

# Exploring the Functional and Geometric Bias of Spatial Relations Using Neural Language Models

王希豪

北京大学中国语言文学系

*wangxihao@pku.edu.cn*

2022 年 10 月 31 日

- 1 Introduction
- 2 Varying targets and landmarks
- 3 Varying spatial relations
- 4 Discussion and conclusion

方位表达式 (Locative expression) 是描述 目标物体 (target object) 和 地标 (landmark) 之间关系的名词短语<sup>a</sup>

<sup>a</sup>空间描述 (spatial description) 是一个范围更大的概念。但在论文中，研究的主要是符合方位表达式结构的数据，所以后续提到的空间描述可以直接当方位表达式来理解。

方位表达式 (Locative expression) 是描述 目标物体 (target object) 和 地标 (landmark) 之间关系的名词短语<sup>a</sup>

<sup>a</sup>空间描述 (spatial description) 是一个范围更大的概念。但在论文中，研究的主要是符合方位表达式结构的数据，所以后续提到的空间描述可以直接当方位表达式来理解。

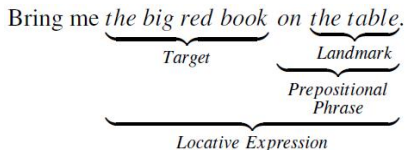


图: 方位表达式实例 [Dobnik et al., 2018]

# Locative expression

方位表达式 (Locative expression) 是描述 目标物体 (target object) 和 地标 (landmark) 之间关系的名词短语<sup>a</sup>

<sup>a</sup>空间描述 (spatial description) 是一个范围更大的概念。但在论文中，研究的主要是符合方位表达式结构的数据，所以后续提到的空间描述可以直接当方位表达式来理解。

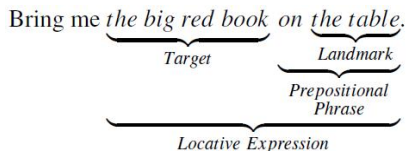


图: 方位表达式实例 [Dobnik et al., 2018]

Locative\_expression := target **relation** landmark

### 几何 (Geometry) 因素 vs. 功能 (Function) 因素

- The umbrella is *over/above* the man.
- The man is *under/below* the umbrella.

## Two Factors

### 几何 (Geometry) 因素 vs. 功能 (Function) 因素

- The umbrella is *over/above* the man.
- The man is *under/below* the umbrella.



图: 几何语义与功能语义比较实验 [Coventry et al., 2001]

# Two Factors

## 几何 (Geometry) 因素 vs. 功能 (Function) 因素

- The umbrella is *over/above* the man.
- The man is *under/below* the umbrella.

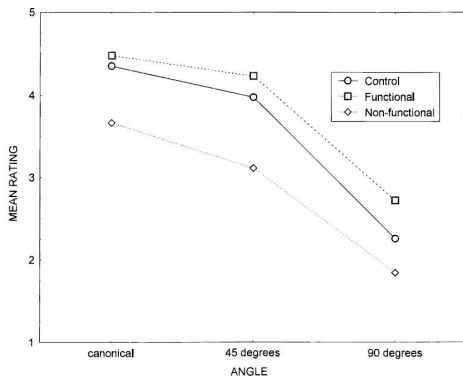


图: 几何语义与功能语义比较实验结果展示 [Coventry et al., 2001]

## 几何 (Geometry) 因素 vs. 功能 (Function) 因素

- The umbrella is *over/above* the man.
- The man is *under/below* the umbrella.

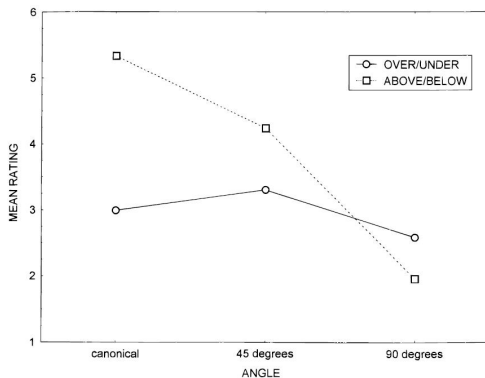


图: 几何语义与功能语义比较实验结果展示 [Coventry et al., 2001]

## 几何 (Geometry) 因素 vs. 功能 (Function) 因素

- The umbrella is *over/above* the man.
- The man is *under/below* the umbrella.

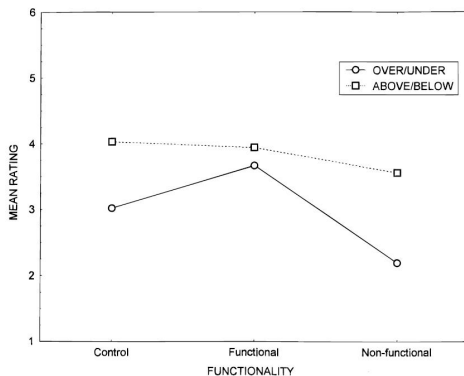


图: 几何语义与功能语义比较实验结果展示 [Coventry et al., 2001]

*Given the distinction between geometric and functional factors in shaping spatial semantics, a useful analysis that would inform the design and creation of computational models of spatial semantics is **to identify the particular semantic bias (geometric/functional) that each spatial term evinces.***

基于几何因素和功能因素在塑造空间语义时的差异，**确定每个空间相关的词语在几何和功能两方面表现出来的特定语义偏差**，对于设计和构建空间语义计算模型将会是有帮助的。

## 理论层面

- 指出了空间关系所表达的基础语义 (grounded semantics) 的复杂性, 即空间关系不只是表达了物体在几何空间上的关系。

## 理论层面

- 指出了空间关系所表达的基础语义 (grounded semantics) 的复杂性, 即空间关系不只是表达了物体在几何空间上的关系。

## 实践层面

- 在自然语言生成领域, 特别是在“机器人定位”等无法利用端到端技术训练的任务上, 可以起到一定的作用。
- 在句法和语义分析中, 可以用于解决 PP-attachment 等困难问题。

## 理论层面

- 指出了空间关系所表达的基础语义 (grounded semantics) 的复杂性, 即空间关系不只是表达了物体在几何空间上的关系。

## 实践层面

- 在自然语言生成领域, 特别是在“机器人定位”等无法利用端到端技术训练的任务上, 可以起到一定的作用。
- 在句法和语义分析中, 可以用于解决 PP-attachment 等困难问题。

## 困难

- 即便是母语者也很难根据语感直接分析出空间关系表达中的几何因素和功能因素, 必须要通过严密的实验和语言学分析才能得出结论。

## 理论层面

- 指出了空间关系所表达的基础语义 (grounded semantics) 的复杂性, 即空间关系不只是表达了物体在几何空间上的关系。

## 实践层面

- 在自然语言生成领域, 特别是在“机器人定位”等无法利用端到端技术训练的任务上, 可以起到一定的作用。
- 在句法和语义分析中, 可以用于解决 PP-attachment 等困难问题。

## 困难

- 即便是母语者也很难根据语感直接分析出空间关系表达中的几何因素和功能因素, 必须要通过严密的实验和语言学分析才能得出结论。目前比较确定的是, *in, on, at, over, under* 偏向于表达功能语义; *above, below, left of, right of* 偏向于表达几何语义。

### Dobnik and Kelleher [2013, 2014]

- 通过计算和比较关系周围“目标-地标”的熵，发现不同关系的选择能力 (selectivity) 是不同的<sup>1</sup>。关系词的熵值越低，选择能力越强，在使用时会偏向某类“目标-地标”。

---

<sup>1</sup>可以理解为在选择对象时更加明确，更容易选取适合的对象。也可以理解为泛用性较弱，能够组合的对象较少。

<sup>2</sup>此处和后文所有提到的“功能类关系”和“几何类关系”实际上都是“倾向于表达功能语义的关系 (Functionally-biased relation)”和“倾向于表达几何语义的关系 (Geometricly-biased relation)”，为了方便进行了一定的简化，但也带来了一定的歧义

### Dobnik and Kelleher [2013, 2014]

- 通过计算和比较关系周围“目标-地标”的熵，发现不同关系的选择能力 (selectivity) 是不同的<sup>1</sup>。关系词的熵值越低，选择能力越强，在使用时会偏向某类“目标-地标”。
  - 前人研究指出的功能类关系 *in*, *on* 和 *over* 都具有相对较低的熵值；而几何类关系 *above* 则有较高的熵值<sup>2</sup>。

---

<sup>1</sup>可以理解为在选择对象时更加明确，更容易选取适合的对象。也可以理解为泛用性较弱，能够组合的对象较少。

<sup>2</sup>此处和后文所有提到的“功能类关系”和“几何类关系”实际上都是“倾向于表达功能语义的关系 (Functionally-biased relation)”和“倾向于表达几何语义的关系 (Geometricly-biased relation)”，为了方便进行了一定的简化，但也带来了一定的歧义

### Dobnik and Kelleher [2013, 2014]

- 通过计算和比较关系周围“目标-地标”的熵，发现不同关系的选择能力 (selectivity) 是不同的<sup>1</sup>。关系词的熵值越低，选择能力越强，在使用时会偏向某类“目标-地标”。
  - 前人研究指出的功能类关系 *in*, *on* 和 *over* 都具有相对较低的熵值；而几何类关系 *above* 则有较高的熵值<sup>2</sup>。
- 作者假设功能类关系具有更强的选择能力，在生成句子时，和“目标-地标”之间的关系更加紧密。于是，利用  $\log$  似然函数计算“目标-地标”和关系之间联系的紧密程度，借此可以为判断生成的空间描述是否合法提供帮助。

---

<sup>1</sup>可以理解为在选择对象时更加明确，更容易选取适合的对象。也可以理解为泛用性较弱，能够组合的对象较少。

<sup>2</sup>此处和后文所有提到的“功能类关系”和“几何类关系”实际上都是“倾向于表达功能语义的关系 (Functionally-biased relation)”和“倾向于表达几何语义的关系 (Geometrically-biased relation)”，为了方便进行了一定的简化，但也带来了一定的歧义

## 数据集: Visual Genome [Krishna et al., 2017]

- 共包含 10.8 万张图片，平均每张图片中有 35 个物体、26 条属性和 21 条关系
  - cup on table
  - girl holding on to bear
  - woman standing on snow

---

<sup>3</sup>这一操作可能会对实验结果产生较大的影响

## 数据集: Visual Genome [Krishna et al., 2017]

- 共包含 10.8 万张图片，平均每张图片中有 35 个物体、26 条属性和 21 条关系
  - cup on table
  - girl holding on to bear
  - woman standing on snow

## 数据处理

- 构建词典，提取空间关系
- 将关系词组看作一个整体，比如将“jump over”记为“jump\_over”
- 合并、简化部分关系，以增加特定关系的出现次数。例如，将“left of”和“to the left of”简化为 left<sup>3</sup>
- 实验只考虑出现次数超过 100 的关系
- 方位表达式前后添加标记：<s> target relation landmark </s>
- 训练集划分成 10 份进行训练，测试集中每个关系都有 20 条实例

---

<sup>3</sup>这一操作可能会对实验结果产生较大的影响

## 概率语言模型 (Probabilistic language model)

$$P(w_{1:T}) = \prod_{t=1}^T P(w_{t+1} | w_{1:t})$$

---

<sup>4</sup>公式中的  $v_{k_{1:t}}$  和  $v_t$  应当是相同的，都表示对句子前  $t$  个词按顺序编码的结果。

## 概率语言模型 (Probabilistic language model)

$$P(w_{1:T}) = \prod_{t=1}^T P(w_{t+1}|w_{1:t})$$

## 基于神经网络的概率语言模型<sup>4</sup>

$$\hat{P}(w_{t+1}|w_{1:t} = v_{k_{1:t}}) = f(v_t; \Theta) = \hat{y}_t$$

---

<sup>4</sup>公式中的  $v_{k_{1:t}}$  和  $v_t$  应当是相同的，都表示对句子前  $t$  个词按顺序编码的结果。

## 概率语言模型 (Probabilistic language model)

$$P(w_{1:T}) = \prod_{t=1}^T P(w_{t+1}|w_{1:t})$$

## 基于神经网络的概率语言模型<sup>4</sup>

$$\hat{P}(w_{t+1}|w_{1:t} = v_{k_{1:t}}) = f(v_t; \Theta) = \hat{y}_t$$

## 损失函数

$$\text{loss}(S) = - \sum_{s \in S} \sum_{t=0}^{|s|} \frac{\log(\hat{y}_t(v_{k_{t+1}}))}{|S| \times |s|}$$

<sup>4</sup>公式中的  $v_{k_{1:t}}$  和  $v_t$  应当是相同的，都表示对句子前  $t$  个词按顺序编码的结果。

## 概率语言模型 (Probabilistic language model)

$$P(w_{1:T}) = \prod_{t=1}^T P(w_{t+1}|w_{1:t})$$

## 基于神经网络的概率语言模型<sup>4</sup>

$$\hat{P}(w_{t+1}|w_{1:t} = v_{k_{1:t}}) = f(v_t; \Theta) = \hat{y}_t$$

## 损失函数

$$\text{loss}(S) = - \sum_{s \in S} \sum_{t=0}^{|s|} \frac{\log(\hat{y}_t(v_{k_{t+1}}))}{|S| \times |s|}$$

## 困惑度 (Perplexity)

$$\text{Perplexity}(S, P) = 2^{\mathbb{E}_S[-\log_2(P(w_{1:T}))]}$$

<sup>4</sup>公式中的  $v_{k_{1:t}}$  和  $v_t$  应当是相同的，都表示对句子前  $t$  个词按顺序编码的结果。

# Overview

- 1 Introduction
- 2 Varying targets and landmarks**
- 3 Varying spatial relations
- 4 Discussion and conclusion

## Hypothesis 1 (Weak)

测试集中，短语的困惑度反映了空间关系在功能和几何语义上的偏差。

# Hypotheses

## Hypothesis 1 (Weak)

测试集中，短语的困惑度反映了空间关系在功能和几何语义上的偏差。

## Hypothesis 2

功能类关系在选择“目标”和“地标”时更具有选择性。

对于数据集中 `<s> target relation landmark </s>` 这样的模式，当其中的 `relation` 是功能类关系时，就更容易预测，也就意味着会有更低的困惑度。

# Hypotheses

## Hypothesis 1 (Weak)

测试集中，短语的困惑度反映了空间关系在功能和几何语义上的偏差。

## Hypothesis 2

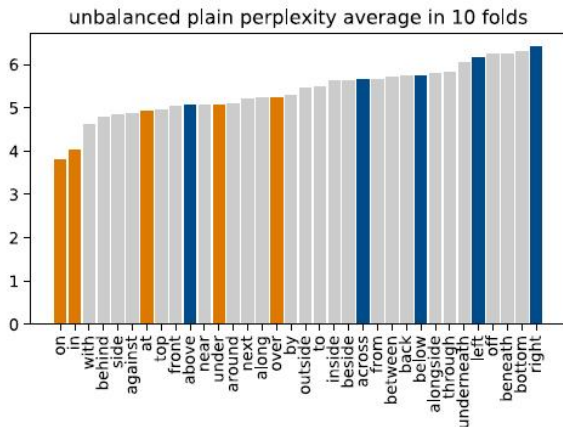
功能类关系在选择“目标”和“地标”时更具有选择性。

对于数据集中 `<s> target relation landmark </s>` 这样的模式，当其中的 relation 是功能类关系时，就更容易预测，也就意味着会有更低的困惑度。

## Hypothesis 3 (Strong)

对于给定的功能类关系，“目标”和“地标”是可预测的。

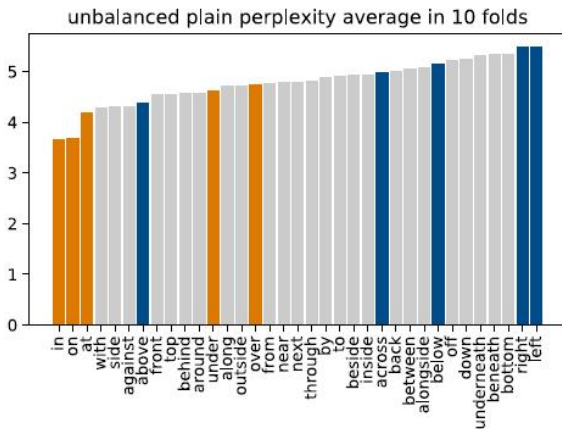
# Results of Language Model 1 based on unbalanced corpus



(a) test-set

图: 空间描述的平均困惑度 (LM1)  
其中, 橘色为功能类关系, 蓝色为几何类关系

# Results of Language Model 1 based on unbalanced corpus



(b) training set

图: 空间描述的平均困惑度 (LM1)

其中, 橘色为功能类关系, 蓝色为几何类关系

## 结论

- 模型在训练集和测试集上产生了较为相似的结果，证明模型具有较强的泛化能力。模型在测试集上的结果更具有代表性

## 结论

- 模型在训练集和测试集上产生了较为相似的结果，证明模型具有较强的泛化能力。模型在测试集上的结果更具有代表性
- (Hypothesis 1) 包含具有相同语义偏差的关系的空间描述，平均困惑度有聚集到一起的趋势

## 结论

- 模型在训练集和测试集上产生了较为相似的结果，证明模型具有较强的泛化能力。模型在测试集上的结果更具有代表性
- (Hypothesis 1) 包含具有相同语义偏差的关系的空间描述，平均困惑度有聚集到一起的趋势
- (Hypothesis 2) 包含功能类关系的空间描述倾向于有更低的困惑度

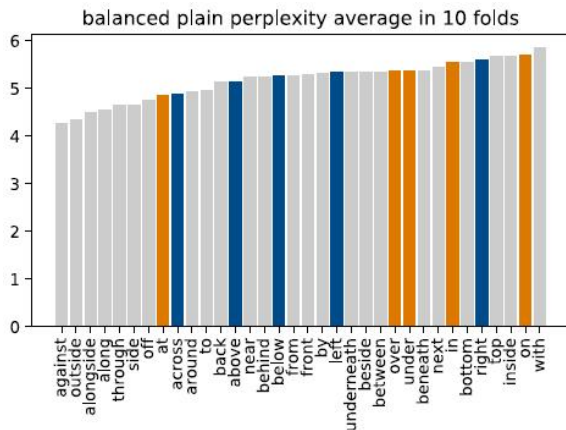
## 结论

- 模型在训练集和测试集上产生了较为相似的结果，证明模型具有较强的泛化能力。模型在测试集上的结果更具有代表性
- (Hypothesis 1) 包含具有相同语义偏差的关系的空间描述，平均困惑度有聚集到一起的趋势
- (Hypothesis 2) 包含功能类关系的空间描述倾向于有更低的困惑度

## 存在的问题

- 关系的困惑度受真实语料中的出现频率影响
  - 所考察的 10 个词的困惑度排名和出现频率排名间的 Spearman 系数达到了 0.9

# Results of Language Model 2 based on balanced corpus

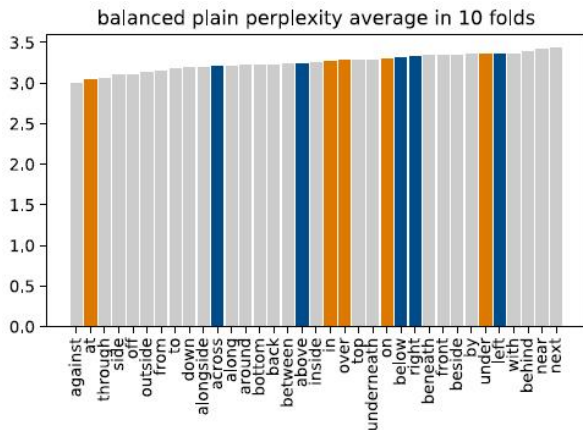


(a) test-set

图: 空间描述的平均困惑度 (LM2)

其中, 橘色为功能类关系, 蓝色为几何类关系

# Results of Language Model 2 based on balanced corpus



(b) training set

图: 空间描述的平均困惑度 (LM2)

其中, 橘色为功能类关系, 蓝色为几何类关系

### 结论

- 由于词典的减小，在训练集上，所有关系的平均困惑度都降低了
- 包含几何类关系的空间描述的平均困惑度降低得更多

## 结论

- 由于词典的减小，在训练集上，所有关系的平均困惑度都降低了
- 包含几何类关系的空间描述的平均困惑度降低得更多

## 解释

- The frequency of using functionally-biased spatial relations are higher in English, this bias and our strong hypothesis for predictability of target-landmark pairs can be expressed with simple joint probabilities which we are estimating with the language model:

$$P(\text{target,relation,landmark}) = P(\text{relation})P(\text{target,landmark}|\text{relation})$$

# Results of Language Model 2 based on balanced corpus

## 结论

- 由于词典的减小，在训练集上，所有关系的平均困惑度都降低了
- 包含几何类关系的空间描述的平均困惑度降低得更多

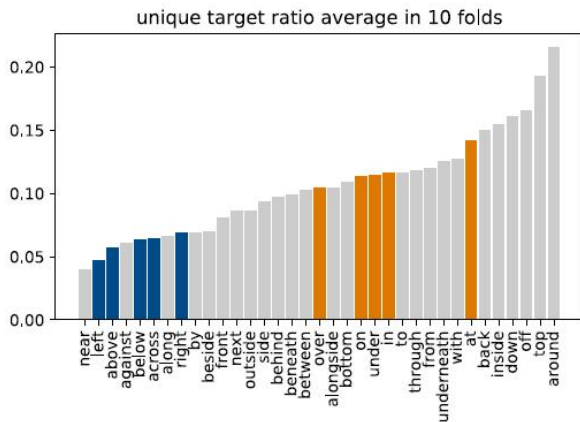
## 解释

- The frequency of using functionally-biased spatial relations are higher in English, this bias and our strong hypothesis for predictability of target-landmark pairs can be expressed with simple joint probabilities which we are estimating with the language model:

$$P(\text{target,relation,landmark}) = P(\text{relation})P(\text{target,landmark}|\text{relation})$$

It is possible that targets and landmarks that occur with these relations are very specific to these relations but infrequent with other relations. When we remove their frequency support provided by the frequency of relations these targets and landmarks become infrequent in the dataset and therefore less probable which on overall results in higher perplexities of phrases with functionally-biased relations. Specificity of targets and landmarks can be a source of these results.

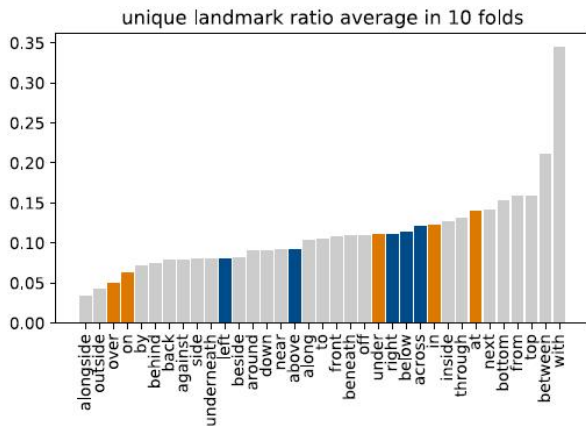
# Results of Language Model 2 based on balanced corpus



(a) targets

图: 各个关系独有的目标和地标占总数的比例

# Results of Language Model 2 based on balanced corpus



(b) landmarks

图: 各个关系独有的目标和地标占总数的比例

## 证据

- (Hypothesis 2) 从目标的角度来看，功能类关系和几何类关系有着明显的差异。功能类关系更容易和独有的目标共现，从而为 LM2 模型带来了更高的困惑度。

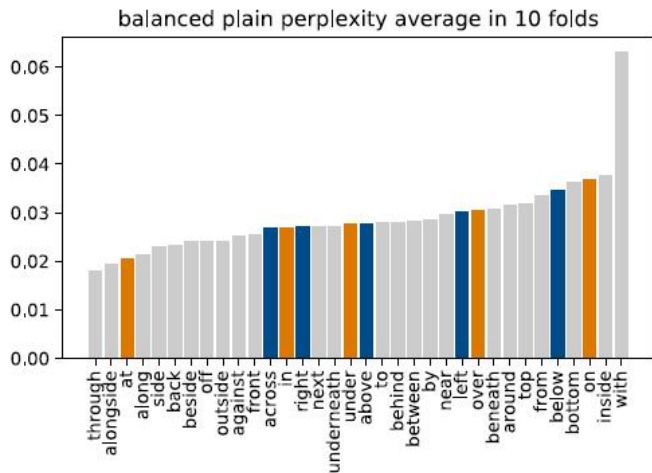
### 证据

- (Hypothesis 2) 从目标的角度来看，功能类关系和几何类关系有着明显的差异。功能类关系更容易和独有的目标共现，从而为 LM2 模型带来了更高的困惑度。
- 地标的独特性和关系倾向之间的关系则较为模糊，因为原始的图像描述数据集在构造的时候，是先选择“目标”再选择“地标”的。

## 证据

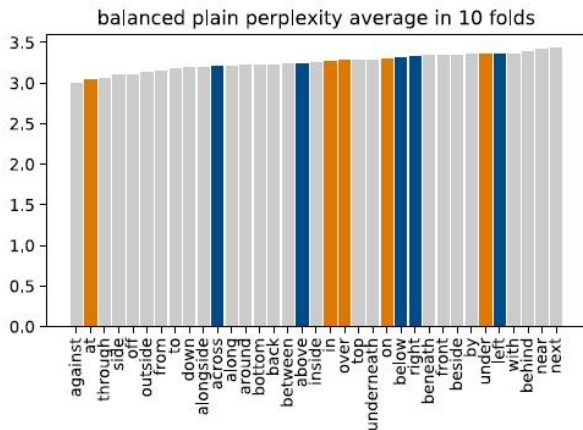
- (Hypothesis 2) 从目标的角度来看，功能类关系和几何类关系有着明显的差异。功能类关系更容易和独有的目标共现，从而为 LM2 模型带来了更高的困惑度。
- 地标的独特性和关系倾向之间的关系则较为模糊，因为原始的图像描述数据集在构造的时候，是先选择“目标”再选择“地标”的。
- 总的来说，“目标”的独特性在功能类关系空间描述的困惑度上有着更大的贡献。

# Results of Language Model 2'



图：空间描述的平均困惑度 (LM2')  
其中，橘色为功能类关系，蓝色为几何类关系

# Results of Language Model 2'



(b) training set

图: 空间描述的平均困惑度 (LM2)

其中, 橘色为功能类关系, 蓝色为几何类关系

### 结论

- 在将 `<s> target relation landmark </s>` 模式更改为 `<s> landmark relation target </s>` 之后，模型的困惑度下降了几个量级。说明地标对于空间关系和目标物体的预测具有更大的贡献
- 但是，概率语言模型的训练是自左向右的，在预测关系时只考虑了目标词，而导致包含更多信息的“地标”总是在预测完空间关系后才出现。

# Overview

- 1 Introduction
- 2 Varying targets and landmarks
- 3 Varying spatial relations**
- 4 Discussion and conclusion

# Hypotheses

## Hypothesis 1 (Weak)

测试集中，短语的困惑度反映了空间关系在功能和几何语义上的偏差。

## Hypothesis 2

功能类关系在选择“目标”和“地标”时更具有选择性。

对于数据集中 `<s> target relation landmark </s>` 这样的模式，当其中的 relation 是功能类关系时，就更容易预测，也就意味着会有更低的困惑度。

## Hypothesis 3 (Strong)

对于给定的功能类关系，“目标”和“地标”是可预测的。

## Hypothesis 4

在方位表达式中，几何类关系和功能类关系相比，更容易被替换。

# Hypotheses

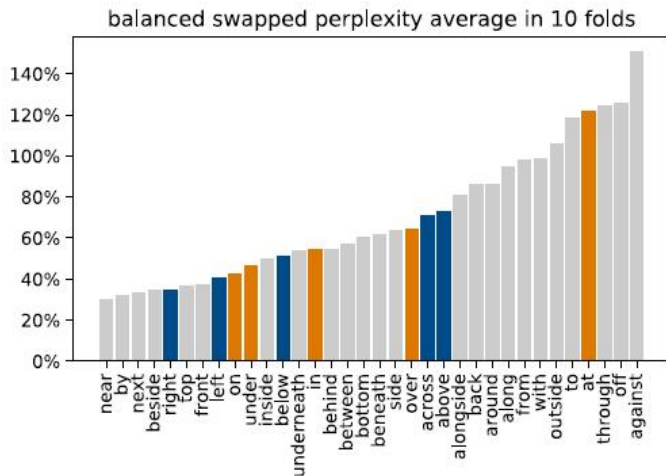


图: 关系替换增加困惑度的百分比  
其中, 橘色为功能类关系, 蓝色为几何类关系

# Overview

- 1 Introduction
- 2 Varying targets and landmarks
- 3 Varying spatial relations
- 4 Discussion and conclusion**

## 总结

- 论文探索了神经网络模型对空间关系的几何和功能特征的识别程度。概率语言模型可以展示空间描述的整体概率信息。

## 总结

- 论文探索了神经网络模型对空间关系的几何和功能特征的识别程度。概率语言模型可以展示空间描述的整体概率信息。
- 分析并在一定程度上证实了“功能类关系对于目标和地标有较高的选择性”的假设：

## 总结

- 论文探索了神经网络模型对空间关系的几何和功能特征的识别程度。概率语言模型可以展示空间描述的整体概率信息。
- 分析并在一定程度上证实了“功能类关系对于目标和地标有较高的选择性”的假设：
  - 在平衡语料上，由于 (i) 功能类关系独有的目标更多，且 (ii) 模型从左向右编码信息，功能类关系的困惑度比几何类关系要高。

## 总结

- 论文探索了神经网络模型对空间关系的几何和功能特征的识别程度。概率语言模型可以展示空间描述的整体概率信息。
- 分析并在一定程度上证实了“功能类关系对于目标和地标有较高的选择性”的假设：
  - 在平衡语料上，由于 (i) 功能类关系独有的目标更多，且 (ii) 模型从左向右编码信息，功能类关系的困惑度比几何类关系要高。
  - 在非平衡语料上，功能类关系独有的目标出现频率更高，模型更容易习得其中的规律，从而使得功能类关系的困惑度更低。

## 总结

- 论文探索了神经网络模型对空间关系的几何和功能特征的识别程度。概率语言模型可以展示空间描述的整体概率信息。
- 分析并在一定程度上证实了“功能类关系对于目标和地标有较高的选择性”的假设：
  - 在平衡语料上，由于 (i) 功能类关系独有的目标更多，且 (ii) 模型从左向右编码信息，功能类关系的困惑度比几何类关系要高。
  - 在非平衡语料上，功能类关系独有的目标出现频率更高，模型更容易习得其中的规律，从而使得功能类关系的困惑度更低。
- 证实了语言模型不仅可以预测关系，还能在一定程度上区分关系包含的语义信息。

## 总结

- 论文探索了神经网络模型对空间关系的几何和功能特征的识别程度。概率语言模型可以展示空间描述的整体概率信息。
- 分析并在一定程度上证实了“功能类关系对于目标和地标有较高的选择性”的假设：
  - 在平衡语料上，由于 (i) 功能类关系独有的目标更多，且 (ii) 模型从左向右编码信息，功能类关系的困惑度比几何类关系要高。
  - 在非平衡语料上，功能类关系独有的目标出现频率更高，模型更容易习得其中的规律，从而使得功能类关系的困惑度更低。
- 证实了语言模型不仅可以预测关系，还能在一定程度上区分关系包含的语义信息。

## 后续研究

- 神经网络模型习得的信息受到数据集质量和大小的影响，后续可以研究数据集对于模型预测结果的影响。

## 总结

- 论文探索了神经网络模型对空间关系的几何和功能特征的识别程度。概率语言模型可以展示空间描述的整体概率信息。
- 分析并在一定程度上证实了“功能类关系对于目标和地标有较高的选择性”的假设：
  - 在平衡语料上，由于 (i) 功能类关系独有的目标更多，且 (ii) 模型从左向右编码信息，功能类关系的困惑度比几何类关系要高。
  - 在非平衡语料上，功能类关系独有的目标出现频率更高，模型更容易习得其中的规律，从而使得功能类关系的困惑度更低。
- 证实了语言模型不仅可以预测关系，还能在一定程度上区分关系包含的语义信息。

## 后续研究

- 神经网络模型习得的信息受到数据集质量和大小的影响，后续可以研究数据集对于模型预测结果的影响。
- 在本文中，将“to the left of”简化为了“left”。对关系进行类似的聚类，可以分析一些尚未研究过的关系的功能和几何性质。

- K. R. Coventry, M. Prat-Sala, and L. Richards. The interplay between geometry and function in the comprehension of over, under, above, and below. Journal of memory and language, 44(3):376–398, 2001.
- S. Dobnik and J. Kelleher. Exploration of functional semantics of prepositions from corpora of descriptions of visual scenes. In Proceedings of the Third Workshop on Vision and Language, pages 33–37, 2014.
- S. Dobnik and J. D. Kelleher. Towards an automatic identification of functional and geometric spatial prepositions. Proceedings of PRE-CogSsci, pages 1–6, 2013.
- S. Dobnik, M. Ghanimifard, and J. Kelleher. Exploring the functional and geometric bias of spatial relations using neural language models. In Proceedings of the First International Workshop on Spatial Language Understanding, pages 1–11, 2018.

- R. Krishna, Y. Zhu, O. Groth, J. Johnson, K. Hata, J. Kravitz, S. Chen, Y. Kalantidis, L.-J. Li, D. A. Shamma, et al. Visual genome: Connecting language and vision using crowdsourced dense image annotations. International journal of computer vision, 123(1):32–73, 2017.