

# 增强因果向量

## 目的

- 充分利用low precision线索词,在保证precision的情况下,提升recall
- 提升语料的coverage,同时得到pattern表示

## 做法

- 使用因果向量,找到因果词对,通过因果词对找到因果短语对
- 因果短语对返回语料,寻找共现上下文(在语法树上的距离较近)
- 同一个短语对可能出现在多种上下文中,给这些上下文分配权重
- 使用pattern,抽取短语对,训练因果向量(约束: 使其结果与已有因果向量相似)

## pattern表示

用串表示,短语对用占位符,(c,e,\$p)三元组

最终 先做动词,看看效果如何,之后再推广到连词上

## 因果词对

利用训练好的因果向量,识别抽取的因果短语对中的因果词对

## 策略

- 过滤低频词: 过滤在因果词表中不存在的词
- 过滤长度大于4的短语: 当短语较长时,预示因果关系是句子级别的,无法简单的在词级别体现; 所以过滤这些短语对,提高识别出的因果词对的精度
- 过滤单字: 在中文里,词语一般以多字出现,同时也能过滤分词错误, 如: "处理 不 当----外交 风波"
- 计算词对因果打分,选取打分最高的词对
- 检查这个词对的词性,约束为[动词,形容词,名词];如果不是,则舍弃

## 预测结果如下

- 19女性接触农药会不会引起不孕症----19 女性 接触 农药----不孕症----m n v n----n----引起 ['接触', '不孕症']
- 跟进服务在以后的合作中我们会根据客户的特点和要求, 不断调整服务方式, 改善服务质量, 使客户满意----改善 服务 质量----客户 满意----v v n----n v----使 ['服务', '满意']
- 张志聪果然与众不同, 他以防风、苏叶、杏仁各药等分为剂, 水煎后温服, 使病人出汗, 小便即通, 水肿全消----水 煎 后 温服----病人 出 汗----n v n d v----n v n----使 ['温服', '病人']
- 长期的漂泊生活, 使桑地诺积累了见识, 练就了机警沉静、深谋远虑的风采----长期 的 漂 泊 生活----桑地诺 积 累 了 见 识----b u v v----n h v u n----使 ['漂泊', '见识']
- 违反以上规定造成损害后果的, 由擅自通过或施工方承担法律责任----违反 以上 规定----损 害 后果----v n d n----v n----造成 ['违反', '后果']

- 棒颈球棒的颈部弯曲，使脚后跟稍微偏离杆身线---棒颈 球棒 的 颈部 弯曲---脚后跟 稍微 偏离 杆身线---n n u n a---n d v n---使 ['弯曲', '偏离']
- 1) 阻挡紫外线顶立节能隔音窗选用特种玻璃，可阻挡99来自太阳光的紫外线对人体皮肤造成的伤害---99 来自 太阳光 的 紫外线---伤害---m v n u n---v---造成 ['紫外线', '伤害']
- 其它文件都可正常上传，但如果同一个文件上传总是出现后台处理失败，可能该视频在切割时造成部分码段丢失出错，建议将文件头尾切割掉几秒，或重新转换下文件格式---该 视频---部分 码段 丢失 出错---r n---m n v v---造成 ['视频', '出错']
- 针对脑瘫患儿的心理障碍，在综合治疗中尊重患者，认真倾听和理解、安慰、鼓励患儿，在矫治肢体功能障碍的同时，激发患儿的积极参与性，提高治疗训练效果也是脑瘫治疗方法，改善患儿心理状态的健康教育，积极与患儿交流，逐渐克服患儿的偏执，依赖心理，建立独立、自信、坚强、向上的信心，培养其自理能力，使其适应社会---培养 其 自理 能力---其 适应 社会---v r v n---r v n---使 ['培养', '适应']
- 同一化学试验中的所有称量，应自始至终使用同一架天平，使用不同天平会造成误差---使用 不同 天平---误差---v a n---n---造成 ['不同', '误差']
- 遇高热、明火及强氧化剂易引起燃烧---高热 、 明火 及 强氧化剂---燃烧---n w p n c n---v---引起 ['高热', '燃烧']
- 液体或雾引起眼刺激或灼伤---液体 或 雾---眼 刺激 或 灼伤---n c n---n v c v---引起 ['液体', '灼伤']
- 天价薪酬已引起了决策高层的关注---天价 薪酬---决策 高层 的 关注---n n---v n u v---引起 ['天价', '关注']

## 短语对(词对的最小子树)

### 共线上下文