汉语隐式篇章关系识别

孙静 李艳翠 周国栋 冯文贺

引言

- 篇章关系是指同一篇章内部，相邻片段或跨度在一定范围内的两个片段之间的语义连接关系，如条件关系、转折关系、因果关系等

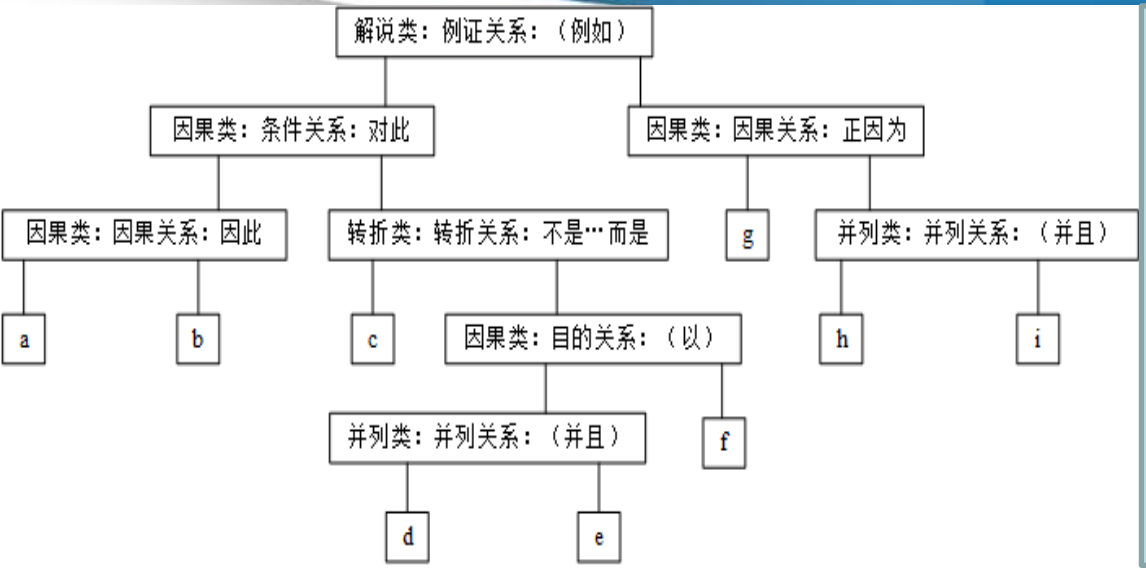
- 篇章关系识别可以分为显式关系识别和隐式关系识别
  - 例1：浦东开发开放是一项振兴上海，建设现代化经济、贸易、金融中心的跨世纪工程，因此大量出现的是以前不曾遇到过的新情况、新问题。
  - 例2：他在两起汽车走私案中触犯刑律，[因此]构成走私罪。

汉语语料库

- 自建的汉语语料库 (Chinese Discourse Treebank，简称CNDB)，CNDB采用树的形式表示汉语的篇章结构，每一段落构建一棵篇章结构树。

```
<P ID="3">
<R ID="2" …… ConnectiveType="隐式关系"…… Connective="例　　如"
RelationType="例证关系" ……ConnectiveAttribute="可添加
" ……Sentence="浦东开发开放……新情况、新问题。对此，浦东不是
简单的采取……纳入法制轨道。|去年初……没有发现一例回扣。"
SentencePosition="1…167|168…230" ChildList="3|8" ……/>
<R ID="3" …… ConnectiveType="显式关系"…… Connective="对此"
RelationType="条件关系" ……ConnectiveAttribute="不可删除" ……
Sentence="浦东开发开放……新情况、新问题。|对此，浦东不是……
纳入法制轨道。" SentencePosition="1…60|61…167"
ChildList="4|5"……/>
……
</P>
```

汉语语料库



标注的CTB6.0中的158个文档(chtb001-chtb0130，chtb0211-chtb0240)，739 段，每段平均3个句子。总共有3398个子句 (叶子节点)， 2331非终节点 (一个非终节点表示一个关系，二元或多元关系)。在CNDB 中，显式关系有453个，隐式关系有1878个，达到了80%

第一层	第二层	第三层	训练实例 (所占比重%)	测试实例 (所占比重%)
因果类	因果关系；推断关系；假设关系；目的关系；条件关系；背景关系	因此，所以，因为，由于…所以，如果…那么	297(18.10)	67(28.27)
转折类	转折关系；让步关系	虽然…但，即使...也	17(1.04)	7(2.95)
并列类	并列关系；顺承关系；递进关系；选择关系；对比关系	同时，并，并且，然后	979(59.66)	123(51.90)
解说类	解说关系；总分关系；例证关系；评价关系	总之，例如	348(21.2)	40(16.88)
总和			1,641	237

特征提取

- 上下文特征(Fcon)
  - 完全嵌入论元模式
  - 共享论元模式
- 词汇特征
  - 词对特征(FWP)
  - 词，词性(FVwp)
- 依存树特征(Fdep)
  - 从每个论元对应的依存树中选择所有拥有被支配者的词和依存类型

实验结果

表2 单个特征及所有特征的总正确率

Feature	Word&pos	Dependency rules	Wordpairs	Context	Acc.(%)
FVwp	+	-	-	-	54.17
FDep	-	+	-	-	54.51
FWP	-	-	+	-	56.25
FCon	-	-	-	+	60.76
FAll	+	+	+	+	62.15

表3 不同特征组的总正确率

Feature	Word&pos	Wordpairs	Dependency rules	Context	Acc.(%)
FVwp	+	-	-	-	54.17
FWP	+	+	-	-	58.68
FDep	+	+	+	-	61.11
FCon	+	+	+	+	62.15

实验结果

表4 4大类别Precision,Recall和F1-measure,"--"代表0.00

类别	Precision-%	Recall-%	F1-measure-%
因果类	50.0	5.88	10.34
并列类	62.78	95.43	75.26
解说类	47.06	19.51	27.18
转折类	--	--	--
All(微平均)	39.96	30.20	28.20

结束语

- 研究了基于汉语篇章语料库(CNDB)中的4大类别(因果类、并列类、解说类和转折类)的隐式篇章关系识别问题
- 自建的汉语篇章语料库(CNDB)，应用词汇特征，上下文特征和依存树特征对隐式关系进行了简单探讨
- 隐式关系占有很高的比重，又因为隐式关系中缺乏连接词，都导致汉语隐式篇章关系识别困难，具有很大的挑战性

谢谢！