

汉语句子隐喻的识别和理解关键技术研究

key techniques for detection and interpretation
of chinese sentence-level metaphor

田展铭 介绍

Structure

- Metaphor detection
 - level of abstraction analysis: 抽象度分析
 - BiLSTM: 双向长短期神经网络 & 多层注意力机制
- Metaphor understanding
 - 合作网模型 -> 双向语义关系
 - 分析两个概念域的互动关系

Metaphor classification

- 名词性隐喻：生活是一场旅途 -> 生活是多彩的
- 动词性隐喻：股票跳水 -> 股票大跌
- 形容词性隐喻：渺茫的歌声 -> 微弱的歌声
- 副词性/介词性/量词性 较少且隐喻性较弱

Level of abstraction

- 抽象度体现了词语指代的事物的抽象程度
 - e.g. 马路 -> 抽象度低 思考 -> 抽象度高
- 三种隐喻结构：
 - 抽象源域 到 具体目标域 e.g. 雕塑是凝固的思想
 - 具体源域 到 具体目标域 e.g. 这孩子真熊
 - 抽象源域 到 抽象目标域 e.g. 知识就是力量
- 抽象度一个很好的隐喻特征，能很好的表现目标域和源域的冲突
- 英语：MRC心理学辞典

$$Rel_{Abs} = \frac{\sum_{i=1}^{|Abs|} Relateness(W, abs_i)}{|Abs|}$$

$$Rel_{con} = \frac{\sum_{j=1}^{|Con|} Relateness(W, con_j)}{|Con|}$$

$$Abs_W = Rel_{abs} - Rel_{con}$$

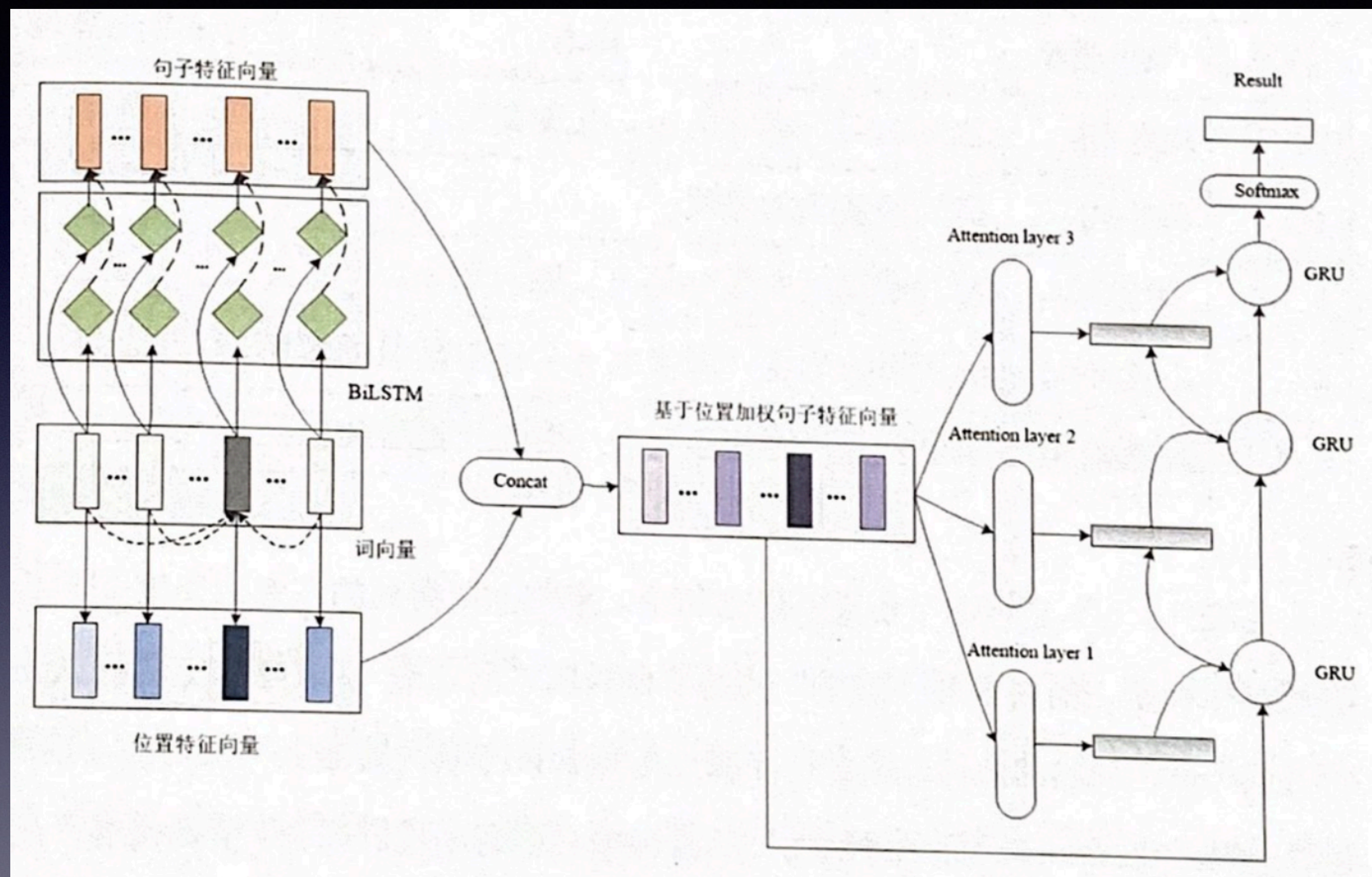
中文词抽象度计算方法

Bidirectional Long-Short Term Memory Neural Network

- 有效处理不定长的句子
- 比起单向阅读，这个模型同时处理向前和向后阅读
- 多层注意力机制运算，防止长句的必要信息忽略

Multi-level attention model

- 寻找当前任务下输入的文本最关键的信息，并且赋予最高的注意力权重
- 最高抽象度的词，系统应该优先关注和被刺激，做为注意词
- 使用多层来提高准确度，测试表明3层表现最佳



BiLSTM:

1. 根据抽象度选注意词
2. 根据注意词位置，对句子特征向量加权
3. 对位置加权的特征向量进行多次注意力权重计算
4. 对隐喻性进行判断

GRU:

1. 利用重制门控和更新门控来提效计算
2. 生成了特征向量E作为二元判断的输入

Softmax:

把特征向量转换为条件概率分布

Cooperation network

- 获取理解结果的两个步骤
 - 筛选两个概念域之间的相似度
 - 根据上下文信息，选出最相关上下文的属性作为结果
- 合作机制/互动理论：
 - 源域若属于目标域的属性集/与目标域属性有相似关系，那么就是存在合作性。
 - 抑制具有弱联结的属性，凸显强联结属性
 - 目标域对自身的特征进行选择/强调/压缩，使得两个概念域相互作用

$$\text{Domain}(\text{Concept}) = \{k_1, k_2, \dots, k_i, \dots, k_D\},$$

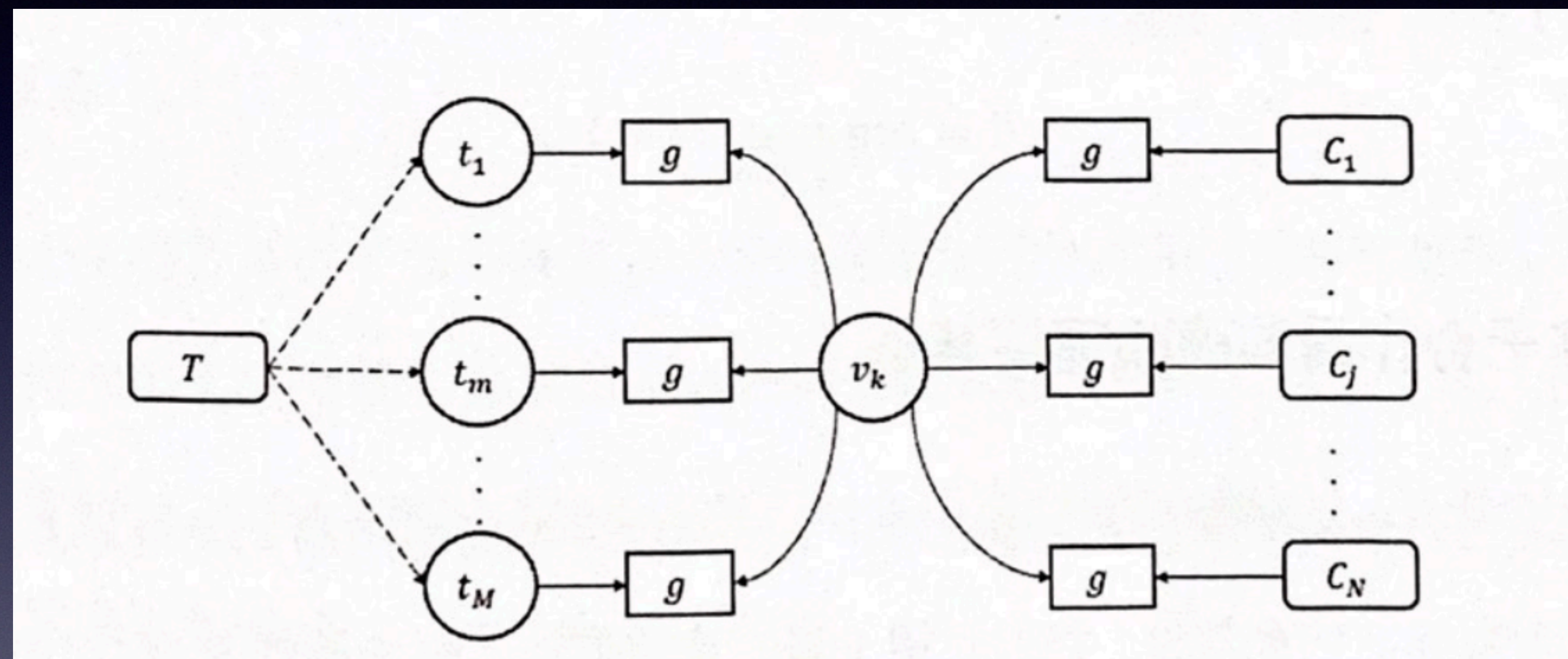
定义一个词的知识域

$$\text{Rel}(A \rightarrow B) = \frac{\text{Sim}[A, \text{Domain}(B)]}{\text{Sim}[B, \text{Domain}(B)]}$$

A能联想到B的数值计算

Cooperation intensity

- $\text{Coop}(A, B) = \text{Coop}(B, A)$
- 如果A和B互相联想的程度高，那么 $\text{Coop}(A, B)$ 会很高
- 如果A和B互相联想的程度低，那么 $\text{Coop}(A, B)$ 会很低
- 如果A很容易联想到B，但B联想不到A，那么 $\text{Coop}(A, B)$ 在两者联想度之间



Process:
对于任何一个源域 V_k
隐喻句上下文 C_1, \dots, C_n
目标域 t_1, \dots, t_m

上下文和目标域经过 g （合作强度函数）和源域进行合作

那么针对每一个源域 V_1, V_2, \dots ,
就构成了合作网

Merits & drawbacks

- 多层注意力机制和合作网
- 从大文本（段落）中判断注意词是难点
- 需要人工标注目标域和源域
- 通过情感分析可以获得更高的准确性