

北京大学学报(自然科学版)  
Acta Scientiarum Naturalium Universitatis Pekinensis  
doi: 10.13209/j.0479-8023.2018.060

# 面向细粒度隐式篇章关系识别的远距离 监督特征学习算法

唐裕婷 李艳斌 刘露 于中华 陈黎<sup>†</sup>

四川大学计算机学院, 成都 610065; <sup>†</sup> 通信作者, E-mail: cl@scu.edu.cn

**摘要** 针对中文细粒度隐式篇章关系识别进行研究, 考虑细粒度篇章关系的方向性特点, 提出一种基于远距离监督的特征学习算法。该算法使用远距离监督的方法, 自动标注显式篇章数据, 然后利用词与连词之间的相对位置信息, 训练各个词的词表达, 将词的修辞功能以及关系的方向性编码到密集词表达中, 将这样的词表达应用到细粒度隐式篇章关系分类器。实验结果表明, 在细粒度隐式篇章关系识别任务中, 提出的方法的分类准确率达到 49.79%,  $F$  值达到 50.69%, 相比未考虑篇章关系方向性的方法有较大提高。

**关键词** 细粒度; 隐式篇章关系; 中文; 词表达; 方向性

## Feature Learning by Distant Supervision for Fine-Grained Implicit Discourse Relation Identification

TANG Yuting, LI Yanbin, LIU Lu, YU Zhonghua, CHEN Li<sup>†</sup>

Department of Computer Science, Sichuan University, Chengdu 610065; <sup>†</sup> Corresponding author, E-mail: cl@scu.edu.cn

**Abstract** Aiming at the identification of Chinese fine-grained implicit discourse relation and taking the directionality characteristic in account, the authors propose a feature learning algorithm based on the distant supervision to label explicit discourse data automatically. The relative position information between conjunction and words are applied to train the intensive word representation. Then the rhetorical function of words and the directionality of relations are encoded into the representation of intensive words, which will be applied to the relation classification of fine-grained implicit discourses. From the experimental studies of the proposed approach, the classification accuracy reaches 49.79% and F1-value reaches 50.69%, which are better than those approaches neglecting the directionality of discourse relations.

**Key words** fine-grained; implicit discourse relation; Chinese; word representation; directionality

作为自然语言处理(natural language processing, NLP)的重要任务之一, 篇章关系识别逐渐受到关注。篇章关系用于描述两个文本单元(子句、句子等)之间的逻辑衔接关系, 这两个文本单元称为基本篇章单元(elementary discourse unit, EDU)。自动识别两个 EDU 之间的篇章关系有助于自然语言处理的很多下游应用, 如问答系统<sup>[1]</sup>、自动文摘<sup>[2]</sup>、机器翻译<sup>[3]</sup>等。

根据篇章单元中是否存在连词, 篇章关系分为显式篇章关系(例 1)和隐式篇章关系(例 2)。由于具有隐式篇章关系的文本单元中不存在对篇章关系具有指示作用的连词, 使得隐式篇章关系识别成为篇章关系识别任务中的主要挑战。

**例 1** [因为]他今天生病了<sub>EDU1</sub>, [所以]没来上课<sub>EDU2</sub>。(显式因果关系)

四川省科技支撑项目(2014GZ0063)资助

收稿日期: 2018-04-15; 修回日期: 2018-08-20; 网络出版时间: 2018-08-22 18:27:59

**例 2** 他生病了 EDU1, 没来上课 EDU2。(隐式因果关系)

**例 3** 没来上课 EDU1, 他生病了 EDU2。

现有的研究通常将隐式篇章关系识别看成分类问题,且基于英文的研究居多。相比英文,中文省略连词的现象更为普遍<sup>[4]</sup>,但由于缺少规模较大的标准标注语料,因此针对中文的研究较少。多数基于中文的研究是围绕 CoNLL 发布的篇章分析 share tasks 来进行的。目前中文的隐式篇章关系识别主要是粗粒度的关系识别,还未见有关细粒度隐式篇章关系识别的研究报道。然而,明确定位每个文本单元的篇章语义角色会更有意义,例如,识别出具有“因果关系”的文本单元中,哪一个文本单元表示原因,哪一个文本单元表示结果,将更有利于问答系统、文本蕴含等任务的研究。本文将这种能表示文本单元的逻辑语义角色的篇章关系称为细粒度篇章关系。相比粗粒度篇章关系,细粒度篇章关系具有方向性,并且属于同种粗粒度关系下的不同细粒度关系具有语义差异(如“原因在前”和“证据在前”,二者都属于因果关系)。例 2 的两个文本单元 EDU1 和 EDU2 具有因果关系,且 EDU1 表示原因,EDU2 表示结果,方向由 EDU1 指向 EDU2。若交换 EDU1 和 EDU2 的位置(例 3),在粗粒度隐式篇章关系中,同样被视为因果关系,但实际上例 3 与例 2 并不相同。在例 3 中,EDU1 表示结果,EDU2 表示原因,方向由 EDU2 指向 EDU1。在细粒度隐式篇章关系中,例 2 被视为“原因在前”关系,例 3 则被视为“结果在前”关系,二者不应该被等同对待。因此,已有的粗粒度隐式篇章识别方法由于没有考虑方向性而不能直接用于细粒度隐式篇章关系识别。

本文针对中文的细粒度隐式篇章关系识别进行研究,借鉴 Braud 等<sup>[5]</sup>的思想,基于现有的标注数据较少的情况,利用远距离监督自动标注大量显式篇章数据,加入 EDU 中与连词的相对位置关系,训练词表达,将其用于细粒度隐式篇章关系分类并取得较好效果。实验结果表明,本文方法在 HIT-CDTB<sup>[6]</sup>的细粒度篇章关系数据上达到 49.79% 的准确率,与 Braud 等<sup>[5]</sup>的方法相比,准确率提高 4%, $F$  值提高 4%。

## 1 相关工作

英文隐式篇章关系识别主要基于宾州篇章树库(The Penn Discourse TreeBank, PDTB)进行粗粒度

的隐式关系识别<sup>[7-9]</sup>。中文的篇章关系标注语料主要是中文篇章树库(Chinese Discourse TreeBand, CDTB)<sup>[5]</sup>。Wang<sup>[10]</sup>采用特征工程的方法,人工提取了一些特征(如词对等),再利用逻辑斯蒂回归算法进行分类。Niko<sup>[11]</sup>提出并验证了,对于隐式篇章关系识别任务,人工构造大量语言学特征并不是必须的,可以使用神经网络自动学习特征。Rönqvist 等<sup>[12]</sup>使用循环神经网络(Recurrent Neural Network, RNN)来编码 EDU,并引入注意力机制,取得了较好的效果。

目前的隐式篇章关系识别研究主要着眼于粗粒度篇章关系识别,对更有应用价值的细粒度篇章关系识别缺乏关注。现有的方法中,人工构造特征,代价大且存在稀疏问题。虽然低维的密集的词表达可以有效解决稀疏问题,但通用词向量只编码了语义信息,与特定任务无关,且标注的数据过少,无法直接训练词表达。因此, Braud 等<sup>[5]</sup>提出,不直接把显式标注数据作为隐式篇章关系识别的训练数据<sup>[13]</sup>,而是用来训练词表达,通过对 EDU 中出现的词与连词的共现情况分析,可以获得不同的词语对不同篇章关系的使用分布情况,将词的修辞功能编码到密集的词表达中,实验证明这种方法相比简单的语义编码更有效。

方向性是细粒度篇章关系的独特性质,对细粒度隐式篇章关系的识别具有重要意义。在粗粒度隐式篇章关系识别中,通常的做法是通过词的向量表达获得篇章单元的向量表达,然后将两个篇章单元的向量表达进行拼接后送入分类器<sup>[5,14]</sup>。这样的方法在细粒度篇章关系识别任务上并不合理,因为篇章关系是有方向性的,如果底层编码不将方向性考虑进去,那么  $A \oplus B$  与  $B \oplus A$  (其中  $A$  和  $B$  表示两个句子向量,  $\oplus$  表示拼接操作)的编码是相同的,但它们在细粒度篇章关系中并不属于同种关系,这对上层分类器性能有负面影响。本文利用大量自动标注的显示篇章数据训练词表达,为了结合细粒度篇章关系的方向性特性,在训练词表达时,引入每个词相对于连词的位置信息。

## 2 模型方法

已公开的中文篇章关系标注语料中,哈尔滨工业大学标注的中文篇章关系语料 HIT-CDTB 提供了细粒度篇章关系标注。本文的语料使用 HIT-CDTB。

本文首先使用远距离监督方法自动标注的大

量显式篇章数据训练词表达,然后将训练的词表达作为细粒度隐式篇章分类器的输入,整体流程如图1所示。

### 2.1 自动标注显式篇章数据

高质量词表达的训练必须基于大量的显式篇章数据,由于 HIT-CDTB 语料中标注的样例较少(7910 条),因此本文需要自动标注大量显式篇章数据。由于训练的词表达是基于和连词的共现统计,因此在自动标注时只需要确定连词和边界,而不需要确定 EDU 间的具体关系。连词识别和 EDU 的确定作为自然语言处理的另一任务并非本文的研究重点,本文采用远距离监督的方式来进行自动标注。远距离监督在关系抽取<sup>[15]</sup>、微博情感分析<sup>[16]</sup>等任务中广泛使用。由于海量数据可以减小噪声的影响以及学习算法具有容噪能力,在这些任务中,将获得的海量文直接本作为训练数据,取得了较好效果。本文使用包含 409416986 个词的新闻语料,使用远距离监督进行自动标注,得到包含 264 个不同连词的共 7399347 个带噪声的显式篇章样例。

#### 2.1.1 确定连词

连词是篇章关系的重要指示线索,但对于连词,并没有一个具体的定义。在 HIT-CDTB 的标注准则中,由于没有限制连词的形式,导致连词中出现较长的短语以及“一九九三年”等确切的时间词,这些词作为连词的候选,被匹配到的可能性较小,且会扩大连词向量空间模型的维度,使得构建出的词表达维度过大。因此,在获取候选连词的时候,对从 HIT-CDTB 中获得的连词进行筛选: 1) 不考虑形如“在……情况下”等的连词; 2) 去除在 HIT-CDTB 中

出现次数为 1 的连词; 3) 去除具体时间。

由于本文基于和连词的共现情况来构造词表达,因此连词的个数决定了最后词表达的维度。连词越多,自动标注得到的显式数据就越多,映射到的关系类型也越多,能减少依赖连词的编码映射不到关系的情况,但是连词越多,也会导致词表达维度过大,这与我们要训练低维密集词表达的初衷违背,因此参考一般词向量的维度,选择 HIT-CDTB 中标注的筛选后的频度前 300 的连词,考虑到用于自动标注的语料可能没有某些候选连词的情况,考虑位置关系后,最后得到的词表达维度应该小于等于  $300 \times 2$ 。

采用远距离监督的方法,直接通过匹配候选连词得到的连词是有歧义的<sup>[17]</sup>,包括两个方面的歧义: 1) 该连词是否具有指示篇章关系的作用; 2) 连词与篇章关系并不是一一对应的。对于第二种歧义,由于本文的自动标注不需要确定到具体的关系,因此可以忽略。对于第一种,现有的工作多通过人工构造特征训练连词分类器<sup>[6,11]</sup>进行消歧。

由于中文连词的歧义现象占少数<sup>[18]</sup>,并且考虑到连词分类器准确率达到 100% 同样会带来噪声,因此为了简化标注过程,本文不对连词进行消歧。本文通过使用大量数据来弥补不进行连词分类带来的噪声。再者,出现在形如“blue and red”这样句子中的“and”,如果考虑其语义,将词语看作一种最基本篇章单元,此处的“and”同样表示两种颜色的并列关系。因此连词的歧义对于学习词与各个篇章关系的关系在语义上是没有影响的。

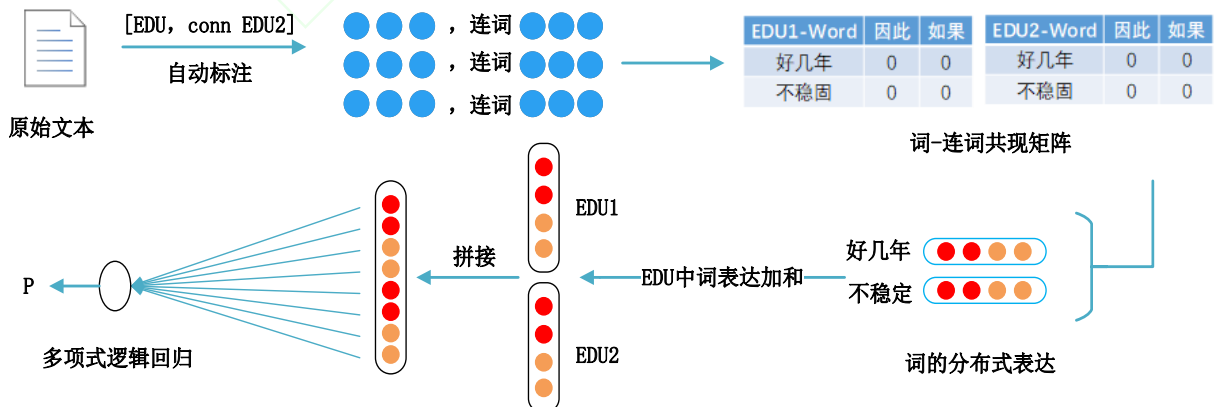


图 1 模型处理框架

Fig. 1 The processing framework of the proposed approach

### 2.1.2 确定 EDU 边界

作为浅层篇章分析的重要任务之一, EDU 边界确定相对较难<sup>[19]</sup>。Braud 等<sup>[5]</sup>先构造特征训练分类器, 定位篇章单元是构成句间关系还是句内关系, 然后通过启发式规则来确定 EDU 边界。Hooda 等<sup>[20]</sup>利用长短期记忆(long short term memory, LSTM)网络来识别 EDU (在 PDTB 里面, 将 EDU 叫做 Argument, 简称 Arg), 将 EDU 的识别看作序列标注问题, 使用 LSTM 捕获特征, 判断输入序列的每个词是属于 Arg1, Arg2, connective, 还是 None, 从而进行 Arg 的确定。由于该任务并不是本文的主要工作, 因此本文采用简单的模板<sup>[21]</sup>, 形如[EDU1, connectives EDU2]来确定 EDU, 即两个 EDU 在同一个句子中。Wu 等<sup>[21]</sup>的研究表明, 在数据量足够的情况下, 基于这样简单的模板是足够的。此外, 本文对海量数据的标注并不需要确定具体关系, EDU 的边界确定是为了限制连词与词的相关应该在某一范围内, 本文将这一范围限制在同一个句子中。基于海量数据可以涵盖大多数词与连词的假设, 使用该模板得到的海量文本对于学习词与连词的关系是足够的。

### 2.2 构建与连词有关的词向量表达

连词向量空间模型是标准向量空间模型的变体<sup>[22]</sup>, 不同连词的个数确定了底层向量空间的维度, 通过统计每个词在不同文本单元中与不同连词的共现情况, 将每个词映射为连词个数维的实值向量。

在细粒度隐式篇章关系识别中, 要考虑篇章关系的方向性, 不能直接应用粗粒度隐式篇章关系识别的方法。如例 4(a) 中, EDU1 是因, EDU2 是果, 关系由 EDU1 指向 EDU2, 而 4(b) 中 EDU1 是果, EDU2 是因, 关系由 EDU2 指向 EDU1。

**例 4(a)** 出生之后小仙女的呼吸平顺 EDU1, 并不需要额外的氧气 EDU2。

**例 4(b)** 今天女足和朝鲜的比赛着实让我激动了一把 EDU1, 比赛之精彩让我出乎意料 EDU2。

因此, 基于上述观察, 假设文本中每个词在不同 EDU 中与连词的共现情况可以反映一个词在关系中的方向偏好, 即若某些词与“因为”共现, 出现在 EDU2 中, 与“所以”共现, 常出现在 EDU1, 那么这些词都常常用于表示原因, 在关系方向中处于指向的那一方。使得这些词的词表达不仅编码了词与因果关系的语义联系, 同样编码了篇章关系的方向性。

本文构造词表达的方法如下。

对 EDU1 和 EDU2 中的词分别构建一个与连词的共现矩阵; 矩阵的每个元素表示对应的文本中其他词和连词的共现权重;  $V_1$  表示出现在 EDU1 中的  $n_1$  个词的集合, 用  $V_2$  表示出现在 EDU2 中的  $n_2$  个词的集合;  $C$  表示  $m$  个连词的集合, 得到构建的两个矩阵  $F_1$  (维度为  $n_1 \times m$ ) 和  $F_2$  (维度为  $n_2 \times m$ )。  $F_1$  表示出现在 EDU1 中的不同词与各个连词的共现情况,  $F_2$  表示出现在 EDU2 中的不同词与各个连词的共现情况, 因此同一个词就可以根据  $F_1$  得到  $m$  维向量表达  $\mathbf{vector}_i$ , 向量的每一维表示该词出现在 EDU1 中与连词的关联; 根据  $F_2$  得到  $m$  维向量表达  $\mathbf{vector}_r$ , 向量的每一维表示该词出现在 EDU2 中与连词的关联。对任一个词, 就可以表示为一个  $2m$  维的向量  $[\mathbf{vector}_i, \mathbf{vector}_r]$ 。共现权重的计算公式见式(1)~(3), 采用文献[5]中效果最好的 PPMI-IDF, 为了减少点互信息对低频词的偏向性, 计算时去掉词频小于等于 5 的词。

$$PPMI_{i,j} = \max(0, \log\left(\frac{P_{i,j}}{P_{i,*}P_{*,j}}\right)), \quad (1)$$

$$IDF_i = \log\left(\frac{m}{\sum_{k=1}^m f_{i,k}}\right), \quad (2)$$

$$weight_{i,j} = PPMI_{i,j} * IDF_i, \quad (3)$$

其中,  $weight_{i,j}$  表示矩阵的第  $i$  行第  $j$  列对应的值, 即词  $w_i$  和连词  $c_j$  的共现权重。  $PPMI_{i,j}$  表示  $w_i$  和  $c_j$  的非负点互信息值(PMI 小于 0 时, 取值为 0), 用于衡量词  $w_i$  和连词  $c_j$  的相关度, 从而映射到和特定关系的相关度。计算见式(1), 其中  $p_{i,j}$  表示  $w_i$  和  $c_j$  共同出现在一个 EDU 中的概率,  $p_{i,*}$  和  $p_{*,j}$  分别表示  $w_i$  和  $c_j$  出现的概率。  $IDF_i$  用来规范化  $PPMI_{i,j}$  的值, 计算见式(2), 其中  $m$  表示不同连词数,  $f_{i,j}$  表示  $w_i$  和  $c$  共同出现在一个 EDU 中的频数, 如表 1 和 2 所示, 得到两个共现矩阵(部分)。

表 1 连词与词(EDU1)PPMI-IDF 关联矩阵示例

Table 1 Illustrative example of PPMI-IDF association matrix between connectives and words of EDU1

词	连词			
	因此	如果	后来	但是
好几年	0.12	0	1.67	0.34
不稳固	0.35	2.08	0	0.20
真的	0.03	0	0	0.01
,	0	0	0	0

表 2 连词与词(EDU2)PPMI-IDF 关联矩阵示例

Table 2 Illustrative example of PPMI-IDF association matrix between connectives and words of EDU2

词	连词			
	因此	如果	后来	但是
好几年	0	0	0.04	0.60
不稳固	0	0	0	1.82
真的	0	0.16	0	0.03
,	0	0	0	0

通过拼接表 1 和表 2 得到的词表达, 每个词就可以表示为 1 个  $2m$  维的向量。如表 1 和 2 所示, 假设总共出现的连词只有 4 个, 则“好几年”可被表示为  $[0.12, 0, 0.47, 0.34, 0, 0, 0.04, 0.60]$  这样一个 8 维向量, 同样“不稳固”可表示为  $[0.35, 2.08, 0, 0.20, 0, 0, 0, 1.82]$ 。可以看到, 像“好几年”这种表时间的词, 更倾向于出现在以“后来”驱动的表时序关系中, 且更可能出现在 EDU1 中, 与“之后”构成先序关系。而“,”这种词, 基本与所有连词共现, 对应在各个维度上的值都是 0, 对我们的模型将不会产生影响。

### 2.3 使用词向量表达

得到每个词的词表达后, 通过加和<sup>[9]</sup>EDU 中所有词的词表达, 得到每个 EDU 的向量表达为  $2m$  维。然后, 将两个 EDU 的向量表达进行拼接, 得到  $4m$  的向量, 作为隐式篇章分类器的输入。为了回避句子长度不同可能带来的影响, 在加和形成每个 EDU 的词向量后用  $L2$  范数进行规范化。

## 3 实验与结果

### 3.1 数据集

HIT-CDTB 是哈尔滨工业大学标注的篇章关系语料, 本文只使用该语料中的分句关系和复句关系样例。其中, 隐式样例 15401 个, 提供了细粒度关系标签的隐式样例 4084 个, 包含共 22 种细粒度关系, 本实验只选择样例数大于 10 的关系类别作为最后的分类标签。该标注语料中存在某些样例缺失对应 EDU 的情况, 在删除缺失样例后, 共得到 3663 个样例, 具体见表 3。

由表 3 可见, 数据存在不均衡现象, 这是由篇章关系的真实分布决定的, 已知的篇章关系的分布是偏斜的, 扩展关系(表 3 中“解释说明”关系是扩展关系的一种)的样例会比较多, 对于每种关系使用相同数量的样例来训练分类器可能会导致错误<sup>[23]</sup>, 因此本文不对样本分布做干涉, 使其保持真实分

布。

表 3 HIT-CDTB 细粒度隐式篇章样例分布

Table 3 Distribution of fine-grained implicit discourse example in HIT-CDTB

细粒度篇章关系	样例数
正向对比	33
证据在先	54
推论在先	34
解释说明	1989
结果在先	154
反向对比	55
实例在后	365
原因在先	724
实例在先	20
相关条件在先	26
先序	154
目的在后	55
总计	3663

### 3.2 实验结果

本文使用多项式逻辑斯蒂来建模分类器, 并与 Braud 等<sup>[5]</sup>没有考虑位置的方法进行对比, 实验结果见表 4 和 5。

通过表 4, 本文在细粒度隐式篇章关系识别任务上达到较好效果, 准确率以及  $F$  值较 Braud 等<sup>[5]</sup>的方法都有很大的提高, 说明在构建词表达时, 简单的将词在不同文本单元中的出现情况考虑进去, 对于细粒度隐式篇章关系识别的有效性。由表 5 可知, 在多数关系识别上, 本文方法较 Braud 等<sup>[5]</sup>的方法有所提高, 但对于样例较少的关系类别, 分类效果并不理想。

表 4 细粒度隐式篇章关系识别结果

Table 4 Result of fine-grained implicit discourse relation classification

方法	Acc/%	F1/%
Braud 2016	46.00	46.20
本文方法	49.79	50.69

表 5 每个关系识别结果

Table 5 Result of per relation identification

关系	P		R		F	
	Braud 2016	本文方法	Braud 2016	本文方法	Braud 2016	本文方法
正向对比	0.468	0.6	0.258	0.224	0.332	0.294
证据在先	0.15	0.2	0.036	0.018	0.056	0.034
推论在先	0	0	0	0	0	0
解释说明	0.6	0.608	0.67	0.724	0.63	<b>0.658</b>
结果在先	0.05	0.1	0.048	0.042	0.048	0.06
反向对比	0.12	0.034	0.082	0.024	0.094	0.028
实例在后	0.226	0.234	0.216	0.234	0.222	<b>0.234</b>
原因在先	0.334	0.362	0.308	0.36	0.322	<b>0.358</b>
实例在先	0.2	0.1	0.024	0	0.044	0.066
相关条件在先	0	0	0	0	0	0
先序	0.2	0.304	0.212	0.168	0.202	0.202
目的在后	0	0.184	0	0.062	0	0.092

为了进一步分析分析原因, 本文进行 5 折交叉验证, 发现对于样例较少的关系类别的预测精确率抖动较大, 可能因为样例个数太少, 分类器并未能进行有效学习, 导致结果不好, 存在随机性, 如图 2 所示, 其中预测精确率始终为 0 的关系不显示。本文对于预测精确率始终为 0 的关系进行考察, 除了该关系的样本数较少外, 在观察了分类混淆矩阵后, 发现“证据在前”关系会被很大可能错分在“原因在前”的类别, “推论在前”关系会被很大可能错分在“解释说明”关系类别, 这些关系都存在语义上的相似性, 当类别细化后, 连词到关系的映射就会改变,

一个连词映射的关系会增多, 因为连词对关系的映射是没有方向性的, 就会使得属于同种粗粒度的细粒度关系间学到的特征可能会比较相似, 再加上数据的不均衡, 从而导致误分类, 但这是由于细粒度篇章关系的语义差异引起的, 并不是本文的研究重点, 因此本文并未对该语义差异做特殊处理。忽略语义差异带来的错误, 本文考虑细粒度篇章关系的方向性, 在细粒度篇章关系识别任务中取得了较好效果。总体而言, 本文初次对中文细粒度隐式篇章关系进行研究, 取得了有意义的结果。

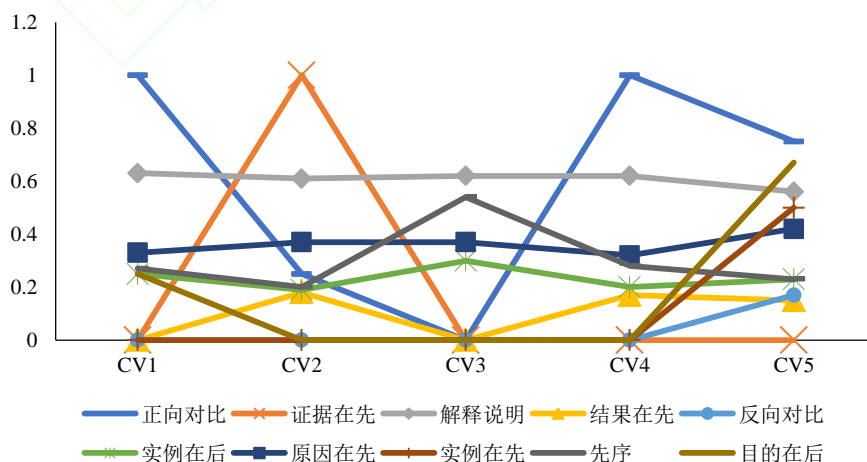


图 2 五折交叉验证各个类别精确率

Fig. 2 The precision of 5-fold cross validation in per relation

## 4 结语

本文针对细粒度隐式篇章关系的识别任务, 基于篇章关系标注语料较少的现状, 利用大量自动标注的连词和与之对应的EDU, 采用基于统计的方法训练低维密集词表达, 利用词与连词的共现关系使得到的词表达能对自身的修辞功能进行编码, 加入词于连词的相对位置信息, 将每个词的关系语义角色偏向性编码到词表达。然后将该词表达用于细粒度隐式篇章关系识别, 取得了较好效果。本文训练词表达的方法高效简单。但点互信息并不能很好的衡量两个词之间的相似性, 且倾向于低频词, 在将来的工作中, 可以考虑使用其他更有效的衡量方式。其次, 本文作为初次尝试, 考虑细粒度篇章关系的方向性时, 采取的方法较简单, 后期可以进一步改善。

### 参考文献

- [1] Sun M, Chai J Y. Discourse processing for context question answering based on linguistic knowledge. *Knowledge-Based Systems*, 2007, 20(6): 511–526
- [2] Yoshida Y, Suzuki J, Hirao T, et al. Dependency-based Discourse Parser for Single-Document Summarization // *Conference on Empirical Methods in Natural Language Processing*. 2014: 1834–1839
- [3] Li J J, Carpuat M, Nenkova A. Assessing the Discourse Factors that Influence the Quality of Machine Translation // *Meeting of the Association for Computational Linguistics*. 2014: 283–288
- [4] Zhou Y, Xue N. PDTB-style discourse annotation of Chinese text // *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers-Volume 1*. Association for Computational Linguistics, 2012: 69–77
- [5] Braud C, Denis P. Learning connective-based word representations for implicit discourse relation identification // *Empirical Methods on Natural Language Processing*. 2016
- [6] 张牧宇, 秦兵, 刘挺. 中文篇章级句间语义关系体系及标注. *中文信息学报*, 2014, 28(2): 28–36
- [7] Lin Z, Kan M Y, Ng H T. Recognizing implicit discourse relations in the Penn Discourse Treebank // *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1-Volume 1*. Association for Computational Linguistics, 2009: 343–351
- [8] Ji Y, Eisenstein J. One vector is not enough: Entity-augmented distributional semantics for discourse relations. *arXiv preprint arXiv:1411.6699*, 2014
- [9] Rutherford A, Demberg V, Xue N. A Systematic Study of Neural Discourse Models for Implicit Discourse Relation // *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*. 2017, 1: 281–291
- [10] Wang J, Lan M. Two end-to-end shallow discourse parsers for English and Chinese in conll-2016 shared task. *Proceedings of the CoNLL-16 shared task*, 2016: 33–40
- [11] Schenk N, Chiarcos C, Donandt K, et al. Do we really need all those rich linguistic features? a neural network-based approach to implicit sense labeling. *Proceedings of the CoNLL-16 shared task*, 2016: 41–49
- [12] Rönqvist S, Schenk N, Chiarcos C. A Recurrent Neural Model with Attention for the Recognition of Chinese Implicit Discourse Relations. 2017: 256–262
- [13] Sporleder C, Lascarides A. Using automatically labelled examples to classify rhetorical relations: An assessment. *Natural Language Engineering*, 2008, 14(3): 369–416
- [14] Qin L, Zhang Z, Zhao H. A stacking gated neural architecture for implicit discourse relation classification // *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*. 2016: 2263–2270
- [15] Angeli G, Premkumar M J J, Manning C D. Leveraging linguistic structure for open domain information extraction // *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. 2015, 1: 344–354
- [16] Magdy W, Sajjad H, El-Ganainy T, et al. Distant Supervision for Tweet Classification Using YouTube

- Labels // ICWSM. 2015: 638–641
- [17] Marcu D. The rhetorical parsing of unrestricted texts: A surface-based approach[J]. *Computational linguistics*, 2000, 26(3): 395–448
- [18] 姬建辉, 张牧宇, 秦兵, 等. 中文篇章级句间关系自动分析. *江西师范大学学报(自然版)*, 2015(2): 124–131
- [19] Xue N, Ng H T, Pradhan S, et al. The conll-2015 shared task on shallow discourse parsing // *Proceedings of the Nineteenth Conference on Computational Natural Language Learning-Shared Task*. 2015: 1–16
- [20] Hooda S, Kosseim L. Argument Labeling of Explicit Discourse Relations using LSTM Neural Networks[J]. *arXiv preprint arXiv:1708.03425*, 2017
- [21] Wu C, Shi X, Chen Y, et al. Improving Implicit Discourse Relation Recognition with Discourse-specific Word Embeddings // *Meeting of the Association for Computational Linguistics*. 2017: 269–274
- [22] Turney P D, Pantel P. From frequency to meaning: Vector space models of semantics[J]. *Journal of artificial intelligence research*, 2010, 37: 141–188
- [23] Pitler E, Louis A, Nenkova A. Automatic sense prediction for implicit discourse relations in text // *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 2-Volume 2*. Association for Computational Linguistics, 2009: 683–691