

基于 Transformer 模型与关系词特征的汉语因果类复句关系自动识别

杨进才 曹元 胡泉 沈显君

华中师范大学计算机学院 武汉 430079

摘要 汉语复句的语义关系丰富而复杂,复句关系自动识别是对复句语义关系的判别,是分析复句所表达意义的重要环节。因果类复句是使用最多的汉语复句,文中以二句式有标因果类复句为研究对象,通过深度学习的方法自动挖掘复句隐含的特征,同时融合了关系词这一语言学研究的显著知识。将 word2vec 词向量与 one-hot 编码的关系词特征结合作为模型的输入,利用卷积神经网络作为前馈层的 transformer 模型来对因果复句关系进行识别。采用文中的方法对因果类复句关系类别进行识别,实验结果的 F1 值达到 92.13%,优于现有的对比模型,表明了该方法的有效性。

关键词: 因果复句;关系识别;词向量;transformer 模型;深度学习

中图法分类号 TP391

Relation Classification of Chinese Causal Compound Sentences Based on Transformer Model and Relational Word Feature

YANG Jin-cai, CAO Yuan, HU Quan and SHEN Xian-jun

School of Computer Science, Central China Normal University, Wuhan 430079, China

Abstract Chinese compound sentences have rich and complicated semantic relations. Recognition of relation category of a Chinese compound sentence is judgment of semantic relations of the sentence, and it is very important to analyze the meaning of compound sentences. Causal compound sentences are the most frequently used sentences in Chinese article. In this paper, a corpus of causal compound sentences with two clauses is taken as research object. A deep learning method is used in order to mine the hidden features of compound sentences automatically. At the same time, this paper integrates significant linguistics knowledge of relational words in the proposed model. Combining word vector of word2vec and relation word feature of one hot coding as input to the model, a transformer model using convolutional neural network as feedforward layer is exploited to identify the relation category of causal compound sentences. Using our model, the F1 value of the experiment reaches 92.13%, which is better than the comparative model.

Keywords Causal compound sentences, Relation classification, Word vector, Transformer model, Deep learning

1 引言

随着互联网的快速发展,网络上出现了非常多的短文本,人们迫切需要计算机能够自动识别短文文本之间的语义关系,为下游的机器翻译^[1]、信息抽取^[2]和自动问答^[3]等自然语言处理任务提供支持。

复句是指含有两个或两个以上分句的句子,是汉语语法的重要实体单位。因果复句是分布范围最广、使用最频繁的复句之一^[4],因果复句是汉语复句研究的不可或缺的对象。随着对现代汉语复句研究不断深入,对因果复句研究的价值也逐渐凸显出来。近年来深度学习方法在图像处理领域应用相对比较成熟,在自然语言处理领域也取得了显著的成效,本文以二句式因果复句为研究对象,利用深度学习模型来对因果复句的关系类别进行自动识别。

2 相关研究

根据复句中是否包含关系词将复句分为有标复句与无标复句。目前对于有标复句类别的自动识别主要是基于规则的

方法和基于机器学习的方法。Yang 等结合汉语复句的句法理论、关系标记搭配理论对二句式非充盈态有标复句进行关系类别的自动标识^[5];利用语义相关度的计算方法来计算两个词语的相关度,从而识别出复句的关系类别^[6]。Huang 等利用决策树算法来提取词性等特征,识别汉语句子间的因果、并列关系^[7]。

对于无标复句关系类别的识别,由于缺少关系词,没有很明显的人工识别特征,多采用深度学习的方法进行无标复句的关系类别的识别^[8]。Schenk 等^[9]利用基于注意力机制的卷积神经网络在中文篇章书库(Chinese Discourse Treebank, CDTB)^[10]上进行无标复句关系的识别。Qin^[11]等将词向量与词性特征结合作为模型的输入,利用卷积神经网络来对无标复句关系进行分类。无标复句的研究仍面临一些困难,即数据标注困难,训练数据量相对较小,各类别的数据分布不均,容易导致模型过拟合,使模型的泛化能力不足。在众多深度学习模型中,Transformer 模型^[12]具有简单的网络架构,其主要结构是注意力机制,在 WMT 2014 英译德和 WMT 2014 上从英语到法语的翻译任务中取得了很好的效果。在自然语

言处理领域的深度学习研究及应用中,除了解决应用领域的原始特征表示与选择合适的深度学习算法两个问题外,如果能够在训练过程中融合已有面向特定应用领域的显然的知识(如人工选取的明显特征规律),对于深度模型而言,依然具有吸引力^[8]。本文探讨在深度学习模型中融合关系词特征,从而实现二句式有标因果复句关系进行自动识别。

3 因果类复句相关概念

3.1 因果复句关系类别概述

汉语复句分类标准众多,文中使用语言学界最有影响的汉语复句三分系统^[13]。三分系统将复句分成因果、并列、转折三大类;将因果类分为因果、假设、条件、目的、推断五小类;将并列类分为并列、连贯、递进、选择四小类;将转折分为转折、让步、假转三小类。文中对因果类复句进行因果、假设、条件、目的、推断的分类。

3.2 有标复句类别判别中的问题

有标复句关系词与复句关系类别有很强的对应关系,但这种对应关系并不是一对一,一个关系词可以对应一种或多种类别。因果复句中部分关系词与关系类别的对应关系如表 1 所列。

表 1 关系词与关系类别对应表

Table 1 Correspondence table of relation words and relation categories

关系词	关系类别
因为、所以、由于、因而...	因果
如果、倘若、若不是...	假设
不管、只要、只有...	条件
以便、以免、借以...	目的
既然、与其、不如...	推断
那、那么、则、如其	假设、推断
便	因果、条件、目的、推断
才	条件、因果
就	假设、因果、条件

依据关系词特征对有标复句的类别识别存在如下问题。

(1) 连用关系词的类别识别

连用关系词是指两个或两个以上的关系词在相邻的位置连续使用,如果分句 1 中连用的两个关系词均可与分句 2 中的关系词搭配,则难以判别分句 1 中哪个关系词与分句 2 中的关系词搭配。

例 1 如果由于环境致癌因素改变了细胞内某些基因,那么增殖出来的后代将可能是癌细胞。(《人民日报》1999 年 5 月 6 日)

在例 1 中出现“如果”“由于”的关系词连用,都可以与“那么”搭配,无法判断是假设关系还是因果关系。

(2) 搭配使用的关系词部分省略

有一些关系词需要与其他标记搭配使用来明确标示复句关系,当单独使用或因搭配标记缺省时,在不同复句中对应不同类别。

例 2 当我的独创产品成为世界一流时,我也自然而然跻身于世界强人之列。(《CCCS 语料库》)

例 3 条件不同,面临的任务也不同。(《CCCS 语料库》)
关系词“也”在例 2 中对应假设关系,在例 3 中对应了假设关系或并列关系。

(3) 跨类别标记关系词的类别识别

有些关系词与关系类别的对应关系可能跨大类,也可能跨小类。在复句三分系统中,有的关系可能既表示因果关系,也可能表示并列或转折关系,这是跨大类的情况。对于文中因果类中的因果、条件、假设、目的、推断五小类,关系词也可能会跨类别。

例 4 她觉得自己好像有必要去跟他说明一下情况,于是亲自去了一趟。(韦君宜《洗礼》)

例 5 过了那林,船便弯进了灵港,于是赵庄便真在眼前了。(鲁迅《社戏》)

例 4 与例 5 中“于是”对应多种关系类别,例 4 是因果关系,属于因果大类,例 5 是连贯关系,属于并列大类,“于是”在这里是跨大类的情况。

因此,仅仅依据关系词不能对复句类别进行准确判别。

4 transformer 模型

4.1 模型结构

transformer 本质是一个 Encoder-Decoder 结构,它是一个由多头注意力机制和前馈神经网络组成的模型,多头注意力层和前馈层后分别都接有残差连接并进行批次归一化处理^[14]。通过多头注意力机制将上下文与远距离的单词结合起来,并行计算处理所有的单词,这样既实现了并行计算,又能捕获全局的语义信息。文中使用的 RM-transformer 模型结构如图 1 所示。

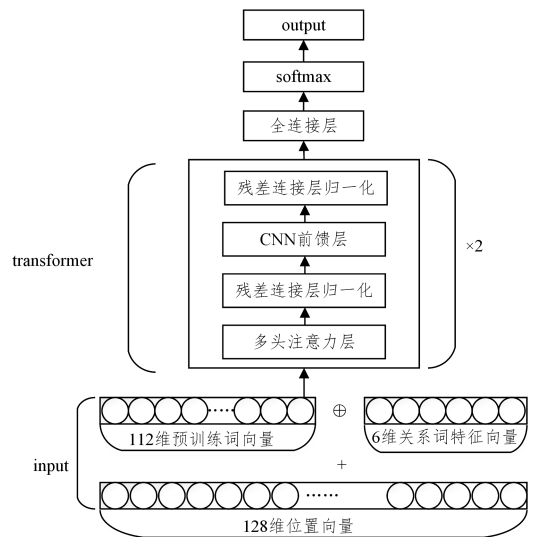


图 1 RM-transformer 结构图

Fig. 1 RM-transformer structure diagram

4.2 模型输入

本文采用预训练的 word2vec 词向量^[15]与关系词特征进行拼接作为模型输入。关系词特征采用 6 维的 one-hot 编码,将词表中所有的词均用 6 维的关系词特征表示。第 1 维用 1 与 0 分别表示是否有关系词,后 5 维对应因果、假设、条件、推断、目的 5 种关系。使用 Gensim 的 word2vec 模型来训练 122 维的词向量,然后与 6 维的关系词特征向量进行拼接,得到一个 128 维的向量。若一个输入句子的长度为 n , W_j ($j=1, \dots, n$) 表示第 j 个词的预训练词向量, R_j ($j=1, \dots, n$) 表示第 j 个词的关系词向量。则每个词的向量 WR 表示如下:

$$WR_j = W_j \oplus R_j \quad (1)$$

其中, \oplus 表示拼接操作。

多头注意力机制可以获取到远距离特征信息,也可以进行并行计算,但无法表示输入句子的位置信息,这里利用 Google 提出的 Position Embedding,将每个位置进行编码,这样多头注意力机制就可以得到每个词的位置信息。位置向量的公式如式(2)与式(3)所示。

$$P_{(j,2i)} = \sin(pos/10000^{2i/d_{model}}) \quad (2)$$

$$P_{(j,2i+1)} = \cos(pos/10000^{2i/d_{model}}) \quad (3)$$

其中, $j(j=1, \dots, n)$ 用来表示单词的位置信息, P_j (奇数位置和偶数位置) 表示第 j 个位置的位置向量, i 是指向量中每个值的索引。 $d_{model} = 128$ 与添加特征后的词向量维度一致。在偶数位置,使用正弦编码;在奇数位置,使用余弦编码。输入模型的向量表示如下:

$$Vector_j = WR_j + P_j \quad (4)$$

其中, $+$ 表示词向量直接相加。

将 128 维的词向量与 128 维的位置向量相加得到的向量 $Vector_j$ 表示作为模型的输入。

4.3 transformer 特征提取

self-attention 就是计算输入每个单词向量的权重,它会随机初始化一组权重矩阵 Q, K, V 。 Q, K, V 是输入模型的词向量与一个随机初始化矩阵相乘得到的结果, d_k 是 Q, K 向量的维度,计算公式如下:

$$attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (5)$$

通过 self-attention 层可以获得输入句子的全局语义信息,在 self-attention 层后需要连接一个前馈层,这里用 CNN 来作为前馈层, CNN 可以捕获句子的局部特征,从而可以获得句子更全面的特征信息。 CNN 采用一维卷积操作,先进行内层卷积操作,内层 filters 的数量采用自己设置的参数,使用 relu 激活函数。再进行内层卷积操作,外层 filters 的数量与词向量的维度一致,保证输入 CNN 的维度与输出的维度一致。进行以上过程的 transformer 特征提取后,将输出的结果输入下一个 transformer 编码器中。特征提取完成后经过一个全连接层和 softmax 层,输出每个类别的概率分布。

5 实验与分析

5.1 实验数据

本文是对二句式有标因果复句的关系类别进行识别,数据集是来自华中师范大学标注的汉语复句语料库(The Corpus of Chinese Compound Sentences,简称 CCCS)和清华新闻分类语料 THUCNews。CCCS 是包含 658447 条的汉语有标复句专用语料库,来源于《人民日报》《长江日报》。清华新闻分类语料 THUCNews 是根据新浪新闻 RSS 订阅频道 2005—2011 年间的历史数据筛选过滤生成,其中包含 14 种类别的短文本新闻语料。我们一共标注了 91646 条二句式有标因果复句,构成语料库简记为 CTCCCS (The Corpus of Two-Sentence Causal Chinese Compound Sentences),数据集各关系类别数据分布如表 2 所列。实验中选择 75% 的数据为训练集,25% 的数据为测试集,数据划分情况如表 3 所列。

表 2 CTCCCS 语料库各关系类别的数据分布

Table 2 Data distribution of relation categories in CTCCCS corpus

关系类别	复句实例个数	比例 / %
因果类	30180	32.93
假设类	31444	34.31
条件类	11576	12.63
目的类	12438	13.57
推断类	6008	6.56

表 3 CTCCCS 数据集划分

Table 3 Data set division of CTCCCS

关系类别	训练集个数	测试集个数
因果类	22635	7545
假设类	31444	7861
条件类	8682	2894
目的类	9328	3110
推断类	4506	1502

5.2 实验参数设置

实验的参数设置主要分为两部分,模型的超参数和训练参数。模型的超参数主要是输入模型中词向量的维度、多头注意力的头数、transformer 编码器的数量、CNN 作为前馈层的 filter 的数量以及卷积核的大小。模型训练使用的是 Adam 优化器,采用随机梯度下降算法,前馈层 CNN 内层卷积使用 relu 激活函数。利用 dropout^[16] 和提前停止训练^[17] 来防止模型的过拟合,模型训练是每 100 步验证一次,若 8 次验证损失都不下降则停止训练。各参数设置详情如表 4 所列。

表 4 参数设置

Table 4 Parameter setting

参数	取值
词向量维度	128
transformer 编码器数量	2
自注意力机制头数	8
序列长度	100
optimizer	Adam
学习率	0.001
epoch	10
batch_size	128
dropout	0.7
filter	128

5.3 实验对比与分析

本文采用的是添加关系词特征的 RM-transformer 模型,为了验证模型的有效性,选取了 CNN, LSTM, biLSTM 等常见的深度学习模型来对比。为了忽略关系词的影响,采用的是添加关系词特征的向量作为模型的输入。同时,为了验证模型所添加的关系词是否对实验结果有影响,将其与不添加关系词特征的 transformer 模型进行比较,验证该特征是否利于因果复句关系类别的识别。实验结果的评价指标是 Accuracy(准确率)、Precision(精确率)、Recall(召回值)和 F1 值,各指标使用的是宏平均(macro-average)计算方法。

实验中对模型各个相同的超参数设置如表 4 所列, CNN 中 filter_sizes 取值为 [3, 4, 5], 多个不同尺寸的卷积核可以捕获不同尺寸的特征,比使用单一尺寸的卷积核效果好。LSTM 和 biLSTM 的层数设置为 1, 隐藏层取值为 128。

实验结果如表 5 所列,本文提出的 RM-transformer 模型对二句式因果复句关系进行自动识别,与 CNN, LSTM, biLSTM 相比较, Accuracy 均有所提高,分别提高了 3.27%, 0.98%, 0.3%。相对于不添加关系词特征的 transformer 模

型, *Accuracy* 提高了 13.74%。模型中使用固定的序列长度, 取值 100, RM-transformer 与 CNN 相比, *Precision*, *Recall* 和 *F1* 值提高得更加明显, 分别是 3.38%, 2.83%, 3%。CNN 将一维文本输入转化为二维的输入, 通过卷积核获取文本的局部语义信息, 虽然使用多个不同尺寸的卷积核来捕获不同尺寸的特征, 但可能对于长距离特征的学习还是有所困难。RM-transformer 通过多头注意力机制进行并行计算, 同时学习全局的特征信息, 再通过 CNN 前馈层来学习序列局部的特征, 从而实现比 CNN 更好的效果。

表 5 各模型实验结果对比

Table 5 Comparison of experimental results of each model

(单位: %)

Algorithm	Accuracy	Precision	Recall	F1
transformer	81.35	82.33	79.05	80.50
CNN	89.79	88.84	89.44	89.13
LSTM	92.08	89.61	92.39	90.87
biLSTM	92.76	91.99	91.96	91.95
RM-transformer	93.06	92.22	92.27	92.13

RM-transformer 与 LSTM 和 biLSTM 相比, *F1* 值分别提高 1.26%, 0.18%, 相比 CNN, *F1* 值提高得并不特别明显。RM-transformer 的 *Recall* 比 LSTM 低 0.12%, LSTM 和 biLSTM 在处理输入文本序列问题上相对比较成熟, 能够学习到序列的远距离特征, 但 LSTM 依赖上文信息, biLSTM 依赖上下文信息, 为了避免信息丢失, 不能把序列切开, 难以实现并行计算。RM-transformer 中的 self-attention 机制使得远距离特征能够直接发生关系, 从而获得全局的特征, RM-transformer 能够达到与 LSTM 和 biLSTM 类似的作用。本文人工标注的数据有限, 在数据量更多的复句数据集上实验, transformer 的并行计算能力或许能有更好的效果。

RM-transformer 与不添加关系词特征的 transformer 模型进行比较, *F1* 值提升了 11.63%, 效果提升明显, 说明关系词对于因果复句的关系判断有非常重要的作用。虽然深度学习模型可以自动挖掘文本的一些语义等特征信息, 但通过添加一些明显的人工识别的特征, 能够让模型的效果更好。

因果复句的各个类别的分类实验结果如表 6 所列。

表 6 因果复句各个类别的分类正确率

Table 6 Classification accuracy of various categories of causal compound sentences

关系类别	正确率/%
因果类	93.25
假设类	92.41
条件类	95.20
目的类	96.85
推断类	83.69

从表 6 看出, 推断类复句的识别正确率明显低于其他类别, 实验中分类错误可能的原因如下: 1) 实验的语料多来自新闻语料, 推断类的复句使用不频繁, 所以收集到的语料偏少, 在训练过程中出现了过拟合。2) 句子中有多个准关系词(可充当关系词的词), 对应不同的类别。例如: “既然/黄静/的行为/属于/维权/过度/, /那/都/是/哪些地方/过/了/度?”, 这个句子中的“既然”表示推断关系, 但“那”既可表示假设关系, 也可表示推断关系, 本应该是推断关系判断为了假设关系。

结束语 文中利用 transformer 模型作为特征提取器来

提取输入序列的特征, 模型中的多头注意力机制可以有效地学习到输入句子的全局的语义信息, 从而实现对二句式因果复句的分类。transformer 模型作为一种深度学习模型, 可以不依赖人工设计的语言特征对句子进行分类, 但在数据量不够的情况下容易过拟合, 影响模型的效果。关系词是识别复句关系的一个明显的特征, 当加入关系词特征, 将关系词特征与词向量一起输入到模型中, 模型识别的 *F1* 值得到了明显的提升, 说明加入关系词特征对模型自动识别起到了很好的效果。

接下来可以将依存句法树^[18]作为输入, 以融入更丰富的句法信息, 利用图卷积神经网络模型^[19]进行有标复句关系类别的识别, 以进一步提高复句类别识别的准确率。

参 考 文 献

- [1] JOTY S, GUZMÁN F, MÁRQUEZ L I, et al. Using Discourse Structure for Machine Translation Evaluation[C]// ACL Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics, 2014: 687-698.
- [2] HUTTUNEN S, VIHAVAINEN A, VON ETTER P, et al. Relevance Prediction in Information Extraction using Discourse and Lexical Features[C]// Proceedings of the 18th Nordic Conference of Computational Linguistics (NODALIDA 2011). 2011: 114-121.
- [3] VERBERNE S, BOVES L, OOSTDIJK N, et al. Evaluating discourse-based answer extraction for why-question answering [C] // ACM, 2007.
- [4] RONG L H, ZHOU G D. An Overview of the Study on Chinese Cause-effect Complex Sentence [J]. Journal of Changchun Normal University, 2011, 30(9): 47-51.
- [5] YANG J C, CHEN Z Z, SHENG X J, et al. Automatic recognition of relation category of non-saturated compound sentences with two clauses [J]. Application Research of Computers, 2017, 34(10): 2950-2953.
- [6] YANG J C, CHEN Z Z, SHENG X J, et al. Word Semantic Relevance Computation and Categories Identification[J]. Application Research of Computers, 2017, 44(5): 280-284.
- [7] HUANG H, CHANG T, CHEN H, et al. Interpretation of Chinese Discourse Connectives for Explicit Discourse Relation Recognition[C]// the 25th International Conference on Computational Linguistics, 2014, 15(1): 632-643.
- [8] XI X F, ZHOU G D. A Survey on Deep Learning for Natural Language Processing[J]. Acta Automatica Sinica, 2016, 42(10): 1445-1465.
- [9] RÖNNQVIST S, SCHENK N, CHIARCOS C. A Recurrent Neural Model with Attention for the Recognition of Chinese Implicit Discourse Relations[EB/OL]. [2017-04-26]. <https://arxiv.org/abs/1704.08092>.
- [10] ZHOU Y, XUE N. The Chinese Discourse TreeBank: a Chinese corpus annotated with discourse relations [J]. Language Resources and Evaluation, 2015, 49(2): 397-431.
- [11] QIN L, ZHANG Z, ZHAO H. Shallow Discourse Parsing Using Convolutional Neural Network [C]// Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, 2016: 70-77.

(下转第 305 页)

局的处理策略。

单个连接词的组合单元跟隐现的“的”同时存在时,系统在转换上会出现错误。如复杂名词短语“膜两侧之间的晶体量之差(the difference in crystal amount between both sides of the film)”,按照目前的识别规则和合并策略,此复杂名词短语的识别结果是“膜/两侧之间/的/晶体量/之/差”。按照目前的调序策略,复杂名词短语的结果是“差 of 晶体量 of 膜 in 两侧”。对照译文,识别的组合单元没有错误,但是调序在“膜”和“两侧”处发生了错误。

结束语 本文提出了基于规则的复杂名词短语的识别和转换的处理方法。通过分析汉英双语对齐的专利文本中复杂名词短语语料的共性,以语义块为切入点,根据汉语语言特征的语义特点、切分功能和组合特点,采用边界感知原则为汉语复杂名词短语中组合单元边界构建了 57 条识别规则并设计了合并策略,得到汉语复杂名词短语的形式化结构。通过对比汉英复杂名词短语的差异,确定了汉英复杂名词短语的转换策略。最后,将识别规则、合并策略和转换策略应用到一个机器翻译系统中。测试结果表明,这种方法可以有效地实现复杂名词短语的识别和转换,提高专利文本中复杂名词短语的机器翻译效果。

在下一步的工作中,我们将通过分析更多的语料,深化对复杂名词短语的识别规则和转换规律的总结,并进一步修改和完善规则,加强对复杂名词短语形式化识别和转换的处理,以提高复杂名词短语识别和转换的准确率,进一步改善机器翻译效果。

参 考 文 献

- [1] 张冬梅,晋耀红. 面向专利机器翻译的要素句蜕识别和转换研究[J]. 计算机科学,2014,41(S1):67-71.
- [12] XING F Y. A Research on Chinese Compound Sentences[M]. Beijing:Commercial Press,2003.
- [13] VASWANI A, SHAZEER N. Attention Is All You Need[C] // 31st Conference on Neural Information Processing Systems, 2017.
- [14] IOFFE S, SZEGEDY C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift [EB/OL]. [2015-03-02]. <https://arxiv.org/abs/1502.03167>.
- [15] MIKOLOV T, SUTSKEVER I, CHEN K, et al. Distributed Representations of Words and Phrases and their Compositionality[J]. Advances in Neural Information Processing Systems, 2013,26:3111-3119.
- [16] SRIVASTAVA N, HINTON G, KRIZHEVSKY A, et al. Dropout: A Simple Way to Prevent Neural Networks from Overfitting[J]. Journal of Machine Learning Research, 2014, 15(1): 1929-1958.
- [17] CARUANA R, LAWRENCE S, GILES L. Overfitting in Neural

- [2] 池毓焕. 多元逻辑组合的汉英对比初探. 第二届 NNC 与语言学研讨会论文集[M]. 北京:海洋出版社,2004:308-312.
- [3] 熊亮. 非句蜕广义对象语义块的分析与处理[D]. 北京:中国科学院声学所,2006.
- [4] 李千驹. 基于 HNC 理论的汉英逻辑组合变换研究[D]. 北京:北京师范大学,2008.
- [5] 李颖,王侃,池毓焕. 面向汉英机器翻译的语义块构成变换[M]. 北京:科学出版社,2009:91-124.
- [6] 詹卫东. 面向中文信息处理的现代汉语短语结构规则研究[M]. 北京:清华大学出版社出版,2000:61-75.
- [7] 李素建,刘群. 汉语语块的定义和获取. 语言计算与基于内容的文本处理[M]. 北京:清华大学出版社,2003:110-115.
- [8] 胡乃全,朱巧明,周国栋. 混合的汉语基本名词短语识别方法[J]. 计算机工程,2009,35(20):199-201.
- [9] 田雪,黄德根. 一种混合的汉语简单名词短语识别方法[J]. 小型微型计算机系统,2017,38(4):749-754.
- [10] 姜亚辉,姬东鸿. 结合半监督与主动学习的复杂名词短语识别[J]. 计算机工程与设计,2015,36(2):498-501,506.
- [11] ZHU Y, JIN Y H. A Chinese-English patent machine translation system based on the theory of hierarchical network of concepts [J]. The Journal of China Universities and Telecommunications, 2012, 19:140-146.
- [12] 刘小蝶,朱筠,晋耀红. 中文专利中有标记并列结构的自动识别研究[J]. 计算机工程,2018,44(6):162-168,175.



LIU Xiao-die, born in 1984, Ph.D, lecturer. Her main research interests include Chinese information processing and corpus linguistics.

(上接第 298 页)

- [12] XING F Y. A Research on Chinese Compound Sentences[M]. Beijing:Commercial Press,2003.
- [13] VASWANI A, SHAZEER N. Attention Is All You Need[C] // 31st Conference on Neural Information Processing Systems, 2017.
- [14] IOFFE S, SZEGEDY C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift [EB/OL]. [2015-03-02]. <https://arxiv.org/abs/1502.03167>.
- [15] MIKOLOV T, SUTSKEVER I, CHEN K, et al. Distributed Representations of Words and Phrases and their Compositionality[J]. Advances in Neural Information Processing Systems, 2013,26:3111-3119.
- [16] SRIVASTAVA N, HINTON G, KRIZHEVSKY A, et al. Dropout: A Simple Way to Prevent Neural Networks from Overfitting[J]. Journal of Machine Learning Research, 2014, 15(1): 1929-1958.
- [17] CARUANA R, LAWRENCE S, GILES L. Overfitting in Neural Nets; Backpropagation, Conjugate Gradient, and Early Stopping [C] // Proceedings of the 13th International Conference on Neural Information Processing Systems(NIPS'00), 2001:381-387.
- [18] CARO L D, GRELLA M. Sentiment analysis via dependency parsing[J]. Computer Standards & Interfaces, 2013,35(5):442-453.
- [19] NIEPERT M, AHMED M, KUTZKOV K. Learning Convolutional Neural Networks for Graphs [EB/OL]. [2016-05-17]. <https://arxiv.org/abs/1605.05273>.



YANG Jin-cai, born in 1967, Ph.D, professor, Ph.D supervisor, is a member of China Computer Federation. His main research interests include advanced database and information system, Chinese information processing, artificial intelligence and natural language processing.