

## 基于多注意力机制的维吾尔语人称代词指代消解

杨启萌<sup>1,2,3</sup> 禹龙<sup>2,3,4</sup> 田生伟<sup>1,2,3</sup> 艾山·吾买尔<sup>2,3,5</sup>

**摘要** 针对深度神经网络模型学习照应语和候选先行语的语义信息忽略了每一个词在句中重要程度,且无法关注词序列连续性关联和依赖关系等问题,提出一种结合语境多注意力独立循环神经网络(Contextual multi-attention independently recurrent neural network, CMAIR)的维吾尔语人称代词指代消解方法.相比于仅依赖照应语和候选先行语语义信息的深度神经网络,该方法可以分析上下文语境,挖掘词序列依赖关系,提高特征表达能力.同时,该方法结合多注意力机制,关注待消解对多层次语义特征,弥补了仅依赖内容层面特征的不足,有效识别人称代词与实体指代关系.该模型在维吾尔语人称代词指代消解任务中的准确率为 90.79%,召回率为 83.25%,F 值为 86.86%.实验结果表明,CMAIR 模型能显著提升维吾尔语指代消解性能.

**关键词** 注意力机制,语境,独立循环神经网络,指代消解

**引用格式** 杨启萌,禹龙,田生伟,艾山·吾买尔.基于多注意力机制的维吾尔语人称代词指代消解.自动化学报,2021,47(6):1412-1421

**DOI** 10.16383/j.aas.c180678

## Anaphora Resolution of Uyghur Personal Pronouns Based on Multi-attention Mechanism

YANG Qi-Meng<sup>1,2,3</sup> YU Long<sup>2,3,4</sup> TIAN Sheng-Wei<sup>1,2,3</sup> AISHAN Wumaier<sup>2,3,5</sup>

**Abstract** The deep neural network model learns the semantic information of anaphora and candidate antecedent, ignores the importance of each word in the sentence, and cannot pay attention to the continuous association and dependence of the word sequence. This paper proposes a Uyghur personal pronoun anaphora resolution method based on contextual multi-attention independent recurrent neural network (CMAIR). Compared with deep neural networks that rely only on the semantic information of anaphora and candidate antecedent, this method can analyze context relations, mine word sequence dependencies, and improve feature expression ability. At the same time, this method combines the multi-attention mechanism, pays attention to the multi-layer semantic features to be resolved, effectively compensates for the lack of content-level features, and effectively recognizes the relationship between personal pronouns and entities. The precision rate of this method in the Uyghur personal pronoun anaphora resolution task is 90.79%, the recall rate is 83.25%, and the F value is 86.86%. The experimental results show that the CMAIR model can significantly improve the performance of Uyghur personal pronoun anaphora resolution.

**Key words** Attention mechanism, context, independently recurrent neural network, anaphora resolution

**Citation** Yang Qi-Meng, Yu Long, Tian Sheng-Wei, Aishan Wumaier. Anaphora resolution of Uyghur personal pronouns based on multi-attention mechanism. *Acta Automatica Sinica*, 2021, 47(6): 1412-1421

收稿日期 2018-10-18 录用日期 2018-12-24  
Manuscript received October 18, 2018; accepted December 24, 2018

国家自然科学基金(61563051, 61662074, 61962057), 国家自然科学基金重点项目(U2003208), 自治区重大科技项目(2020A03004-4), 新疆自治区科技人才培养项目(QN2016YX0051)资助

Supported by National Natural Science Foundation of China (61563051, 61662074, 61962057), Key Program of National Natural Science Foundation of China (U2003208), Major Science and Technology Projects in the Autonomous Region (2020A03004-4), Xinjiang Uygur Autonomous Region Scientific and Technological Personnel Training Project (QN2016YX0051)

本文责任编辑 张民

Recommended by Associate Editor ZHANG Min

1. 新疆大学软件学院 乌鲁木齐 830008 2. 新疆大学软件工程技术重点实验室 乌鲁木齐 830046 3. 新疆大学信号与信息处理重点实验室 乌鲁木齐 830046 4. 新疆大学网络中心 乌鲁木齐 830046 5. 新疆大学信息科学与工程学院 乌鲁木齐 830046

1. School of Software, Xinjiang University, Urumqi 830008

指代 (Anaphora) 作为一种常见的语言现象,广泛存在于自然语言的表达之中.它对语言的简化表达、主题的突出性描述和语言表达连贯性起着重要的作用.对指代成分准确无歧义的消解有助于机器分析和语篇理解<sup>[1]</sup>.在语言学中,指代词称为照应语 (Anaphor),用于指向另一个语言单位,被指代词称为先行语 (Antecedent),用于被指向的语言单位,指代消解 (Anaphora Resolution) 就是确定照应语所指代的先行语的过程<sup>[2]</sup>.维吾尔语人称代词指代消

2. Key Laboratory of software engineering technology, Xinjiang University, Urumqi 830046 3. Key Laboratory of Signal and Information Processing, Xinjiang University, Urumqi 830046 4. Network Center, Xinjiang University, Urumqi 830046 5. College of formation Science and Technology, Xinjiang University, Urumqi 830046

解是研究人称代词与句中名词和名词性短语的指代关系, 图 1 给出维吾尔语人称代词指代消解例句。

如果一个人称代词存在指代关系, 那么它与相关文本中的一个或多个提及的名词或名词性短语关联。本句中“ئوسمان (吾斯英)”和“ئۇنى (他)”存在指代关系, “بۇياگې رەيخانگۈل (布葛热汗)”和“ئۇنى (他)”不存在指代关系。

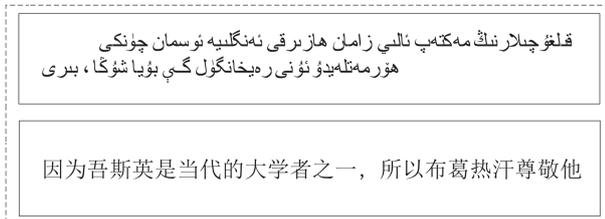


图 1 维吾尔语人称代词指代消解例句

Fig. 1 The example of Uyghur personal pronoun anaphora resolution

近年来, 随着深度学习技术在语音识别、计算机视觉、图像识别等领域的重大突破, 学者们尝试将深度学习模型应用于自然语言处理任务中。例如 Kim 使用卷积神经网络 (Convolutional neural network, CNN) 进行句子建模, 解决情感分类任务<sup>[3]</sup>, Irsoy 等使用循环神经网络 (Recurrent neural network, RNN) 进行意见分析<sup>[4]</sup>, Tai 等使用长短时记忆网络 (Long short term memory network, LSTM) 解决情感分类问题<sup>[5]</sup>, 这些基于深度学习的方法在自然语言处理任务中取得了比以往研究更好的分类效果。

指代消解作为自然语言处理一个重要子任务, 深度学习模型在指代消解中得到广泛的研究。这些研究关注照应语和候选先行语的语义信息, 应用大量的神经网络模型进行候选先行语预测<sup>[6-8]</sup>。目前的研究主要针对中文和英文等具有充足语料库的语种, 对维吾尔语等小语种的研究不够深入, 针对小语种的研究无论是语料标注还是实体识别都需要掌握多级语法知识、语义知识, 甚至相应语言领域知识, 在当前自然语言处理的研究阶段, 要获取和学习研究中所需知识仍比较困难。人称代词指代消解作为指代消解任务更细粒度的一个分支, 不仅依赖照应语和候选先行语特征信息, 还要关注距离特征和上下文语境信息。例如句子:

چۈنكى ئوسمان ئەنگلىيە ھازىرقى زامان ئالىي مەكتەپ قەلغۇچىلارنىڭ بىرى، شۇڭا بۇياگې رەيخانگۈل ئۇنى ھۆرمەتلەيدۇ (因为吾斯英是当代的大学者之一, 所以布葛热汗尊敬他)

我们普遍认为与照应语距离越近的候选先行语存在指代关系概率越大, 根据候选先行语“ئوسمان (吾斯英)”、“مەكتەپ (学者)”和“بۇياگې رەيخانگۈل (布葛热

汗)”, 很难推断“ئوسمان (吾斯英)”是否为照应语“ئۇنى (他)”正确的先行语, 在这种情况下, 研究者会错误地判断“بۇياگې رەيخانگۈل (布葛热汗)”为“ئۇنى (他)”的先行语, 因为文本中“بۇياگې رەيخانگۈل (布葛热汗)”与“ئۇنى (他)”距离更近。但是, 候选先行语“ئوسمان (吾斯英)”才是照应语“ئۇنى (他)”正确的先行语。所以, 人称代词指代消解应该充分考虑候选先行语距离特征和更深层次的语境信息。

针对以上问题, 本文提出基于多注意力机制的深度学习模型应用于维吾尔语人称代词指代消解任务。注意力机制最早应用于视觉图像领域, 目的是在神经网络训练过程中将注意力集中到图像的特定部分。Mnih 等首次提出在 RNN 模型上使用注意力机制进行图像分类<sup>[9]</sup>, 验证了注意力机制在图像处理领域的有效性。随后 Bahdanau 等将注意力机制应用于机器翻译任务<sup>[10]</sup>, 模型取得了显著的效果并使注意力机制成功应用于自然语言处理领域。随着研究不断深入, Yin 等提出基于注意力机制的卷积神经网络进行句子建模<sup>[11]</sup>, Wang 等使用基于注意力机制的 LSTM 解决细粒度的情感分析问题<sup>[12]</sup>, 这些方法的提出验证了注意力机制结合深度学习模型的有效性。

本文结合三种注意力机制来构造 CMAIR 模型: 1) 词向量注意力机制: 突出照应语和候选先行语的重要程度, 2) 距离注意力机制: 有效表达不同词语对消解结果的贡献度, 3) 词性注意力机制: 是句中词和词性的关联。同时, 本文结合 IndRNN 构造语境特征, 并将这 4 种特征作为模型的输入, 可以从多个层面学习词语级和句子级特征信息, 提高指代消解性能。

本文提出的 CMAIR 模型使用 3 种注意力机制和语境特征构建模型, 使模型关注多种特征信息。最后, 将 CMAIR 模型在维吾尔语数据集进行实验, 实验结果表明, 本方法取得了比以往研究更好的分类效果。本文的主要贡献有 5 方面:

- 1) 提出一种结合语境的多注意力独立循环网络 (CMAIR) 应用在维吾尔语人称代词指代消解任务中, 实验取得了当前研究最好效果;
- 2) 本文首次考虑结合语境的深度学习模型, 学习词序关联和依赖关系, 结合语境信息增强了特征表达力;
- 3) 本文提出的词向量、词性、距离注意力机制能从三方面获取更深层次的特征信息, 弥补了单注意力机制仅关注内容层面信息的不足;
- 4) CMAIR 模型结构简单, 具有很强的鲁棒性和泛化能力, 无需额外的句法分析和语义依存分析相关知识;
- 5) 提出一种距离识别算法, 能准确计算待消解

对距离, 识别不同词在句中重要程度, 使 CMAIR 可以充分利用距离信息.

## 1 相关工作

### 1.1 指代消解

指代消解作为自然语言处理重要子任务, 是细粒度的文本分类任务, 更加关注照应语和候选先行语深层次语义特征, 一直以来都得到众多学者的研究和关注. Soon 等首次给出利用机器学习的方法进行指代消解基本步骤, 其思想是把指代消解任务看作二分类问题, 从已标注好的语料库中提取语料的各类词法、语法、语义特征作为消解框架的输入并利用分类器完成指代消解任务<sup>[13]</sup>; Ng 等对 Soon 等研究进行了改进, 抽取 53 个特征, 指代消解的效果有显著地改善<sup>[14]</sup>; Yang 等提出将特征方法和聚类相结合, 模型消解性能有较大的提高<sup>[15]</sup>; 这些方法通过传统的机器学习构造文本语法、句法特征, 能充分利用文本内容层面特征, 有效进行指代关系识别. 此外, 指代消解在多种语言已得到广泛的研究.

近年来, 随着深度学习在自然语言处理领域的广泛应用, 许多研究者也利用深度学习方法解决指代消解任务. Chen 等针对此问题提出一种神经网络模型, 在他们的工作中, 照应语和候选先行语用前馈神经网络编码<sup>[16]</sup>. Clark 等尝试对指代消解的一种神经网络 Mention Rank 模型的启发式损失函数中的超参数利用强化学习方式优化, 提出一种奖励衡量机制, 跟其他方式相比结果突出<sup>[17]</sup>. Iida 等提出一种句内主题指代关系并利用多列卷积神经网络 (Multi-column CNN) 来预测指代关系, 在多个基准测试中取得了很好的结果<sup>[8]</sup>. Yin 等建立一个深度记忆网络获取零代词和先行语语义关联信息<sup>[18]</sup>. 这些方法使用深度学习技术来解决指代消解问题, 可以在不需要依存树、句法关系的情况下通过深度神经网络更好地学习文本的深层次特征信息, 取得了比传统机器学习方法更好的分类效果.

针对维吾尔语指代消解问题, 目前已有学者进行研究. 李敏等提出一种基于栈式自编码深度学习的算法进行维吾尔语名词短语指代消解<sup>[19]</sup>. 田生伟等挖掘维吾尔语隐含的上下文语义层面特征, 提出利用 BiLSTM 的深度学习机制进行基于深层语义信息的维吾尔语人称代词指代消解, 维吾尔语人称代词指代消解的 F 值达到 76.86%, 证明了该方法的有效性<sup>[20]</sup>. 李冬白等通过堆叠多层 RBM 网络和一层 BP 网络构建 DBN 深度网络学习模型, 经过维吾尔语语料库测试, F 值达到 83.81%, 该方法能有效提升维吾尔语人称代词指代消解性能, 推动了维吾尔语指代消解的研究<sup>[21]</sup>.

本文提出的 CMAIR 模型与现有技术的主要区

别在于多注意力机制和 IndRNN 的应用, 我们将指代消解作为多注意力机制和 IndRNN 环境下关注词语重要程度和学习文本深层次语境特征的过程. 在 IndRNN 模型的帮助下, CMAIR 模型学习以顺序的方式对候选先行语进行分类, 从而做出全局最优的决策.

### 1.2 独立循环神经网络

独立循环神经网络 (IndRNN) 是由 Li 等提出的一种新型网络结构, 不仅可以解决传统 RNN 所存在的梯度消失和梯度爆炸问题, 还学习长期依赖关系; 此外, 借助 ReLU 等非饱和和激活函数, 训练之后 IndRNN 更具鲁棒性, 并且通过堆叠多层 IndRNN 还可以构建比现有 RNN 更深的网络<sup>[22]</sup>. 实验结果表明, 与传统的 CNN 和 LSTM 相比, 使用 IndRNN 可以在各种任务中取得更好的结果. IndRNN 基本结构如图 2 所示.

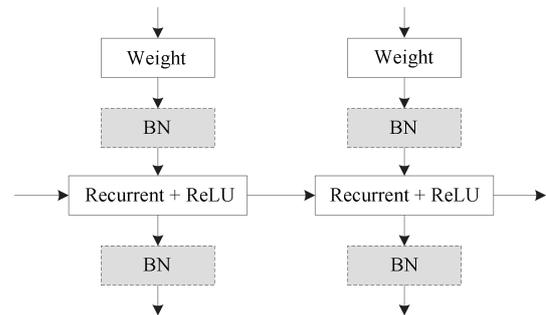


图 2 IndRNN 结构图

Fig. 2 The structure diagram of IndRNN

其中 weight 和 Recurrent + ReLU 表示每一时间步对输入的处理步骤, ReLU 是激活函数, 为了加速训练, 每层之后插入了批标准化 (Batch normalization, BN), 通过堆叠这种基础架构, 可以建立一个深度 IndRNN 网络. 其用公式可以表示为:

$$h_t = \sigma(Wx_t + \mu \odot h_{t-1} + b) \quad (1)$$

其中循环权重  $\mu$  是一个向量,  $\odot$  表示 Hadamard 积 (对应元素相乘). 同一图层中的每个神经元都与其他神经元不相连, 通过叠加两层或更多层的 IndRNN, 可以将神经元连接. 对于第  $n$  个神经元, 隐藏层  $h_{n,t}$  可以通过以下公式得到:

$$h_{n,t} = \sigma(w_n x_t + u_n h_{n,t-1} + b_n) \quad (2)$$

其中  $w_n$  和  $u_n$  分别表示第  $n$  行的输入权重和当前权重, 每个神经元只接收当前状态隐藏层和输入信息. IndRNN 中的每个神经元都有独立的时空特征, 随着参数被共享, RNN 被视为多层感知器. 与传统的 RNN 不同, IndRNN 将随时间独立地聚合空间模式.

由于权重层用来处理输入, 可以将其自然地扩展到其他多个图层中加深处理过程. IndRNN 结构简单, 也可以很容易地添加到不同的网络结构中.

## 2 任务描述

对于分词后长度为  $n$  的句子, 保留句中名词、名词短语和人称代词, 去除其他词语, 得到  $s = \{w_1, w_2, \dots, w_i, \dots, a_i, \dots, w_n\}$ , 其中  $w_i$  为第  $i$  个分词,  $a_i$  为人称代词. 分别把  $s$  中词语逐个向前匹配得到待消解对, 例如  $a_i$  逐个向前匹配得到  $\{a_i w_{i-1}, a_i w_{i-2}, a_i w_{i-3}, \dots, a_i w_1\}$ , 本文的任务是针对  $a_i$  确定其正确对应的先行语, 完成维吾尔语指代消解, 例如句子:

چۈنكى ئوسمان ئەنگلىيە ھازىرقى زامان ئالىي مەكتەپ  
قىلغۇچىلارنىڭ بىرى، شۇڭا بۇياڭ رەيخانگۈل ئۇنى ھۆرمەتلەيدۇ

(因为吾斯英是当代的大学者之一, 所以布葛热汗尊敬他)

其中“ئۇنى (他)”为人称代词, 充当句中照应语成分, “ئوسمان (吾斯英)”、“مەكتەپ (学者)”和“بۇياڭ رەيخانگۈل (布葛热汗)”为名词或名词短语, 充当句中候选先行语成分. 本文将分词后的词语映射为一个连续值多维词向量矩阵  $\mathbf{R}^{K \times V}$ , 其中  $K$  为词典大小, 即数据集中所有分词的数量,  $V$  为每个词向量的维度, 即把一个词映射为  $V$  维的词向量  $x^i \in \mathbf{R}^V$ . 对于上述例句, 本文将照应语分别与句中候选先行语逐个匹配获得指代链  $\{\langle \text{ئوسمان (吾斯英)}, \text{ئۇنى (他)} \rangle, \langle \text{مەكتەپ (学者)}, \text{ئۇنى (他)} \rangle, \langle \text{بۇياڭ رەيخانگۈل (布葛热汗)}, \text{ئۇنى (他)} \rangle\}$ , 对于长度为  $n$  的句子, 本文通过词典  $\{x_1, x_2, x_3, \dots, x_n\}$  提取指代链中候选先行语和照应语之间的特征信息来判断指代链指代关系.

## 3 多注意力 IndRNN 模型

本文通过三种注意力机制充分利用文本语义信息, 输入 CNN 进行卷积核池化操作提取文本局部特征, 并结合 IndRNN 挖掘深层次语境信息从而有效完成指代消解任务. 如图 3 所示, CMAIR 模型主要由 5 部分组成:

1) 注意力矩阵输入层: 存储三种注意力机制的注意力特征向量信息.

2) CNN 层: 将 3 种注意力特征向量和 hand-crafted 特征向量合并为输入特征矩阵, 依次进行卷积和池化操作, 为了使网络模型结构简单化, 本文选用一层卷积的方式进行输入数据提取局部特征, 并使用最大池化进一步降低数据计算复杂度, CNN 层可以保留文本丰富的重要局部特征.

3) IndRNN 层: 将长度为  $n$  的句子  $s = \{w_1, w_2, w_3, \dots, w_i, \dots, w_n\}$  的词向量依次输入 In-

dRNN 中, 其中  $w_i$  为第  $i$  个分词, IndRNN 可以学习文本长期依赖关系, 挖掘更深层次语境特征, 用于后续的处理过程.

4) 全连接层: 用于结合模型中 CNN 层提取的局部特征和 IndRNN 提取的语境特征, 形成输入特征向量的综合表示, 并作为输出层的输入.

5) 输出层: 本文选用 *Softmax* 分类器进行分类, 从而确定照应语和候选先行语指代关系. 分类结果中, 对于一个照应语 *Anaphor* 和  $n$  个候选先行语 *Antecedent*,  $\text{Antecedent}_k (k = 1, 2, \dots, n)$  为照应语 *Anaphor* 正确对应的先行语.

### 3.1 方法概述

为了更好地识别候选先行语和人称代词指代关系, 本文使用 3 种注意力机制使模型从不同层面关注实体重要信息, 并加入 IndRNN 学习文本长期依赖关系, 挖掘文本语境信息, 获取语境特征, 结合 3 种注意力机制作为模型的输入, 并在全连接层加入语境特征能有效减少模型复杂度提高指代消解性能.

如表 1 所示, 对于长度为  $n$  的句子  $s = \{w_1, w_2, w_3, \dots, w_i, \dots, w_n\}$ , 其中  $w_i$  为第  $i$  个分词, 分别对  $w_i$  标注句中成分, 并将每个词语映射为  $k$  维向量, 对于句子  $s$  依次按时序把  $w_i$  输入 IndRNN 学习词语长期依赖关系并获取语境特征, 取指代链中候选先行语和照应语两两匹配形成待消解对, 再从词典提取待消解对中候选先行语和照应语词向量、词性向量和距离向量构建 3 种注意力机制.

表 1 词语句中成分标注

Table 1 Component labeling of words in sentences

标注词语	中文翻译	句中成分 1	句中成分 2
چۈنكى	因为		
ئوسمان	吾斯英	先行语	
ئەنگلىيە	是		
ھازىرقى	当代的		(- 先行语
مەكتەپ	大学者	先行语	
بىرى	之一		先行语 -)
شۇڭا	所以		
بۇياڭ رەيخانگۈل	布葛热汗	先行语	
ھۆرمەتلەيدۇ	尊敬		
ئۇنى	他	照应语	

1) 词向量注意力机制: 提取待消解对中候选先行语、照应语词向量, 并分别与其他词向量矩阵做运算操作得到词向量注意力特征矩阵, 这是待消解对中对于文本内容层面的注意力.

2) 词性注意力机制: 是词和词性关联的属性, 是对词向量注意力的进一步补充, 可以让模型从另一

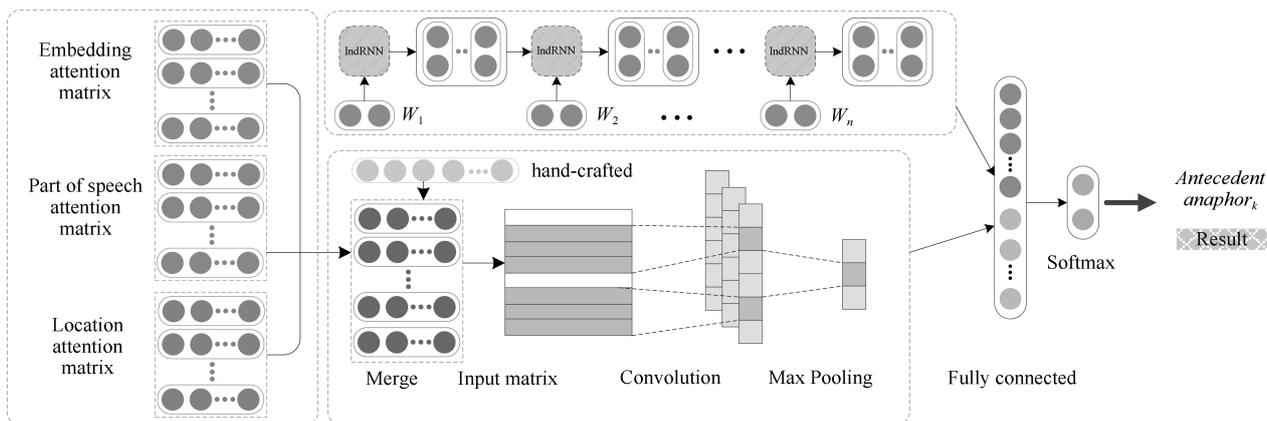


图 3 多注意力机制 IndRNN 模型框架图

Fig. 3 IndRNN model framework with multiple attention mechanisms

方面关注待消解对,从而学到更多隐藏信息.

3) 距离注意力机制: 该方法是融合待消解对候选先行语与照应语间隔词语词向量的方式作为待消解对距离表示. 结合其他两种注意力机制可以充分表述句中每个词语的重要程度.

4) 语境特征: 是句子级的特征表达和注意力机制的补充, 可以学习句中词连续性和长期依赖关系.

融合 3 种注意力机制作为模型的输入, 可以从不同角度关注待消解对特征信息, 在全连接层加入语境特征可以更准确表达待消解对关系, 并减小网络复杂度, 从而可以有效识别指代关系, 提高指代消解性能.

### 3.2 词向量注意力机制

注意力机制是使模型在训练过程中高度关注特定信息, 以达到更好的分类效果. 对于人称代词指代消解任务, 文本内容层面信息最为重要, 多方面关注分析照应语和候选先行语语义信息可提高指代消解性能.

本文提出词向量注意力机制用于维吾尔语人称代词指代消解任务. 对于完整句子  $s = \{w_1, w_2, w_3, \dots, w_i, \dots, w_n\}$ ,  $w_i$  为第  $i$  个分词, 提取词典中  $w_i$  词向量矩阵, 再根据式 (3)、(4) 和其他词向量矩阵做运算操作获取词向量注意力特征矩阵  $C^T$ .

$$e_i^t = f_{ATT}(Z_{t-1}, v_i) \quad (3)$$

$$C_{i,i}^T = \frac{\exp(e_i^t)}{\sum_{j=1}^M \exp(e_j^t)} \quad (4)$$

其中  $v_i$  为分词  $w_i$  词向量,  $Z_{t-1}$  为分词  $w_i$  在上一时刻  $t-1$  的上下文信息,  $f_{ATT}$  为 *softmax* 打分函数, 矩阵  $C^T$  表示每一个词语的重要程度 (即概率), 注意力特征矩阵  $C^T$  可以改写成下式:

$$C_{i,i}^T = \alpha \times \frac{\exp(e_i^t)}{\sum_{j=1}^M \exp(e_j^t)} \quad (5)$$

其中  $\alpha$  为可调参数, 表示每一个词重要程度. 利用得到的注意力特征矩阵和  $w_i$  词向量矩阵运算即可得到模型输入矩阵:

$$Input_i^t = x_i \oplus C_{i,i}^T \quad (6)$$

其中  $\oplus$  表示拼接操作, 本文采用注意力特征矩阵和原词向量简单拼接操作构建模型输入矩阵.

### 3.3 词性注意力机制

照应语和候选先行语内容层面分析是指代消解的关键, 对于候选先行语和照应语在词典中出现频率较低的情况下仅使用词向量注意力机制会降低指代消解准确率. 针对此问题提出一种基于词性注意力机制用于弥补仅关注文本内容层面信息的不足.

针对句子  $s = \{w_1, w_2, w_3, \dots, w_i, \dots, w_n\}$ , 本文对句中每个分词词性进行重新标注, 标注结果如表 2 所示.

标注结果是标注词语和词性的组合, 是标注词语和词性的关联, 对于长度为  $n$  的句子, 可以将标注结果表示为式 (7) 的形式, 其中  $w_i$  是第  $i$  个词,  $c_i$  是词性,  $\oplus$  为拼接操作.

$$Z_{1,n} = w_1 \oplus c_1, w_2 \oplus c_2, w_3 \oplus c_3, \dots, w_n \oplus c_n \quad (7)$$

对于候选先行语是名词短语情况, 由于名词短语包含多个词语, 处理方式和以上不同. 针对这种情况, 本文提取名词短语中所有词语的词向量注意力矩阵, 并按照式 (8) 的方式获取名词短语词性注意力特征矩阵:

$$Z = \alpha \times \frac{\sum_{i=1}^n f_{ATT}(w_i \oplus c_i)}{n} \quad (8)$$

$\alpha$  是名词短语重要程度, 可以手动设置也可以在模型训练过程中自动学习.

和词向量处理方式一样, 提取所有句中标注结果映射为词性向量矩阵  $\mathbf{R}^{K \times V}$ , 其中  $K$  为词典大小,  $V$  为每一个词性标注结果映射的词向量维度, 即为词性注意力矩阵. 指代消解对中提取候选先行语和照应语词性注意力矩阵并按照式 (3)、(4) 获取词性特征注意力矩阵.

表2 词性标注  
Table 2 Part of speech tagger

标注词语	中文翻译	词性	标注结果
چۈنكى	因为	C_ 连词	چۈنكى/C
ئوسمان	吾斯英	NH_ 人名	ئوسمان/NH
ئەنگلىيە	是	ZD_ 系动词	ئەنگلىيە/ZD
ھازىرقى	当代的	ADJ_ 形容词	ھازىرقى/ADJ
مەكتەپ	大学者	N_ 名词	مەكتەپ/N
بىرى	之一	ADJ_ 形容词	بىرى/ADJ
شۇڭا	所以	C_ 连词	شۇڭا/C
بۇيا گې رەيخانگول	布葛热汗	NH_ 人名	بۇيا گې رەيخانگول/NH
ھۆرمەتلەيدۇ	尊敬	V_ 动词	ھۆرمەتلەيدۇ/V
ئۇنى	他	RP_ 人称代词	ئۇنى/RP

### 3.4 距离注意力机制

先行语和照应语的距离隐藏着重要信息, 对提高指代消解性能起着关键作用. 本文提出两种距离计算方式, 第一种方式为对于候选先行语  $w_i$  和照应语  $w_n$  提取所有间隔词语  $w_{i+1:n-1}$ , 提取方式如图 4 所示, 并从词典提取  $W$  中所有词向量并通过算法 1 中式 (9) 运算作为  $w_i$  与  $w_n$  距离表示, 其中  $\alpha$  为每个词的重要程度.

第二种方式为直接计算待消解对词语间隔个数并映射为  $K$  维词向量并结合词向量、词性注意力特征矩阵和语境特征作为模型的输入, 本文使用算法 1 计算不同情况下待消解对距离表示.

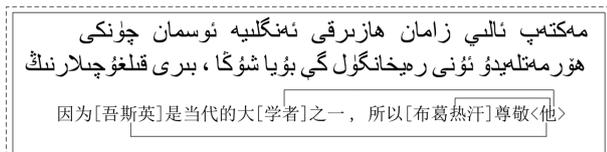


图4 距离计算方式举例

Fig. 4 Example of distance calculation

我们普遍认为待消解对中候选先行语和照应语距离越近存在指代关系的概率越大. 对于图 4 中的例子, 我们认为 “ئۇنى (他)” 和 “بۇيا گې رەيخانگول (布葛

热汗)” 存在指代关系的概率大于 “ئوسمان (吾斯英)”, 因为 “بۇيا گې رەيخانگول (布葛热汗)” 在句中距离 “ئۇنى (他)” 最近, 但是 “ئوسمان (吾斯英)” 才是 “ئۇنى (他)” 正确的先行语. 针对此问题, 本文提出一种距离识别算法用来准确识别待消解对中候选先行语和照应语的距离. 其计算过程如算法 1 所示.

算法 1. 距离识别算法.

输入. 长度为  $n$  的句子  $s$

输出. 所有指代消解对距离  $L$

步骤 1. 将照应语的位置的值置为 0, 所有候选先行语的位置置为  $n$ ,  $n$  为句子的长度;

步骤 2. 定义工作指针  $p$  从照应语的位置向前扫描;

步骤 3. 若扫描对象是词语  $w_i$  (非候选先行语), 则将  $w_i$  加入集合  $L$ , 若扫描对象是标点符号, 加入集合  $L$ , 执行步骤 4 继续扫描;

步骤 4. 若扫描对象是候选先行语, 按照式  $\varphi_1$  计算距离, 其中  $len(L)$  为集合  $L$  的长度,  $f_{ATT}$  为词典中词语  $L_i$  的词向量,  $\alpha$  为  $L$  中每一个词语  $L_i$  的重要程度,  $num(punc)$  为集合  $L$  中包含标点符号的数量;

$$\varphi_1 = \frac{\sum_{i=1}^{len(L)} \alpha \times f_{ATT} L_i}{len(L)} \times \min\{5, num(punc)\} \quad (9)$$

步骤 5. 将集合  $L$  置空并执行步骤 2 继续扫描, 若照应语和句中所有候选先行语匹配完成, 停止扫描.

### 3.5 语境

句子级的特征隐藏着重要信息, 对指代消解性能的提升有重要的作用. 本文从句子层面关注文本深层次信息获取语境特征. 对于长度为  $n$  的句子  $s = \{w_1, w_2, w_3, \dots, w_i, \dots, w_n\}$ , 把每一个分词  $w_i$  映射为  $k$  维向量  $w_i \in \mathbf{R}^k$  作为独立循环神经网络的输入, 相比于传统的 RNN 和 LSTM 网络 IndRNN 可以保留长期记忆, 处理长序列, 训练得到的模型具有鲁棒性, 在处理长序列依赖问题方面的性能优于传统的 RNN 和 LSTM 模型.

### 3.6 hand-crafted 特征

针对维吾尔语人称代词指代消解, 本文结合维吾尔语语言现象和语法特征, 如表 3 所示选择 7 个 hand-crafted 特征用于消解任务.

### 3.7 多注意力机制模型

卷积神经网络可以提取文本局部特征, 获取输入向量更多的隐藏信息, 从而更好地识别指代关系. 卷积层可以利用多卷积核获取输入结构化特征, 卷

表 3 hand-crafted 特征  
Table 3 The feature of hand-crafted

照应语词性		词性一致		单复数一致		性别一致			先行语语义角色			照应语语义角色			存在嵌套	
人称代词	非人称代词	是	否	是	否	是	否	未知	施事者	受事者	无	施事者	受事者	无	是	否
1	0	1	0	1	0	1	0	0.5	1	0.5	0	1	0.5	0	0	1

积操作得到特征向量图:

$$F = f(w \cdot x + b) \quad (10)$$

上式中  $x$  为 3 种注意力特征矩阵,  $w$  为对应权重矩阵,  $b$  为偏置项. 对于有  $m$  个卷积核的卷积操作其得到的特征图表示如下:

$$A = [c_1, c_2, c_3, \dots, c_m] \quad (11)$$

池化操作可以进一步处理特征图, 保留重要特征, 减少模型复杂度. 对于有  $n$  个窗口的池化层下采样后的结果表示如下:

$$\tilde{A} = [\tilde{a}_1, \tilde{a}_2, \tilde{a}_3, \dots, \tilde{a}_n] \quad (12)$$

模型采用 *sigmoid* 激活函数, 如式 (13) 所示, 其中  $z$  为卷积层或池化层的输出.

$$f(z) = \frac{1}{1 + \exp(-z)} \quad (13)$$

本文中池化层和 IndRNN 层的输出作为全连接层的输入, 并使用 *softmax* 分类器得到分类结果, CNN 和 IndRNN 的特征合并与分类如式 (14)、(15) 所示.

$$A^c = \text{Merge}(O_{cnn}, O_{IndRNN}) \quad (14)$$

$$y = \text{softmax}(W_a A^c + b) \quad (15)$$

其中  $W_a$  为全连接层权重矩阵,  $A^c$  为合并特征矩阵, 本文使用反向传播进行模型训练并使用最小化交叉熵来优化模型, 交叉熵函数为:

$$\text{loss} = \frac{1}{N} \sum_{i=1}^N y^{(i)} \log_2 \hat{y}^{(i)} + (1 - y^{(i)}) \log_2 (1 - \hat{y}^{(i)}) \quad (16)$$

其中,  $N$  为输入样本数,  $y^{(i)}$  为第  $i$  个样本实际类别,  $\hat{y}^{(i)}$  则为模型预测类别.

## 4 多注意力 IndRNN 模型

将本文提出的方法在维吾尔语数据集上进行实验, 解决维吾尔语人称代词指代消解任务. 通过在不同神经网络进行对比实验, 证明了 CMAIR 模型的有效性. 在实验过程中, 本文使用 Word2Vec 的 Skip-Gram 算法训练生成词典<sup>[23]</sup>. 每一个词对词向量维度为 150 维. 在词典处理中使用均匀分

布  $U(-0.01, 0.01)$  的方式随机初始化词典中未登陆词, 并采用去停用词的方式进行词典优化, 提高词向量搜索效率. 由于维吾尔语语法特征无需对其进行分词操作, 词性信息在数据集中进行标注. 对于名词短语本文采用标注关键词的方式, 例如名词短语 ھازىرقى زامان ئالىي مەكتىپ قىلغۇچىلارنىڭ بىرى (当代大学者之一), 其中 مەكتىپ (学者) 为短语关键词, 在提取词向量时给予关键词较大的权重, 并采用短语中每一个词向量累加求均值的方式表示名词短语词向量矩阵和词性向量矩阵.

### 4.1 实验数据

目前, 中文和英文指代消解任务已经非常成熟, 标注语料完善. 维吾尔语作为小语种, 指代消解相关研究正在起步阶段, 还未见有标注数据集相关报道. 因此, 本文针对维吾尔语人称代词指代消解任务进行了数据采集和标注.

本文数据集均摘自新华网维吾尔文频道和天山网, 筛选出含有人称代词相关数据, 在维吾尔语专家的指导下完成语料标注工作. 在本次实验中, 采用 427 篇标注语料. 共包含数据实例 44 571 条, 其中存在指代关系的正例为 11 831 条, 不存在指代关系的负例为 32 740 条.

### 4.2 超参数

本文使用 3 种注意力机制结合语境特征完成实验. 实验中注意力特征向量经 CNN 提取更高层次语义特征, 词序列输入 IndRNN 获取语境特征, 再由 *Softmax* 分类器得到最终分类结果. 实验结果显示, 在表 4 所示参数设置时, 分类效果最佳. 本文采用随机初始化参数并使用梯度下降最小化目标函数<sup>[24]</sup>. 另外, 应用 Dropout 在 CNN 和 IndRNN 的输出层以减少模型参数, 加速模型训练和防止过拟合<sup>[25]</sup>.

表 4 实验参数设置

Table 4 Hyper parameters of experiment

Parameter	Parameter description	Value
$t$	Training epochs	50
$b$	Batch	100
$d$	Dropout rate	0.5
$l$	IndRNN layers	3
$k$	Kernel Size	3

### 4.3 评测指标

为了衡量指代消解结果的质量, 本文选用准确率(P)、召回率(R)、F值(F)三种重要指标评测指代消解实验结果, 准确率为正确消解的个数与实际消解个数的比率, 反应模型的准确程度, 召回率是正确消解的对象数与应消解对象数的比率, 反应模型的完备性, F值是准确率和召回率的调和平均值, 计算方式为  $F = 2 \times P \times R / (P + R)$ .

### 4.4 实验结果与分析

为了验证 CMAIR 模型的有效性, 本文将模型实验结果与以往学者的研究结果做比较. 实验结果如表 5 所示, 从表中可以看出, 本文提出的 CMAIR 模型取得了最好的分类效果(F 值为 86.86%), 其中相比于田生伟等实验结果, F 值提高 10%, 相比于李冬白等的实验结果 F 值提高 3.61%, 说明该方法能有效提高维吾尔语人称代词指代消解性能, 证明了 CMAIR 模型的有效性.

表 5 与以往研究对比 (%)

Table 5 Compared with previous studies (%)

Model	P	R	F
Tian	82.33	72.07	76.86
Li	88	80	83.81
CMAIR	90.79	83.25	86.86

为了研究和对比指代消解在不同模型和 CMAIR 在不同情况下的性能, 提出如下 3 种对比实验: 1) 3 种注意力机制的有效性; 2) 词向量和词性向量维度对 CMAIR 性能影响; 3) hand-crafted 特征与语境特征对 CMAIR 性能影响.

将本文提出的 3 种注意力机制使用以下模型在相同的数据集上进行实验:

1) CNN: 基本的卷积神经网络结构, 输入为 hand-crafted 特征;

2) ATT-CNN-1: 单注意力的卷积神经网络结构, 将词向量注意力特征矩阵与 hand-crafted 特征作为模型输入, 形成单注意力机制;

3) ATT-CNN-2: 两种注意力的卷积神经网络结构, 输入为词向量注意力特征矩阵、词性注意力特征矩阵和 hand-crafted 特征;

4) ATT-CNN-3: 多注意力机制的卷积神经网络模型, 输入为 3 种注意力特征矩阵与 hand-crafted 特征;

本文将提出的 4 组实验在维吾尔语数据集上进行对比, 分析维吾尔语人称代词指代消解性能. 表 6 给出 4 组实验在不同模型下的实验结果.

从表 6 结果可以看出, 本文提出的 3 种注意力机制在维吾尔语数据集上实验都取得了不错的分

类结果. 其中 ATT-CNN-3 在维吾尔语人称代词指代消解任务中性能达到最优. 分析结果可知, 加入词向量注意力机制的 ATT-CNN-1 比只加入 hand-crafted 特征的 CNN 模型 F 值提升 3.97%, 这说明普通的 CNN 没有针对待消解对提取更多的特征信息, 所以无法准确识别照应语和候选先行语指代关系. 证明了词向量注意力特征矩阵含有丰富的语义信息, 可以准确表达上下文关系, 在模型训练过程中突出待消解对词语重要性, 从而提高指代消解性能.

表 6 不同模型消解性能对比 (%)

Table 6 Comparison of different model anaphora resolution performance (%)

Model	P	R	F
CNN	75.47	74.16	74.81
ATT-CNN-1	80.14	77.46	78.78
ATT-CNN-2	82.37	78.80	80.55
ATT-CNN-3	83.02	79.61	81.27

加入词向量、词性注意力机制的 ATT-CNN-2 相比于单注意力机制的 ATT-CNN-1 模型 F 值提高 1.77%, 这说明加入词性注意力特征能在模型训练过程中高度关注照应语和候选先行语词和词性关联信息, 弥补了仅依赖文本内容信息的不足, 从而提高指代消解性能, 这也说明了词性注意力机制在维吾尔语人称代词指代消解的有效性.

对比 ATT-CNN-3 和其他三种模型实验结果可以看出, 加入 3 种注意力机制的 ATT-CNN-3 模型的三类评测指标都达到最优, 相比于 CNN、ATT-CNN-1 和 ATT-CNN-2, 其 F 值分别提高 6.46%、2.49% 和 0.72%, 说明加入多注意力机制的模型能从不同层面关注待消解对信息, 从而能更好地识别指代关系, 验证了多注意力机制在维吾尔语人称代词指代消解任务中的有效性.

为了进一步对比本文提出的 CMAIR 模型在维吾尔语人称代词指代消解任务中的分类效果, 训练不同维度词向量分析词向量维度对分类结果影响. 分别使用 10 维、50 维、100 维、150 维和 200 维训练生成词向量矩阵和词性向量矩阵, 实验结果如图 5 所示.

使用不同维度词向量和词性向量构造三种注意力特征并在 CMAIR 模型进行 5 倍交叉验证实验, 从图 5 结果可以看出, 在向量维度为 0 时, 即模型中不使用注意力机制, 其 F 值最低, 加入 10 维向量后分类效果有小幅提升, 可以看出低维度向量不能充分表达文本语义信息, 使模型在训练过程中没有学习待消解对高维特征. 在加入 50 维向量构建注意力特征作为模型输入时, 模型性能有大幅度提升, 可以说明高维向量可以表达文本深层次语义信息, 提

高模型学习能力. 当向量维度在 50 到 150 时, 模型 F 值仍有提升, 但提升效果不是很显著, 并在 150 维时达到最优. 当向量维度达到 200 维时, 分类效果有所下降, 这说明当向量维度达到一定阈值时, 会产生不能充分表达待消解对信息和过拟合现象, 导致分类结果产生波动. 合理的选用向量维度对分类结果有重要的影响, 所以本文选取 150 维作为词向量和词性向量的维度.

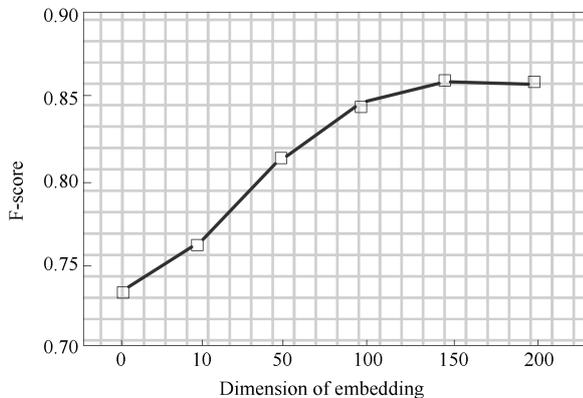


图 5 不同维度词向量分类 F-score 比较

Fig. 5 Comparison of different dimension word vector classification F-score

hand-crafted 作为指代消解分类重要信息, 对分类结果起着重要作用, 本文在以往研究的基础上添加了语境特征, 为了探究 hand-crafted 和语境特征的有效性, 分别移除 hand-crafted 和语境特征实验, 实验结果如表 7 所示, 从表中可以看出, 去除 hand-crafted, 仅包含注意力特征和语境特征的情况下 ( $V_{attention} + V_{context}$ ) 其 F 值相比于 CMAIR 模型降低了 5.55%, 本文选用的 hand-crafted 是结合维吾尔语语言现象和语法特征提出的反应待消解对在规则和知识方面的表示与联系.

表 7 不同特征类型对指代消解性能影响 (%)

Table 7 The effect of different feature types on the anaphora resolution (%)

特征类型	P	R	F
$V_{attention} + V_{context}$	83.29	79.43	81.31
$V_{hand-crafted} + V_{attention}$	86.81	80.24	83.40
CMAIR	90.79	83.25	86.86

分析实验结果可知, 去除 hand-crafted 的指代消解性能明显下降, 证明了 hand-crafted 的引入对提高指代消解性能起着关键作用. 分析表 7 中去除语境特征的实验结果  $V_{hand-crafted} + V_{attention}$  可以看出相比于去除 hand-crafted 的实验 F 值提升了 2.09%, 这是由于相比于 hand-crafted 特征, 语境特征含有丰富的上下文语境信息, 能在模型训练

过程中学习词序关系和词依赖关系, 是句子连续分布式向量表示, 另外, IndRNN 可以保留长期记忆, 处理长序列, 对模型性能提升有关键性作用. 相比于 CMAIR 实验结果, 去除语境特征的 F 值降低 3.46%, 这说明多注意力机制虽能从不同层面关注文本特征, 但缺少文本连续性的表述, 输入词序列的连续性和依赖性也对模型性能的提升有较大的作用. 本文通过引入 hand-crafted 信息充分体现了维吾尔语语言和语法特征, 语境特征的引入能关注句子级的特征信息, 确认了待消解对中照应语和候选先行语的联系, 从而取得了更好的分类效果.

## 5 结束语

在维吾尔语人称代词指代消解任务中, 如今大部分的研究都是结合 hand-crafted 和深度学习的方式进行, 这种方式仅依赖原始输入特征, 不能突出每个词在句中的重要程度和关注词语多层次特征. 本文基于注意力机制和语境特征的研究, 提出一种多注意力机制的维吾尔语人称代词指代消解方法, 该模型结合 3 种注意力机制, 从待消解对中候选先行语和照应语不同层面分析语义信息. 同时, 提出一种距离识别算法, 能有效识别待消解对距离, 使模型能充分学习距离信息. 此外, 本文在以往的研究基础上加入语境特征, 用于输入词序列并学习句中每一个词的词序关联和依赖关系. 实验结果表明, 加入多注意力机制和语境特征的 CMAIR 模型能显著提升维吾尔语人称代词指代消解性能.

## 致谢

感谢王县县、王欢欢、牛苗、祁青山等同学在本文提供帮忙和支持, 在此谨向他们致以诚挚的谢意.

## References

- Zelenko D, Aone C, Tibbetts J. Coreference resolution for information extraction. In: Proceedings of the 2004 ACL Workshop on Reference Resolution and its Applications. Barcelona, Spain: ACL, 2004. 9–16
- Deemter K V, Kibble R. On coreferring: Coreference in muc and related annotation schemes. *Computational Linguistics*, 2000, **26** (4): 629–637
- Kim Y. Convolutional neural networks for sentence classification. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing. Doha, Qatar: ACL, 2014. 1746–1751
- Irsay O, Cardie C. Opinion mining with deep recurrent neural networks. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing. Doha, Qatar: ACL, 2014. 720–728
- Tai K S, Socher R, Manning C D. Improved semantic representations from tree-structured long short-term memory networks. In: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics. Beijing, China: ACL, 2015. 1556–1566

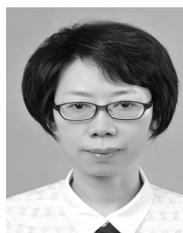
- 6 Chen C, Ng V. Chinese zero pronoun resolution with deep neural networks. In: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics. Berlin, Germany: ACL, 2016. 778–788
- 7 Chen C, Ng V. Deep reinforcement learning for mention-ranking coreference models. In: Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. Texas, USA: ACL, 2016. 2256–2262
- 8 Iida R, Torisawa K, Oh J H. Intra-sentential subject zero anaphora resolution using multi-column convolutional neural network. In: Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. Texas, USA: ACL, 2016. 1244–1254
- 9 Mnih V, Heess N, Graves A. Recurrent models of visual attention. In: Proceedings of the Advances in Neural Information Processing Systems. Montreal, Canada: NIPS, 2014. 2204–2212
- 10 Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate [Online], available: <https://arxiv.org/pdf/1409.0473v6.pdf>, December 27, 2018
- 11 Yin W, Schütze H, Xiang B, Zhou B. Abcnn: Attention-based convolutional neural network for modeling sentence pairs. In: Proceedings of the 2016 Transactions of the Association for Computational Linguistics. Texas, USA: ACL, 2016. 259–272
- 12 Wang Y, Huang M, Zhao L. Attention-based lstm for aspect-level sentiment classification. In: Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. Texas, USA: ACL, 2016. 606–615
- 13 Soon W M, Ng H T, Lim D C Y. On coreferring: A machine learning approach to coreference resolution of noun phrases. *Computational Linguistics*, 2001, **27** (4): 521–544
- 14 Ng V, Cardie C. Improving machine learning approaches to coreference resolution. In: Proceedings of the 40th Annual Meeting on Association for Computational Linguistics. Pennsylvania, USA: ACL, 2002. 104–111
- 15 Yang X, Zhou G, Su J, Tan C L. Coreference resolution using competition learning approach. In: Proceedings of the 41th Annual Meeting on Association for Computational Linguistics. Sapporo, Japan: ACL, 2003. 176–183
- 16 Chen C, Ng V. Chinese zero pronoun resolution: an unsupervised approach combining ranking and integer linear programming. *Springer Verlag*, 2014, **36** (5): 823–834
- 17 Clark K, Manning C D. Deep reinforcement learning for mention-ranking coreference models [Online], available: <https://arxiv.org/pdf/1609.08667.pdf>, December 27, 2018
- 18 Yin Q, Zhang Y, Zhang W, Liu T. Chinese zero pronoun resolution with deep memory network. In: Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. Texas, USA: ACL, 2016. 606–615
- 19 Li Min, Yu Long, Tian Sheng-Wei, Turglm Ibrahim, Zhao Jian-Guo. Coreference resolution of uyghur noun phrases based on deep learning. *Acta Automatica Sinica*, 2017, **43** (11): 1984–1992  
(李敏, 禹龙, 田生伟, 吐尔根·依布拉音, 赵建国. 基于深度学习的维吾尔语名词短语指代消解. *自动化学报*, 2017, **43** (11): 1984–1992)
- 20 Tian Sheng-Wei, Qin Yue, Yu Long, Turglm Ibrahim, Feng Guan-Jun. Anaphora resolution of uyghur personal pronouns based on Bi-LSTM. *Acta Electronica Sinica*, 2018, **46**(7): 1691–1699  
(田生伟, 秦越, 禹龙, 吐尔根·依布拉音, 冯冠军. 基于 Bi-LSTM 的维吾尔语人称代词指代消解. *电子学报*, 2018, **46**(7): 1691–1699)
- 21 Li Dong-Bai, Tian Sheng-Wei, Yu Long, Turglm Ibrahim, Feng Guan-Jun. Deep learning for pronominal anaphora resolution in uyghur. *Journal of Chinese Information Processing*, 2017, **31** (4): 80–88  
(李冬白, 田生伟, 禹龙, 吐尔根·依布拉音, 冯冠军. 基于深度学习的维吾尔语人称代词指代消解. *中文信息学报*, 2017, **31** (4): 80–88)
- 22 Li S, Li W, Cook C, Zhu C, Gao Y. Independently recurrent neural network (indrnn): Building A longer and deeper rnn. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Utah, USA: IEEE, 2018. 5457–5466
- 23 Mikolov T, Chen K, Corrado G, Dean J. Efficient estimation of word representations in vector space [Online], available: <https://arxiv.org/pdf/1301.3781.pdf>, December 27, 2018
- 24 Duchi J, Hazan E, Singer Y. Adaptive subgradient methods for online learning and stochastic optimization. *Journal of Machine Learning Research*, 2011, **12** (7): 2121–2159
- 25 Hinton G E, Srivastava N, Krizhevsky A, Sutskever I, Salakhutdinov R R. Improving neural networks by preventing co-adaptation of feature detectors [Online], available: <https://arxiv.org/pdf/1207.0580.pdf>, December 27, 2018



杨启萌 新疆大学博士研究生. 主要研究方向为自然语言处理.

E-mail: yqm\_xju@163.com

(YANG Qi-Meng Ph. D. candidate at Xinjiang University. His main research interest is natural language processing.)



禹龙 新疆大学教授. 主要研究方向为计算机智能技术与计算机网络. 本文通信作者. E-mail: yul\_xju@163.com

(YU Long Professor at Xinjiang University. Her research interest covers computer intelligence technology and computer networks. Corresponding author of this paper.)



田生伟 新疆大学教授. 主要研究方向为自然语言处理和计算机智能技术.

E-mail: tianshengwei@163.com

(TIAN Sheng-Wei Professor at Xinjiang University. His research interest covers natural language processing and computer intelligence technology.)



艾山·吾买尔 新疆大学副教授. 主要研究方向为自然语言处理及机器翻译.

E-mail: Hasan1479@xju.edu.cn

(AISHAN Wumaier Associate professor at Xinjiang University. His research interest covers natural language processing and machine translation.)