doi:10.3772/j.issn.2095-915x.2023.02.010

基于长短时记忆和条件随机场藏文分词模型

于永斌[1] 陆瑞军[1] 尼玛扎西[2] 群诺[2] 王昊[1] 唐倩[1] 彭辰辉[1] 项秀才让[2]

开放科学（资源服务）标识码（OSID）

1. 电子科技大学 成都 610054；
2. 西藏大学 拉萨 850000

摘要：[目的 / 意义] 本文提出基于长短时记忆 (Long short-term memory，LSTM) 神经网络和条件随机场 (Conditional Random Field, CRF) 的藏文分词模型。[方法 / 过程] 引入注意力机制，获取更多特征信息 , 提升模型关注上下文信息与当前音节之间联系；提出一种音节扩展方法，获取更多的输入特征信息与语料信息，增强模型单音节特征信息以获取更多语义信息的能力。[局限] 本文在西藏大学数据集 12261 条的基础上，扩充至 74384 条，形成 Tibetan-News 数据集。[结果 / 结论] 实验结果表明，在模型中加入注意力机制并使用音节扩展方法后，模型在 Tibetan-News 数据集上的精确率、召回率和 F1 分别提升 2.9%、3.5% 和 3.2%。基于本文模型的分词系统已在工程上应用推广。

关键词：藏文分词；长短时记忆网络；条件随机场；注意力机制

中图分类号：G35 TP391

Tibetan Word Segmentation Model Based on LSTM and CRF

YU Yongbin[1] LU Ruijun[1] NYIMA Tashi[2] QUN Nuo[2] WANG Hao[1] TANG Qian[1] PENG Chenhui[1] XIANGXIU Cairang[2]

1. University of Electronic Science and Technology of China, Sichuan 610054, China;
2. Tibet University, Tibet 850000, China

Abstract: [Objective/Significance] This paper proposes a deep recurrent neural network Tibetan word separation model based on Long short-term memory (LSTM) and Conditional Random Field (CRF). [Methods/Processes] The soft attention mechanism is applied to improve the ability to extract the context information of Tibetan text sequences, and the syllable expansion method is applied to improve the single syllable and semantic feature. [Limitations] Based on the Tibetan University's dataset, this paper constructs the Tibetan-News dataset from 12261 sentences to 74384 sentences. [Results/Conclusions] The experimental results

基金项目 科技创新 2030-“新一代人工智能”重大项目 - 藏语言文字自动识别技术研发和应用（2022ZD0116100）。

作者简介 于永斌（1978-），博士，副教授，主要研究方向为非线性电路与系统、人工智能神经网络、遗传算法等、超大规模集成电路的物理设计、现代控制理论及其应用等；陆瑞军（1997-），硕士，研究方向为机器翻译与自然语言处理；尼玛扎西（1964-），博士，教授，研究方向为藏文信息处理，计算机科学技术，E-mail：nmzx@utibet.edu.cn；群诺，教授，研究方向为自然语言处理和计算机网络；王昊（1998-），硕士，研究方向为深度学习与计算机视觉；唐倩（1997-），硕士，研究方向为忆阻神经网络与深度学习；彭辰辉（1997-），硕士，研究方向为深度学习与自然语言处理；项秀才让，研究人员，研究方向为藏文信息处理。

引用格式 于永斌 , 陆瑞军 , 尼玛扎西 , 等 . 基于长短时记忆和条件随机场藏文分词模型 [J]. 情报工程 , 2023, 9(2): 108-116.

show that, compared with the Tibetan word segmentation models of LSTM and CRF, the accuracy, recall and F1 of the Tibetan word segmentation models based on soft attention LSTM and CRF on Tibet-News dataset are respectively Up 2.9%, 3.5% and 3.2%. The segmentation system based on this paper is already applied in engineering field.

Keywords: Word segmentation; Attention Mechanism; Long Short Term Memory Network; Conditional Random Field

1 研究背景

藏文信息处理的基础任务[1]是藏文分词，分词是将连续字符按照一定规则重组并生成词序列的行为。分词研究可分为基于字典的分词方法和基于字的分词方法[2]。基于字典的藏文分词方法以词典为主要划分依据。基于字的分词方法是目前较流行的基于统计的分词方法。其基于给定的每个字的标签划分。

近年随着文本数据的增多，神经网络模型在分词中广泛应用[3-4]深度学习在中文分词中的应用已得到了很好的发展[5-10]，于藏文中应用较少[6]。在现有研究中，基于循环神经网络（RNN）的藏文分词模型是最常用的模型之一，该模型使用长短时记忆网络（LSTM）作为特征提取层，条件随机场（CRF）作为解码层。但该模型存在不能有效关注上下文信息和当前音节的标签联系的问题。

针对目前基于 RNN 模型在藏文分词中存在的问题，本文提出了基于注意力机制 LSTM 和 CRF 的藏文分词模型。本模型以连续藏文文本作为输入，使用音节扩展、并引入软注意力机制。

本文的主要贡献如下：

1、在基于LSTM和CRF的藏文分词模型中，引入注意力机制使模型能够更为有效地关注上下文信息与当前标签之间的联系。

2、在模型输入部分使用音节扩展方法，使得输入音节包含上下文信息，使得模型能够获取到更多的语义信息。

接下来，本文将依次介绍相关工作、模型结构、实验结果与结论。

2 相关工作

长短时记忆网络（LSTM）[11]广泛应用于自然语言处理任务，如分词模型[12]、序列标注[13]等，并取得了很好的结果。LSTM 可以捕捉潜在的长距离依赖关系[11]，能有效缓解长期依赖问题。条件随机场[14-16]（CRF）解决了最大熵马尔可夫模型的标注偏差问题。但 LSTM 得到分词标签后不具备标签之间的转移关系，基于 CRF 的分词方法准确率不高。

使用双向 LSTM-CRF 模型进行序列标记任务，较 LSTM、双向 LSTM、CRF、LSTM-CRF 相比效果最佳[17-18]。门控循环单元（GRU）[19]较 LSTM 参数量更少，但效果相当。卷积神经网络（CNN）提取给定单词的字符级表示[20-21]，模型准确率较高。然而双向 LSTM-CRF 和 GRU 不能有效关注上下文信息与当前输入的联系，CNN 则无法关注输入序列的顺序。

doi:10.3772/j.issn.2095-915x.2023.02.010

注意力机制可在机器翻译任务上对翻译和局部翻译单元之间的对应关系同时建模[22]。软注意力机制（Soft Attention）与硬注意力机制相比，Soft Attention 的注意力焦点更加发散[23]。而通过双向注意机制，可将上下文特征及对应句法知识结合起来，所得模型达到当年最高水平[24]。

音节扩展思想源于自然语言序列处理，其细节详见 3.2 节。

3 模型结构

每句藏文文本，经过编码及音节拓展后输入到长短时记忆网络（LSTM），该部分用于将输入映射为隐藏层。其输出作为解码器部分条件随机场（CRF）的输入。通过上述结构可有效保存语句前后标签的信息。

基于 LSTM -CRF 的藏文分词改进模型如图 1 所示。

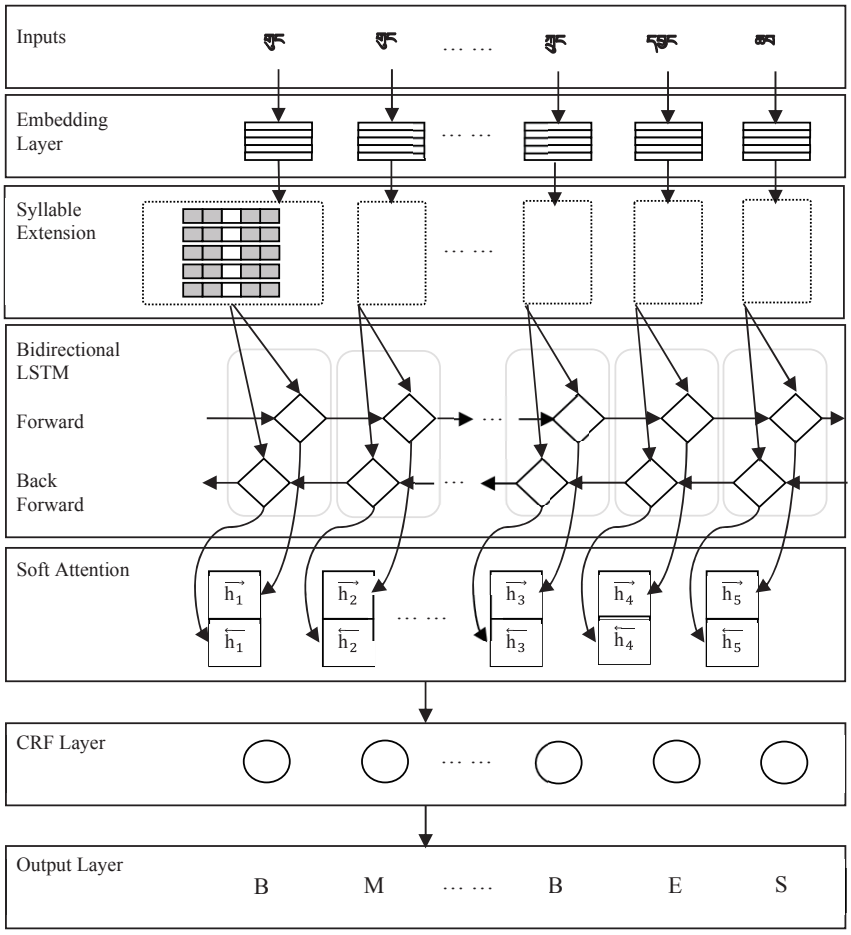

图 1 改进基于 LSTM 和 CRF 的藏文分词模型

3.1 LSTM

长短时记忆网络（LSTM）[11] 可以缓解序列中长期依赖的问题，这使得 LSTM 在较长的序列中比普通循环神经网络（RNN）具有更好的性能，LSTM 结构如图 2 所示。

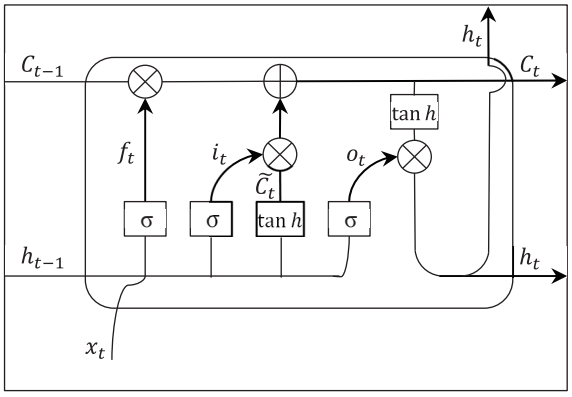

图 2 LSTM 模型结构 [11]

LSTM 模型是主要由遗忘门 f_t，输入门 i_t 和输出门 o_t 组成。其输入是前一个时刻的输出 h_{t-1} 以及当前时刻的输入 x_t，遗忘门输出见式（1）：

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

输入门 i_t 和临时状态 $\tilde{C}_t$ 的输出见式（2）、式（3）：

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

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

细胞状态 C_t 由临时状态 $\tilde{C}_t$、遗忘门值 f_t 和记忆值 i_t 计算，具体见式（4）：

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

最后是计算输出门值 o_t 和隐含状态值 h_t，具体见式（5）、式（6）：

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

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

在式（1）至式（6）中，W_i 为每一个时刻输入的权重，W_p 为前一个时刻输出的权重，W_0 为当前时刻的输出权重。b 为偏置，x_t 为 t 时刻的输入。

3.2 音节扩展

在数据集中，对当前输入音节，同句话中其余位置的音节越靠近当前音节，则两者关联性越强。通过将音节拼接可使神经网络提取到更多输入特征。音节扩展的具体表示如图 3 所示：

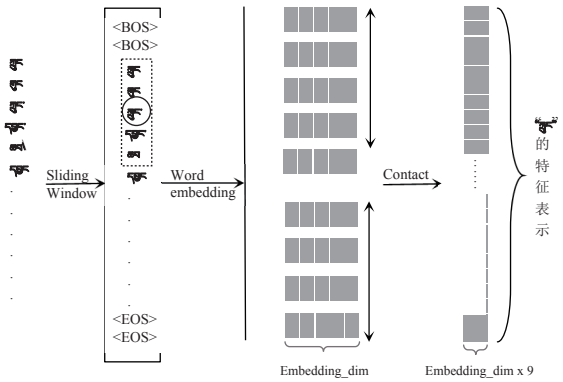

图 3 音节扩展方法特征表示

音节拓展具体步骤如下：

1、在藏语语句 $X=(x_1, x_2, \cdots, x_i, \cdots, x_n)$ 的句首和句尾分别添加两个特殊音节 <*BOS*> 和 <*EOS*> 作为起始音节和结束音节（<*BOS*> 和 <*EOS*> 没有特殊意义）。

2、以当前输入音节为中心向两侧各扩展两个音节单元。

3、使用 1*5 的滑动窗口，将窗口内音节与相邻音节组成的词条拼接作为当前音节的输入。

3.3 注意力机制

如图 1 所示，在改进模型中，软注意力机制（Soft Attention）层与双向长短时记忆网络（Bidirectional LSTM）层相连，使用 Soft Attention 可帮助模型建立目标与其上下文相关隐含特征之间的联系，辅助模型更好区分目标的分词标签。将 Bidirectional LSTM 层提取的藏文序列特征向量输入注意力机制，得到的上下文注意力（Attention）向量。软注意力机制的具体实现表示见图 4：

doi:10.3772/j.issn.2095-915x.2023.02.010

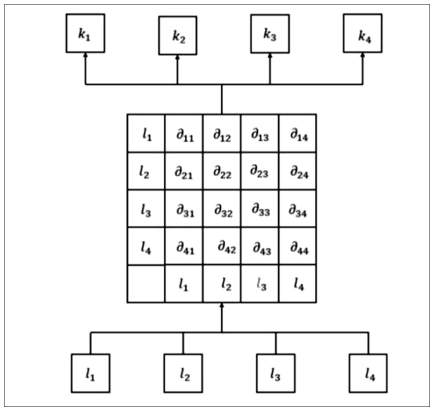

图 4 Soft Attention 示意图

具体步骤如下：

1、计算输入藏文序列中第个音节和第个音节的相关度。其中 m 为句子的长度，$f()$ 表示相似度计算方法，o_i、o_j 为 Bidirectional LSTM 的输出的第 i 个和第 j 个位置特征。具体计算公式如（7）所示：

$$s_{ij}=f(o_i, o_j),\ i,j=1,2,\cdots,m \quad (7)$$

2、将第一步得到的相关度进行归一化，计算公式见式（8）：

$$\alpha_i=soft\ max(s_{ij}),\ i=1,2,\cdots,m \quad (8)$$

3、将第二步中得到的输入序列中所有音节对当前位置音节的影响权重进行加权计算，最终可得到当前位置音节的 Attention 向量，具体见算式（9）：

$$A=\alpha_i o_i+\cdots+\alpha_m o_m,\ i=1,2,\cdots,m \quad (9)$$

3.4 损失函数

基于长短时记忆网络（LSTM）的藏文分词算法最后通过条件随机场（CRF）来得到最终结果。在此模型中，使用的是线性链条件随机场。对于数据集中的任意一个句子序列 $X=(x_1, x_2, \cdots, x_i, \cdots, x_n)$，其中，$X$ 表示为句子，x_i 表示为句子中的第 i 个音节。如果序列的长度为 n，标签个数为 m，那么共有 m^n 个序列结果，即 m^n 个 $Y=(y_1, y_2,\cdots, y_i, \cdots, y_n)$，$y_i$ 表示为第 i 个音节的标签。计算出每个可能的序列结果的得分 $score(Y)$，求出任一标签序列的概率，选择概率最大的作为标注结果。模型的损失函数为 $loss=-logP(Y|X)$，其中 $P(Y|X)$ 计算见公式（10）、公式（11）：

$$score(Y)=\sum_{i=1}^{n}(e_i[y_i])+\sum_{i=2}^{n}(R[y_{i-1}, y_i]) \quad (10)$$

$$logP(Y\mid X)=score(Y)-log(\sum_{Y} e^{score(Y)}) \quad (11)$$

4 实验结果

本文实验均基于构建的 Tibetan-News 数据集。数据集共计 74384 句，涵盖经济、文化等新闻。数据集中单句最长为 160 个音节，单句最短为 2 个音节。数据集划分见表 1：

表 1 数据集划分情况

数据集名称	数据集类别	语句数（条）
Tibetan-News	训练集	52068
	测试集	22316

下面对 Tibetan-News 数据集构建流程进行介绍。该数据集的构建分为四个步骤：藏文语料获取、藏文数据集数据处理、藏文数据集的构建及藏文数据集的去重。

藏文语料通过爬取中国西藏网相关新闻获得。获得的藏文语料需进行文本预处理，预处理的具体包括去除无效标签、编码转换、文档切分、基本纠错、去除空白标点等步骤。

在数据集构建中，为减少人工标注的工作量，需构建藏文分词模型对预处理的文本进行预分词。为使数据集可用于分词任务，需要根

据空格和“.”获取每个音节的标签。所述标签为 *B*、*M*、*E* 或 *S*，其中 *B* 表示起始字标签，*M* 表示中间字标签，*E* 表示结束字标签，*S* 表示单个字标签。随后模型利用条件随机场对藏文语料进行预分词，使用维特比算法 [25] 降低回溯的代价。最终使用 Simhash[26] 去重算法，去除数据集中相同语句。

本文使用音节扩展的方法提升模型的效果。为证明音节扩展有效，在双向长短时条件和随机场（Bidirectional LSTM CRF）模型、双向条件门控循环单元随机场（Bidirectional GRU CRF）模型和条件随机场（CRF）模型中均使用了该方法，不同模型使用音节扩展方法的 F1 曲线如图 5 实验结果见表 2。

表 2 使用音节扩展方法对比实验结果

模型名称	分类条件	P(%)	R(%)	F1(%)
Bidirectional LSTM CRF	单音节输入	89.2	90.0	89.6
	音节扩展	89.9	92.5	91.2
Bidirectional GRU CRF	单音节输入	89.9	90.9	90.4
	音节扩展	90.6	92.8	91.7
CRF	音节扩展	86.2	85.8	86.0
	单音节输入	83.8	82.6	83.2

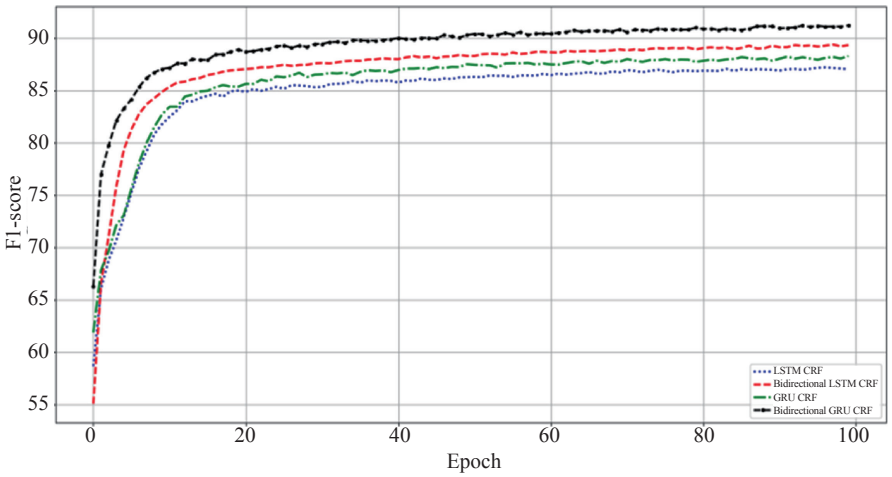

图 5 使用注意力机制和音节扩展方法 F1 曲线图

在表中可看出，使用音节扩展后，基于 CRF 的分词方法精确率、召回率和 F1 分别提高了 2.4%、3.2% 和 2.8%，基于 Bidirectional LSTM CRF 的藏文分词方法精确率、召回率和 F1 分别提升了 0.7%、2.5% 和 1.6%。

编码器部分嵌入注意力机制的方法来提升模型的效果。使用软注意力（Soft Attention）可帮助获取藏文语句中的上下文信息及在目标中建立上下文相关的隐含特征的联系，辅助模型更好地区分目标的分词标签，提升模型提取特征的能力。为了证明注意力机制有效，对上述模型进行了改进，即在长短时记忆网络（LSTM）之后通过 Soft Attention 使得模型关注上下文信息。LSTM 嵌入注意力机制的 F1 曲线如图 6 所示，具体的实验结果见表 3。

表 3 使用注意力机制对比实验结果

模型名称	P(%)	R(%)	F1(%)
Bidirectional LSTM CRF	89.2	90.0	89.6
Attention Bidirectional LSTM CRF	91.3	91.5	91.4
Bidirectional GRU CRF	89.9	90.9	90.4
Attention Bidirectional GRU CRF	91.5	92.5	92.0

doi:10.3772/j.issn.2095-915x.2023.02.010

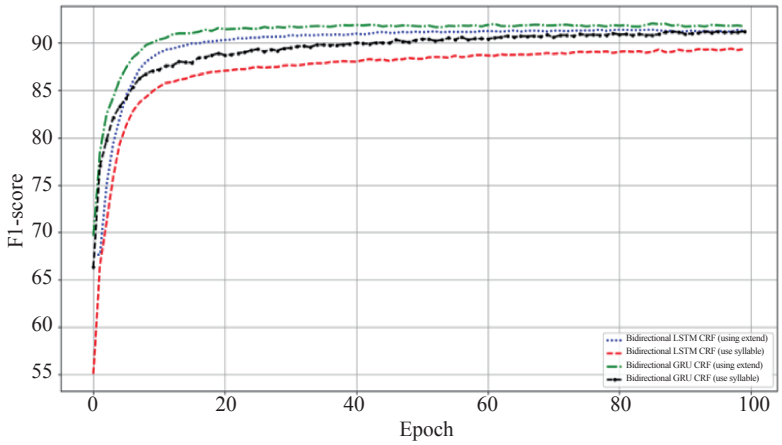

图 6 不同模型融合注意力机制 F1 曲线对比图

从表中可以看出，在模型中嵌入注意力机制之后，基于 Bidirectional LSTM CRF 的藏文分词方法精确率、召回率和 F1 分别提升了 2.1%、1.5% 和 1.8%，基于双向门控循环单元的分词方法精确率、召回率和 F1 分别提升了 1.6%、1.6% 和 1.6%。通过实验数据对比，证明在模型中嵌入注意力机制的方法能够有效地提升模型的效果。

表 4 使用注意力机制和音节扩展实验结果对比

模型名称	P(%)	R(%)	F1(%)
Bidirectional LSTM CRF（单音节输入）	89.2	90.0	89.6
Bidirectional LSTM CRF（音节扩展）	89.9	92.5	91.2
Attention Bidirectional LSTM CRF（单音节输入）	91.3	91.5	91.4
Attention Bidirectional LSTM CRF（音节扩展）	92.1	93.5	92.8
Bidirectional GRU CRF（单音节输入）	89.9	90.9	90.4
Bidirectional GRU CRF（音节扩展）	90.6	92.8	91.7
Attention Bidirectional GRU CRF（单音节输入）	91.5	92.5	92.0
Attention Bidirectional GRU CRF（音节扩展）	92.8	93.6	93.2

为了能够更好地提升模型的效果，本节将音节扩展方法和注意力机制嵌入模型。不同模型使用音节扩展方法和嵌入注意力机制后 F1 曲线如图 7 所示，具体的实验结果见表 4。

从表中可看出，使用音节扩展方法和嵌入注意力机制可分别提高 LSTM 网络和 CRF 分词模型精确率、召回率和 F1-Measure 至 2.9%、3.5% 和 3.2%，使用音节扩展方法和嵌入注意力机制可分别提高基于注意力机制的 GRU CRF 藏文分词模型精确率、召回率和 F1 至 12.9%、2.7% 和 2.8%。对比实验结果发现，使用音节扩展的方法和嵌入注意力机制的方法能够有效地提升模的效果。

在表 5 中，[27-29] 的结果是通过其代码在 Tibetan-News 数据集上进行复现，可以看出本文提出的模型相比 Lattice+Word Emb 模型精确率、召回率和 F1 分数分别提高了 1.9%、3.0% 和 2.5%，相比于 WCON mode 模型精确率、召回率和 F1 分数分别提高了 2.0%、2.3% 和 2.2%。此外本文还给出了现有可查最新的藏文分词模型的分词结果。最新发表论文的模型结果为 Ensemble，F1 指标为 92.3%，由于数据集不一样，模型的结果不能够衡量模型的优异性。但是通过表中可以看出，现有第四章 基于长短时记忆网络和条件随机场的藏文分词算法的基于深度神经网络的藏文分词模型是基于长短时记忆网络和条件随机场的藏文分词模型。

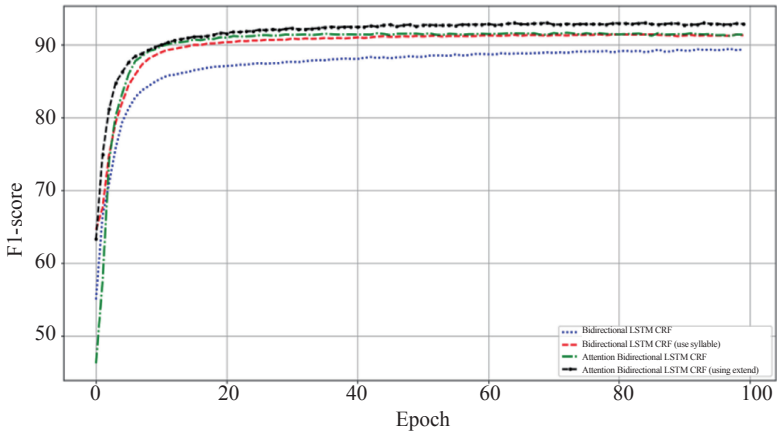

图 7 不同模型使用音节扩展方法和注意力机制 F1 曲线对比图

表 5 提出模型和主流模型的实验结果

模型名称	P(%)	R(%)	F1(%)
Lattice+Word Emb[27]	90.9	90.6	90.7
WCON model[28]	90.8	90.8	90.8
BiLSTM-CRF[29]	93.21	92.10	92.65
Ensemble	92.3	92.4	92.3
Attention Bidirectional GRU CRF(ours)	92.8	93.6	93.2

5 分词示例

见图 8 中展示的不同模型分词结果。图中前三行为藏文原语及藏文语句的真实标签。其后以此为 LSTM 和 CRF 的藏文分词模型、基于 GRU 和 CRF 的藏文分词模型和第四章使用音节扩展方法和融合自注意力改进模型的分词结果。

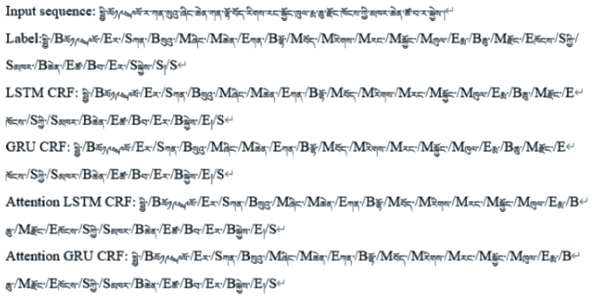

图 8 不同模型分词示例

6 结论

本文基于长短时记忆网络（LSTM）和条件随机场（CRF）的藏文分词模型，融入注意力机制使得模型能够有效的对上下文信息和当前音节之间的联系。此外，使用音节扩展的方法使得模型能够获取到更多的特征信息。实验结果表明，融入注意力机制和使用音节扩展方法能够有效的提升模型的性能。

致谢

本文得到“工业与信息化部人工智能产业创新任务揭榜挂帅项目——汉藏英神经机器翻译系统”资助，感谢审稿专家的指导！

参 考 文 献

[1] 关白 . 信息处理用藏文分词单位研究 [J]. 中文信息学报 , 2010, 24(3): 124-128.

[2] 游治勇 . 基于字的分词方法的研究与实现 [D]. 电子科技大学 , 2015.

[3] Jiang T, Yu H, Jam Y. Tibetan word segmentation system based on conditional random fields[C]. 2011 IEEE 2nd International Conference on Software Engineering and Service Science. IEEE, 2011: 446-448.

[4] Li B H, Liu H D, Long C J, et al. Tibetan word segmentation based on deep learning[J]. Computer Engineering and Design, 2018, 39(1): 194-198.

[5] Devlin J, Chang M W, Lee K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding[J]. arXiv preprint arXiv:1810.04805, 2018.

doi:10.3772/j.issn.2095-915x.2023.02.010

[6] 桑杰端珠 , 才让加 . 神经网络藏文分词方法研究 [J]. 青海科技 , 2018, 25(6): 7.
[7] Xipeng Qiu, Hengzhi Pei, Hang Yan, and Xuanjing Huang.Multi-criteria chinese word segmentation with transformer[J]. arXiv preprint arXiv:1906.12035, 2019.
[8] Wang C, Chen J, Wu X. Dictionary Chinese Word Segmentation research a method combined with CRFs[C]. 5th International Conference on Computer Sciences and Convergence Information Technology. IEEE, 2011: 962-965.
[9] Wang C, Chen J, Wu X. Dictionary Chinese Word Segmentation research a method combined with CRFs[C]. 5th International Conference on Computer Sciences and Convergence Information Technology. IEEE, 2011: 962-965.
[10] Chen X, Qiu X, Zhu C, et al. Long short-term memory neural networks for chinese word segmentation[C]. Proceedings of the 2015 conference on empirical methods in natural language processing. 2015: 1197-1206.
[11] Hochreiter, Sepp, Schmidhuber, et al. Long short-term memory.[J]. Neural Computation, 1997.
[12] Jia L, Jiang T, Meng J H, et al. Tibetan Text Classification Method Based on BiLSTM Model[C].
[13] Greff K, Srivastava R K, Koutník J, et al. LSTM: A search space odyssey[J]. IEEE Transactions on Neural Networks and Learning Systems, 2016, 28(10): 2222-2232.
[14] Sarawagi S, Cohen W W. Semi-markov conditional random fields for information extraction[J]. Advances in Neural Information Processing Systems, 2004:17.
[15] Arnab A, Zheng S, Jayasumana S, et al. Conditional random fields meet deep neural networks for semantic segmentation: Combining probabilistic graphical models with deep learning for structured prediction[J]. IEEE Signal Processing Magazine, 2018, 35(1): 37-52.
[16] Zhuo J, Cao Y, Zhu J, et al. Segment-level sequence modeling using gated recursive semi-markov conditional random fields[C]. Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2016: 1413-1423.
[17] Huang Z, Xu W, Yu K. Bidirectional LSTM-CRF models for sequence tagging[J]. arXiv preprint arXiv:1508.01991, 2015.
[18] Wang L, Yang H. Tibetan word segmentation method based on BiLSTM_ CRF model[C]. 2018 International Conference on Asian Language Processing (IALP). IEEE, 2018: 297-302.
[19] Chung J, Gulcehre C, Cho K H, et al. Empirical evaluation of gated recurrent neural networks on sequence modeling[J]. arXiv preprint arXiv:1412.3555, 2014
[20] Ma X, Hovy E. End-to-end sequence labeling via bi-directional lstm-cnns-crf[J]. arXiv preprint arXiv:1603.01354, 2016.
[21] Yuan H, Wang J, Zhang X. 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. 2018: 1058-1062.
[22] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio.Neural Machine Translation by Jointly Learning to Align and Translate[C]. International Conference on Learning Representations.
[23] XUK, BAJ, KIROSR, et al. Show, attend and tell: Neural image caption generation with visual attention[C]. International conference on machine learning. 2015: 2048-2057.
[24] Tian Y, Song Y, Ao X, et al. Joint Chinese word segmentation and part-of-speech tagging via two-way attentions of auto-analyzed knowledge[C]. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. 2020: 8286-8296.
[25] Andrew Viterbi.Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. [J]. IEEE Transactions on Information Theory, 1967, 13(2): 260–269.
[26] Charikar M S. Similarity estimation techniques from rounding algorithms[C]. Proceedings of the thiry-fourth annual ACM symposium on Theory of computing. 2002: 380-388.
[27] Yang J, Zhang Y, Liang S. Subword encoding in lattice LSTM for Chinese word segmentation[J]. arXiv preprint arXiv:1810.12594, 2018
[28] Higashiyama S, Utiyama M, Sumita E, et al. Incorporating word attention into character-based word segmentation[C]. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers). 2019: 2699-2709
[29] Cuo, Zhuoma R, Cai Z, et al. Cross-Domain Tibetan Word Segmentation Based on Deep Learning[J]. Journal of Physics: Conference Series, 2020