

本文责任编辑 赵铁军
Recommended by Associate Editor ZHAO Tie-Jun
1. 新疆大学软件学院 乌鲁木齐 830008 2. 新疆大学网络中心 乌鲁木齐 830046 3. 电子科技大学信息与软件工程学院 成都 610054
1. School of Software, Xinjiang University, Urumqi 830008
2. Network Center, Xinjiang University, Urumqi 830046 3. School of Information and Software Engineering, University of Electronic Science and Technology of China, Chengdu 610054

译文: 2017 年 1 月 1 日时间 11 时左右, 在南京雨花西路和共青团路交叉口, 一辆货车往右转弯过程中导致一辆电动车刮倒, 电动车上母亲当场身亡.

事件句 2 يولدىن ئۆتۈپ كېتىۋاتقان كىشىلەر دەرھال ئايالنىڭ قۇچقىدىكى 9 ئايلىق بوۋاقتى يېقىن ئەتراپتىكى نەنجىڭ شەھەرلىك 1-دوختۇرخانىغا ئېلىپ بارغان.

译文: 过路人立刻把女子怀里 9 个月的婴儿送往附近南京市第一医院.

以上两个事件出现在同一篇章的不同位置, 下划线词为触发词. 表 1 和表 2 分别列出了两个事件句对应事件元素抽取情况.

表 1 事件句 1 中的元素
Table 1 Arguments in event sentence 1

元素	对应内容	译文
Time-Arg	2017-1 يىلى ئاينىڭ 1-كۈنى سائەت 11 لەر ئەتراپىدا	2017 年 1 月 1 日时间 11 时左 右
Place-Arg	نەنجىڭ شەھىرىدىكى يۇخۇا غەربىي يولى بىلەن ئىتتىپاق يولىنىڭ كېسىشىش نۇقتىسىدا	南京雨花西路和 共青团路交叉口
Wrecker-Arg		
Suffer-Arg	ئانا	母亲
Tool-Arg	بىر يۈك ماشىنىسى	一辆货车

表 2 事件句 2 中的元素
Table 2 Arguments in event sentence 2

元素	对应内容	译文
Agent-Arg	يولدىن ئۆتۈپ كېتىۋاتقان كىشىلەر	过路人
Artifact-Arg	ئايالنىڭ قۇچقىدىكى 9 ئايلىق بوۋاقتى	女子怀里 9 个月 的婴儿
Tool-Arg		
Origin-Arg		
Destination-Arg	نەنجىڭ شەھەرلىك 1- دوختۇرخانىغا	南京市第一医院
Time-Arg		

对事件 1 和事件 2 以及表 1 和表 2 进行分析, 事件 1 中 4 个元素可在当前事件句中获得, 缺失一个元素; 事件 2 中只有 Agent-Arg、Artifact-Arg 和 Destination-Arg 三个元素在当前事件句中获得, 其他角色均缺失. 通过观察和分析可以发现, 事件 2 的缺失角色 Origin-Arg 和 Time-Arg 与事件 1 中

Place-Arg 和 Time-Arg 所对应的内容一致, 即事件 2 的 2 个缺失元素可以从事件 1 中得到填充. 通过填充, 使事件 2 语义更完整, 表达更加清晰.

主要贡献如下:

- 1) 将填充问题转换成二分类问题, 对维吾尔语事件缺失元素进行填充.
- 2) 提出了注意力机制的独立循环神经网络和胶囊网络的并行模型.
- 3) 充分考虑事件元素类型, 将事件元素分为 4 类, 按照事件元素类型相同的规则构建样本.

1 相关研究

目前, 事件抽取技术在自然语言处理领域受到广泛关注. 早期, 采用模式匹配算法^[2-3] 和浅层机器学习方法^[4] 进行信息抽取. 而随着深度学习^[5] 的提出, 事件抽取技术的研究有了进一步的进展^[6-7]. Chen 等^[8] 引入词表模型捕获词汇语义线索, 使用动态池化卷积神经网络 (Dynamic multi-pooling convolutional neural network, DMCNN), 在 ACE2005 语料上抽取事件本体. Chang 等^[9] 在双向 LSTM (Long short-terin memary) 基础上对隐藏状态进行池化, 从而对英文文本进行事件抽取. Zeng 等^[10] 提出了一种卷积双向 LSTM 模型, 从原始文本中捕获句子级和词汇信息, 对事件触发词和元素标记. 田生伟等^[11] 将词向量作为双向 LSTM 模型的输入, 挖掘给定事件句隐藏的上下文语义信息, 结合事件触发词建立注意力机制对维吾尔语事件时序关系识别. 然而这些学者的研究都是基于单模型的, 对特征的提取有一定的局限性. 黎红等^[12] 提出 DCNNs-LSTM 的顺序处理模型, 将特征依次输入 DCNNs 和 LSTM 完成维吾尔语突然事件的识别. 该方法采用两种模型顺序组合的方式对特征进行处理, 所以在特征传输过程中会丢失部分特征.

以上都是对事件抽取的研究, 对事件缺失元素填充的论文较少, Gupta 等^[13] 使用规则和统计学习的方法只对缺失的 Time 元素进行填充. Huang 等^[14] 利用上下文和领域相关文档的信息, 针对 Target、Place 等几类进行填充. 侯立斌等^[15] 提出了采用基于规则和机器学习的方法, 对中文跨事件的缺失事件元素识别和填充. 赵文娟等^[16] 提出了基于句法依存分析的角色填充思路和技术, 以“森林火灾”事件为例, 用最大熵算法对填充过程进行了说明. 以上研究都没有充分考虑事件和事件元素的上下文语义特征, 而且集中在汉语和英语等大语种, 对于维吾尔语事件缺失元素填充研究很少.

事件抽取研究主要集中在触发词识别、事件类型分类、元素识别等任务上. 事件缺失元素填充的研究主要是从其他事件中获得事件元素, 对当前事

件元素进行补充. 缺失元素填充是在触发词识别和元素识别等前提任务的基础上进行研究. 这为缺失元素填充提供了便利, 但是由于这些前提任务识别率的局限性, 为缺失元素填充任务带来了一些级联错误. 另外, 由于构成事件元素的单词个数不统一, 这也为缺失元素填充任务带来一定的难度.

基于上述问题, 提出结合注意力机制^[17-18] 独立循环神经网络和胶囊网络的并行模型, 来解决的维吾尔语缺失元素填充问题. 将富含上下文语义信息的词向量做为胶囊网络的输入, 获取其位置信息和局部特征, 将 18 项事件间特征和候选元素特征, 结合注意力机制, 经过独立 RNN 的处理, 进一步获取规则特征. 两个并行的模型同时处理各自更加擅长处理的特征, 会减少有效特征在传输过程中的丢失. 然后将两类有效特征融合进行分类, 最终完成维吾尔语事件缺失元素的填充.

2 预备知识

2.1 维吾尔语语言特点

维吾尔语是典型的黏着性语言, 在构词形式和语法形式上, 都是通过在词根或词干上结合词缀或词尾的方式实现的.

维吾尔语中的动词在不同的语态结构中, 形式上会发生变化, 主要体现在动词尾部变形加上一定的词缀. 动词通过添加词缀表现不同的感情色彩、时态等, 其中词干包含词的词汇意义, 词缀和词尾都有固定的意义. 例如, 词干“پارتلاش”(爆炸), 加入词缀“دى”后变为“پارتلىدى”(已爆炸), 表示过去发生或者已经完成的动作. 而当添加“دىمۇ”时, 变为“پارتلىدىمۇ”(爆炸了吗), 表示对当前爆炸事件的疑问.

维吾尔语中“格语法”是一种特殊的语言形式, 名词或名词短语有主格、属格、向格、宾格、位格、从格和界限格等多种格属性. 根据名词和名词短语的不同和上下文的变化, 会在名词或名词短语后面附加上不同的格后缀. 例如, 在名词或名词短语后加“نىڭ”变为属格, 表示人或事物的领属关系. 格语法在语法形式上具备独立性, 语法意义上具备稳定性.

本文是基于维吾尔语事件的研究, 涉及到事件触发词和事件元素, 其中, 事件触发词多为动词, 而事件元素一般为名词或名词短语. 为了保证维吾尔语的语言特色, 本文在预处理阶段未对语料进行词干提取, 保留了词缀信息的完整性. 在编码阶段, 利用词嵌入技术, 充分挖掘了维吾尔语文本的深层语义信息.

2.2 事件定义

定义 1. 事件: 指在特定的环境和时间下发生, 由若干角色参与, 表现出动作特征的一件事情^[19].

如事件句 1 和事件句 2 所示, 分别描述的是交通事故事件和运输事件.

定义 2. 事件触发词: 指在事件描述中能清晰表达事件所发生事件的词汇, 亦称指示词. 事件句 1 中的“تۈرۈلۈپ كېتىشىگە”(刮倒)”和事件句 2 中的“ئېلىپ بارغان”(送往), 分别为事件句 1 和事件句 2 的触发词.

定义 3. 事件元素: 指描述事件具体信息的文本短语, 包括参与者、时间和地点等.

定义 4. 缺失元素: 每类事件有对应的元素, 若元素对应的内容不存在, 即为缺失元素. 缺失元素分可填充缺失元素和不可填充缺失元素. 表 1 中 Wrecker-Arg (肇事者) 和表 2 中的 Tool-Arg (运送工具) 为不可填充元素, 表 2 中 Origin-Arg (源地) 和 Time-Arg (时间) 对应的内容在事件句中没有出现, 但可以通过事件句 1 相应内容进行填充, 为可填充缺失元素. 本文缺失元素识别的基础上进行, 即只对可填充元素进行研究 (后文所提缺失元素均为可填充元素).

定义 5. 候选元素: 除缺失元素所在事件句之外的其他事件中的所有元素. 候选元素应与缺失元素类型相同, 本文将所有的事件元素分为 4 类: 时间、人物、地点和其他.

定义 6. 元素对: 指维吾尔语文本中所有缺失元素按照元素类型相同的组对规则与候选元素进行组对后的元素对.

2.3 任务描述

缺失元素填充, 是指对于可以被填充的元素如何从其他事件描述中选择合适的元素进行填充. 本文将填充问题转换为二分类问题, 即按照一定的规则将缺失元素与候选元素两两组对, 并判断当前缺失元素是否可以被候选元素填充, 从而构成正负样例. 具体步骤如下:

步骤 1. 将实验语料中的每篇语料所提取出来事件放入事件列表 EL_i ($i = 1, 2, \dots, N$) 中, N 为实验语料的总数.

步骤 2. 循环遍历 EL_i 中每一个事件, 将缺失的元素放入到列表 EM 中.

步骤 3 (样本 1). 循环遍历 EM , 将缺失元素与所有候选元素两两组对, 构成元素对 $\langle m, c \rangle$, 然后判断 m 是否可以被 c 填充, 若可以被 c 填充, 则为正例, 标签 y 为 1; 若不可以被 c 填充, 则为负例, 标签 y 为 0. 构成元素对 $\langle m, c; y \rangle$, 将其放入到元素对集合 AL 中.

步骤 3 (样本 2). 循环遍历 EM , 按照定义 6 将缺失元素与候选元素两两组对, 构成元素对 $\langle m, c \rangle$, 然后判断 m 是否可以被 c 填充, 若可以被

c 填充, 则为正例, 标签 y 为 1; 若不可以被 c 填充, 则为负例, 标签 y 为 0. 构成元素对 $\langle m, c; y \rangle$, 将其放入到元素对集合 AL 中.

步骤 4. 循环步骤 2 和步骤 3, 直至得到所有的元素对, 并将其放入到集合 AL 中.

3 模型

本文提出了一个 Att_IndRNN_CapsNet 模型, 用于维吾尔语事件缺失元素填充. 模型分为输入层, 联合处理层, 融合分类层 3 层. 将输入层的特征到联合处理层, 其中事件及候选元素的 18 项特征作为 Att_IndRNN 的输入, 首先建立注意力机制, 再通过独立 RNN 获取事件和元素的特征. 事件触发词和候选元素的词向量输入胶囊网络, 挖掘上下文语义信息. 其次, 将两个层的输出传入到融合分类层进行特征融合和分类, 进而完成事件缺失元素的填充. 图 1 描述了整个维吾尔语事件缺失元素填充的模型结构.

3.1 输入层

输入层由两部分构成, 第一部分是将事件触发词和候选元素通过 word2vec 映射成富含上下文语义信息的低维向量 α , $\alpha \in \mathbf{R}^M$, M 是词向量维度; 第二部分是事件和事件元素构成的内部规则特征 h . 词向量 α 和规则特征 h 作为联合处理层的输入.

3.2 联合处理层

1) Att_IndRNN 层

从输入层接收的规则特征 h 经过本层处理. 首先计算注意力权重, 公式如下:

$$t_i = \tanh(w_h h_i + b_h) \quad (1)$$

$$a_i = \frac{\exp(t_i)}{\sum_{j=1}^L \exp(t_j)} \quad (2)$$

其中, $h_i \in \mathbf{R}$ 为第 i 个手动特征, w_h 和 b_h 分别为权重和偏置. a_i 表示特征 h_i 的注意力权重. 根据注意力权重向量 $a \in \mathbf{R}^L$, 对 h 进行注意力加权, 得到经过注意力机制的输出 p , 计算公式如下:

$$p = \sum_{i=1}^L a_i h_i \quad (3)$$

循环神经网络^[20] 在序列学习问题中获得广泛应用, 并且成果显著. 然而由于常见的梯度消失和梯度爆炸问题, 循环神经网络通常难以训练. 本文引入独立循环神经网络 (IndRNN)^[21], 这种新型的 RNN 能有效解决网络收敛时的梯度爆炸和消失问题. 在 IndRNN 中, 循环输入用 Hadamard 乘积处理. 将经过注意力机制的输出向量 p 输入到 IndRNN, 进一步挖掘事件和事件元素的规则特征 k .

$$k_t = \sigma(W p_t + \mu \odot k_{t-1} + b) \quad (4)$$

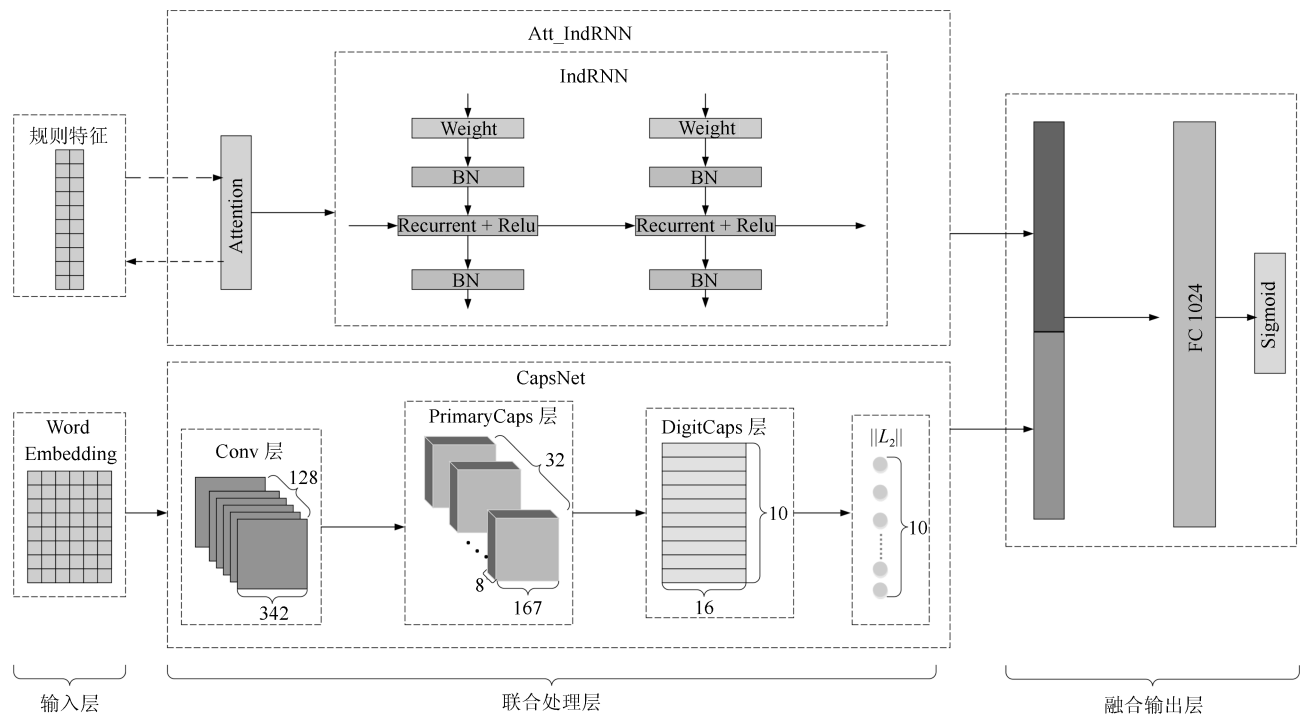


图 1 模型结构图

Fig. 1 Model structure

其中, $p_t \in \mathbf{R}^M$ 和 $k_t \in \mathbf{R}^N$ 分别代表时间步长 t 的输入和隐藏状态, $W \in \mathbf{R}^{N \times M}$, $u \in \mathbf{R}^{N \times N}$ 和 $b \in \mathbf{R}^N$ 分别为当前输入的权重, 循环输入以及神经元偏置值, $\odot$ 是 *Hadamard* 乘积, N 是当前层中神经元的个数. 每层中每个神经元与其他神经元不相连, 彼此相互独立, 神经元的连接可以通过叠加两层或者更多层的 IndRNN 来实现. 对于第 n 个神经元, 隐藏层 $k_{n,t}$ 可以通过以下公式得到:

$$k_{n,t} = \sigma(w_n p_t + u_n k_{n,t-1} + b_n) \quad (5)$$

其中, w_n 和 u_n 分别表示第 n 行的输入权重和循环权重. 每个神经元仅在前一时间步从输入和它自己的隐藏状态中接收信息. 也就是说, IndRNN 中的每个神经元独立地处理一种类型的时空模型. 即随着时间的推移 (即通过 u) 独立地聚集空间模式 (即通过 w). 不同神经元之间的相关性可以通过两层或多层的堆叠来加以利用. 在这种情况下, 下一层的每个神经元处理上一层所有神经元的输出.

2) CapsNet 层

胶囊神经网络 (Capsule network, CapsNet)^[22] 是由 Hinton 在 2017 年 10 月份首次提出的新型深度学习网络架构, 用于图像的分类, 其不同于卷积神经网络的矢量输入输出, 胶囊层的输入输出均为向量, 并且采用动态路由算法来更新胶囊参数. 而 Zhao 等^[23] 首次将胶囊网络用于文本建模, 验证了胶囊网络在文本处理任务上的有效性. 故本文引入胶囊网络处理从输入层接收的语义特征 α , 得到富含上下文语义信息的局部特征 q .

在胶囊网络中, 激活函数 *Squashing* 既保留了输入向量的方向, 又将输入向量的模压缩到 (0, 1) 之间. 输出 v_j 的计算公式如下:

$$v_j = \frac{\|s_j\|^2}{1 + \|s_j\|^2} \frac{s_j}{\|s_j\|} \quad (6)$$

其中, v_j 是胶囊 j 的矢量输出, s_j 是总输入矢量.

胶囊网络第一层为一个激活函数为 ReLU 的卷积层, 除了第一层胶囊外, 所有胶囊的总输入 s_j 是对下面层胶囊的所有预测向量 $\hat{u}_{j|i}$ 的加权求和, 它是通过下面层胶囊的输出 u_i 乘以权重矩阵 W_{ij} 而得到的. 公式如下:

$$s_j = \sum_i c_{ij} \hat{u}_{j|i} \quad (7)$$

$$\hat{u}_{j|i} = W_{ij} u_i \quad (8)$$

其中, c_{ij} 是在动态路由过程中确定的耦合系数, 表示每一个低层胶囊与其相对应的高层胶囊之间的权重. 对于每个胶囊 i 来说, 所有的权重 c_{ij} 的总和为

1. c_{ij} 由采用的动态路由算法中的 *softmax* 函数决定, 计算公式如下:

$$c_{ij} = \frac{\exp(b_{ij})}{\sum_k \exp(b_{ik})} \quad (9)$$

其中 b_{ij} 是胶囊 i 与胶囊 j 的对数概率, 用来更新 c_{ij} , 并将其初始化为 0, 在路由迭代过程中, b_{ij} 会不断更新, 更新公式如下:

$$b_{ij} = b_{ij} + \hat{u}_{j|i} \cdot v_j \quad (10)$$

3.3 融合输出层

本层将联合处理层得到的语义特征 q 和规则特征 k 完成融合, 具体操作是将规则特征 k 拼接到语义特征 q 之后, 得到融合特征 m , 公式如下:

$$m = q \oplus k \quad (11)$$

其中, $\oplus$ 表示特征的拼接.

之后, 将融合层的特征 m 输入到一个全连接层, 得到输出 f , 再通过 *sigmoid* 函数得到结果 $S(f)$. *sigmoid* 公式如下:

$$S(f) = \frac{1}{1 + e^{-f}} \quad (12)$$

其中, $S(f)$ 是一个值为 0 到 1 的概率. 若 $S(f)$ 大于阈值, 则分类结果判定为 1, 即当前样本为正例, 表示缺失元素可被该候选元素填充; 否则, 分类结果判定为 0, 即当前样本为负例, 表示缺失元素不可被该候选元素填充.

4 实验

4.1 实验数据

目前, 国际上有 MUC (仅有 English 语料) 和 ACE (有 Arabic、Chinese 和 English3 种语料) 两种用于事件抽取的语料库, 未发现关于维吾尔语事件抽取的语料. 实验选取天山网、人民网等维吾尔语网页作为语料来源, 利用网络爬虫下载网页, 经去重、去噪处理后筛选出包含事件描述的新闻报道文本作为实验语料, 参照 ACE 标注体系, 在实验组维吾尔专家指导下对语料进行标注.

本实验共标注了 210 篇语料, 按照样本 2 构建方法, 共生成了 3696 条样本数据. 统计发现其中包含 1777 条可被当前候选元素填充的样本, 有 1919 条不可被当前候选元素填充的样本.

4.2 特征提取

特征的选择对模型的实验效果有很大的影响, 根据实验组维吾尔语语言专家的意见, 选取了以下 18 项事件特征和元素特征.

1) 事件类别: 反映了事件所属类型, 若缺失事件与候选事件类别相同, 特征取 1; 否则, 特征取 0.

2) 事件子类别: 进一步定义了事件所属类别, 与事件类别类似, 若缺失事件与候选事件类别相同, 特征取 1; 否则, 特征取 0.

3) 事件间间隔事件数目: 对实验语料进行统计可知, 缺失事件与候选事件对间隔为 $[0, 3]$ 的约占 81%. 所以在此范围内, 特征取 1; 否则, 特征取 0.

4) 事件对前后关系: 缺失事件与候选事件在文档中的前后关系, 若缺失事件在前, 特征为 1; 否则, 特征为 0.

5) 事件极性: 有 Positive 和 Negative 两种, 描述了事件为肯定事件还是否定的事件. 若缺失事件与候选事件极性相同, 特征取 1; 否则, 特征取 0.

6) 事件时态: 用来描述事件是过去发生的、正在发生的还是将来发生的. 若缺失事件与候选事件时态相同, 特征取 1; 否则, 特征取 0.

7) 依存关系: 两个事件的触发词出现在一个句子中, 则认定两个事件具有依存关系. 若缺失事件与候选事件具有依存关系, 则特征取 1; 否则, 特征取 0.

8) 和缺失事件类别相同的事件分布: 统计整个文档中各个类别事件的个数, 与缺失事件类别相同的事件个数作为特征.

9) 和缺失事件子类别相同的事件分布: 统计整个文档中各个类别事件的个数, 与缺失事件子类别相同的事件个数作为特征.

10) 和候选事件类别相同的事件分布: 统计整个文档中各个类别事件的个数, 与候选事件类别相同的事件个数作为特征.

11) 和候选事件子类别相同的事件分布: 统计整个文档中各个子类别的个数, 与候选事件子类别相同的事件个数作为特征.

12) 共指关系: 若两个事件为共指关系, 缺失元素被候选元素填充的概率更大. 缺失事件与候选事件具有共指关系, 则特征取 1; 否则, 特征取 0.

13) 缺失事件与候选元素的距离: 统计缺失事件的触发词与候选元素的间隔词的个数, 个数为 $[0, 50]$ 的占 69%. 若在此范围内, 则特征取 1; 否则, 特征取 0.

14) 缺失元素在缺失事件中担当的角色: 缺失事件的所有元素 MA , 判断当前缺失元素在 MA 中的位置 (即索引), 作为特征.

15) 候选元素在候选事件中担当的角色: 候选事件的所有元素 CA , 判断当前候选元素在 CA 中的位置 (即索引), 作为特征.

16) 和缺失元素类型相同的元素在文档中分布: 根据定义 5, 元素类型分为 4 类, 统计文档中各个类

型元素的个数, 与缺失元素类型相同的元素个数作为特征.

17) 候选元素语义角色: 语义角色是指名词和动词组成语义结构后, 名词在该语义结构中所担任的角色. 所有语义角色构成一个列表, 候选元素的语义角色在列表中的位置 (即索引) 作为特征.

18) 候选元素语义类别: 语义类别可以很好地反应候选元素的信息, 根据实验组维吾尔语语言专家意见, 将维吾尔语具有语义的语义类别划分为“Hum_ 人类”、“Loc_ 地点”等 14 个类别, 所有语义类别构成一个列表, 候选元素的语义类别在列表中的位置 (即索引) 作为特征.

为了提高模型的识别效果, 在以上 18 项特征的基础上, 引入了 Word Embedding. 本文选用了基于神经网络训练的 Word Embedding, 包含了丰富的上下文信息, 充分表示缺失事件触发词、候选事件的触发词和候选元素在文本中的语音信息, 同时避免了维数灾难^[24]. 本文使用 Mikolov 等^[25] 提出的 word2vec 工具进行训练. 为了更准确地获取每个词在低维空间中语义的分布情况, 在原有实验语料的基础上进行了扩充, 选取天山网、人民网等维语版网页作为语料来源, 利用网络爬虫下载网页, 进行去重、去噪处理之后获取不限题材且未标注的文本作为生语料.

在处理候选元素时, 会遇到候选元素的词汇个数不统一的情况, 经统计元素词汇个数为 $[1, 5]$ 的占 70.87%, 设元素个数为 N , 若 $N \geq 5$, 本文会取当前获选元素的后 5 个词; 若 $N < 5$ 时, 会进行补零处理, 即 N 个词转换成词向量, $5-N$ 个与词向量维度相同的零向量, 合并组成元素词向量.

4.3 实验设计

本文实验测评方式采用 MUC 标准, 即准确率 P , 召回率 R 和 $F1$ 考察缺失元素填充性能. 其中, 衡量模型整体性能的 $F1 = (2 \times P \times R) / (P + R)$.

模型的参数设置对维吾尔语事件缺失元素填充效果起着关键的作用. 最优参数设置如表 3 所示.

表 3 模型最优参数表

Table 3 Optimal parameters

参数	值
lr	0.005
$lrdr$	0.1
bs	16
ep	50
dr	0.3
opt	adam

其中, lr 表示训练过程中的学习率; $lrdr$ 学习率衰

减率; *bs* 表示每一次迭代批处理样本数; *ep* 表示模型训练最优的迭代次数; *dr* 表示训练过程中的丢码率; *opt* 表示模型优化器算法。

为探索不同角度下维吾尔语事件缺失元素填充效果, 本文设计了如下 6 个实验: 1) 样本构建方法对实验性能的影响; 2) 本文模型与其他模型实验性能对比; 3) 词向量维度对实验性能的影响; 4) 不同种类特征对实验性能的影响; 5) 独立特征与融合特征对实验性能的影响; 6) IndRNN 层数的影响。

1) 样本构建方法对实验性能的影响

为了验证不同样本构建方法对实验性能的影响, 本文选取两种样本构建方法。样本 1 为不基于元素种类相同规则的构建方法。样本 2 为第 2.3 节构建样本方法。实验结果如表 4 所示。

表 4 不同样本对实验性能的影响 (%)

Table 4 Hyper parameters of experiment (%)

样本种类	<i>P</i>	<i>R</i>	<i>F1</i>
样本 1	85.76	80.6	83.1
样本 2	86.94	84.14	85.52

由表 4 知, 不加入元素种类相同的规则的样本 1, *P* 为 85.76%, *R* 为 80.6%, *F1* 为 83.1%。与样本 1 相比, 样本 2 的 *P*, *R* 和 *F1* 分别提高了 1.18%, 3.54% 和 2.42%。由此证明了样本 2 中基于规则建方法的有效性, 本文中之后的实验均采用样本 2 的构建方法。

2) 与其他模型实验性能对比

为了验证本文提出模型的有效性, 将本文模型与单独的模型、注意力机制的单独模型、组合模型和注意力机制的组合模型做对比。各个模型均在各自最优参数下进行实验, 以确保对比实验结果的有效性。

由图 2 和表 5 可知, 相比于 IndRNN、CapsNet 和 IndRNN_CapsNet, 加入注意力机制后的模型, *F1* 值分别提高了 0.98%、0.61% 和 2.96%, 这是由于注意力机制可以从众多特征中选择出对当前任务目标更关键的信息, 从而提高了模型的识别能力。与单独模型相比, IndRNN_CapsNet 的 *F1* 值分别提高了 3.27% 和 0.01%, 这是因为 IndRNN_CapsNet 同时提取了局部特征和全局语义信息。而本文模型比 IndRNN_CapsNet 的 *F1* 值提高了 2.96%, 结合了注意力机制和组合模型的两种优点, 使模型达到最优。

3) 词向量对实验性能的影响

词向量富含词汇语义和上下文关系信息, 但不同维度的词向量蕴含的语义信息不同。本文分别选择 10, 30, 50, 100 和 150 维作为词向量的维度进行

词实验。实验结果如表 6 所示。

由表 6 可知, 随着词向量维度的增加, 词向量蕴含的语义信息越多, 在 50 维时 *F1* 值最大, 模型性能达到最优。当维度增加为 100 维和 150 维时, *F1* 值分别下降了 1.52% 和 4.4%, 这是因为高维度的词向量虽然含有更多的语义信息, 但同时也会增加一些无用的干扰信息和噪音, 从而影响模型的性能。

表 5 本文模型与其他模型实验性能对比 (%)

Table 5 Comparison between our model and other models (%)

模型	<i>P</i>	<i>R</i>	<i>F1</i>
IndRNN	77.6	81.06	79.29
CapsNet	79.51	85.84	82.55
Att_IndRNN	78.13	82.54	80.27
Att_CapsNet	81.63	84.74	83.16
IndRNN_CapsNet	84.17	81.02	82.56
Att_IndRNN_CapsNet	86.94	84.14	85.52

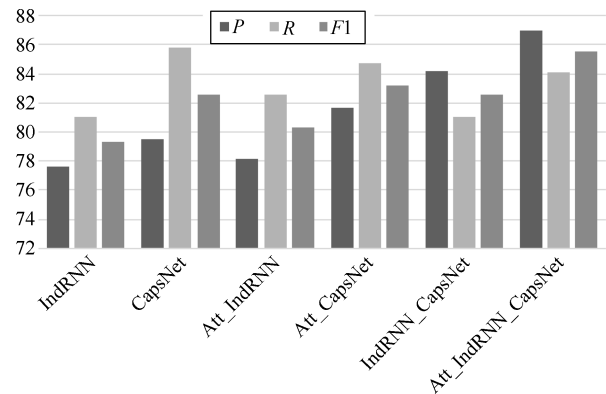


图 2 模型对比图

Fig. 2 Comparison between our model and other models

4) 语义特征和规则特征对实验性能的影响

第 4.2 节抽取的 18 项特征主要是事件类型、事件对位置关系以及元素分布情况等基于事件和元素的规则特征, 并没有考虑事件和元素的语义信息。为了探索语义特征对事件缺失元素填充的影响, 将去掉语义特征的模型与包含两类特征的模型做对比。此外, 本节还探索了规则特征对填充性能的影响, 将只包含语义特征的模型与包含全部特征的模型做对比。为了探究维吾尔语语言特点对实验的影响, 本文将去除动词后缀和格后缀的维吾尔语文本训练的词向量称为语义特征 A, 将原始文本训练的词向量称为语义特征 B, 将两种不同的语义特征做对比。实验结果如表 7 所示。

表 6 词向量对实验性能的影响 (%)

Table 6 Influence of word vector dimension (%)

维度	P	R	$F1$
10	78.3	83.58	80.85
30	81.44	84.27	82.83
50	86.94	84.14	85.52
100	84.45	83.55	84
150	80.17	80.09	81.12

表 7 不同种类特征对实验性能的影响 (%)

Table 7 Influence of different kinds of features (%)

特征	P	R	$F1$
语义特征 A	77.84	81.92	79.83
语义特征 B	78.85	83.66	81.18
规则特征	74.66	77.87	76.23
语义特征 A + 规则特征	81.17	86.44	83.72
语义特征 B + 规则特征	86.94	84.14	85.52

由表 7 可知, 在去掉语义特征 A, 只包含规则特征条件下, 其 P 、 R 和 $F1$ 与包含两类特征的 P 、 R 和 $F1$ 相比分别降低了 6.51%, 8.57%, 7.49%; 去掉语义特征 B, 只包含规则特征条件下, 其 P 和 R 与包含两类特征的 P 和 R 相比分别降低了 12.28%, 6.27%, 反映整体性能的 $F1$ 值降低了 9.29%。实验结果说明了对语义特征引入的有效性, 这是因为事件缺失元素填充与语义有关, 规则特征只是考虑了事件和元素的内部结构特征, 缺乏对事件句的语义信息考虑。而与仅包含语义特征 A 相比, 包含两类特征的模型的 P 、 R 和 $F1$ 值分别上升了 3.33%, 4.52% 和 3.89%; 与仅包含语义特征 B 相比, 包含两类特征的模型的 P 、 R 和 $F1$ 值分别上升了 8.09%, 0.48% 和 4.34%, 由此证明了规则特征在维吾尔语缺失元素填充上的有效性。包含动词后缀和格后缀信息的语义特征 B 与不包含动词后缀和格后缀信息的语义特征 A 相比, 在单独作为特征和与规则特征结合两种情况下, 衡量模型整体性能的 $F1$ 值均有提高, 分别提高了 1.35%, 1.8%, 这是因为维吾尔语中动词后缀和格后缀是表达语法信息和时态信息的部分, 如果去掉则无法更深层次挖掘维吾尔语文本的语义信息。

5) 独立特征与融合特征对实验性能的影响

不同模型对不同种类的特征有着不同的学习能力。为探究没有进行融合处理的独立特征和融合特征对缺失元素填充的影响, 本节设计了将独立特征分别输入 Att_IndRNN 和 CapsNet 中, 有 Att_IndRNN_h_CapsNet_w 和 Att_IndRNN_w_Caps-

Net_h 两种情况。另外可以将两种特征融合之后, 再分别 Att_IndRNN 和 CapsNet 中。其中, CapsNet_w 表示将语义特征输入到胶囊网络中, w 表示语义特征, h 表示规则特征。

由表 8 可知, Att_IndRNN 处理规则特征且 CapsNet 处理语义特征的模型, 与 Att_IndRNN 处理语义特征且 CapsNet 处理规则特征的模型相比, P 提高了 10.34%, 衡量整体性能的 $F1$ 值提高了 4.11%, 这是因为语义特征富含上下文全局特征, 经过胶囊网络后可以获取其局部信息, 而结合注意力机制的独立循环神经网络处理规则特征, 可以获取全局特征。而将两种特征融合之后再分别输入到两个模型, 与 Att_IndRNN 处理规则特征且 CapsNet 处理语义特征的模型相比, P 降低了 4.24%, 衡量整体性能的 $F1$ 值降低了 2.37%, 说明了特征直接融合再经过模型处理, 虽然特征更加丰富, 但是同时也会使噪声增加。

表 8 独立特征与融合特征对实验性能的影响 (%)

Table 8 Influence of independent features and fusion features (%)

模型	P	R	$F1$
Att_IndRNN _{h+w} _CapsNet _{h+w}	82.7	83.61	83.15
Att_IndRNN _w _CapsNet _h	76.6	86.87	81.41
Att_IndRNN _h _CapsNet _w	86.94	84.14	85.52

6) 独立循环神经网络层数的影响

模型中的隐含层直接影响模型对维吾尔语事件缺失元素的填充能力。当隐含层过少时, 模型不能很好地学习到更高阶特征。而隐含层过多时, 模型可能出现过拟合现象。为了探究独立循环神经网络层数对实验性能的影响, 本节设计了 1 层、2 层和 3 层独立循环神经网络 3 种情况, 实验结果如表 9。

由表 9 可知, 2 层隐含层与 1 层隐含层相比, P 增加了 4.96%, R 增加了 0.18%, 衡量模型整体性能的 $F1$ 值增加了 2.56%, 当隐含层增加到 3 层时, P 、 R 和 $F1$ 值分别降低了 4.38%, 2.76% 和 3.56%。这是因为模型经过多层独立循环神经网络所提取的高阶特征不同。针对维吾尔语事件缺失元素填充, 独立循环神经网络层数为 2 时, 模型性能最优。

表 9 独立循环神经网络层数对实验性能的影响 (%)

Table 9 Influence of the number of IndRNN (%)

层数	P	R	$F1$
1	81.89	83.96	82.96
2	86.94	84.14	85.52
3	82.56	81.38	81.96

5 结束语

事件缺失元素填充的研究有助于事件抽取技术的发展. 现有的研究主要针对英语、汉语等大语种, 对于维吾尔语事件缺失元素填充的研究很少, 且现有的研究未考虑事件和事件元素的上下文语义关系. 针对以上不足, 本文提出了结合注意力机制的独立循环神经网络和胶囊网络并行的维吾尔语事件缺失元素填充模型. 该方法将由词向量构成的语义特征和事件触发词与事件元素构成的规则特征分别输入两个不同的模型, 即胶囊网络和注意力机制的独立循环神经网络, 之后将两个模型输出的两种特征融合, 作为分类器的输入进而完成事件缺失元素填充. 此外, 本文还将事件元素分为 4 类, 按照元素类型相同的规则进行匹配, 来构建样本, 从而进一步提升了模型的性能.

致谢

感谢杨启萌、祁青山、牛苗、王欢欢等同学对本文提供的帮助和支持, 在此谨向他们致以诚挚的谢意!

References

- Mann G S, Yarowsky D. Multi-field information extraction and cross-document fusion. In: Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics. Ann Arbor, USA: ACL, 2005. 483–490
- Jiang Ji-Fa. A Research about the Pattern Acquisition for Free Text IE [Ph. D. dissertation], Graduate School of Chinese Academy of Sciences (Institute of Computing Technology), 2004
(姜吉发. 自由文本的信息抽取模式获取的研究 [博士学位论文], 中国科学院研究生院 (计算技术研究所), 2004)
- Patwardhan S, Riloff E. Effective information extraction with semantic affinity patterns and relevant regions. In: Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning. Prague, Czech Republic: ACL, 2007. 717–727
- Chen C, Ng V. Joint modeling for chinese event extraction with rich linguistic features. In: Proceedings of COLING 2012. Mumbai, India: ACL, 2012. 529–544
- Yu Kai, Jia Lei, Chen Yu-Qiang, Xu Wei. Deep learning: yesterday, today, and tomorrow. *Journal of Computer Research and Development*, 2013, **50**(9): 1799–1804
(余凯, 贾磊, 陈雨强, 徐伟. 深度学习的昨天, 今天和明天. 计算机研究与发展, 2013, **50**(9): 1799–1804)
- Lin Yi-Ou, Lei Hang, LI Xiao-Yu, Wu Jia. Deep learning in NLP: Methods and applications. *Journal of University of Electronic Science and Technology of China*, 2017, **46**(6): 913–919
(林奕欧, 雷航, 李晓瑜, 吴佳. 自然语言处理中的深度学习: 方法及应用. 电子科技大学学报, 2017, **46**(6): 913–919)
- Xi Xue-Feng, Zhou Guo-Dong. A survey on deep learning for natural language processing. *Acta Automatica Sinica*, 2016, **42**(10): 1445–1465
(奚雪峰, 周国栋. 面向自然语言处理的深度学习研究. 自动化学报, 2016, **42**(10): 1445–1465)
- Chen Y B, Xu L H, Liu K, Zeng D L, Zhao J. Event extraction via dynamic multi-pooling convolutional neural networks. In: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing. Beijing, China: ACL, 2015. 167–176
- Chang C Y, Teng Z, Zhang Y. Expectation-regulated neural model for event mention extraction. In: Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. San Diego, USA: ACL, 2016. 400–410
- Zeng Y, Yang H H, Feng Y S, Wang Z, Zhao D Y. A convolution BiLSTM neural network model for Chinese event extraction. In: Proceedings of the International Conference on Computer Processing of Oriental Languages. Cham, Switzerland: Springer, 2016. 275–287
- Tian Sheng-Wei, Hu Wei, Yu long, Turglm Ibrahim, Zhao Jian-Guo, Li Pu. Temporal relation identification of Uyghur event based on Bi-LSTM with attention mechanism. *Journal of Southeast University (Natural Science Edition)*, 2018, **48**(03): 393–399
(田生伟, 胡伟, 禹龙, 吐尔根·依布拉音, 赵建国, 李圃. 结合注意力机制的 Bi-LSTM 维吾尔语事件时序关系识别. 东南大学学报 (自然科学版), 2018, **48**(03): 393–399)
- Li Hong, Yu Long, Tian Sheng-Wei, Turglm Ibrahim. Uyghur emergency event extraction based on DCNNs-LSTM model. *Journal of Chinese Information Processing*, 2018, **32**(6): 52–61
(黎红, 禹龙, 田生伟, 吐尔根·依布拉音, 赵建国. 基于 DCNNs-LSTM 模型的维吾尔语突发事件识别研究. 中文信息学报, 2018, **32**(6): 52–61)
- Gupta P, Ji H. Predicting unknown time arguments based on cross-event propagation. In: Proceedings of the ACL-IJCNLP 2009 Conference Short Papers. Suntec, Singapore: ACL, 2009. 369–372
- Huang R, Riloff E. Peeling back the layers: detecting event role fillers in secondary contexts. In: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1. Portland, USA: ACL, 2011. 1137–1147
- Hou Li-Bin, Li Pei-Feng, Zhu Qiao-Ming, Qian Pei-De. Using cross-event inference to fill missing event argument. *Computer Science*, 2012, **39**(7): 200–204
(侯立斌, 李培峰, 朱巧明, 钱培德. 基于跨事件理论的缺失事件角色填充研究. 计算机科学, 2012, **39**(7): 200–204)
- Zhao Wen-Juan, Liu Zhong-Bao, Wang Yong-Fang. Research on event role annotation based on syntactic dependency analysis. *Information Science*, 2017, **35**(07): 65–69
(赵文娟, 刘忠宝, 王永芳. 基于句法依存分析的事件角色填充研究. 情报科学, 2017, **35**(07): 65–69)
- Yin Q Y, Zhang Y, Zhang W N, Liu T, Wang W. Zero pronoun resolution with attention-based neural network. In: Proceedings of the 27th International Conference on Computational Linguistics. Santa Fe, USA: ACL, 2018. 13–23

- 18 Feng Chong, Kang Li-Qi, Shi Ge, Huang He-Yan. Causality extraction with GAN. *Acta Automatica Sinica*, 2018, **44**(05): 811–818
(冯冲, 康丽琪, 石戈, 黄河燕. 融合对抗学习的因果关系抽取. 自动化学报, 2018, **44**(05): 811–818)
- 19 Fu Jian-Feng. Research on Event-Oriented Knowledge Processing [Ph.D. dissertation], Shanghai University, China, 2010
(付剑锋. 面向事件的知识处理研究 [博士学位论文], 上海大学, 中国, 2010)
- 20 Lakew S M, Cettolo M, Federico M. NA Comparison of Transformer and Recurrent Neural Networks on Multilingual Neural Machine Translation [Online], available: <https://arxiv.org/pdf/1806.06957.pdf>, June 20, 2018
- 21 Li S, Li W, Cook C, Zhu C, Gao Y B. Independently recurrent neural network (indrnn): building a longer and deeper RNN. In: *Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition*. Salt Lake City, USA: IEEE, 2018. 5457–5466
- 22 Sabour S, Frosst N, Hinton G E. Dynamic routing between capsules. In: *Proceedings of the Advances in Neural Information Processing Systems*. Long Beach, USA: Curran Associates, Inc., 2017. 3856–3866
- 23 Zhao W, Ye J B, Yang M, Lei Z Y, Zhang S F, Zhao Z. Investigating Capsule Networks with Dynamic Routing for Text Classification [Online], available: <https://arxiv.org/pdf/1804.00538.pdf>, September 3, 2018
- 24 He Yu, Pan Da, Fu Guo-Hong. Chinese explanatory opinionated sentence recognition based on auto-encoding features. *Acta Scientiarum Naturalium Universitatis Pekinensis*, 2015, **51**(2): 234–240
(贺宇, 潘达, 付国宏. 基于自动编码特征的汉语解释性意见句识别. 北京大学学报 (自然科学版), 2015, **51**(2): 234–240)
- 25 Mikolov T, Sutskever I, Chen K, Corrado G S, Distributed representations of words and phrases and their compositionality. In: *Proceedings of the Advances in Neural Information Processing Systems*. Lake Tahoe, USA: Curran Associates, Inc., 2013. 3111–3119



王县县 新疆大学硕士研究生. 主要研究方向为自然语言处理.

E-mail: sjzwangxianxian@163.com

(WANG Xian-Xian Master student 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.

E-mail: tianshengwei@163.com

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



王瑞锦 电子科技大学讲师. 主要研究方向为量子通信安全, 大数据分析及安全. E-mail: wrj8882003@163.com

(WANG Rui-Jin Lecturer at University of Electronic Science and Technology of China. His research interest covers quantum communication security, big data analysis and security.)