

关于期权的基础知识及套利过程

曾铨煜

资管研发四部

2017 年 7 月 21 日

大纲

- ① 概述
- ② 中文开放式关系抽取实现方式
- ③ 实验
- ④ 结果分析
- ⑤ 错误分析

大纲

- ① 概述
- ② 中文开放式关系抽取实现方式
- ③ 实验
- ④ 结果分析
- ⑤ 错误分析

传统关系抽取

- ⊗ 优点：提取关系基于匹配规则，速度快
- ⊗ 缺点：规则需要预先定义，适用范围窄

开放式关系抽取

- ⊗ 优点：应用范围广，只需要语料库作为输入数据即可进行处理
- ⊗ 缺点：初始信息少、语义类别难确定、依赖于训练语料

传统关系抽取

- ⊗ 优点：提取关系基于匹配规则，速度快
- ⊗ 缺点：规则需要预先定义，适用范围窄

开放式关系抽取

- ⊗ 优点：应用范围广，只需要语料库作为输入数据即可进行处理
- ⊗ 缺点：初始信息少、语义类别难确定、依赖于训练语料

大纲

- ① 概述
- ② 中文开放式关系抽取实现方式
- ③ 实验
- ④ 结果分析
- ⑤ 错误分析

分词与依存句法分析

提取候选关系

大纲

- ① 概述
- ② 中文开放式关系抽取实现方式
- ③ 实验
- ④ 结果分析
- ⑤ 错误分析

评估标准

精确率

$$Precision = \frac{|T|}{|A|} \times 100\%$$

召回率

$$Recall = \frac{|T|}{|S|} \times 100\%$$

F1 值

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

评估标准

精确率

$$Precision = \frac{|T|}{|A|} \times 100\%$$

召回率

$$Recall = \frac{|T|}{|S|} \times 100\%$$

F1 值

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

评估标准

精确率

$$Precision = \frac{|T|}{|A|} \times 100\%$$

召回率

$$Recall = \frac{|T|}{|S|} \times 100\%$$

F1 值

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

人工标注关系文件

- ⊛ 中文文本语句内的关系数量
- ⊛ 谓词短语
- ⊛ 论元内容
- ⊛ 论元数目

评估程序

- ① 以语义模式重复频数阈值 t^{sem} 作为自变量，运行中文开放式关系抽取系统处理语料库
- ② 按语句将机器提取的关系与人工标注的关系进行比对并统计、输出数据

大纲

- ① 概述
- ② 中文开放式关系抽取实现方式
- ③ 实验
- ④ 结果分析
- ⑤ 错误分析

ZORE 基于“燃规 3”语料库的运行结果

结论

体现了先前提及的对召回率 Recall 与精确度 Precision 的平衡性权衡，F1 值最大处关系抽取系统性能最好。

ZORE 与 DPM 基于“燃规 3”语料库的运行结果对比

结论

推断是由于该语料库是法规，文本结构较为碎片化，以上下文环境进行分析的兼类词处理难以顺利工作导致的。

ZORE 与 DPM 基于“燃规 3”语料库的运行结果对比

结论

可以推断缺少对兼类词处理不是 ZORE 算法提升性能所遇到的瓶颈。

大纲

- ① 概述
- ② 中文开放式关系抽取实现方式
- ③ 实验
- ④ 结果分析
- ⑤ 错误分析

错误分析

- ① t^{sem} 语义模式最低频率限制导致遗漏正确关系
- ② 自然语言处理带来的错误，如将论元识别为“、构筑物”而不是“建、构筑物”
- ③ 关系模式选择错误，导致抽取出错误的关系

谢谢!