

# 基于双向门限递归单元神经网络的维吾尔语形态切分

哈里旦木·阿布都克里木, 程 勇, 刘 洋, 孙茂松

(清华大学 计算机科学与技术系, 智能技术与系统国家重点实验室, 清华信息科学与技术国家实验室(筹), 北京 100084)

**摘要:** 以维吾尔语为代表的低资源、形态丰富语言的信息处理对于满足“一带一路”语言互通的战略需求具有重要意义。这类语言通过组合语素来表示句法和语义关系, 因而给语言处理带来严重的数据稀疏问题。该文提出基于双向门限递归单元神经网络的维吾尔语形态切分方法, 将维吾尔词自动切分为语素序列, 从而缓解数据稀疏问题。双向门限递归单元神经网络能够充分利用双向上下文信息进行切分消歧, 并通过门限递归单元有效处理长距离依赖。实验结果表明, 该方法相比主流统计方法和单向门限递归单元神经网络获得了显著的性能提升。该方法具有良好的语言无关性, 能够用于处理更多的形态丰富语言。

**关键词:** 双向门限递归单元; 神经网络; 维吾尔语; 形态切分

中图分类号: TP391.2

文献标志码: A

文章编号: 1000-0054(2017)01-0001-06

DOI: 10.16511/j.cnki.qhdxxb.2017.21.001

## Uyghur morphological segmentation with bidirectional GRU neural networks

ABUDUKELIMU Halidanmu, CHENG Yong, LIU Yang, SUN Maosong

(State Key Laboratory of Intelligent Technology and Systems,  
Tsinghua National Laboratory for Information Science and  
Technology, Department of Computer Science and Technology,  
Tsinghua University, Beijing 100084, China)

**Abstract:** Information processing of low-resource, morphologically-rich languages such as Uyghur is critical for addressing the language barrier problem faced by the One Belt and One Road (B&R) program in China. In such languages, individual words encode rich grammatical and semantic information by concatenating morphemes to a root form, which leads to severe data sparsity for language processing. This paper introduces an approach for Uyghur morphological segmentation which divides Uyghur words into sequences of morphemes based on bidirectional gated recurrent unit (GRU) neural networks. The bidirectional GRU exploits the bidirectional context to resolve ambiguities and model long-distance dependencies using the gating mechanism. Tests show that this approach significantly outperforms conditional random fields and unidirectional GRUs. This approach is language-independent and can be applied to all morphologically-rich languages.

**Key words:** bidirectional gated recurrent unit; neural network; Uyghur; morphological segmentation

“一带一路”国家重大战略涉及 60 余个国家、50 余种国家通用语言和 200 余种民族语言, 覆盖 44 亿人口, 语言屏障问题逐渐成为推动深度合作与交流的重要阻碍。“一带一路”所涉及的绝大多数语言都是形态丰富语言。与汉语和英语等孤立语和屈折语不同, 以维吾尔语为代表的形态丰富语言通过词干和词缀多种组合在词汇层面表示丰富的句法和语义关系, 因而给语言技术处理带来严重的数据稀疏问题。目前, 中国面向形态丰富语言的信息处理技术相对而言严重滞后于汉语和英语, 远远无法满足“一带一路”语言互通的战略需求。因此, 深入开展以维吾尔语为代表的形态丰富语言的处理技术研究具有重要的意义。

维吾尔语在语素的组合上具有高度的灵活性, 虽然词干和词缀的数量有限, 但是理论上可以组合生成无限的词语。由于绝大多数维吾尔词语在语料库中只出现一次, 如何通过将维吾尔词语切分成语素序列(即维吾尔语形态切分)来缓解数据稀疏问题成为维吾尔语信息处理的关键任务之一。

传统的维吾尔语形态切分方法主要分为 2 类: 基于规则的方法<sup>[1]</sup>和基于统计的方法。基于规则的方类方法的优点在于人类专家能够观察和总结形态切分的规律并以规则的形式将知识传授给计算机, 但缺点是人工成本高。基于统计的方法可分为半监督学习方法和有监督学习方法 2 类。以 Morfessor 为代表的半监督学习方法利用最小描述长度准则或

收稿日期: 2016-07-08

基金项目: 国家自然科学基金重点项目(61331013);

国家“八六三”高技术项目(2015AA015407)

作者简介: 哈里旦木·阿布都克里木(1978—), 女, 博士研究生。

通信作者: 刘洋, 副教授, E-mail: liuyang2011@tsinghua.edu.cn

极大似然准则从未标注和已标注数据中自动发现语素并实现切分<sup>[2]</sup>。有监督学习方法则将形态切分视作序列标注问题,通过条件随机场(conditional random field, CRF)<sup>[3]</sup>等概率图模型进行建模,利用特征函数刻画形态切分的规律<sup>[4-8]</sup>。尽管基于统计的方法相比基于规则的方法而言不仅节省了人工成本,而且显著提升了切分准确率,但仍面临着2个挑战:1)语素的离散表示导致数据稀疏;2)特征设计难以覆盖所有语言现象。

近年来,深度学习在自然语言处理中获得了成功的应用。相比传统基于统计的方法,深度学习不仅能够通过采用连续表示的方式缓解数据稀疏问题,而且能够自动从数据中学习特征表示,缓解了人工特征设计难以保证覆盖面的问题。本文将深度学习引入维吾尔语形态切分,利用双向门限递归单元(gated recurrent unit, GRU)<sup>[9]</sup>神经网络建立序列标注模型,不仅充分利用了待标语素左侧和右侧的上下文信息,而且通过门限机制实现对长距离依赖的建模。实验结果表明,本文提出的方法显著提升了维吾尔语形态切分的准确率。

## 1 研究背景

### 1.1 维吾尔语形态切分

维吾尔语属于阿尔泰语系,是一种典型的黏着语,通过在词干上添加词缀来实现丰富的句法和语义功能。例如表1中,维吾尔语词干“زaman”是指名词“时代”,在词干上增加一个语素,并生成一个新的名词“زامانۇى”(即“现代”);进一步增加词缀将产生一个动词“زامانۇلاش”(即“现代化”)。以此类推,表达更加丰富句法和语义关系的词可以通过不断附加词缀来获得。表1中的最后一个词“زامانۇلاشتۇرۇلمايمىز”(即“我们不能被现代化”)实际上已经在词汇层面表示通常意义下句子层面的信息。

表1 维吾尔语词示例

维吾尔语词	汉语译文
زaman	时代
زامانۇى	现代
زامانۇلاش	现代化
زامانۇلاشتۇر	使现代化
زامانۇلاشتۇرۇل	被现代化
زامانۇلاشتۇرۇلما	不能被现代化
زامانۇلاشتۇرۇلمايمىز	我们不能被现代化

虽然维吾尔语的词干和词缀的数量是有限的,但是组合方式却是千变万化的,从而产生大量的词汇。绝大多数维吾尔语词在语料库中只出现一次,这给概率估计带来严重的数据稀疏问题,因此给维吾尔语处理技术带来极大的挑战。

维吾尔语形态切分的目标是将维吾尔语词切分为语素的序列。例如,将维吾尔语词“مۇكەممەلەش”(即“完全地”)切分为2个语素:

$$\begin{array}{rcc} \text{مۇكەممەلەش} & = & \text{لەش} + \text{مۇكەممەل} \\ \text{完全地} & & \text{地} \quad \quad \text{完全} \end{array}$$

这是维吾尔语信息处理的关键任务之一。

### 1.2 现有方法所面临的挑战

现有的基于统计的方法主要采用2种策略实现维吾尔语形态切分:第一种策略以 Morfessor 为代表,主张采用半监督学习的方法从未标注数据和已标注数据中自动发现语素<sup>[2]</sup>;另一种策略则将形态切分视作序列标注问题,利用条件随机场进行有监督学习,通过设计特征来刻画形态切分的规律<sup>[4-8]</sup>。尽管基于统计的方法在维吾尔语切分上取得了较好的准确率,但仍然面临以下2个挑战:

1)语素的离散表示导致数据稀疏:即使将词切分为语素,不少语素仍可能只在语料库上出现1次,从而给半监督学习带来数据稀疏问题。

2)特征设计难以覆盖所有语言现象:有监督学习依赖人工设计特征函数来刻画形态切分规律。由于形态切分的复杂性,无法保证所设计特征能覆盖所有语言现象。不仅如此,特征模板往往在训练数据上实例化为数以百万计的特征函数,给特征权重的训练带来挑战。

因此,维吾尔语形态切分亟需新的思想和方法来进一步提升准确性。

## 2 基于深度学习的维吾尔语形态切分

### 2.1 面向序列标注的循环神经网络

近年来,循环神经网络(recurrent neural network, RNN)在形态切分、词性标注和中文分词等序列标注任务上取得了成功的应用<sup>[10-14]</sup>。相比传统的统计方法,循环神经网络具有以下优点:

1)利用连续表示缓解数据稀疏问题:利用连续、稠密的实数向量表示词语或最小意义单元,利用向量空间中的距离度量词语和最小意义单元之间的关联性。

2)自动从数据中学习特征表示:循环神经网络能够自动从数据中学习刻画语言规律的特征表

示, 无需人工设计特征, 也在一定程度上缓解了覆盖所有语言现象的问题。

图 1 给出了维吾尔语形态切分的神经网络架构。

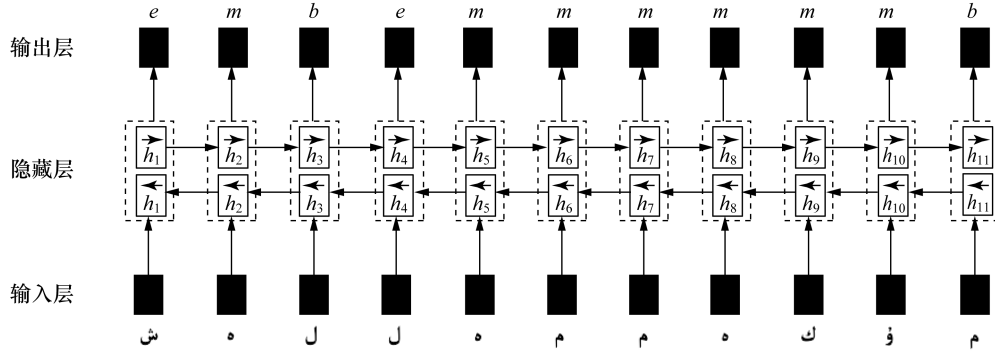


图 1 面向维吾尔语切分的双向门限递归单元神经网络

本文采用的标记集是  $\{b, m, e, s\}$ ,  $b$  表示语素的开始字符,  $m$  表示语素的中间字符,  $e$  表示语素的结束字符,  $s$  表示单字符语素<sup>[4]</sup>。图 1 所示的标记序列将维吾尔词“مۆكەممەلەش”切分成 2 个语素: “مۆكەممەل” (即“完全”) 和 “ەش” (即“地”)。

令  $X = (x_1, x_2, \dots, x_t, \dots, x_T)$  为维吾尔语字符序列,  $Y = (y_1, y_2, \dots, y_t, \dots, y_T)$  为相应的标记序列。每个标记的取值范围为标记集即  $y_t \in \{b, m, e, s\}$ 。

在 RNN 中, 每个标记的输出概率分布取决于相应的隐状态:

$$P(y_t) \propto \exp(f(y_t, h_t)). \quad (1)$$

其中:  $f(\cdot)$  是一个非线性函数;  $h_t$  是第  $t$  个字符对应的隐状态, 通常定义为

$$h_t = g(h_{t-1}, x_t), \quad (2)$$

$g(\cdot)$  是一个非线性激活函数。

### 2.2 单向门限递归单元

从式(2)可以看出, RNN 在第  $t$  个字符产生的隐状态  $h_t$  包含了从第 1 个字符至第  $t$  个字符的全部历史信息, 从而在理论上能够处理长距离依赖关系。然而, RNN 在训练过程中存在“梯度消失”或“梯度爆炸”问题, 使得当前隐状态主要受邻近隐状态的影响, 受远距离隐状态的影响趋近于零<sup>[15]</sup>。

为了解决这个问题, 更复杂的非线性激活函数不断被提出, 其中最具代表性的是长短时记忆(long short-term memory, LSTM)<sup>[16]</sup> 和门限递归单元(gated recurrent unit, GRU)<sup>[9,17]</sup>。2 种激活函数的基本思想比较接近, 都是通过门限机制来控制

给定一个维吾尔词“مۆكەممەلەش”(即“完全地”), 神经网络以维吾尔字符序列作为输入, 输出每个字符的标记。

信息在循环神经网络中的传播, 使神经网络具备选择性“记忆”或“遗忘”特定信息的能力。由于 GRU 架构更加简单, 目前在自然语言处理中获得了广泛的应用, 因此本文采用 GRU。

图 2 给出了 GRU 的示意图, 其形式化描述如下:

$$h_t = (1 - z_t) h_{t-1} + z_t \tilde{h}_t. \quad (3)$$

其中  $z_t$  表示处理第  $t$  个字符时的更新门(update gate), 用于实现历史隐状态信息  $h_{t-1}$  和当前隐状态信息  $\tilde{h}_t$  的线性插值。 $z_t$  越大, GRU 越倾向于“遗忘”历史隐状态信息而“记忆”当前隐状态信息。

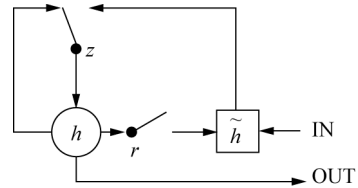


图 2 门限递归单元

当前隐状态信息的定义如下:

$$\tilde{h}_t = \tanh(\mathbf{W}x_t + \mathbf{U}(r_t \odot h_{t-1})). \quad (4)$$

其中:  $\mathbf{W}$  和  $\mathbf{U}$  是权重矩阵;  $\odot$  是元素级乘法运算符;  $r_t$  表示处理第  $t$  个字符时的重置门(reset gate), 用于平衡当前输入信息  $x_t$  和历史隐状态信息  $h_{t-1}$  对生成当前隐状态信息  $\tilde{h}_t$  的影响。 $r_t$  越小,  $x_t$  对于生成  $\tilde{h}_t$  的影响越大,  $h_{t-1}$  所起的作用则越小。

因此, 通过更新门和重置门, GRU 能够缓解原始 RNN 所面临的“梯度消失”或“梯度爆炸”问题, 有效处理维吾尔语形态切分中的长距离依赖关系。

### 2.3 双向门限递归单元

虽然从左往右的单向门限递归单元能够充分利用历史信息,解决长距离依赖关系问题,但是并没有考虑当前字符右边的历史信息。以中文为例,给定一个未切分的中文句子“南京市长江大桥”,如果从左往右扫描句子,句子的歧义性将很有可能导致生成错误的切分“南京/市长/江大桥”。如果从右向左扫描句子,则可能生成正确的切分“南京/市/长江/大桥”。

因此,充分利用双向上下文对于消解切分歧义具有重要的意义。这种现象不仅出现在汉语中,在维吾尔语中也非常普遍。因此,本文通过引入双向门限递归单元<sup>[17]</sup>来对维吾尔语形态切分中双向上下文信息进行建模。双向递归门限单元最早在神经机器翻译中使用<sup>[17]</sup>,双向递归神经网络<sup>[18]</sup>和双向LSTM<sup>[19]</sup>也曾在语音识别等任务中获得成功应用。

图1中,双向门限递归单元神经网络以维吾尔语字符序列作为输入,在输入层首先利用GRU从左往右生成正向隐状态序列 $\{\vec{h}_1, \vec{h}_2, \dots, \vec{h}_t, \dots, \vec{h}_T\}$ ,然后再从右往左生成逆向隐状态序列 $\{\overleftarrow{h}_1, \overleftarrow{h}_2, \dots, \overleftarrow{h}_t, \dots, \overleftarrow{h}_T\}$ 。对于第 $t$ 个字符,正向隐状态和逆向隐状态联合起来组成该字符的隐状态:

$$h_t = (\vec{h}_t, \overleftarrow{h}_t), \quad (5)$$

最后神经网络使用式(1)计算输出的标记。

## 3 实验

### 3.1 实验设置

由于目前维吾尔语形态切分缺乏公开的标注数据集,本文手工建立了一个维吾尔语形态切分数据集。首先从天山网(<http://uy.ts.cn/>)下载维吾尔语网页,从中提取19 629个维吾尔语词,然后进行人工切分和校对,获得8 227种语素类型。这些语素类型在数据集中出现47 763次。

表2中,本文将数据集随机切分成训练集、开发集和测试集。训练集用于训练模型参数,开发集用于优化模型超参数,测试集用于测试方法的切分性能。本文发表后将会公开该数据集。

表2 维吾尔语形态切分语料库

数据集	语素类型数量	语素数量	词语数量
训练集	6 315	42 882	17 629
开发集	959	2 435	1 000
测试集	953	2 446	1 000

本文对以下方法进行了对比实验:基于最小描述长度准则的半监督学习方法 Morfessor、CRF、前向门限递归单元(FGRU)神经网络、后向门限递归单元(BGRU)神经网络以及双向门限递归单元(BiGRU)神经网络。

对于Morfessor方法,本文采用形态分析开源工具包Morfessor 2.0<sup>[21]</sup>,使用其默认设置完成训练、调试和测试。

对于CRF方法,本文采用条件随机场开源工具包CRF++(<https://sourceforge.net/projects/crffpp/>)。特征定义参考文[10-11]中对基准系统CRF的特征定义。区别是文[10-11]中的数据集是基于句子的,因此既考虑了字级特征,也考虑了词级特征;而本文的数据集是基于词的,因此只考虑了字级特征。

对于基于GRU的3种方法(即FGRU、BGRU和BiGRU),本文在神经机器翻译开源工具包GroundHog(<https://github.com/lisa-groundhog/GroundHog>)的基础上进行扩展,实现了维吾尔语形态切分功能。GRU的隐层大小设置为1 000。采用GroundHog自带的训练算法在图形处理器(GPU)上进行训练,训练速度为每小时完成1万次迭代。

本文使用标准的评价指标准确率、召回率和 $F$ 值来衡量形态切分性能。

### 3.2 实验结果

#### 3.2.1 维吾尔语形态切分的数据稀疏性

首先对维吾尔语形态切分的数据稀疏程度进行定量分析。

表3给出了维吾尔语词语在数据集上出现的频度和比例。可以发现8 589个维吾尔语词语在数据集中只出现1次,占有所有词语的45.4%;高频词所占比例非常低,如出现5次的词语仅占3.3%。低频词过多会导致数据稀疏问题,使得模型难以准确估计低频事件的概率。

表3 维吾尔语词语频度和比例

词语频度	词语数量	比例/%
1	8 589	45.4
2	2 880	15.2
3	1 488	7.8
4	977	5.1
5	636	3.3
>5	4 416	23.2

进一步分析维吾尔语语素在数据集上出现的频度和比例,如表4所示。维吾尔语素比词语更加稀疏,出现1次的语素所占比例达到55.9%,出现超过5次的语素所占比例仅为15.1%。这说明即使把维吾尔语词语切成语素序列,依然面临着严重的数据稀疏问题。

表4 维吾尔语语素频度和比例

语素频度	语素数量	比例/%
1	3 576	55.9
2	915	13.6
3	498	7.4
4	306	4.6
5	231	3.4
>5	1 013	15.1

因此,通过采用连续、稠密的实数向量作为维吾尔语语素的表示来缓解数据稀疏问题是非常重要的。

### 3.2.2 向量维度的影响

由于向量的维度作为一个超参数决定了模型的表达能力,从而直接影响了维吾尔语形态切分的性能。为此,本文在开发集上考察了不同向量维度对BiGRU形态切分准确率、召回率和F值的影响。

表5显示当向量维度设为300的时候,BiGRU取得了最高的准确率、召回率和F值。这是因为当维度过低时模型的表达能力较弱,而维度过高时训练数据不足以充分训练过多的参数。因此,本文在后续实验中将向量维度固定为300。

表5 向量维度对BiGRU形态切分性能的影响

向量维度	准确率/%	召回率/%	F
100	96.37	96.42	97.53
200	96.79	96.10	97.67
300	97.08	96.73	97.85
400	96.16	96.70	97.59
500	96.05	96.06	97.52

### 3.2.3 对比实验结果

表6列出了各种方法在测试集上的准确率、召回率和F值。

Morfessor作为一种基于最小描述长度准则的半监督学习方法,在维吾尔语形态切分上并没有取得良好的效果,可能的原因是Morfessor采用离散

表示,在语素层面仍然面临着严重的数据稀疏问题。

表6 对比实验结果

方法	准确率/%	召回率/%	F
Morfessor	73.10	73.40	73.10
CRF	94.20	95.23	96.80
FGRU	89.48	86.76	84.84
BGRU	84.76	81.74	90.66
BiGRU	96.74	97.39	97.81

CRF作为一种基于概率图模型的有监督学习方法,相比Morfessor而言性能取得大幅度的提升,原因在于手工设计的特征函数能够较好地刻画维吾尔语形态切分的规律。

单向门限递归单元神经网络(即FGRU和BGRU)相比Morfessor而言性能取得了显著的提升,但是没有超过CRF的。FGRU和BGRU产生了差异性非常大的形态切分结果,并且具有较高的互补性。相对而言,BGRU要优于FGRU。

BiGRU综合了FGRU和BGRU的优点,能够充分利用双向上下文信息,取得了最高的准确率、召回率和F值。

表7给出了维吾尔语形态切分的实例。给定一个维吾尔语词“ھەيدىگەنلىكى”(即形容词“驾驶的”),正确的切分结果应该是:“ھەيدى”(即动词“驾驶”)、“گەنلىك”(即形动词附加成分)和“ى”(即第三人称)。FGRU从正向进行切分,错误地切成了形动词“ھەيدىگەنلىك”与第三人称“ى”2个语素。与之相反,BGRU从反向进行切分,虽然正确切出了动词“ھەيدى”,却错误地将剩下的部分切为形动词完成时“گەنلىكى”。由此可见,2个方向具有互补性。因此,BiGRU能够综合2个方向的优点并产生正确的切分结果。

表7 维吾尔语形态切分实例分析

方法	结果
FGRU	ھەيدىگەنلىك + ى
BGRU	ھەيدى + گەنلىكى
BiGRU	ھەيدى + گەنلىك + ى

## 4 结论

本文为维吾尔语形态切分提出了基于双向门限递归单元神经网络的方法,通过连续表示缓解维吾尔语在词汇和语素层面所面临的数据稀疏问题,根

据门限机制对维吾尔语字符之间的长距离依赖进行建模,充分利用双向上下文信息进行切分歧义消解。实验结果表明,本文方法在性能上显著超过了以 Morfessor 和 CRF 为代表的主流统计方法。

### 参考文献 (References)

- [1] Orhun M, Tanguã C, Adah E. Rule based analysis of the Uyghur nouns [J]. *International Journal on Asian Language Processing*, 2009, **19**(1): 33 - 43.
- [2] Sami V, Peter S, Arne G et al. Morfessor 2.0: Python Implementation and Extensions for Morfessor Baseline, ISBN 978-952-60-5501-5 [R]. Helsinki: Aalto University, 2013.
- [3] Lafferty J, McCallum A, Pereira F. Conditional random fields: probabilistic models for segmenting and labeling sequence data [C] // Proceedings of the 18th International Conference on Machine Learning. Williamstown, MA, USA: Morgan Kaufmann, 2001: 282 - 289.
- [4] Ruokolainen T, Kohonen O, Virpioja S et al. Supervised morphological segmentation in a low-resource learning setting using conditional random fields [C]// Proceeding of the Seventeenth Conference on Computational National Language Learning. Sofia, Bulgaria: Association for Computational Linguistics, 2013: 8 - 9.
- [5] Aisha B, SUN Maosong. A statistical method for Uyghur tokenization [C]// International Conference on Natural Language Processing and Knowledge Engineering. Dalian: IEEE, 2009: 24 - 27.
- [6] 买热哈巴·艾力,姜文斌,王志洋,等. 维吾尔语词法分析的有向图模型 [J]. *软件学报*, 2012, **23**(12): 3115 - 3129.  
Aili M, JIANG Wenbin, WANG Zhiyang, et al. Directed graph model of Uyghur morphological analysis [J]. *Journal of Software*, 2012, **23**(12): 3115 - 3129. (in Chinese)
- [7] Wumaier A, Tian S. Conditional random fields combined FSM stemming method for Uyghur [C]// International Conference on Computer Science and Information Technology. Beijing: IEEE, 2009: 8 - 11.
- [8] Ablimit M, Kawahara T, Pattar A, et al. Stem-affix based Uyghur morphological analyzer [J]. *International Journal of Future Generation Communication and Networking*, 2016, **9**(2): 59 - 72.
- [9] Chung J, Gulcehre C, Cho K, et al. Empirical evaluation of gated recurrent neural networks on sequence modeling [Z/OL]. (2014-12-11). <https://arxiv.org/abs/1412.3555>.
- [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. Lisbon, Portugal: Association for Computational Linguistics, 2015: 17 - 21.
- [11] Yao Y, Huang Z. Bi-directional LSTM recurrent neural network for Chinese word segmentation [Z/OL]. (2016-02-16). <http://arxiv.org/abs/1602.04874>.
- [12] Morita H, Kawahara D, Kurohashi S. Morphological analysis for unsegmented languages using recurrent neural network language model [C]// Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. Lisbon, Portugal: Association for Computational Linguistics, 2015: 17 - 21.
- [13] Wang L, Cao Z, Xia Y, et al. Morphological segmentation with window ISTM neural networks [C]// Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence. Phoenix, AZ, USA: Association for the Advancement of Artificial Intelligence, 2016: 2842 - 2848.
- [14] Wang P, Qian Y, Soong F, et al. Part-of-speech tagging with bidirectional long short-term memory recurrent neural network [Z/OL]. (2015-10-21). <http://arxiv.org/abs/1510.06168>.
- [15] Bengio Y, Simard P, Frasconi P. Learning long-term dependencies with gradient descent is difficult [J]. *IEEE Transactions on neural networks*, 1994, **5**(2): 157 - 166.
- [16] Hochreiter S, Schmidhuber J. Long short-term memory [J]. *Neural Computation*, 1997, **9**(8): 1735 - 1780.
- [17] Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate [Z/OL]. (2014-09-01). <https://arxiv.org/abs/1409.0473>
- [18] Schuster M, Paliwal K. Bidirectional recurrent neural networks [J]. *IEEE Transactions on signal processing*, 1997, **45**(11): 2673 - 2681.
- [19] Graves A, Jaitly N, Mohamed A. Hybrid speech recognition with deep bidirectional ISTM [C]// 2013 IEEE Workshop on Automatic Speech Recognition and Understanding. Olomouc, Czech: IEEE, 2014: 8 - 12.