

# 关系词非充盈态复句的特征融合 CNN 关系识别方法<sup>①</sup>



杨进才, 汪燕燕, 曹元, 胡金柱

(华中师范大学 计算机学院, 武汉 430079)

通讯作者: 杨进才, E-mail: jcyang@mail.ccnu.edu.cn

**摘要:** 汉语文章中复句占多数, 复句关系类别的识别是对复句分句之间的语义关系的甄别, 是分析复句语义的关键。在关系词非充盈态复句中, 部分关系词缺省, 因此, 不能通过关系词搭配的规则来对非充盈态复句进行类别识别, 且通过人工分析分句的特征进行类别识别费时费力。本文以二句式非充盈态复句为研究对象, 采用在卷积神经网络中融合关系词特征的 FCNN 模型, 尽可能减少对语言学知识和语言规则的依赖, 通过学习自动分析两个分句之间语法语义等特征, 从而识别出复句的关系类别。使用本文提出的方法对复句关系类别识别准确率达 97%, 实验结果证明了该方法的有效性。

**关键词:** 关系词非充盈态; 复句语义关系; 关系词; 特征融合; 卷积神经网络

引用格式: 杨进才, 汪燕燕, 曹元, 胡金柱. 关系词非充盈态复句的特征融合 CNN 关系识别方法. 计算机系统应用, 2020, 29(6): 224-229. <http://www.c-s-a.org.cn/1003-3254/7369.html>

## Relation Classification of Non-Saturated Chinese Compound Sentence via Feature Fusion CNN

YANG Jin-Cai, WANG Yan-Yan, CAO Yuan, HU Jin-Zhu

(School of Computer, Central China Normal University, Wuhan 430079, China)

**Abstract:** In Chinese essay, compound sentences are the majority. Recognition of relation category is screening for semantic relation of clauses in a compound sentence, and it is the key to analyze the meaning of the whole compound sentences. In a non-saturated compound sentence, the relation words are absent. So, the non-saturated compound sentence can not be classified by the features of the relation word collocation. In this work, an unbalanced corpus of non-saturated compound sentences with two clauses is taken as the research object. This study proposes a convolutional neural network for relation classification that automatically learns features from two clauses and minimizes the dependence on pre-existing natural language processing tools and language rules. The model fuses the features of relation to improve the performance. The experimental results show that the accuracy is 97% and that the proposed model outperforms the best baseline systems with sentence level features.

**Key words:** compound sentence with non-saturated relation word; semantic relation of compound sentence; relation mark; fusion feature; convolutional neural network

语义分析是自然语言处理 (Natural Language Processing, NLP) 领域中的一项基础任务, 是自然语言理解的基础。在统计机器翻译<sup>[1]</sup>、信息抽取<sup>[2]</sup>、自动问

答<sup>[3]</sup>和计算机辅助评估<sup>[4]</sup>等下游任务中, 通过语义分析获取句子级别上的丰富的语义信息, 推动自然语言处理领域的发展。

① 基金项目: 国家社科基金 (19BY092)

Foundation item: National Social Science Foundation of China (19BY092)

收稿时间: 2019-09-10; 修改时间: 2019-10-10, 2019-10-21; 采用时间: 2019-10-30; csa 在线出版时间: 2020-06-10

汉语文章中有 2/3 的句子是复句,因而复句在现代汉语中占十分重要的地位.复句通常由两个及两个以上意义密切相关的分句组成.关系词(也称关系标记)是用来连接各个分句构成复句的语法成分,但在汉语复句中,分句中的关系词可以部分省略或全部省略<sup>[5]</sup>.

关系词非充盈态复句是指关系词在分句中没完全显式出现的复句.复句的关系类别是指两个分句之间语义关系类别,一个含有多个分句的复句可以划分为多个含两个分句式的复句.因此,本文选取含有两个分句的二句式复句作为研究对象.

二句式复句定义为一个二元组 $\langle c1, c2 \rangle$ ,其中  $c1$ ,  $c2$  为分句.

非充盈态复句满足:  $(r(c1)=\Phi \wedge r(c2) \neq \Phi) \vee (r(c1) \neq \Phi \wedge r(c2) = \Phi)$

复句的类别与关系词搭配有很强的关系,一个关系词与不同的关系词搭配,构成不同的类别.例句 1 中,关系词“既”与“也”搭配,复句关系类别为并列关系.例句 2 中,关系词“如果”与“也”搭配,类别为因果关系.对于一个每个分句的关系词显式出现的充盈态复句,根据关系词搭配的可以判断出对应的类别.对于例 3,例 4 和例 5,仅分句 2 中出现关系词“也”,对这样的关系词非充盈态复句,无法根据关系词的搭配直接判断复句的类别.

例 1: 这种电话既是用于紧急情况下的报警,也用于遇到一般困难时的求助.

例 2: 如果技术掌握得当,阔叶树移栽也有成功的实例.

例 3: 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.

例 4: 他听课,也不打声招呼.

例 5: 条件不同,面临的任务也不同.

## 1 相关工作

对于复句关系类别识别的方法分为两类,第一类是借助关系词识别和关系词搭配的研究分析复句的语义关系. Huang HH 等基于半监督学习方法探索关系标记的概率分布,发现成对的关系词连用对分析篇章的语义关系起到的强提示<sup>[6]</sup>.胡金柱、舒江波等将词性和关系标记搭配理论相结合,提出正向选择算法用于关系词识别<sup>[7]</sup>.胡金柱、陈江曼等分析了关系词连用情况,提出一种连用关系词识别算法<sup>[8]</sup>.胡金柱、胡泉等

对在关系词识别过程中的规则解析算法进行了研究,提出了包含匹配算法<sup>[9]</sup>.

第二类是结合句法树的词法、句法特征,采用机器学习的方法识别复句关系类别.李艳翠、孙静等基于已标注的清华汉语树库,采用最大熵、决策树和贝叶斯方法判断准关系词是否为关系词以及复句关系类别的识别<sup>[10]</sup>.周文翠和袁春风<sup>[11]</sup>选取主语、谓语等相关特征,使用支持向量机模型识别并列复句.杨进才、陈忠忠等结合汉语复句的句法理论和分句间的语义信息,提出了基于语义相关度的非充盈态的二句式复句关系类别识别<sup>[12]</sup>.Huang HH 等利用决策树算法提取词性、句子长度、关系标志特征,来识别汉语句子间的因果、并列关系<sup>[13]</sup>.

上述方法需要依赖良好的语言学知识,通过人工总结规则以及发掘特征,不利于发现隐藏的特征,影响识别系统的自动化.

本文以复句整体为处理对象,在神经网络模型中输入关系词特征,以自动地从复句分句中发现与关系词相关的蕴含的语义关系特征.

## 2 方法

### 2.1 复句分类体系

汉语复句关系分类体系很多,著名汉语语言学家邢福义先生的复句关系“三分系统”<sup>[5]</sup>是目前广泛使用的主流分类体系.三分系统将复句分成因果、并列、转折 3 大类,其中每一大类又分成多个小类(如表 1 所示).

表 1 复句三分系统

第一层	第二层
因果类复句	因果句
	推断句
	假设句
	条件句
	目的句
并列类复句	并列句
	连贯句
	递进句
	选择句
转折类复句	转折句
	让步句
	假转句

$R\langle c1, c2 \rangle$  为复句关系类别.  $R\langle c1, c2 \rangle \in \{\text{因果, 并列, 转折}\}$ , 因果 = {因果|推断|假设|条件|目的}, 并

列={并列|连贯|递进|选择}, 转折={转折|让步|假设}. 本文采用复句三分系统作为类别标识标准.

### 2.2 模型结构

使用的 Shallow CNN 模型结构如图 1 所示. 首先, 利用查询词表, 提取词对应的词向量, 将整个复句转换

成句子向量并将其作为卷积层的输入<sup>[14]</sup>; 同时, 将关系词特征输入到我们的模型中. 经过卷积和 maxpooling 操作, 不同窗口大小的滤波器获取不同的特征映射, 最后这些特征连接起来, 作为全连接层和 Softmax 层的输入, 输出各个类别概率分布.

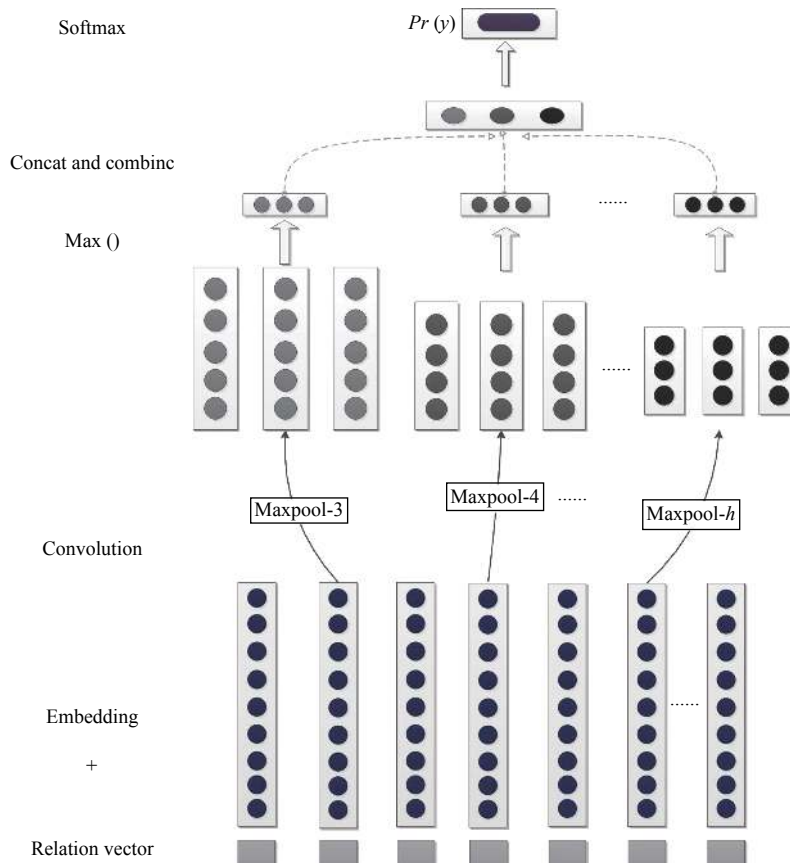


图 1 SCNN 的结构图

### 2.3 词向量的获取

一个关系词在不同的语境下, 对应的复句关系类别可能不同, 关系词与类别是一对多的关系. 经过对大规模地语料统计, 我们考察了关系词在关系类别、词性、搭配情况以及在搭配中处于前呼还是后应等属性, 统计了 499 个准关系词的上述特征属性, 构建关系词库.

图 2 展示的是部分关系词与复句关系类别的对应. 根据关系表采用 onehot 离散化每一个词, 得到每一个词对应的关系向量,  $R_{ri} \in R^l$  中 (其中  $l$  是关系类别数), 若用  $X_i \in R^k$  表示每个词在预训练词向量中对应的向量, 则每个词对应的特征向量记为  $Y_i$ :

$$Y_i = X_i \oplus R_{ri}$$

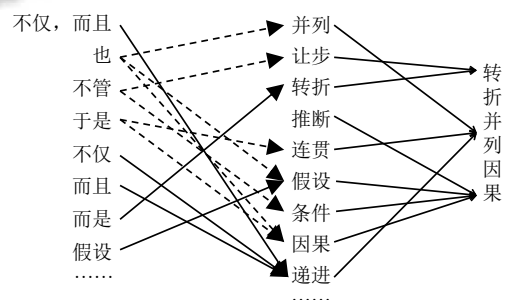


图 2 关系词——关系类别对应

一个长度为  $n$  的句子表示为:

$$Y_{1:n} = Y_1 \oplus Y_2 \oplus \dots \oplus Y_n \quad (1)$$

其中,  $\oplus$  是连接操作, 用  $Y_{i:itj}$  代表特征  $Y_i, Y_{i+1}, \dots, Y_{itj}$  之间的拼接操作.

## 2.4 卷积核最大池化操作

卷积操作是滤波器  $w \in R^{nk}$ , 一个大小为  $h$  个单词的窗口的滤波器产生新特征. 例如, 一个特征  $c_i$  是由大小为  $Y_{i:i+h-1}$  的窗口生成的. 句子级的特征向量利用滤波器矩阵  $[W_1, W_2, \dots, W_{n-h+1}]$  进行卷积操作生成, 这些句子级的向量被转化为特征映射:

$$c_i = [\text{relu}(W \cdot Y_{1:h} + b), \dots, \text{relu}(W \cdot Y_{n-h+1} + b)] \quad (2)$$

这里的下标  $[i+h-1]$  表示卷积窗口的始末下标. 此外, 在 embedding 层和滤波器矩阵之间应用卷积操作, 即利用不同窗口大小的卷积核作用于整个句子, 以获取不同的特征映射表示, 得到复句的不同特征表示. 接下来, 在上一步得到的特征映射中使用 max-pooling 操作<sup>[15]</sup>, 即  $\hat{c}_i = \max(\hat{c}_i)$ , 这样就得到了特征映射中最重要特征  $\hat{c}$ .

## 2.5 拼接操作

在特征  $\hat{c}$  上增加上标, 将不同滤波器  $[\hat{c}^1, \hat{c}^2, \dots, \hat{c}^h]$  连接起来形成句子表示向量  $v$ , 如式 (3) 所示:

$$v = \hat{c}^1 \oplus \hat{c}^2 \oplus \dots \oplus \hat{c}^h \quad (3)$$

然后将得到的特征通过全连接层 Softmax, 输出最终的概率分布.

## 2.6 模型训练

在训练过程中, 我们有两种类型的词向量: 一种是通过训练数据训练词向量, 另一种是结合关系词特征得到的词向量.

CNN 网络训练是基于梯度的 Adadelta 优化器<sup>[16]</sup>和反向传播算法<sup>[17]</sup>. 同时, 我们结合了 early stopping<sup>[18]</sup>和 dropout<sup>[19]</sup>来防止过拟合.

交叉熵函数用来证明可以加速反向传播算法, 提供良好的网络性能和相对较短的停滞期, 尤其对分类任务<sup>[20]</sup>. 为了构建目标函数, 应该考虑交叉熵损失函数和 L2 正则化项. 我们使用 ReLU 作为模型的激活函数, 可以产生与 Sigmoid 函数接近的结果, 而 ReLU 不存在指数计算, 求导计算量小, 且收敛得更快, 同时缓解了过拟合的情况. 模型参数调节列表如表 2 所示.

## 3 实验与分析

### 3.1 数据集

汉语复句语料库 (the Corpus of Chinese Compound Sentence, CCCS), 可访问 <http://218.199.196.96:8080/jiansuo/TestFuju.jsp> 获取. 语料库是包含 658 447 条的

汉语有标复句专用语料库. 来源于《人民日报》、《长江日报》. 我们从中选取了 24 706 条二句式非充盈态有标复句, 构成语料库简记为 NCCST. NCCST 语料库中, 各关系类别的数据分布如表 3 所示, 3 种关系类别的数据分布是不平衡的, 这会影响模型的训练效果, 易造成过拟合, 模型难收敛. 本文采用过采样技术<sup>[15]</sup>处理不平衡样本数据.

表 2 模型参数调节列表

参数	说明	取值范围	取值
Optimizer	模型优化器	Adadelta, RMSprop	Adadelta
Batch_size	梯度下降时, 每个 batch 样本数	8, 16, 32, 64	32
Dropout	每次更新参数时随机断开输入神经元的比例, 防止过拟合	[0.05, 1]	在输入层之后使用的是 0.5, Softmax 层之后使用的是 1
Filter_sizes	卷积层中选择不同的卷积核大小	[2,3,4],[3,4],[3,4,5]	卷积窗口使用的是 [3,5,5]

表 3 数据统计分布表

关系类别	训练集	测试集	总计	比例 (%)
转折类	4857	2083	6940	28.09
因果类	11 657	4995	16 652	67.04
并列类	781	333	1114	4.51

### 3.2 实验对比与分析

决策树算法 (C5.0)<sup>[13]</sup>: 作为传统机器学习模型的代表, 在各种分类问题上取得不错的效果. 本次对比的对象中, Huang HH 等使用了词性、句子长度和关系连接词等特征, 用 C5.0 模型训练.

Semantic\_relevance 模型<sup>[12]</sup>: 该模型是基于汉语复句语义相关度计算, 主要考虑了词语共现、关系词搭配距离、词间距等因素. 根据计算结果判别复句语义关系类别.

我们进行了 3 个实验, 第 1 个是预训练词向量作为 CNN 模型的输入; 第 2 个实验在 CNN 中引入外部知识, 加入了关系词特征, 得到融合模型 FCNN; 第 3 个是对比实验, 与文献<sup>[13]</sup>中, Huang HH 等用决策树算法进行的汉语句子的关系分类进行对比. 此外, 我们也进行了是否使用停用词表 (停用词表不包括所有的关系词) 的对比实验, 实验结果表明其对结果影响很小, 几乎可以忽略不计.

表 4 的结果显示, Shallow CNN 比文献<sup>[12]</sup>和文

献[13]中的方法在正确率分别高出 5.16% 和 1.96%, 说明神经网络利用词向量能够捕获基本的词法、语义相似度一些信息. Shallow CNN 通过卷积运算自动学习词法方面的特征, 将低级特征组合成高级特征, 得到整个复句的语义表示; 同时使用 max 池化操作捕获复句关系词特征, 对复句的关系类别识别起到显著的作用.

表 4 识别正确率的对比结果

模型	正确率 (%)
Semantic_relevance	89
C5.0	92.20
Shallow CNN	94.16

## 4 讨论

为了保证实验结果的公平, 所有模型使用的是同一数据集. 表 5 分别展示了 C5.0、shallow CNN 和 FCNN 模型在 Precision、Recall 和 Macro-F1 不同的表现结果.

表 5 不同评价指标的实验结果比较 (%)

方法	Precision	Recall	Macro-F1
C5.0	88.80	72.80	80.00
Shallow CNN	90.74	88.71	96.84
FCNN	90.56	91.07	97.16

结果表明, 相比较于 C5.0, Shallow CNN 和 FCNN 在 Macro-F1 上高出 16.84% 以上, 这表明我们的方法在复句关系分类方面要好于 C5.0 算法. C5.0 算法是决策树算法的一种, 该算法需要人工构建特征, 分析分句之间的语言学特征比如词性、关系连接词等, 从而构建对应的特征矩阵. 因此, 特征的选取对 C5.0 算法起到关键的作用, 而特征的好坏依赖于语言学知识和规则的总结. 本文提出的 Shallow CNN 模型直接利用 CNN 做特征提取, 得到句子的特征表示. 值得一提的是, FCNN 在 Shallow CNN 的基础上加了关系词特征, 它在召回率和 Macro-F1 上比 Shallow CNN 好, 分别提高了 2.36% 和 0.32%, 这表明关系词提供了一些有利于复句关系类别分类的信息. FCNN 模型融合关系词与关系类别之间映射关系这一特征, 首先 CNN 利用 embedding 层获取复句的语义特征表示, 然后融合对关系类别识别起到关键作用的关系词特征, 进一步强化神经网络自动学习该特征, 从而更加高效地识别关系类别.

我们对标识结果进行统计分析, 发现 FCNN 模型不仅对于一个关系词对应单种关系类别的识别效果好,

并且对于像“也”这样的关系词, 它对应多个关系类别 (如图 2 所示), FCNN 也要比 Shallow CNN 的识别效果要好. 对于文中的例句, FCNN 能够识别出例 3 是并列关系, 例 4 是转折关系, 例 5 是因果关系. 模型中融合了关系词特征, 从侧面说明了关系词特征对复句类别的识别起到了提示作用.

我们进一步分析使用预训练词向量初始化的影响. 如图 3 所示, 相比较于使用  $(-1, 1)$  的均匀分布初始化词向量, 而使用训练语料初始化词向量的 Micro-F1 值大约提高了 24%.

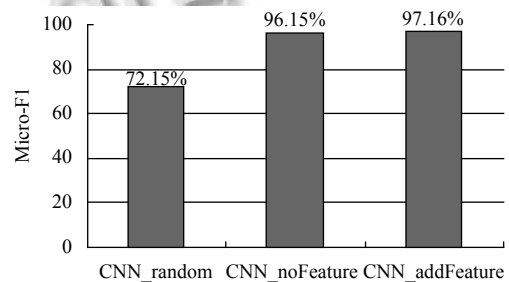


图 3 词向量不同对模型的影响

依据上述分析, 一方面, 基于特征的方法会很大程度地依赖特征提取的质量, 并且耗时耗力; 另一方面, 特征提取又需借助于 NLP 解析工具, 这也会带来解析工具已经存在的传播误差, 影响系统性能. 因此, Shallow CNN 和融合特征的 FCNN, 减少对先验的语言学知识和大量的手动提取的特征的依赖, 总体上要优于传统的机器学习的方法.

## 5 总结

本文中, 我们利用卷积神经网络模型学习了两个分句之间有效的关系表示, 进一步训练 Shallow CNN 分类器, 更好地处理复句的两个分句的语义关系, 并且该模型对复句的关系分类是简单高效的. 当加入关系词特征时, 系统的效果得了提升, 我们模型已远远优于仅依赖手动提取特征的模型. 自动学习特征能得到很好的结果, 可用于替代利用语言学知识和现有的 NLP 工具设计的特征.

在接下来的工作中, 我们将重点探索无标复句的关系类别识别. 并在依存句法树上构建神经网络模型, 融合更丰富的语法与结构特征<sup>[21]</sup>.

## 参考文献

- 1 Meyer T. Disambiguating temporal-contrastive discourse connectives for machine translation. Proceedings of the ACL 2011 Student Session. Stroudsburg, PA, USA. 2011. 46–51.
- 2 Huttunen S, Vihavainen A, Yangarber R. Relevance prediction in information extraction using discourse and lexical features. Proceedings of the 18th Nordic Conference of Computational Linguistics (NODALIDA 2011). Riga, Latvia. 2011. 114–121.
- 3 Sun MY, Chai JY. Discourse processing for context question answering based on linguistic knowledge. Knowledge-Based Systems, 2007, 20(6): 511–526. [doi: 10.1016/j.knosys.2007.04.005]
- 4 Mohler M, Bunescu R, Mihalcea R. Learning to grade short answer questions using semantic similarity measures and dependency graph alignments. Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA, USA. 2011. 752–762.
- 5 邢福义. 汉语复句研究. 北京: 商务印书馆, 2001.
- 6 Huang HH, Chang TW, Chen HY, *et al.* Interpretation of Chinese discourse connectives for explicit discourse relation recognition. Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers. Dublin, Ireland. 2014. 632–643.
- 7 胡金柱, 陈江曼, 杨进才, 等. 基于规则的连用关系标记的自动标识研究. 计算机科学, 2012, 39(7): 190–194. [doi: 10.3969/j.issn.1002-137X.2012.07.043]
- 8 胡金柱, 胡泉, 舒江波. 复句关系词自动识别中规则解析的包含匹配算法研究. 华中师范大学学报(自然科学版), 2014, 48(5): 643–649.
- 9 胡金柱, 舒江波, 胡泉, 等. 复句关系词自动识别中规则的表达方法研究. 计算机工程与应用, 2016, 52(1): 127–132. [doi: 10.3778/j.issn.1002-8331.1401-0214]
- 10 李艳翠, 孙静, 周国栋, 等. 基于清华汉语树库的复句关系词识别与分类研究. 北京大学学报(自然科学版), 2014, 50(1): 118–124.
- 11 周文翠, 袁春风. 并列复句的自动识别初探. 计算机应用研究, 2008, 25(3): 764–766. [doi: 10.3969/j.issn.1001-3695.2008.03.034]
- 12 杨进才, 陈忠忠, 沈显君, 等. 二句式非充盈态有标复句关系类别的自动标志. 计算机应用研究, 2017, 34(10): 2950–2953. [doi: 10.3969/j.issn.1001-3695.2017.10.016]
- 13 Huang HH, Chen HH. Contingency and comparison relation labeling and structure prediction in Chinese sentences. Proceedings of the 13th Annual Meeting of the Special Interest Group on Discourse and Dialogue. Stroudsburg, PA, USA. 2012. 261–269.
- 14 LeCun Y, Bottou L, Bengio Y, *et al.* Gradient-based learning applied to document recognition. Proceedings of the IEEE, 1998, 86(11): 2278–2324. [doi: 10.1109/5.726791]
- 15 Collobert R, Weston J, Bottou L, *et al.* Natural language processing (almost) from scratch. The Journal of Machine Learning Research, 2011, 12: 2493–2537.
- 16 Zeiler MD. ADADELTA: An adaptive learning rate method. arXiv: 1212.5701, 2012.
- 17 Rumelhart DE, Hinton GE, Williams R. Learning representations by back-propagating errors. Cognitive Modeling, 1988, 5(3): 1.
- 18 Caruana R, Lawrence S, Giles L. Overfitting in neural nets: Backpropagation, conjugate gradient, and early stopping. Proceedings of the 13th International Conference on Neural Information Processing Systems. Cambridge, UK. 2001. 381–387.
- 19 Srivastava N, Hinton G, Krizhevsky A, *et al.* Dropout: A simple way to prevent neural networks from overfitting. The Journal of Machine Learning Research, 2014, 15(1): 1929–1958.
- 20 Turian J, Ratinov L, Bengio Y. Word representations: A simple and general method for semi-supervised learning. Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA, USA. 2010. 384–394.
- 21 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.