091–1107
- [9] 李文宽, 刘培玉, 朱振方, 等. 基于卷积神经网络和贝叶斯分类器的句子分类模型. 计算机应用研究, 2020, 40(2): 19–22
- [10] 柔特, 才让加. 基于循环卷积神经网络的藏文句类识别. 中文信息学报, 2019, 33(12): 76–82
- [11] Ban M, Cai Z, Cai R, et al. Tibetan interrogative sentence recognition and classification based on phrase features. MATEC Web of Conferences, 2021, 336(4): 06017
- [12] 毛尔盖·桑木旦. 藏文语法概论. 西宁: 青海民族出版社, 2005
- [13] 格桑居冕. 实用藏文语法. 成都: 四川民族出版社, 2004
- [14] Mikolov T, Chen K, Corrado G, et al. Efficient estimation of word representations in vector space [EB/OL]. (2013–09–07)[2021–04–08]. <https://arxiv.org/abs/1301.3781>
- [15] Pennington J, Socher R, Manning C. GloVe: global vectors for word representation // Conference on Empirical Methods in Natural Language Processing. Doha, 2014: 1532–1543
- [16] Joulin A, Grave E, Bojanowski P, et al. Bag of tricks for efficient text classification // Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics (Volume 2, Short Papers). Valencia, 2017: 427–431
- [17] Zhou C, Sun C, Liu Z, et al. A C-LSTM neural network for text classification. Computer Science, 2015, 1(4): 39–44
- [18] Peng Z, Wei S, Tian J, et al. Attention-based bidirectional long short-term memory networks for relation classification // Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers). Berlin, 2016: 207–212
- [19] Lai S, Xu L, Liu K, et al. Recurrent convolution neural networks for text classification // Proceedings of the 29th AAAI Conference on Artificial Intelligence. Austin, 2015: 2267–2273