

# “隐含”关系

李佳政

## “因果”关系

- 句子间有因果关系，互相证明。
- 有明确的指示词 => “so”、“because”、“consequently”。
- 相邻的两个句子，或在同一句之内。

## “隐含”关系

- 表达的语义之间存在关联，相似或相反
- 不存在指示词
- 距离不确定，有可能相邻，也可能距离较远。

# “因果”关系

- 句子间有因果关系，互相证明。
- 有明确 [yīn hán] 隐含。
- 相邻的 编辑

📷 本词条缺少**名片图**，补充相关内容使词条更完整，还能快速升级，赶紧来**编辑**吧！

“隱含

隐含是一个汉语词语，读音yǐnhán，隐约含有、暗中包含的意思

中文名	隐含	释 义	心中隐含着无限羞愧
拼 音	yǐnhán	性 质	汉语词汇

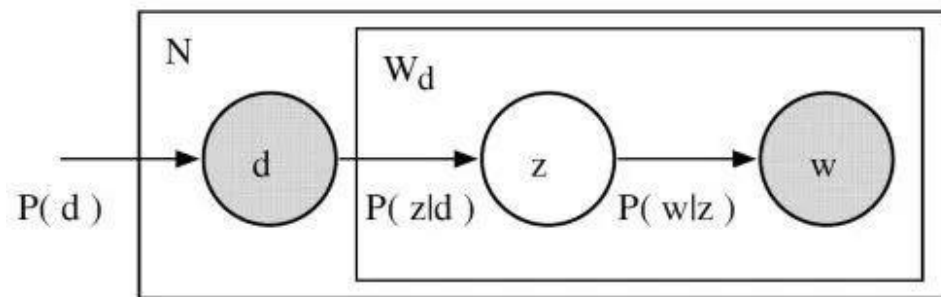
- 表达的语义之间存在关联，相似或相反
- 不存在指示词
- 距离不确定，有可能相邻，也可能距离较远。

# 句子间“隐含”关系

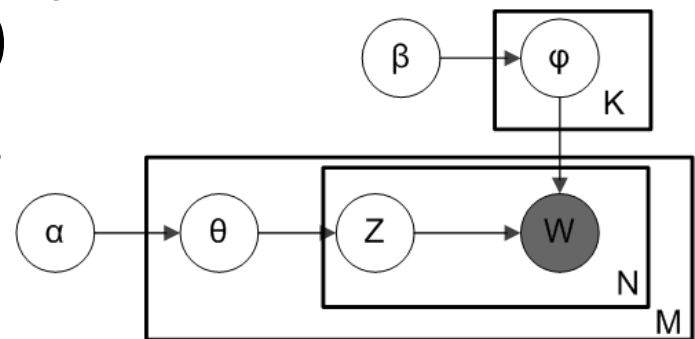
1. 词的语义相似度
2. 句子间的关系
3. 分类与回归任务
4. 其余相关研究

# 词是最小的语义单元

1. 编辑距离。
2. LSI, latent semantic index, 解决了一义多词的问题, 核心SVD。
3. pLSA, 用概率模型替代SVD。



4. LDA, latent Dirichlet allocation, pLSA的贝叶斯版本, 最终得到一个主题下单词概率分布。采用BOW(bag of words)模型, 不关注语序, 具体的主题名称需要手动指定。



# 词是最小的语义单元

- 5. 词向量, word embedding。主流方法: word2vec、Glove, 使用向量空间余弦相似度计算, Glove效果要好一些, 非神经网络模型。
- 6. WordNet, 由人工标注, 可解决一词多义问题, 基于字典。  
由于层级表示, 面向对象思想, 可以进行父类泛化。

# 句子(短文本)语义相似度

1. 基于词向量。
  - 1) 对句子内的词向量进行线性求和(平均), 余弦相似度。
  - 2) 输入词向量, 使用CNN对词向量的各维度进行学习。
2. 编辑距离。
3. 手工特征转换: USA => the United States
4. BM25算法。

# 推理

- 文本蕴含, Textual Entailment

任务描述: 两个句子, 判断关系。

- 1) A soccer game with multiple males playing.  
Some men are playing a sport. **Entailment**
- 2) A black race car starts up in front of a crowd of people.  
A man is driving down a lonely road. **Contradiction**
- 3) An older and younger man smiling.  
Two men are smiling and laughing at the cats playing on the floor. **Neutral**



# 文本蕴含推理方法

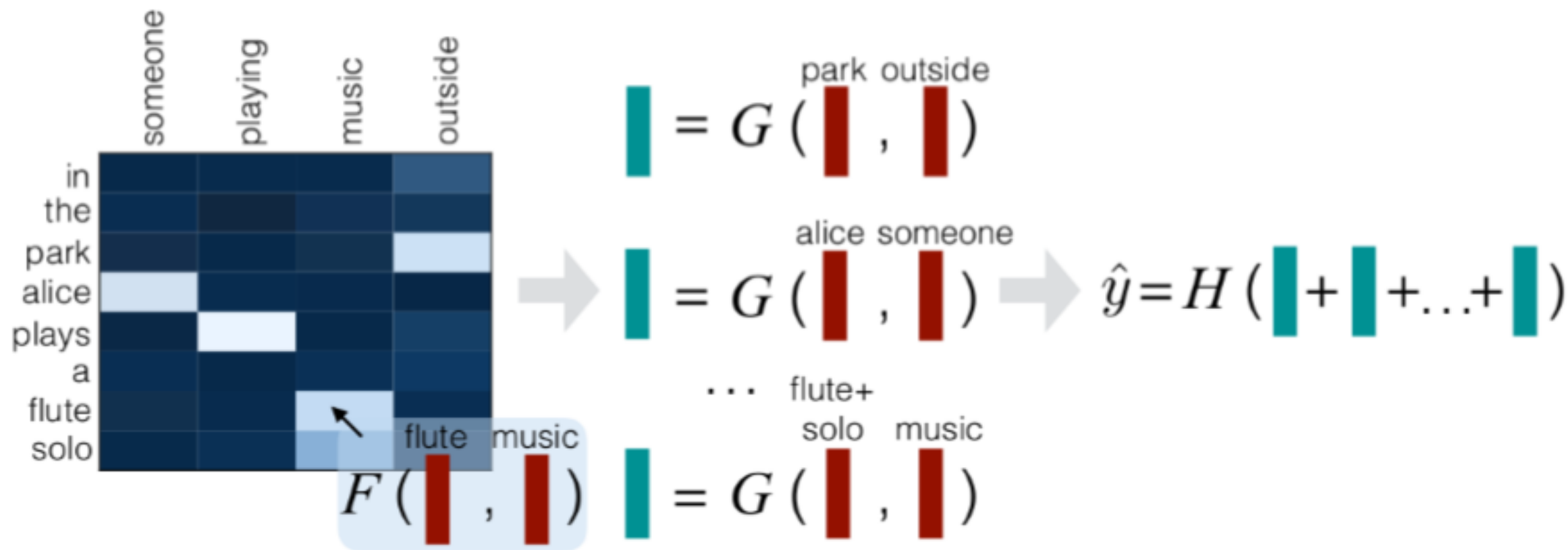
1、 基于特征的方法，训练集准确率很高99.7，测试集低78.2。

N-gram特征。

2、 句子编码方法，均为神经网络方法。

双向RNN，自注意力机制……，单模型测试集最佳效果88.8，集成模型最佳90.1。

# 文本蕴含推理方法



# 科学论文利用“隐含”关系

- 筛选候选句子
- 结合语义相似度及句子间表达含义的关系分类得到句子的评分
- 利用textrank算法得到句子得分，排序得到抽取式摘要

Summarization is one of the key features of human intelligence. It plays an important role in understanding and representation. With rapid and continual expansion of texts, pictures and

(3) *Closed system*. The process of automatic summarization is closed, does not interact with other processes in cyberspace or social space. Actually, reviews and comments in cyberspace are open, easily available, and valuable for composing and improving summaries.

# 其他可能应用的研究

- \* 谓词逻辑推理（知识表示）

$$\text{Father}(x, y) \wedge \text{Father}(y, z) \Rightarrow \text{Grandfather}(x, z)$$

Cyc项目，作为外部知识来源，过于庞大，无法继续开发。

- \* 句子的作用分类，如开篇点题、承上启下、总结作用、首尾呼应，采用分类器分类，增加相应句子的评分。

- \* 明斯基的框架(Frame)

心理学研究，回答“金丝雀会飞吗？”要比“鸟会飞吗？”时间长，父类属性传递，促成了面向对象思想与WordNet。