

文章编号:1007-5321(2019)03-0114-06

DOI:10.13190/j.jbupt.2018-148

# 基于卷积神经网络的词义消歧

张春祥<sup>1</sup>, 赵凌云<sup>2</sup>, 高雪瑶<sup>2</sup>

(1. 哈尔滨理工大学 软件与微电子学院, 哈尔滨 150080; 2. 哈尔滨理工大学 计算机科学与技术学院, 哈尔滨 150080)

**摘要:** 为了提高词义消歧性能,提出了一种基于卷积神经网络的消歧方法. 以歧义词为中心,向左右两侧连续扩展4个邻接词汇单元,选取其中的词形、词性和语义类作为消歧特征. 以消歧特征为基础,使用卷积神经网络来确定歧义词的语义类别. 利用 SemEval-2007: Task#5 的训练语料和哈尔滨工业大学语义标注语料来优化卷积神经网络. 使用 SemEval-2007: Task#5 的测试语料来测试词义消歧分类器的性能,所提方法的消歧平均准确率有提高. 实验结果表明,该方法在词义消歧中是可行的.

**关键词:** 词义消歧; 卷积神经网络; 消歧特征; 语义类别

中图分类号: TP391.2

文献标志码: A

## Word Sense Disambiguation Based on Convolution Neural Network

ZHANG Chun-xiang<sup>1</sup>, ZHAO Ling-yun<sup>2</sup>, GAO Xue-yao<sup>2</sup>

(1. School of Software and Microelectronics, Harbin University of Science and Technology, Harbin 150080, China;

2. School of Computer Science and Technology, Harbin University of Science and Technology, Harbin 150080, China)

**Abstract:** In order to improve the performance of word sense disambiguation (WSD), a disambiguation method based on convolution neural network (CNN) is proposed. Ambiguous word is viewed as center and four adjacent word units around its left and right sides are extended. Word, part-of-speech and semantic categories are extracted as disambiguation features. Based on disambiguation features, CNN is used to determine semantic categories of ambiguous words. Training corpus of SemEval-2007: Task#5 and semantic annotation corpus from Harbin Institute of Technology are used to optimize CNN classifier. Testing corpus of SemEval-2007: Task#5 is used to test the performance of WSD classifier. Average disambiguation accuracy of the proposed method is improved. Experiments show that this method is feasible in WSD.

**Key words:** word sense disambiguation; convolution neural network; disambiguation features; semantic categories

机器翻译、自动文摘、信息检索和文本分类的性能与词义消歧密切相关. 杨陟卓、Koppula 和 Duque 等<sup>[1-3]</sup>将词义消歧转化为图求解问题. 杨安等<sup>[4]</sup>利用无标注文本构建词向量模型,结合特定领域的关

键词信息来进行词义消歧. 李赞等<sup>[5]</sup>研究了语义相关词之间的主题聚类特性,选择两两相关的多个词得到相关词集合,并计算平均语义相关度. Faisal 等<sup>[6]</sup>提出了一种基于支持向量机 (SVM, support

收稿日期: 2018-07-11

基金项目: 国家自然科学基金项目 (61502124, 60903082); 中国博士后科学基金项目 (2014M560249); 黑龙江省普通高校基本科研业务费专项资金项目 (LGYC2018JC014); 黑龙江省自然科学基金项目 (F2015041, F201420)

作者简介: 张春祥 (1974—), 男, 教授, 硕士生导师.

通信作者: 高雪瑶 (1979—), 女, 副教授, 硕士生导师, E-mail: xueyao\_gao@163.com.

vector machine)的词义消歧方法. 王桐等<sup>[7]</sup>给出了一种基于 WordNet 的概念语义相似度计算方法. 李源和何径舟等<sup>[8-9]</sup>使用最大熵模型来进行词义消歧,提高了消歧性能. Gutierrez 等<sup>[10]</sup>提出了一种个性化的 PageRank 算法,结合词义频率来处理词语歧义问题. Jimeno 和 Huang 等<sup>[11-12]</sup>利用神经网络来确定歧义词汇的语义类别. Tripodi 等<sup>[13]</sup>提出了一种基于博弈论的词义消歧方法,根据分布信息来计算语义相似度,解决了词语歧义问题. Iacobacci 和 Kang 等<sup>[14-15]</sup>将词嵌入应用于词义消歧过程中,提高了消歧性能. Zhan 和 Hun 等<sup>[16-17]</sup>从语义词典中挖掘语言学知识为词义消歧提供帮助.

提出了一种将歧义词左右邻接的 4 个词汇单元的词形、词性和语义类作为消歧特征,使用卷积神经网络进行消歧的方法.

### 1 词义消歧特征的提取

以歧义词为中心,从其左右 4 个邻接词汇单元中提取词形信息、词性信息和语义类信息作为消歧特征,来判别歧义词的真实含义. 从左右 4 个邻接词汇单元中,共抽取了 12 个消歧特征. 当歧义词左边或右边的邻接词汇单元数量不足 2 时,所对应位置的消歧特征值置为 NULL. 从而保证每个歧义词都有 12 个消歧特征.

对于包含歧义词“中医”的汉语句,其消歧特征的提取过程如下.

汉语句:政府大力提倡继续深化中医医疗机构的发展.

分词结果:政府 大力 提倡 继续 深化 中医 医 疗 机 构 的 发 展.

词性标注:政府/n 大力/d 提倡/v 继续/v 深化/v 中医/n 医疗/n 机构/n 的/u 发展/n . /wj

语义标注:政府/n/D 大力/d/K 提倡/v/H 继续/v/I 深化/v/I 中医/n/D 医疗/n/H 机构/n/D 的/u/K 发展/n/I . /wj/ -1

从汉语句中,提取歧义词左右 4 个邻接词汇单元,分别为“继续/v/I”、“深化/v/I”、“医疗/n/H”和“机构/n/D”. 词汇单元的词形用  $W$  来表示;词汇单元的词性用  $P$  来表示;词汇单元的语义类用  $S$  来表示. 从 4 个邻接词汇单元中,共抽取了 12 个消歧特征,分别为:  $W_{l2}$  = 继续、 $P_{l2}$  = v、 $S_{l2}$  = I、 $W_{l1}$  = 深化、 $P_{l1}$  = v、 $S_{l1}$  = I、 $W_{r1}$  = 医疗、 $P_{r1}$  = n、 $S_{r1}$  = H、 $W_{r2}$  = 机构、 $P_{r2}$  = n 和  $S_{r2}$  = D. 以哈尔滨工业大学

语义标注语料为基础,统计出消歧特征  $x$  的出现频度  $F(x)$ .

### 2 基于卷积神经网络的词义消歧

歧义词  $c$  有  $r$  种语义类  $s_1, s_2, \dots, s_r$ . 在包含歧义词  $c$  的句子中,一共提取了  $m$  个特征向量. 经过二进制转化后,每个向量对应于一个  $n$  位二进制数,从而构成了一个  $m \times n$  的二进制特征矩阵. 人工语义类别有  $r$  种,经过二进制转化后,每个语义类别对应一个  $r$  位二进制数.

对于歧义词  $c$ ,使用  $M = (F(W_{l2}), F(P_{l2}), F(S_{l2}), F(W_{l1}), F(P_{l1}), F(S_{l1}), F(W_{r1}), F(P_{r1}), F(S_{r1}), F(W_{r2}), F(P_{r2}), F(S_{r2}))$  来表示其消歧特征向量. 将频度  $F(x)$  转换成二进制数,用  $\text{binary}(F(x))$  来表示. 因此,  $v_1 = \text{binary}(F(W_{l2}))$ 、 $v_2 = \text{binary}(F(P_{l2}))$ 、 $v_3 = \text{binary}(F(S_{l2}))$ 、 $v_4 = \text{binary}(F(W_{l1}))$ 、 $v_5 = \text{binary}(F(P_{l1}))$ 、 $v_6 = \text{binary}(F(S_{l1}))$ 、 $v_7 = \text{binary}(F(W_{r1}))$ 、 $v_8 = \text{binary}(F(P_{r1}))$ 、 $v_9 = \text{binary}(F(S_{r1}))$ 、 $v_{10} = \text{binary}(F(W_{r2}))$ 、 $v_{11} = \text{binary}(F(P_{r2}))$ 、 $v_{12} = \text{binary}(F(S_{r2}))$ . 将  $v_1, v_2, v_3, v_4, v_5, v_6, v_7, v_8, v_9, v_{10}, v_{11}$  和  $v_{12}$  作为行,构造特征矩阵  $V = (v_1^T, v_2^T, v_3^T, v_4^T, v_5^T, v_6^T, v_7^T, v_8^T, v_9^T, v_{10}^T, v_{11}^T, v_{12}^T)^T$ .

在卷积层中,通常设置大小为  $h \times n$  的卷积核,其目的是提取更完整的消歧特征. 其中:  $h$  表示卷积核可以在多少个相邻消歧特征上滑动;  $n$  表示消歧特征向量的维数. 卷积核以步长 1 在特征矩阵  $V$  上滑动,每滑动一次就进行一次卷积操作,提取消歧特征,直至卷积核遍历完整个特征矩阵  $V$ . 特征矩阵  $V = (v_1^T, v_2^T, \dots, v_m^T)^T$ . 根据式(1)计算卷积所提取的特征值.

$$a_i = R(e) = R(w * v_{i:i+h-1}^T + b) \tag{1}$$

其中:  $w$  为卷积核,表示卷积层的权重矩阵;  $b$  表示偏置值;  $e$  表示卷积层的净激活;  $a_i$  表示特征映射中的第  $i$  个特征值;  $R(\cdot)$  表示激活函数,使用 Re-lu 激励函数,如

$$R(e) = \max(0, e) \tag{2}$$

当卷积核遍历完高度为  $m$  的特征矩阵  $V$  后,共提取了  $m - h + 1$  个特征值. 然后,构造特征向量  $Z$ ,如

$$Z = (a_1, a_2, \dots, a_{m-h+1})^T \tag{3}$$

针对歧义词“中医”而言,其特征矩阵的卷积操作过程如图 1 所示.

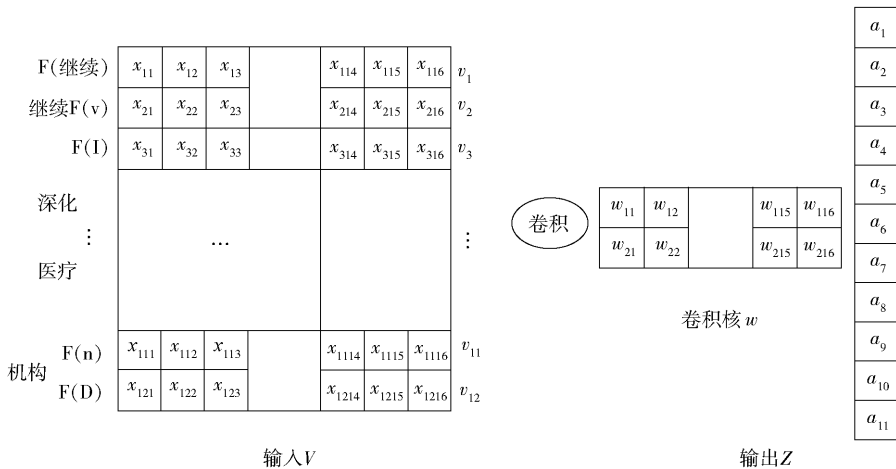


图1 卷积操作过程

歧义词“中医”的消歧特征向量为 $(F(\text{继续}), F(v), F(I), F(\text{深化}), F(v), F(I), F(\text{医疗}), F(n), F(H), F(\text{机构}), F(n), F(D))$ . 消歧特征的频度转换为一个16位二进制数,即 $v_i = (x_{i1}, \dots, x_{ij}, \dots, x_{i16}), i=1, 2, \dots, 12, x_{ij} \in \{0, 1\}$ . 构造二进制特征矩阵 $V = (x_{ij})_{12 \times 16}$ ,其中, $i=1, 2, \dots, 12, j=1, 2, \dots, 16$ . 设置卷积核为 $w_{2 \times 16}$ . 从 $V$ 中选取大小为 $2 \times 16$ 的子矩阵与卷积核 $w$ 进行卷积,即对应元素相乘再相加,所得到的特征值为特征向量 $Z$ 中的 $a_1$ . 卷积核以步长1进行滑动,再次从特征矩阵 $V$ 中选取大小为 $2 \times 16$ 的子矩阵与卷积核 $w$ 进行卷积,得到 $Z$ 中的 $a_i (i=2, 3, \dots, 11)$ . 在卷积操作之后,得到特征向量 $Z = (a_1, a_2, \dots, a_{11})$ .  $v_1$ 和 $v_2$ 构成的子矩阵 $(v_1^T, v_2^T)^T$ 经过卷积操作后得到 $a_1$ .  $v_i$ 和 $v_{i+1}$ 构成的子矩阵 $(v_i^T, v_{i+1}^T)^T$ 经过卷积操作后得到 $a_i (i=2, 3, \dots, 11)$ .

在池化层中,对输入的特征向量 $Z$ 进行下采样,提取最重要的特征. 采用最大池化方法获取特征向量 $Z$ 中的最大特征值,如

$$a' = \max(a_1, a_2, \dots, a_{m-h+1}) \quad (4)$$

对于不同的卷积核,获取的特征信息是不同的. 设置多个卷积核,提取多个不同的特征值. 若有 $t$ 个卷积核,则可以提取 $t$ 个最大特征值. 池化层的输出为 $D = (a'_1, a'_2, \dots, a'_t)$ . 其中 $a'_i$ 表示第 $i$ 个卷积核所对应的最大特征值.

利用softmax函数将多个神经元的输出映射到 $(0,1)$ 区间上,从而实现歧义词的语义分类. 其计算过程如

$$Q = f(Y) = f(WD + B) \quad (5)$$

其中: $Q = (P(s_1|c), \dots, P(s_i|c), \dots, P(s_r|c))$ 表

示输出; $Y = (y_1, y_2, \dots, y_i)$ 表示净激活; $W$ 和 $B$ 表示softmax层的参数; $f(\cdot)$ 为softmax函数; $P(s_i|c)$ 表示歧义词 $c$ 在语义类别 $s_i$ 下的出现概率,计算过程如

$$P(s_i|c) = \frac{\exp(y_i)}{\sum_{l=1}^r \exp(y_l)} \quad (6)$$

在歧义词 $c$ 的语义类概率分布 $P(s_1|c), \dots, P(s_i|c), \dots, P(s_r|c)$ 中,选取概率最大的语义类 $s$ 作为预测的语义类别,如式(7)所示.

$$s = \arg \max_{i=1,2,\dots,r} (P(s_i|c)) \quad (7)$$

为每个歧义词构造一个卷积神经网络. 在确定歧义词的语义类别时,使用所对应的卷积神经网络. 因此,歧义词 $c$ 所对应的卷积神经网络的输出层维度是固定的,为 $c$ 的语义类别数 $r$ .

歧义词“中医”的卷积神经网络结构如图2所示,主要包括输入层、卷积层、池化层和输出层.

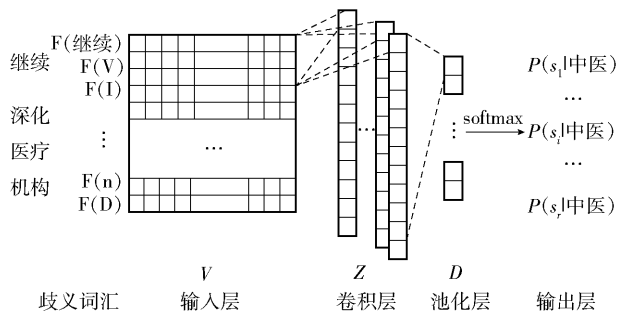


图2 卷积神经网络的结构

对包含歧义词“中医”的汉语句子进行分词、词性标注和语义类标注. 选取其左右邻接的4个词汇单元,即“继续/v/I”、“深化/v/I”、“医疗/n/H”和

“机构/n/D”. 抽取词形、词性和语义类,共得到 12 个消歧特征. 每个消歧特征对应于一个 16 位二进制数,形成特征矩阵  $V_{12 \times 16}$ . 设置多个  $2 \times 16$  的卷积核,经过卷积操作之后,得到特征向量  $Z$ . 对特征向量  $Z$  进行池化操作,得到一维向量  $D$ . 最后,利用 softmax 函数来计算歧义词“中医”在语义类别  $s_i$  下的概率分布  $P(s_i | \text{中医})$ ,其中,  $i = 1, 2, \dots, r$ . 具有最大概率的类别即为歧义词“中医”的语义类别.

### 3 基于卷积神经网络的词义消歧算法

采用 CNN 来确定歧义词语义类别的过程主要包括 CNN 模型训练和语义分类. CNN 模型的训练过程包括前向传播和反向传播. 语义分类就是前向传播过程. CNN 模型的训练过程如下.

输入: 包含歧义词  $c$  的训练样本的消歧特征向量  $M = (F(W_{L2}), F(P_{L2}), F(S_{L2}), F(W_{L1}), F(P_{L1}), F(S_{L1}), F(W_{R1}), F(P_{R1}), F(S_{R1}), F(W_{R2}), F(P_{R2}), F(S_{R2}))$ , 人工标注语义类  $hs$ ;

输出: 优化的参数  $\theta = \{w, b, W, B\}$ .

1) 初始化迭代次数  $k$  和参数集  $\theta = \{w, b, W, B\}$ ;

2) 将  $M$  转换为二进制特征矩阵  $V$ ;

3) 建立网络模型;

for( $t = 1$ ;  $t \leq k$ ;  $t++$ ) {

执行前向传播过程:

① 利用式(1)~(3)计算卷积层的输出特征向量  $Z$ .

② 根据式(4)得到池化层的最大特征值,多个特征值组成一维特征向量  $D$ .

③ 根据式(5)计算歧义词  $c$  在语义类别  $s_i$  下的概率分布  $P(s_i | c)$  ( $i = 1, 2, \dots, r$ ).

执行反向传播过程:

④ 根据式(8)计算误差. 其中:  $t_1, t_2, \dots, t_r$  为人工标注语义类  $hs$  的 One-hot 表示;  $\lambda$  表示正则化系数.

$$J(W, B) = - \sum_{i=1}^r t_i \log(P(s_i | c)) + \lambda \|W\|^2 \quad (8)$$

⑤ 根据式(9)计算输出层的灵敏度  $\gamma$ . 其中:  $\circ$  表示对应位置上的 2 个元素相乘.

$$\gamma = \frac{\partial J}{\partial Q} \circ f'(Y) \quad (9)$$

⑥ 根据输出层的灵敏度  $\gamma$ , 利用式(10)计算池

化层的灵敏度  $\beta$ , 其中 1 表示池化层激活函数的导数.

$$\beta = W^T \gamma \circ 1 \quad (10)$$

⑦ 根据池化层的灵敏度  $\beta$ , 利用式(11)计算卷积层的灵敏度  $\delta$ , 其中  $L(\cdot)$  表示上采样操作.

$$\delta = L(\beta) \circ R'(e) \quad (11)$$

⑧ 根据输出层的灵敏度  $\gamma$  和卷积层的灵敏度  $\delta$ , 利用式(12)~(15)更新参数  $\theta$ , 其中:  $a$  表示学习率,  $\text{rot180}(\cdot)$  表示旋转  $180^\circ$ .

$$W = W - \alpha(\gamma D^T) \quad (12)$$

$$B = B - \alpha \gamma \quad (13)$$

$$w = w - \alpha(\delta \text{rot180}(V)) \quad (14)$$

$$b = b - \alpha \sum_{i,j} (\delta[i,j]) \quad (15)$$

执行反向传播过程④~⑧来优化 CNN 的参数.

语义分类过程:

将包含歧义词  $c$  的测试样本的消歧特征向量转换为二进制数,并输入到 CNN 的卷积层. 使用优化后的参数  $\theta$  作为 CNN 的参数,根据式(1)、(2)、(3)得到卷积层的输出特征向量  $Z$ . 以  $Z$  为基础,设置多个卷积核,根据式(4)得到池化层的最大特征值,形成一维特征向量  $D$ . 根据式(5)计算歧义词  $c$  在语义类别  $s_i$  下的概率分布  $P(s_i | c)$ , 其中  $i = 1, 2, \dots, r$ . 利用式(7)确定歧义词  $c$  的语义类别.

### 4 实验分析

在实验中,使用了 SemEval-2007: Task#5 的训练语料和测试语料,比较了叶贝斯(Bayes)模型、深度信念网络(DBN, deep belief network)、SVM 和所提出方法的消歧效果,共进行了 4 组实验. 实验 1 选取歧义词左右邻接的 2 个词汇单元的词形作为消歧特征,采用 Bayes 模型来确定其语义类别. 实验 2 选取歧义词左右邻接的 4 个词汇单元的词形、词性和语义类作为消歧特征,利用 DBN 来确定其语义类别. 实验 3 选取歧义词左右邻接的 4 个词汇单元的词形、词性和语义类作为消歧特征,利用 SVM 来确定其语义类别. 实验 4 选取歧义词左右邻接的 4 个词汇单元的词形、词性和语义类作为消歧特征,利用 CNN 来确定其语义类别. 利用 SemEval-2007: Task#5 的训练语料来分别优化 Bayes 模型、DBN、SVM 和 CNN. 分别用优化后的 Bayes 模型、DBN、SVM 和 CNN 对 SemEval-2007: Task#5 的测试语料进行语义



分类.

采用 SemEval-2007 提供的平均准确率  $p_a$  来评价词义消歧效果,其计算过程如

$$p_i = \frac{m_i}{n_i}, \quad p_a = \frac{\sum_{i=1}^N p_i}{N} \quad (16)$$

其中: $N$  为所有目标歧义词的数目; $m_i$  为第  $i$  个歧义词正确分类的测试句子数; $n_i$  为包含第  $i$  个歧义词的所有测试句子数; $p_i$  为第  $i$  个歧义词的消歧准确率; $p_a$  为词义消歧的平均准确率.

在实验中,选取了 28 个具有代表性的歧义词进行测试,分别为“表面”、“菜”、“单位”、“动摇”、“儿女”、“镜头”、“开通”、“气息”、“气象”、“使”、“推翻”、“望”、“眼光”、“震惊”、“中医”、“本”、“补”、“成立”、“队伍”、“赶”、“旗帜”、“日子”、“天地”、“挑”、“长城”、“吃”、“动”和“叫”.测试语料的消歧准确率如表 1 所示.

从表 1 可以看出,实验 2、实验 3 和实验 4 的消歧平均准确率比实验 1 高,其原因是:在实验 1 中,仅采用了歧义词左右 2 个词汇单元的词形作为消歧特征.在实验 2、实验 3 和实验 4 中,采用了歧义词左右 4 个邻接词汇单元的词形、词性和语义类作为消歧特征.相对于词形特征而言,词性和语义类具有更强的语言现象覆盖能力和判别消歧能力.实验 3 的消歧平均准确率比实验 2 的要高.由此可知,SVM 的语义分类能力要好于 DBN.实验 4 的消歧平均准确率比实验 2 和实验 3 的都高.由此可知,相对于 DBN 和 SVM 而言,CNN 更适合于词义消歧问题.

根据表 1 画出的折线图来比较 4 组实验的消歧准确率,如图 3 所示.

从图 3 中可以看出:对于绝大多数歧义词而言,基于 CNN 的词义消歧分类器的性能要好于基于 Bayes 模型的词义消歧分类器、基于 DBN 的词义消歧分类器和基于 SVM 的词义消歧分类器.由此可

表 1 测试语料的消歧准确率				%
歧义词汇	实验 1 (Bayes)	实验 2 (DBN)	实验 3 (SVM)	实验 4 (CNN)
表面	50.0	61.1	44.4	82.3
菜	33.3	55.6	83.3	72.2
单位	47.1	58.8	58.8	82.3
动摇	75.0	62.5	68.7	93.7
儿女	50.0	70.0	35.0	94.9
镜头	60.0	53.3	53.3	53.3
开通	70.0	70.0	65.0	85.0
气息	64.3	71.4	71.4	64.2
气象	37.5	62.5	81.2	87.5
使	62.5	62.5	68.7	81.2
推翻	50.0	70.0	60.0	60.0
望	69.2	76.9	76.9	92.3
眼光	71.4	71.4	71.4	71.4
震惊	71.4	71.4	78.5	92.8
中医	37.5	62.5	62.5	87.5
本	68.0	60.0	52.0	98.9
补	40.0	50.0	65.0	64.9
成立	59.3	63.3	51.8	66.6
队伍	36.4	45.5	50.0	59.0
赶	27.8	50.0	50.0	55.5
旗帜	55.6	55.6	55.5	72.2
日子	46.9	46.9	50.0	50.0
天地	72.0	40.0	68.0	63.9
挑	50.0	42.9	42.8	57.1
长城	14.3	38.1	76.1	71.4
吃	60.9	43.5	47.8	52.1
动	60.0	50.0	50.0	50.0
叫	56.4	30.0	48.7	70.0
平均准确率	53.4	56.9	60.2	72.5

知,基于卷积神经网络的消歧方法更适合于词义消歧问题.

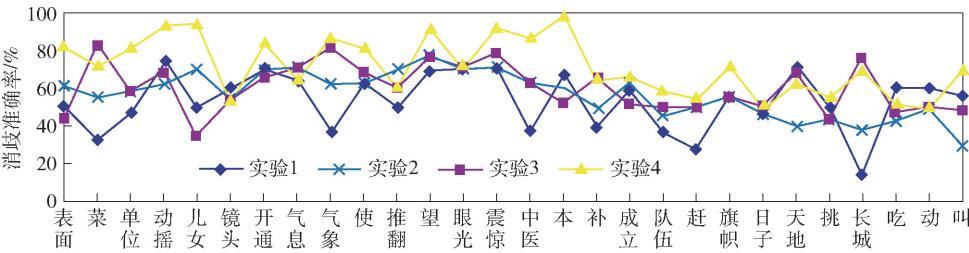


图 3 4 组实验的测试语料的消歧准确率

## 5 结束语

选取歧义词左右邻接的4个词汇单元的词形、词性和语义类作为消歧特征,利用卷积神经网络来确定其真实含义。使用SemEval-2007: Task#5的训练语料和哈尔滨工业大学语义标注语料来优化CNN分类器。实验结果表明,相对于贝叶斯模型、深度信念网络和支持向量机而言,所提方法的消歧性能有所提升。

## 参考文献:

- [1] 杨绍卓, 黄河燕. 基于异构关系网络图的词义消歧研究[J]. 计算机研究与发展, 2013, 50(2): 437-444.  
Yang Zhizhuo, Huang Heyan. WSD method based on heterogeneous relation graph[J]. Journal of Computer Research & Development, 2013, 50(2): 437-444.
- [2] Koppula N, Rani B P, Rao K S. Graph based word sense disambiguation[C] // The 1<sup>st</sup> International Conference on Computational Intelligence and Informatics. Hyderabad: Springer, 2017: 665-670.
- [3] Duque A, Martinezromo J, Araujo L. Can multilinguality improve biomedical word sense disambiguation? [J]. Journal of Biomedical Informatics, 2016, 64(1): 320-332.
- [4] 杨安, 李素建, 李芸. 基于领域知识和词向量的词义消歧方法[J]. 北京大学学报(自然科学版), 2017, 53(2): 204-210.  
Yang An, Li Sujian, Li Yun. Word sense disambiguation based on domain knowledge and word vector model[J]. Acta Scientiarum Naturalium Universitatis Pekinensis, 2017, 53(2): 204-210.
- [5] 李赞, 黄开妍, 任福继, 等. 维基百科的中文语义相关词获取及相关度分析计算[J]. 北京邮电大学学报, 2009, 32(3): 109-112.  
Li Yun, Huang Kaiyan, Ren Fuji, et al. Wikipedia based semantic related Chinese words exploring and relatedness computing[J]. Journal of Beijing University of Posts and Telecommunications, 2009, 32(3): 109-112.
- [6] Faisal E, Nurifan F, Sarno R. Word sense disambiguation in Bahasa Indonesia using SVM[C] // The 3<sup>rd</sup> International Seminar on Application for Technology of Information and Communication. Semarang: IEEE, 2018: 239-243.
- [7] 王桐, 王磊, 吴吉义, 等. WordNet 中的综合概念语义相似度计算方法[J]. 北京邮电大学学报, 2013, 36(2): 98-106.  
Wang Tong, Wang Lei, Wu Jiyi, et al. Semantic similar-

ity calculation method of comprehensive concept in WordNet[J]. Journal of Beijing University of Posts and Telecommunications, 2013, 36(2): 98-106.

- [8] 李源, 翟宏森, 刘凤娇, 等. 汉语复句中基于依存关系与最大熵模型的词义消歧方法研究[J]. 计算机与数字工程, 2018, 46(1): 78-82.  
Li Yuan, Zhai Hongsen, Liu Fengjiao, et al. Research on word sense disambiguation method based on dependency relation and maximum entropy in Chinese complex sentences[J]. Computer & Digital Engineering, 2018, 46(1): 78-82.
- [9] 何径舟, 王厚峰. 基于特征选择和最大熵模型的汉语词义消歧[J]. 软件学报, 2010, 21(6): 1287-1295.  
He Jingzhou, Wang Houfeng. Chinese word sense disambiguation based on maximum entropy model with feature selection[J]. Journal of Software, 2010, 21(6): 1287-1295.
- [10] Gutierrez Y, Vazquez S, Montoyo A. Spreading semantic information by word sense disambiguation [J]. Knowledge-Based Systems, 2017, 132(1): 47-61.
- [11] Jimeno Y A. Word embeddings and recurrent neural networks based on long-short term memory nodes in supervised biomedical word sense disambiguation[J]. Journal of Biomedical Informatics, 2017, 73(1): 137-147.
- [12] Huang Zhehuang, Chen Yidong. An improving SRL model with word sense information using an improved synergetic neural network model[J]. Journal of Intelligent & Fuzzy Systems, 2016, 31(3): 1469-1480.
- [13] Tripodi R, Pelillo M. A game-theoretic approach to word sense disambiguation[J]. Computational Linguistics, 2017, 43(1): 31-70.
- [14] Iacobacci I, Pilehvar M T, Navigli R. Embeddings for word sense disambiguation: an evaluation study[C] // The 54<sup>th</sup> Annual Meeting of the Association for Computational Linguistics. Berlin: ACL, 2016: 897-907.
- [15] Kang M Y, Kim B, Lee J S. Word sense disambiguation using embedded word space[J]. Journal of Computing Science and Engineering, 2017, 11(1): 32-38.
- [16] Zhan Jingwen, Chen Yanmin. Research of Chinese word sense disambiguation based on HowNet[C] // The 2011 International Conference on Artificial Intelligence and Computational Intelligence. Taiyuan: Springer, 2011: 477-482.
- [17] Hung C, Chen S J. Word sense disambiguation based sentiment lexicons for sentiment classification [J]. Knowledge-Based Systems, 2016, 110(1): 224-232.