人工智能语言模型在语言心理学中的应用

工作坊成员

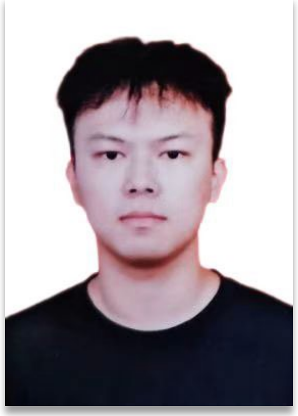
丁鼐，研究员，浙江大学

ding_nai@zju.edu.cn

邹家杰，博士后，浙江大学

jiajiezou@zju.edu.cn

关注语言的认知神经编码，相关文章发表于NeuroImage、Nature Communications等期刊及ACL等人工智能顶级会议



刘威，博士研究生，浙江大学

liuweizju@zju.edu.cn

关注语言模型的可解释性，相关文章发表于EMNLP，AAAI等人工智能顶级会议

陈泓桦，博士研究生，浙江大学

honghuachen@zju.edu.cn

关注语言的认知神经编码与语言模型的可解释性。

问卷调研

需求1(66%): GPT等大模型的基本原理

需求2(92%):自动分析语料特征

需求3(82%): 筛选或者生成语料

Workshop目的：以实用主义为主, 介绍AI工具的具体使用方法

问卷调研

需求1(66%): GPT等大模型的基本原理

需求2(92%):自动分析语料特征

需求3(82%): 筛选或者生成语料

方式(71.3%)：介绍AI工具的具体使用方法 ✓

方式(28.7%)：介绍现有研究 ✗

Workshop目的：以实用主义为主, 介绍AI工具的具体使用方法

目录

1. 概述
    - 代码平台
    - 模型结构
    - 语言任务

需求1：GPT等大模型的基本原理

2. 特征计算
    - 词汇特征
    - 语义特征
    - 句法特征

需求2：自动分析语料特征

3. 其他案例
    - 可控文本生成
    - 群体认知

需求3：筛选或者生成语料

其它心理语言学应用案例

代码平台

1. 概述
    - 代码平台
    - 模型结构
    - 语言任务

2. 特征计算
    - 词汇特征
    - 语义特征
    - 句法特征

3. 其他案例
    - 可控文本生成
    - 群体认知

代码平台 概述

1. 传统语言特征计算工具包: **hanlp**, spacy, nltk, CoreNLP

2. 语言模型搭建、加载、训练平台: **huggingface**

代码平台　特征计算

1. 传统语言特征计算工具包: **hanlp**, spacy, nltk, CoreNLP

HanLP

面向生产环境的前沿多语种自然语言处理技术

极速安装　使用指南

He H, et al. EMNLP, 2021.

代码平台 特征计算

Hanlp: 中文友好, 提供大量开箱即用的语言特征计算接口

中文分词
将文本切分为独立语义单位。支持粗分、细分双重标准，高低优先级自定义词典，多语种

词性标注
给每个单词标注语法类别，支持多种词性标注集：CTB、PKU、863、UD、NPCMJ，多语种

命名实体识别
识别文本中的专有名词，支持多种规范：PKU、MSRA、OntoNotes，黑白名单词典、多语种

依存句法分析
分析单词语法上的依赖，支持多种句法体系：SD、UD、PMT，投射/非投射，多语种

成分句法分析
分析句子在语法上的递归构成，并将其表示为树形结构。支持可视化、多语种

语义依存分析
分析句子中单词之间的语义关系图。支持多标准：CSDP、DM、PAS、PSD，支持多语种

语义角色标注
分析句子的谓词论元结构。支持可视化、多语种

抽象意义表示
将句子的意义表示为以概念为节点的单源有向无环图的语言学框架，支持多语种

代码平台　特征计算

Hanlp: 中文友好, 提供大量开箱即用的语言特征计算接口

词性标注

请输入一段中文文本：

HanLP为生产环境带来次世代最先进的多语种NLP技术。
我的希望是希望张晚霞的背影被晚霞映红。

48/800

粗分　对比1.x　CTB　PKU　863　词性标注

此页内容

简介

词性标注（Part-of-Speech tagging、POS）是一种标注句子中每个单词的词性（也称词类、语法类别）的任务。HanLP支持CTB、PKU、863、NPCMJ、Universal Dependencies等词性标注集。

调用方法

特征计算

Hanlp等工具包使用简单, 但可自定义程度差:

1. 不支持完成复杂语言任务, 如推理等

2. 不支持自定义模型结构

3. 不支持分析模型内部参数

语言模型

2. 语言模型搭建、加载、训练平台: **huggingface**

Hugging Face

Transformers

- 模型搭建
- 模型训练
- 模型推理
- …

Datasets

- 数据集加载
- 多线程预处理
- 高效缓存
- …

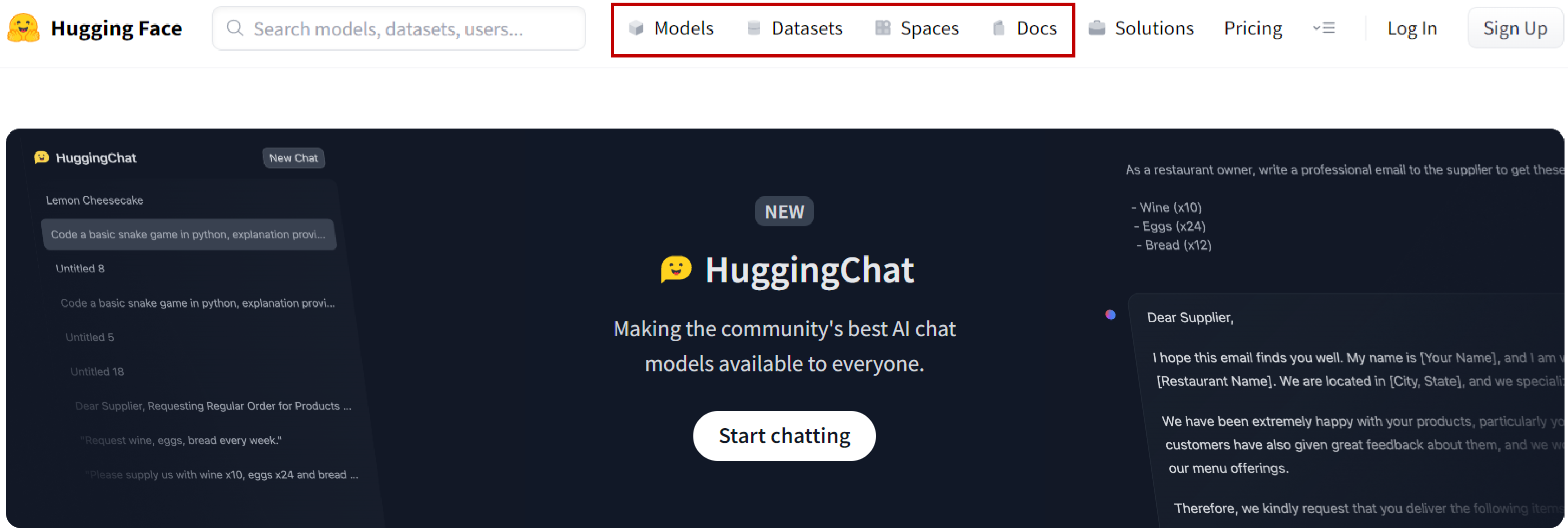
Tokenizers

- 快速分词
- 格式转换
- 自定义分词器
- …

Wolf T, et al. EMNLP, 2020.

代码平台　语言模型

Huggingface提供大量的开源模型和数据集，以及中文文档



https://huggingface.co/

代码平台　语言模型

模型路径

facebook/**bart-large-cnn** like 406

Summarization PyTorch TensorFlow JAX Rust Transformers cnn_dailymail English bart text2text-generation Eval Results

AutoTrain Compatible arxiv:1910.13461 License: mit

Model card Files and versions Community 31 Train Deploy Use in Transformers

模型描述

Edit model card

BART (large-sized model), fine-tuned on CNN Daily Mail

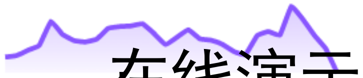
BART model pre-trained on English language, and fine-tuned on CNN Daily Mail. It was introduced in the paper BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension by Lewis et al. and first released in [this repository (https://github.com/pytorch/fairseq/tree/master/examples/bart).

Disclaimer: The team releasing BART did not write a model card for this model so this model card has been written by the Hugging Face team.

Model description

BART is a transformer encoder-encoder (seq2seq) model with a bidirectional (BERT-like) encoder and an autoregressive (GPT-like) decoder. BART is pre-trained by (1) corrupting text with an arbitrary noising function, and (2) learning a model to reconstruct the original text.

Downloads last month
1,345,544



在线演示

Hosted inference API

Summarization Examples

The tower is 324 metres (1,063 ft) tall, about the same height as an 81-storey building, and the tallest structure in Paris. Its base is square, measuring 125 metres (410 ft) on each side. During its construction, the Eiffel Tower surpassed the Washington Monument to become the tallest man-made structure in the world, a title it held for 41 years until the Chrysler Building in New York City was finished in 1930. It was the first structure to reach a height of 300 metres. Due to the addition of a broadcasting aerial at the top of the tower in 1957, it is now taller than the Chrysler Building by 5.2 metres (17 ft). Excluding transmitters, the Eiffel Tower is the second tallest free-standing structure in France after the Millau Viaduct.

Compute

模型结构

1. 概述
   - 代码平台
   - 模型结构
   - 语言任务
2. 特征计算
   - 词汇特征
   - 语义特征
   - 句法特征
3. 其他案例
   - 可控文本生成
   - 群体认知

语言模型的定义：语言模型是预测句子出现概率的模型

通俗解释: 语言模型可以判断一句话有多么常见

P(“我喜欢吃巧克力”) > P(“我喜欢吃桌子”)

P(“猫追老鼠”) > P(“猫老鼠追”)

语言模型的定义：语言模型是预测句子出现概率的模型

通俗解释: 语言模型可以判断一句话有多么常见

语言模型的目标

- 计算一个句子的出现概率:

$$P(S) = P(w_1, w_2, \dots, w_n) = P(w_1) * P(w_2|w_1) * \cdots * P(w_n|w_1, w_2, \dots)$$

- $P(w_n|w_1, w_2, \dots)$ 给定上文之后，一个词汇出现的概率

- **语言模型的实际功能：根据上文或上下文预测下一个词**

模型结构 概述

基于神经网络的语言模型:

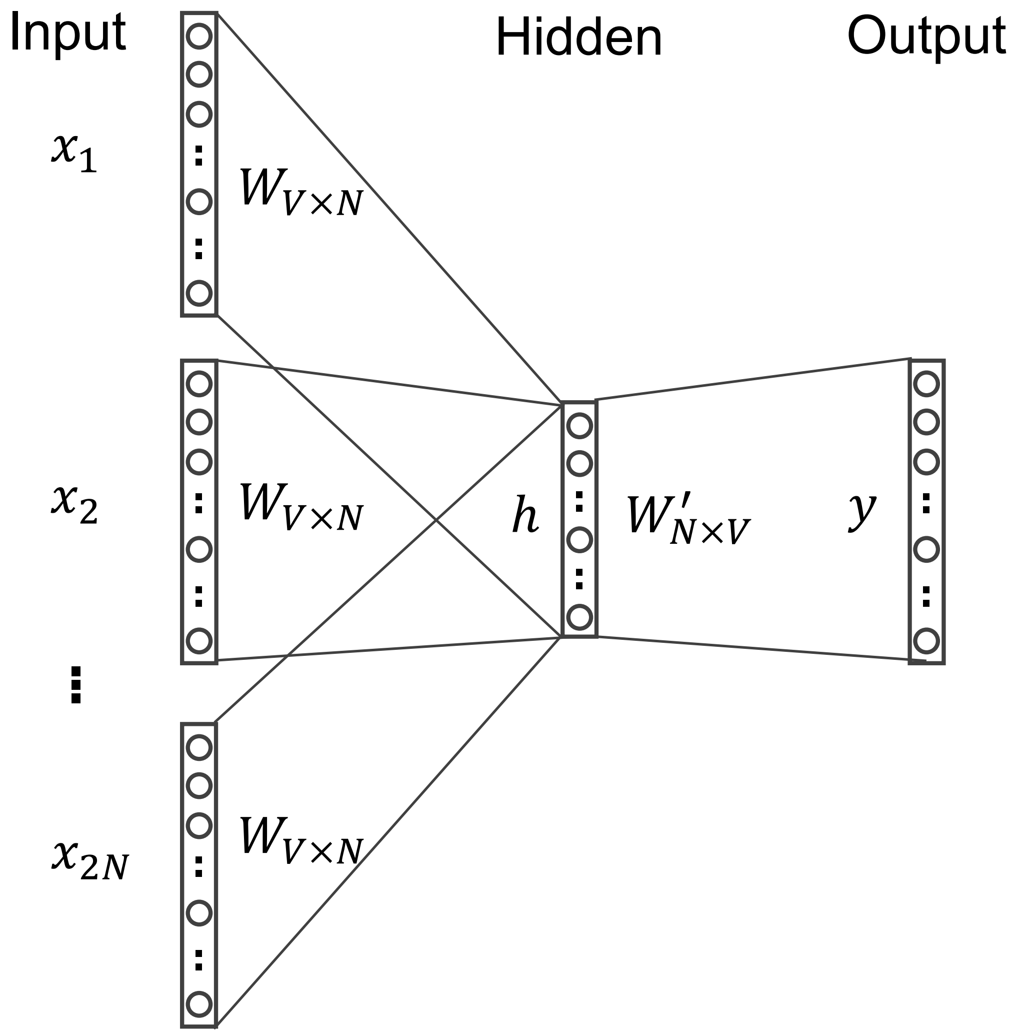
1. 词袋模型: **CBOW**, skip-gram …

2. 循环神经网络模型: **RNN-LM**, LSTM-LM …

3. Transformer模型: **GPT**, BERT…

模型结构　词袋模型

Continuous Bag Of Word (CBOW)



Task: 用前N个和后N个邻近词预测当前词 (N<5)

Input: $2N$个邻近词$x_1, \dots, x_{2N}$

Output: 当前词y

Mikolov, Tomas et al. *arXiv preprint*, 2013.

词袋模型

Continuous Bag Of Word (CBOW)

语料: 欢迎参加工**作**坊

输入 (词向量xW) $\xrightarrow{\text{平均}}$ 隐藏层 (h) $\xrightarrow{\text{分类}}$ 输出 (y)

工	1	2	1	1
坊	1	1	0	2

→

1	1.5	0.5	1.5

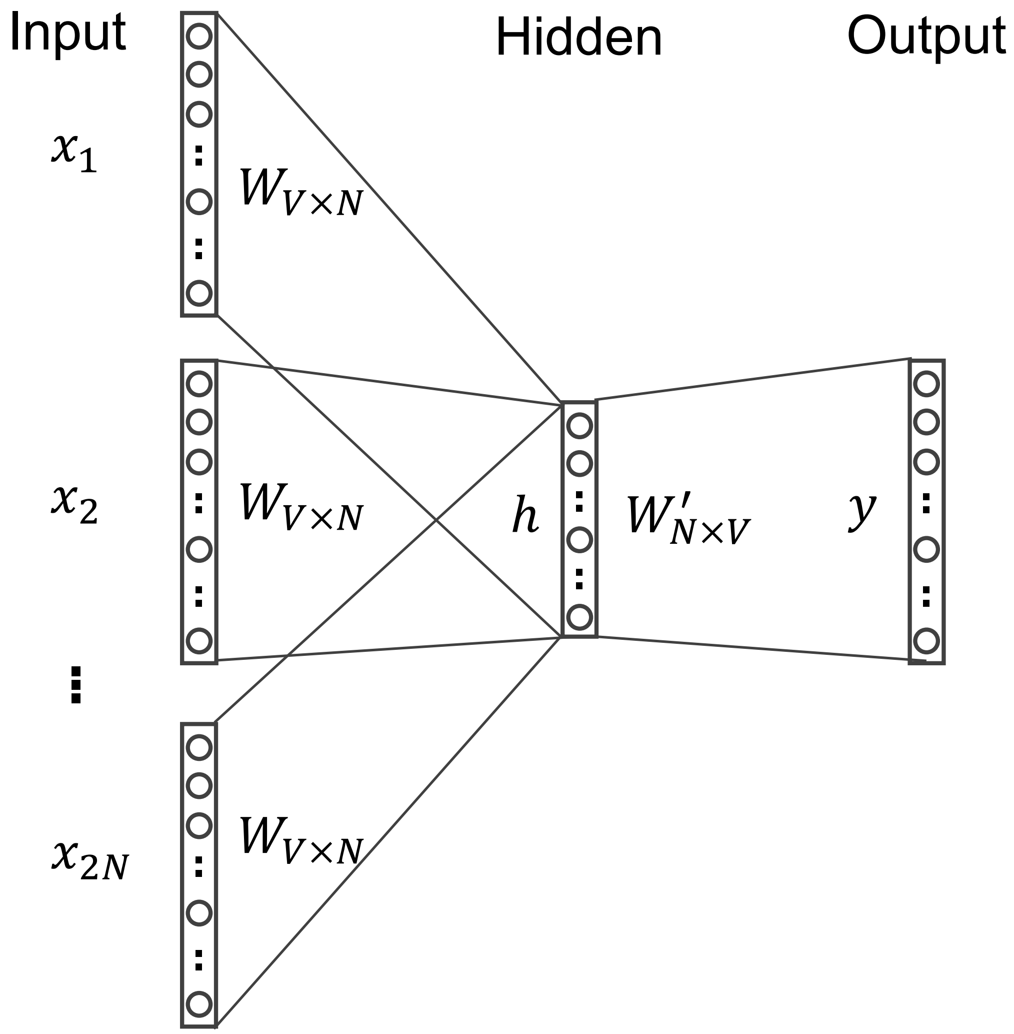
→

作	2	1	0	1

Mikolov, Tomas et al. *arXiv preprint*, 2013.

模型结构　词袋模型

Continuous Bag Of Word (CBOW)



特点:

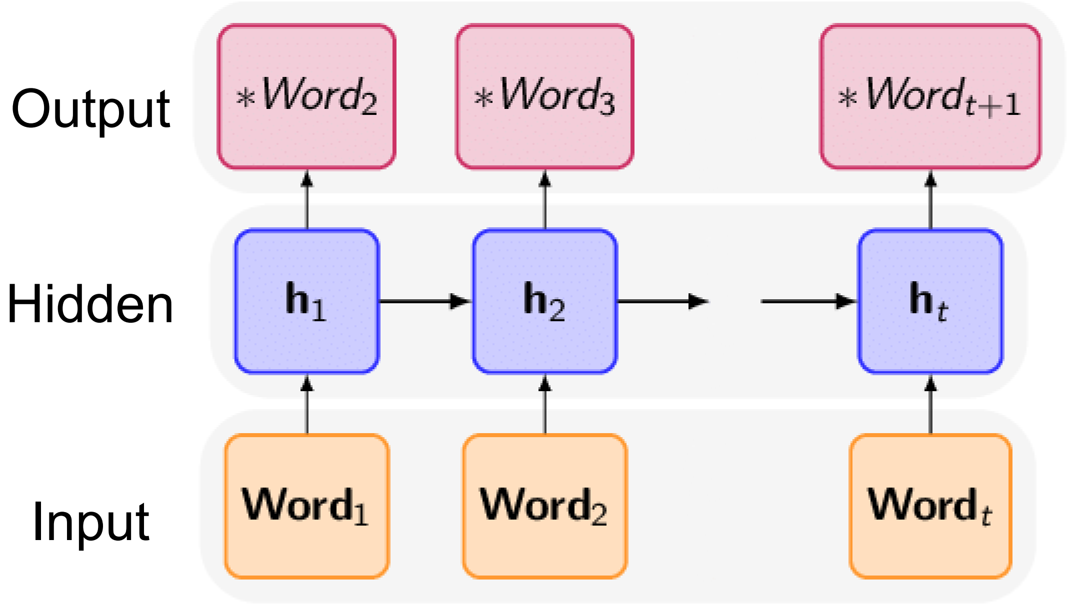
1. 通过映射矩阵W, CBOW可以将单词映射为包含语义信息的词向量 (word2vec)

2. CBOW**不考虑输入单词的顺序信息**。

Mikolov, Tomas et al. *arXiv preprint*, 2013.

模型结构 循环神经网络模型

Recurrent Neural Network (RNN)



Task: 使用历史信息和当前词来预测下一个词

Input: 第t个单词$word_t$, t-1时刻的历史信息h_{t-1}

Output: 第t+1个单词$word_{t+1}$

h_t可连接任务分类器, 完成其他任务 (词性标注、情感分析等)

Mikolov, Tomas, et al. Interspeech. 2010.

循环神经网络模型

Recurrent Neural Network (RNN)

语料: 欢迎参加工作**坊**

输入 (词向量$word_t$) $\xrightarrow{\text{融合历史信息}}$ 隐藏层 (h_t) $\xrightarrow{\text{分类}}$ 输出 (词向量$word_{t+1}$)

作	2	1	0	1

→

1.2	0.5	1.4	2.1

→

坊	1	1	0	2

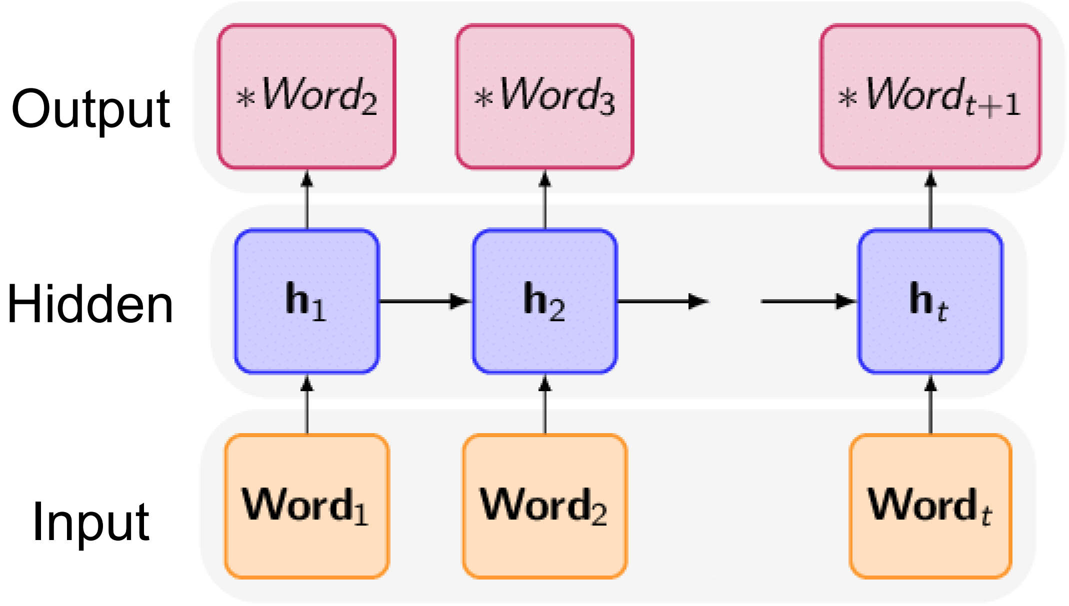
隐藏层 (h_{t-1})

0.8	1.7	0.5	0.2

↗

模型结构 循环神经网络模型

Recurrent Neural Network (RNN)



特点:

1. 模型拥有短期记忆能力，能够记住前N个单词的历史信息 (由于梯度消失等问题，N一般小于10)。

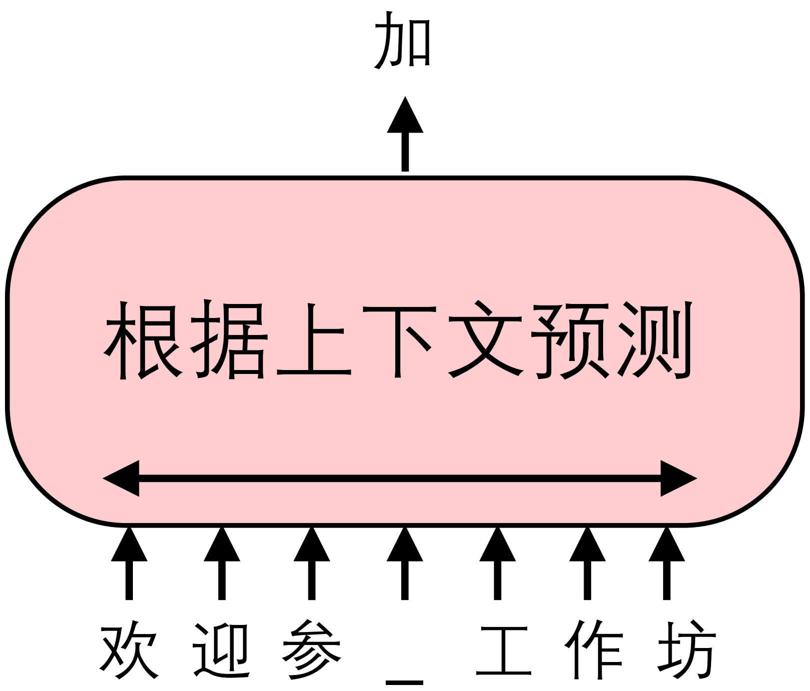
2. 模型只能串行计算，训练速度慢。

Mikolov, Tomas, et al. Interspeech. 2010.

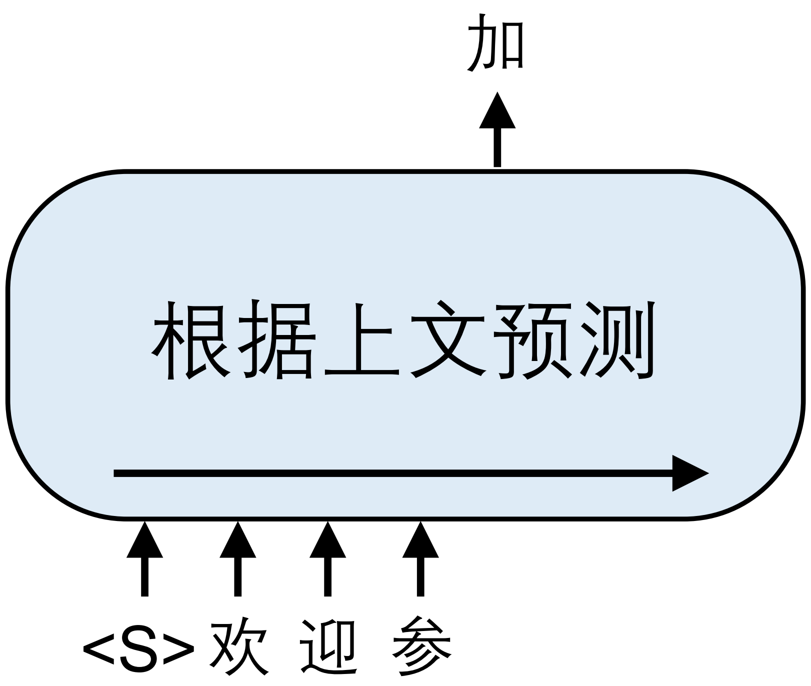
模型结构 Transformer模型

模型分类

上下文(非增量式)模型: BERT等



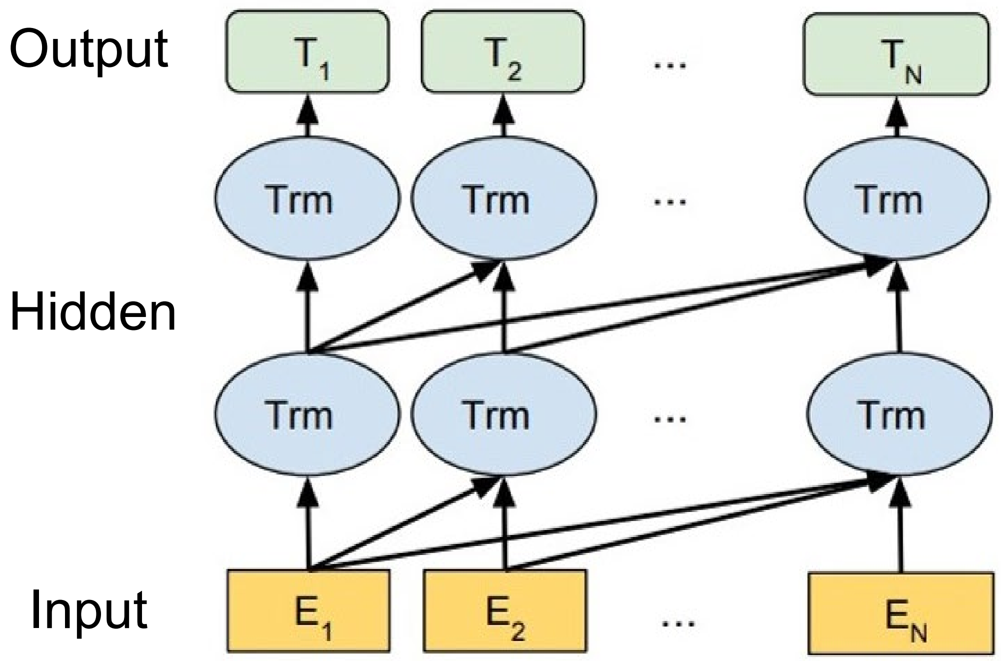
上文(增量式)模型: GPT等



Lewis M, et al. ACL. 2020

模型结构 Transformer模型

Generative Pre-trained Transformer (GPT)



Task: 利用**所有前文**预测下一个词

Input: 第1,…,N个单词$E_1, \ldots, E_N$

Output: 第N+1个单词T_N

GPT的隐藏层可连接任务分类器，完成其他任务 (词性标注、情感分析等)

Radford A, et al. OpenAI. 2018

模型结构　Transformer模型

Generative Pre-trained Transformer (GPT)

语料: 欢迎参加工作**坊**

输入$(E_1, \dots, E_N)$ $\xrightarrow{\text{自注意力机制}}$ 隐藏层(Trm) $\xrightarrow{\text{分类}}$ 输出T_N

欢	3	1	2	6
⋮				
工	1	2	1	1
作	2	1	0	1

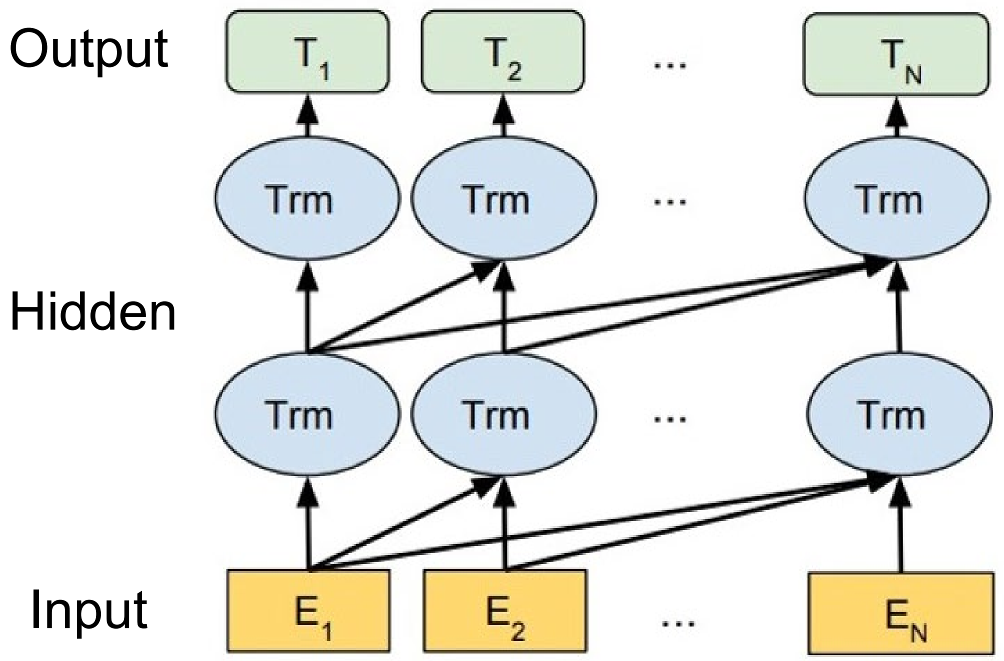
1.2	0.5	1.4	2.1

坊	1	1	0	2

Radford A, et al. OpenAI. 2018

模型结构　Transformer模型

Generative Pre-trained Transformer (GPT)



特点:

1. 可以同时考虑所有前文对当前词的影响, 实现长距离依赖 (最长可达15K)。

2. 可以并行计算，训练速度快。

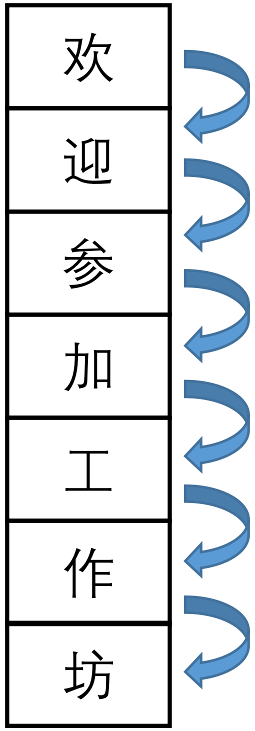
Radford A, et al. OpenAI. 2018

发展趋势

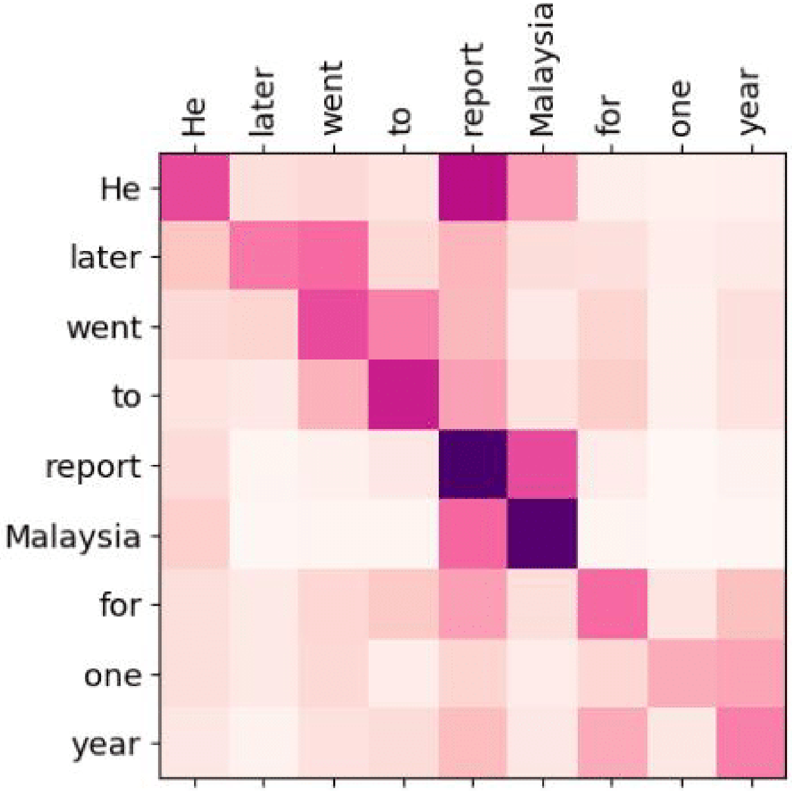
趋势: 更高效的融合上下文信息

词袋模型

循环神经网络模型



Transformer模型



语言任务

1. 概述
    - 代码平台
    - 模型结构
    - 语言任务
2. 特征计算
    - 词汇特征
    - 语义特征
    - 句法特征
3. 其他案例
    - 可控文本生成
    - 群体认知

语言任务　任务分类

其他语言任务可粗略分为两类：**文本分类** 和 **文本生成**

文本序列 -> 类别标签

- 自然语言推理
- 多项选择
- **词性标注**
- **主题分析**
- 指代消解
- 意图识别
- 关系抽取
- **句法解析**

…

文本序列 -> 文本序列

- 机器翻译
- 对话系统
- 风格迁移
- 条件续写
- **可控文本生成**
- 自动文摘
- 语法纠正
- 开放域问答

…

如何使用语言模型完成其他语言任务?

预训练+微调

训练任务: 预测词

训练语料：维基百科、人民日报等

上下文学习 (in context learning)

给定任务描述和示例, chatGPT等模型能在不修改参数的前提下从上下文中学习任务。

将中文翻译为英文	根据关键词生成文本	← *task description*
吃晚饭 => eat dinner	经济 => 全球GDP增长率为2.7%	← *example*
工作坊 =>	娱乐 =>	← *prompt*

思维链 (Chain of Thought)

在示例中提供推理过程有助于模型正确回答问题。

(a) Few-shot

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?
A: The answer is 11.

Q: A juggler can juggle 16 balls. Half of the balls are golf balls, and half of the golf balls are blue. How many blue golf balls are there?
A:

(Output) The answer is 8. ✗

(b) Few-shot-CoT

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?
A: Roger started with 5 balls. 2 cans of 3 tennis balls each is 6 tennis balls. 5 + 6 = 11. The answer is 11.

Q: A juggler can juggle 16 balls. Half of the balls are golf balls, and half of the golf balls are blue. How many blue golf balls are there?
A:

(Output) The juggler can juggle 16 balls. Half of the balls are golf balls. So there are 16 / 2 = 8 golf balls. Half of the golf balls are blue. So there are 8 / 2 = 4 blue golf balls. The answer is 4. ✓

Wei J, et al. *NIPS*, 2022.

语言任务 代码演示

如何使用huggingface平台完成主题分析任务

1. 从huggingface官网上找到对应的任务模型

语言任务　代码演示

如何使用huggingface平台完成主题分析任务

2. 加载模型

```python
from transformers import AutoTokenizer, AutoModelForSequenceClassification, pipeline
model_path = 'uer/roberta-base-finetuned-chinanews-chinese'
tokenizer = AutoTokenizer.from_pretrained(model_path)
model = AutoModelForSequenceClassification.from_pretrained(model_path)
```

3. 模型计算

```python
text_classification = pipeline('sentiment-analysis', model=model, tokenizer=tokenizer)
res = text_classification('欢迎参加工作坊！')[0]

print(f"\nInput: {text}\nPrediction: {res['label']}, Score: {res['score']:.3f}")
```

✓ 0.9s　Python

```
Input: 欢迎参加工作坊!
Prediction: culture, Score: 0.723
```

小结

1. 语言模型的基本目标: 预测词

- 不同结构的语言模型, 主要区别在于如何融合上下文。

2. 通过预训练+微调、上下文学习等范式, 语言模型能够学会其他语言任务

- 预训练+微调: 常见语言任务 / 有足够标注数据
- 上下文学习: 缺少标注数据 / 有多任务设置需求

词汇特征

1. 概述
    - 代码平台
    - 模型结构
    - 语言任务

2. 特征计算
    - 词汇特征
    - 语义特征
    - 句法特征

3. 其他案例
    - 可控文本生成
    - 群体认知

特征计算

为什么大家需要掌握特征的计算方法？

特征计算

为什么大家需要掌握特征的计算方法？

Comment from a reviewer:

The paper compares the DNN-derived features to word-related features such as frequency and surprisal.

Word frequency is estimated from the BNC corpus. **The BNC corpus contains 100 Million words, whereas DNN was trained on several Billions of words.**

Surprisal is derived from the same corpus and derived using a trigram model. **Trigram models are now far surpassed by DNN-based language models.**

This means that the predictive power of the word-related features is likely to be **underestimated**. The authors could draw on **(1) frequency estimated from the corpora used for DNN (BookCorpus + Wikipedia), (2) estimate surprisal from a strong off-the-shelf model such as GPT-2.**

特征计算

为什么大家需要掌握特征的计算方法？

Comment from a reviewer:

The paper compares the DNN-derived features to word-related features such as frequency and surprisal.

Word frequency i… …ontains 100 **Million words, wh… …ls.**

Surprisal is derive… …nodel. **Trigram models are now f…**

标准化、自主化、先进化

This means that th… …y to be **underestimated**. The authors could draw on **(1) frequency estimated from the corpora used for DNN (BookCorpus + Wikipedia), (2) estimate surprisal from a strong off-the-shelf model such as GPT-2.**

词汇特征

1. 概述
   - 代码平台
   - 模型结构
   - 语言任务
2. 特征计算
   - 词汇特征
   - 语义特征
   - 句法特征
3. 其他案例
   - 可控文本生成
   - 群体认知

a 词频

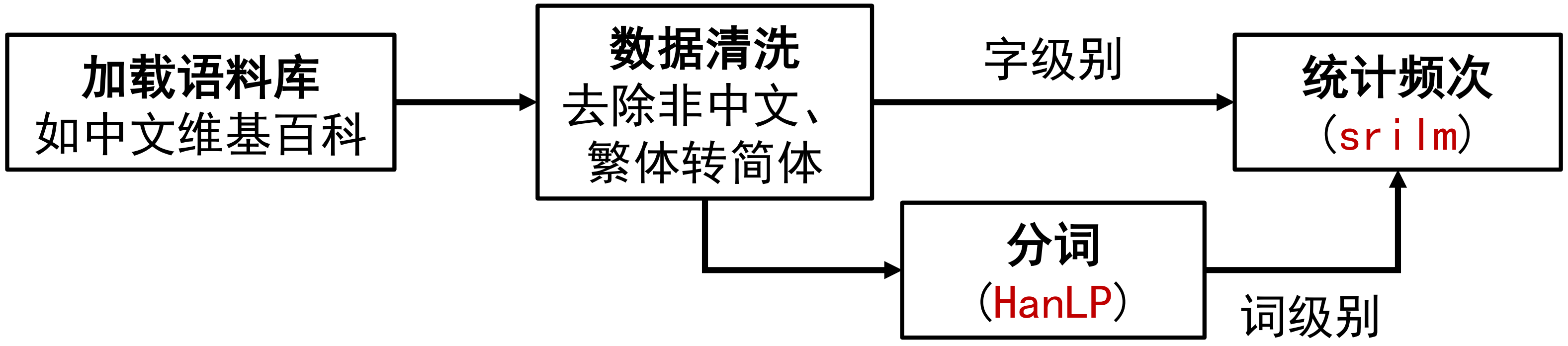
b 转移概率

c 词性

特征计算方法

- 词频 ■ 转移概率 ■ 词性

计算流程



中文语料库集合: https://github.com/InsaneLife/ChineseNLPCorpus

HanLP: https://github.com/InsaneLife/ChineseNLPCorpus

srilm: http://www.speech.sri.com/projects/srilm/

词汇特征　特征计算方法

- 词频　■ 转移概率　■ 词性

计算代码实操

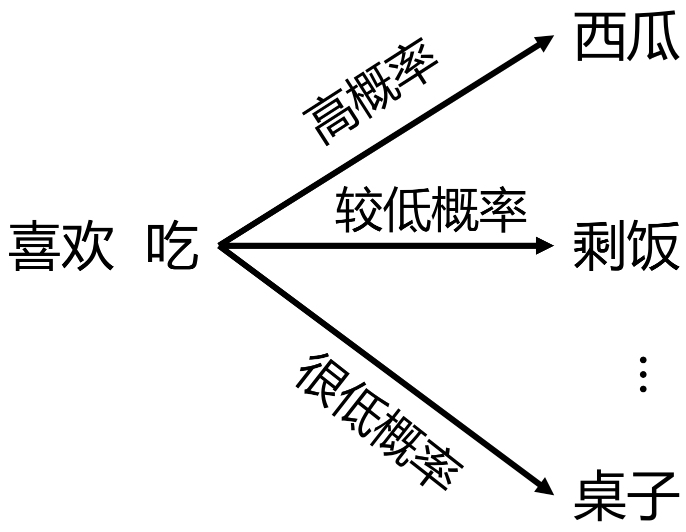
特征计算方法

■ 词频 ■ 转移概率 ■ 词性

喜欢 吃 ⟶ ?

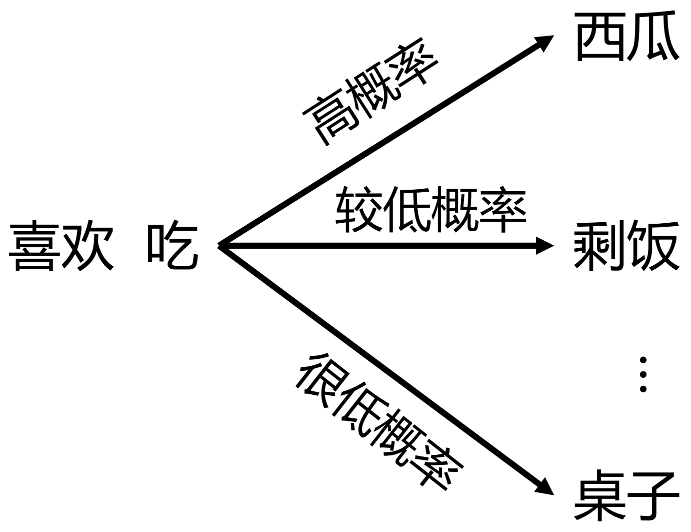
词汇特征 特征计算方法

■ 词频 ■ 转移概率 ■ 词性



词汇特征 特征计算方法

■ 词频　■ 转移概率　■ 词性

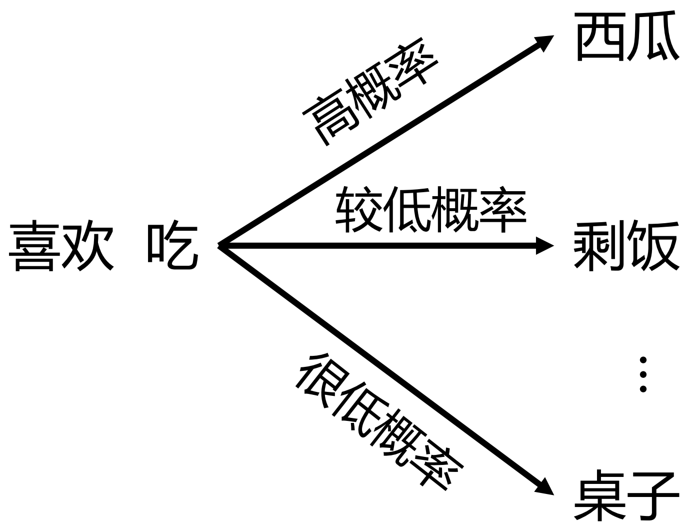


转移概率 (3-gram)

P(西瓜 | 吃, 喜欢)

词汇特征　特征计算方法

■ 词频　■ **转移概率**　■ 词性



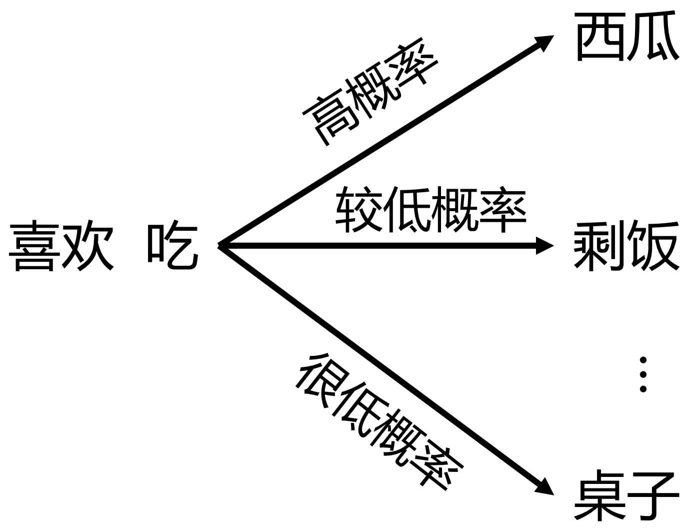
转移概率 (3-gram)

P(西瓜 | 吃, 喜欢)

惊异度 (surprisal): 后文出现特定词的惊异程度

s = - log(P(西瓜 | 吃, 喜欢))

特征计算方法



转移概率 (3-gram)

P(西瓜 | 吃, 喜欢)

惊异度 (surprisal): 后文出现特定词的惊异程度

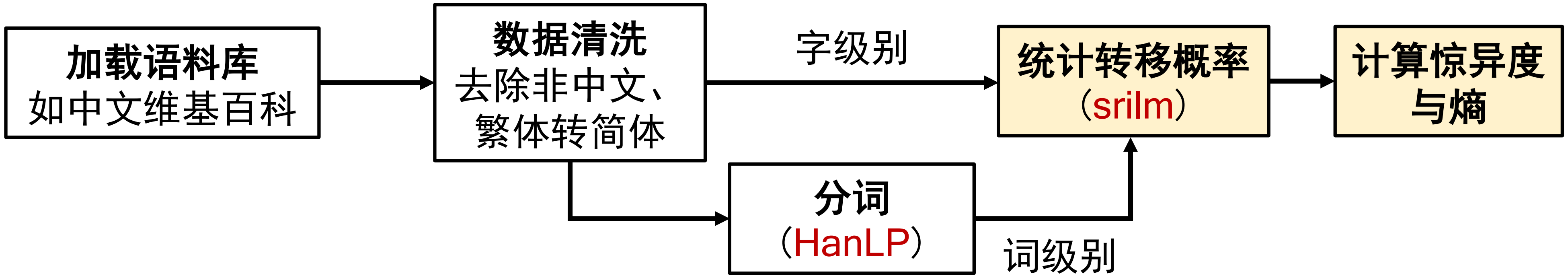
s = - log(P(西瓜 | 吃, 喜欢))

熵 (entropy): 后文的不确定度

$e = -\sum_{i=1}^{I} P(\mathrm{w}_i|\text{吃, 喜欢}) \times \log(P\,(w_i|\text{吃, 喜欢})$

特征计算方法

计算流程 (基于N-gram)



特征计算方法

■ 词频　　■ 转移概率　　■ 词性

计算代码实操 (基于N-gram)

特征计算方法

■ 词频 ■ 转移概率 ■ 词性

N-gram的局限性

- 基于词串匹配，可考虑的词串长度有限 (一般小于5)
- 无法捕获语义相似等信息

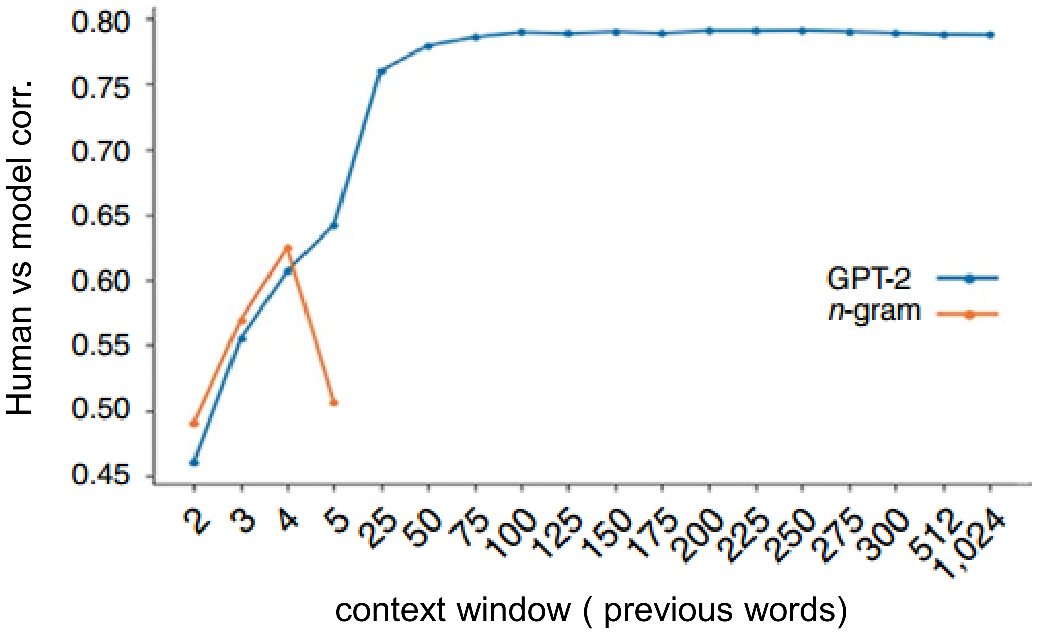
词汇特征 特征计算方法

■ 词频 ■ 转移概率 ■ 词性

N-gram的局限性

- 基于词串匹配，可考虑的词串长度有限 (一般小于5)
- 无法捕获语义相似等信息

N-gram、GPT与人一致程度比较

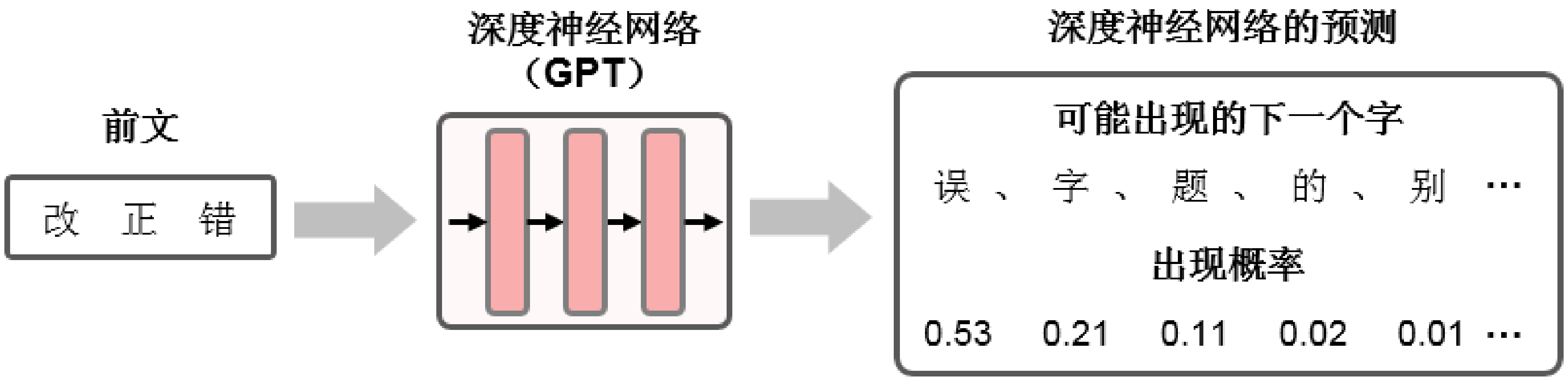


Goldstein, A. et al. *Nat. Neurosci.*, 2022.

特征计算方法

■ 词频　■ 转移概率　■ 词性

计算流程 (基于GPT)



特征计算方法

■ 词频 ■ 转移概率 ■ 词性

计算流程 (基于GPT)

选择模型（HuggingFace） → 读取下一个字的概率 → 计算惊异度与熵

计算代码实操 (基于GPT)

词汇特征

特征计算方法

- 词频
- 转移概率
- 词性

这个门被锁了，锁很难被打开。

VV NN

特征计算方法

■ 词频　　■ 转移概率　　■ 词性

计算流程

分词（HanLP） → 词性标注（HanLP）

计算代码实操

语义特征

1. 概述
    - 代码平台
    - 模型结构
    - 语言任务
2. 特征计算
    - 词汇特征
    - 语义特征
    - 句法特征
3. 其他案例
    - 可控文本生成
    - 群体认知

a 共现矩阵

b 词向量

c 上下文有关的语义表征

特征计算方法

■ 共现矩阵 ■ 词向量 ■ 上下文有关

计算流程

例句

小明喜欢吃西瓜。小明喜欢打篮球。小明经常去花店。

上下文关系 (size=1)

小明喜欢

。小明喜欢

。小明经常

“小明”的共现向量

频次	小明	。	喜欢	吃	西瓜	打	篮球	经常	去	花	店
小明	0	2	2	0	0	0	0	1	0	0	0

语义特征　特征计算方法

■ 共现矩阵　■ 词向量　■ 上下文有关

计算流程

开源词向量

https://github.com/Embedding/Chinese-Word-Vectors

https://nlp.stanford.edu/projects/glove/

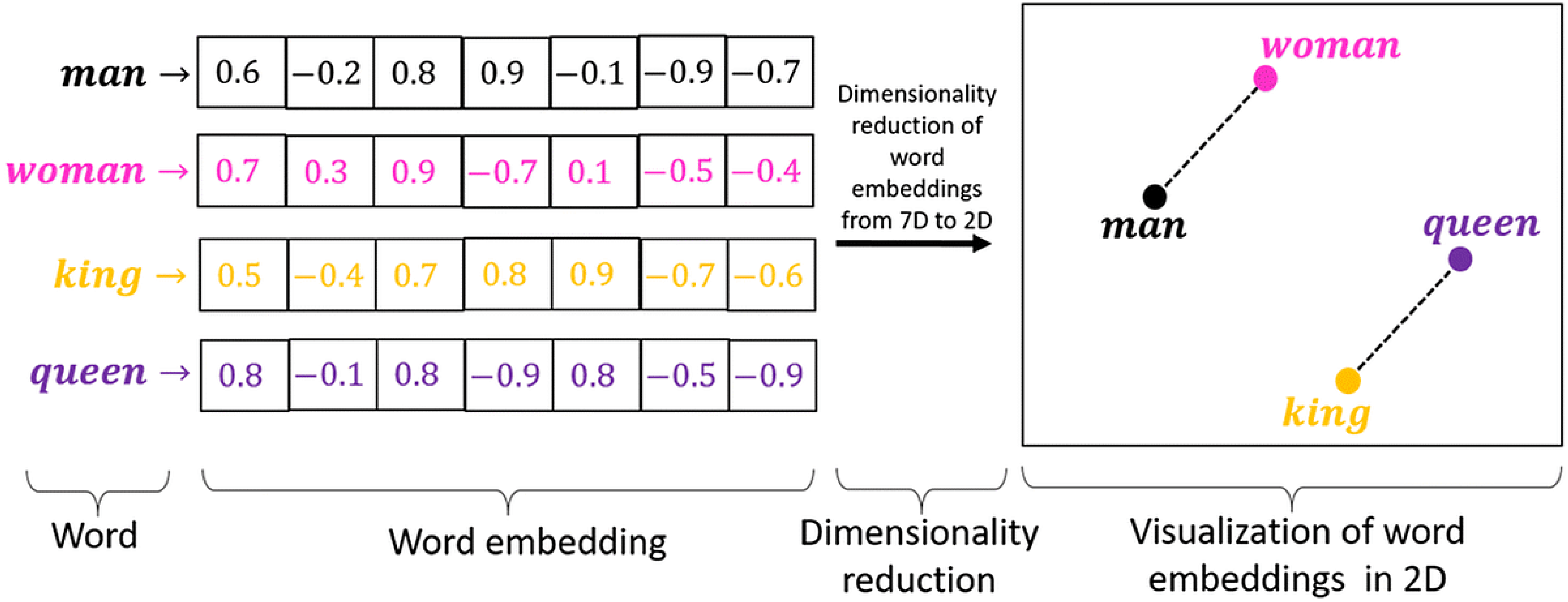
https://drive.google.com/file/d/0B7XkCwpI5KDYNlNUTTlSS21pQmM/edit?resourcekey=0-wjGZdNAUop6WykTtMip30g

Corpus	Context Features			
	Word	Word + Ngram	Word + Character	Word + Character + Ngram
Baidu Encyclopedia 百度百科	300d	300d	300d	300d / PWD: 5555
Wikipedia_zh 中文维基百科	300d	300d	300d	300d
People's Daily News 人民日报	300d	300d	300d	300d
Sogou News 搜狗新闻	300d	300d	300d	300d
Financial News 金融新闻	300d	300d	300d	300d
Zhihu_QA 知乎问答	300d	300d	300d	300d
Weibo 微博	300d	300d	300d	300d
Literature 文学作品	300d	300d / PWD: z5b4	300d	300d / PWD: yenb
Complete Library in Four Sections 四库全书*	300d	300d	NAN	NAN
Mixed-large 综合 Baidu Netdisk / Google Drive	300d 300d	300d 300d	300d 300d	300d 300d

语义特征　特征计算方法

■ 共现矩阵　■ 词向量　■ 上下文有关

计算流程



https://medium.com/@hari4om/word-embedding-d816f643140

特征计算方法

■ 共现矩阵 ■ 词向量 ■ 上下文有关

计算代码实操

语义特征　特征计算方法

■ 共现矩阵　■ 词向量　■ 上下文有关

计算流程

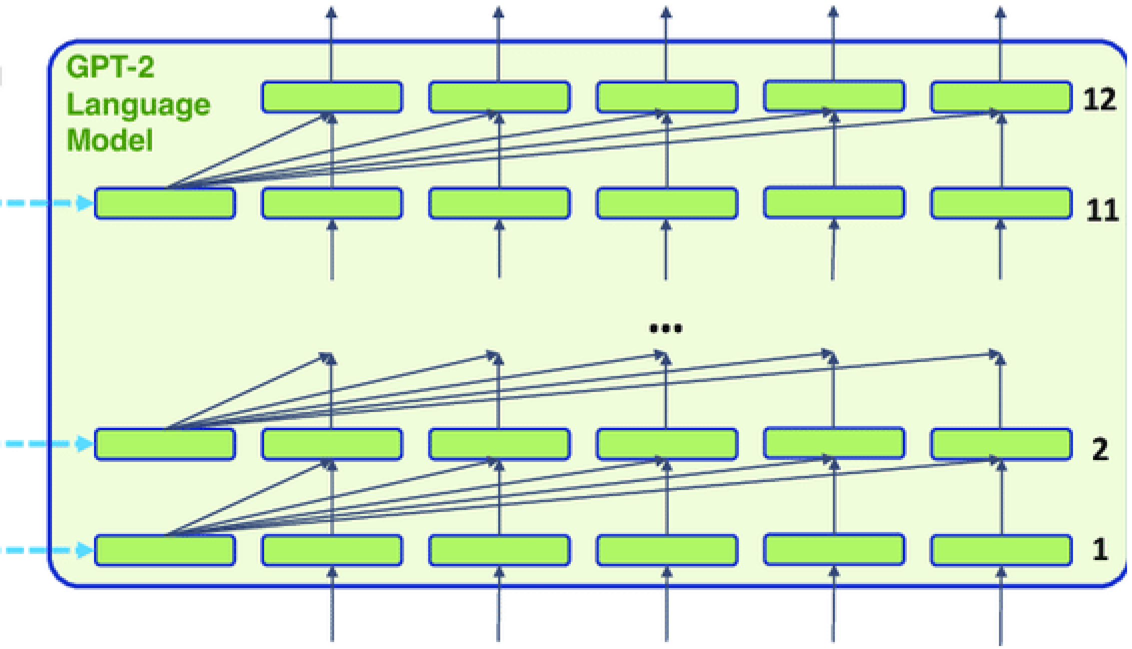
模型输出

第11层激活

⋮

第2层激活

第1层激活



模型输入

语义特征

特征计算方法

- 共现矩阵
- 词向量
- 上下文有关

计算流程

选择模型（HuggingFace） → 读取模型隐藏层表征

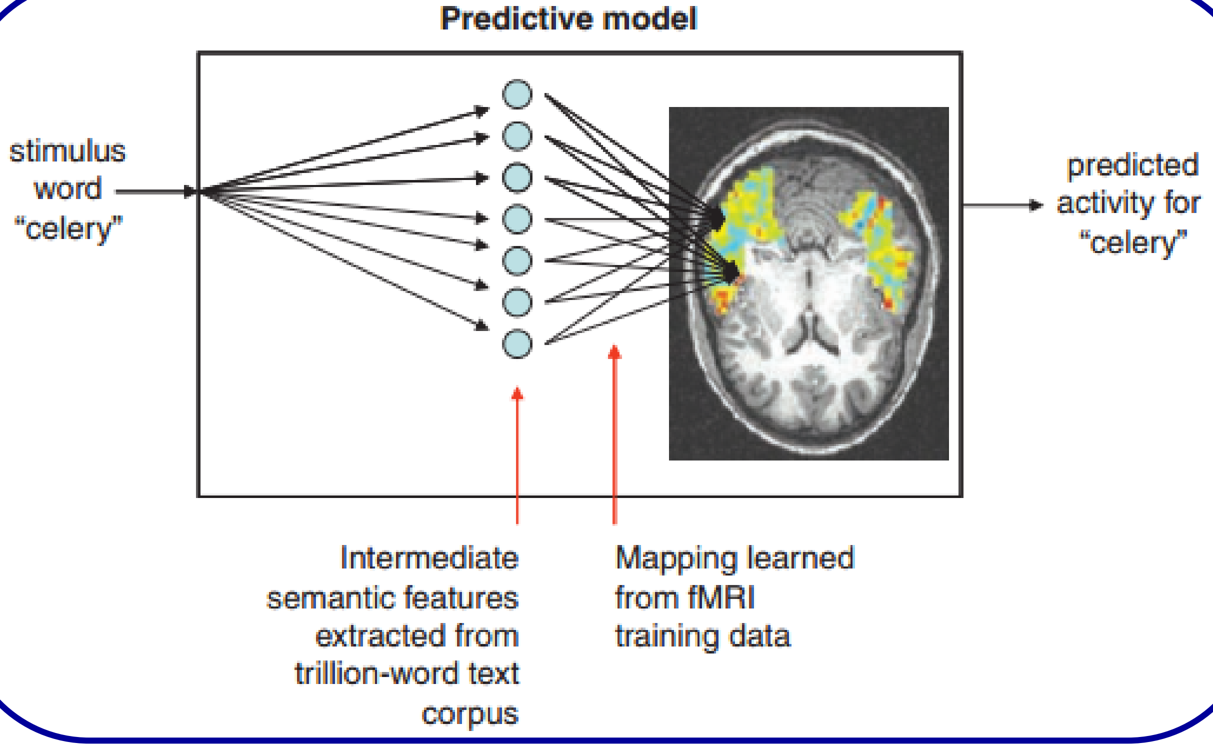
计算代码实操

语义特征 研究案例

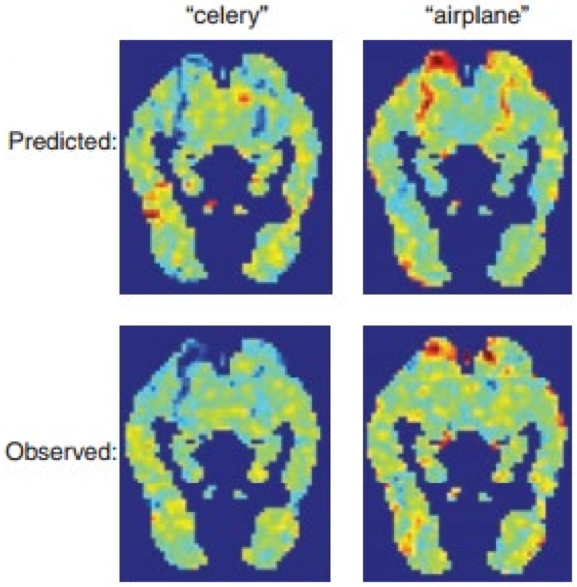
■ 共现矩阵 ■ 词向量 ■ 上下文有关

案例1：基于共现矩阵预测词的神经响应

如何应用



研究结果



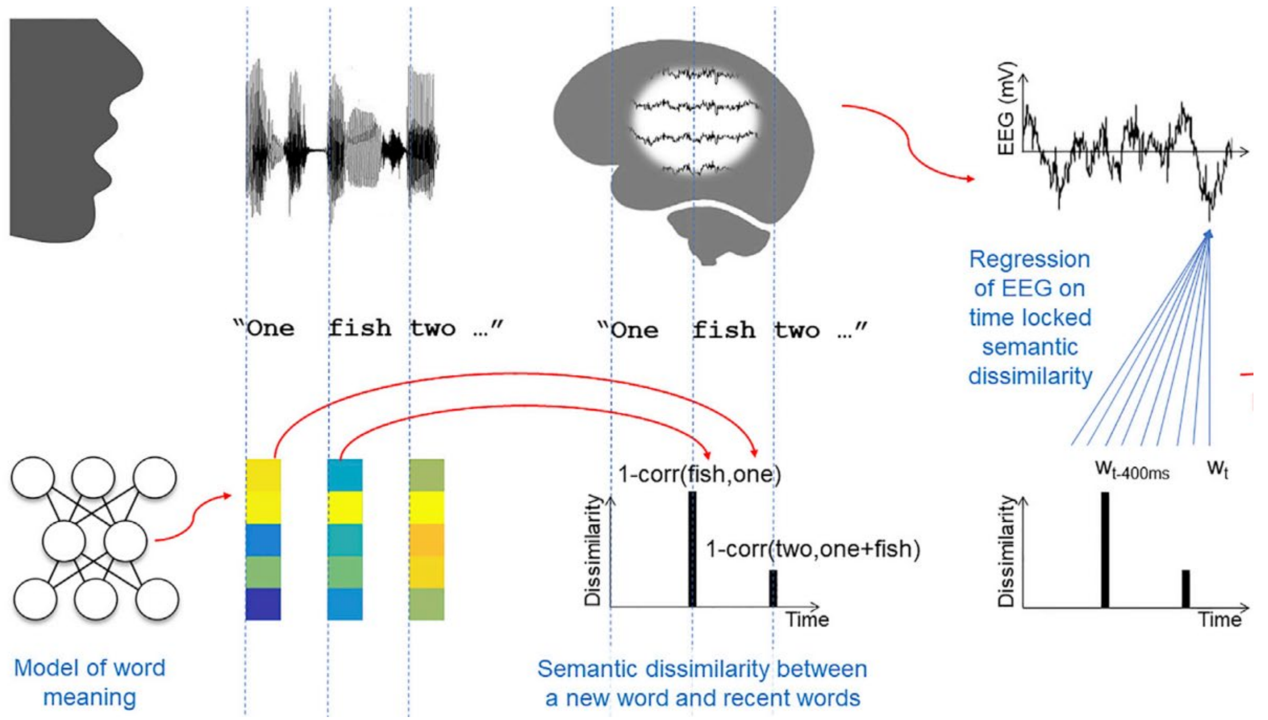
Mitchell, T. et al. *Science*, 2008.

语义特征　研究案例

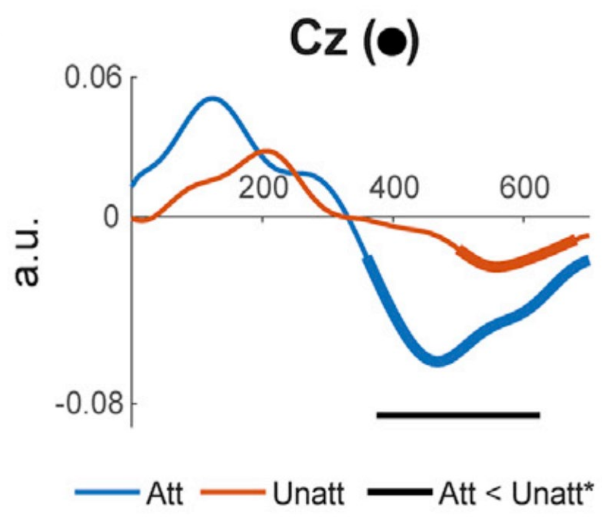
■ 共现矩阵　■ 词向量　■ 上下文有关

案例2：连续语音语义特征的神经编码

如何应用



研究结果



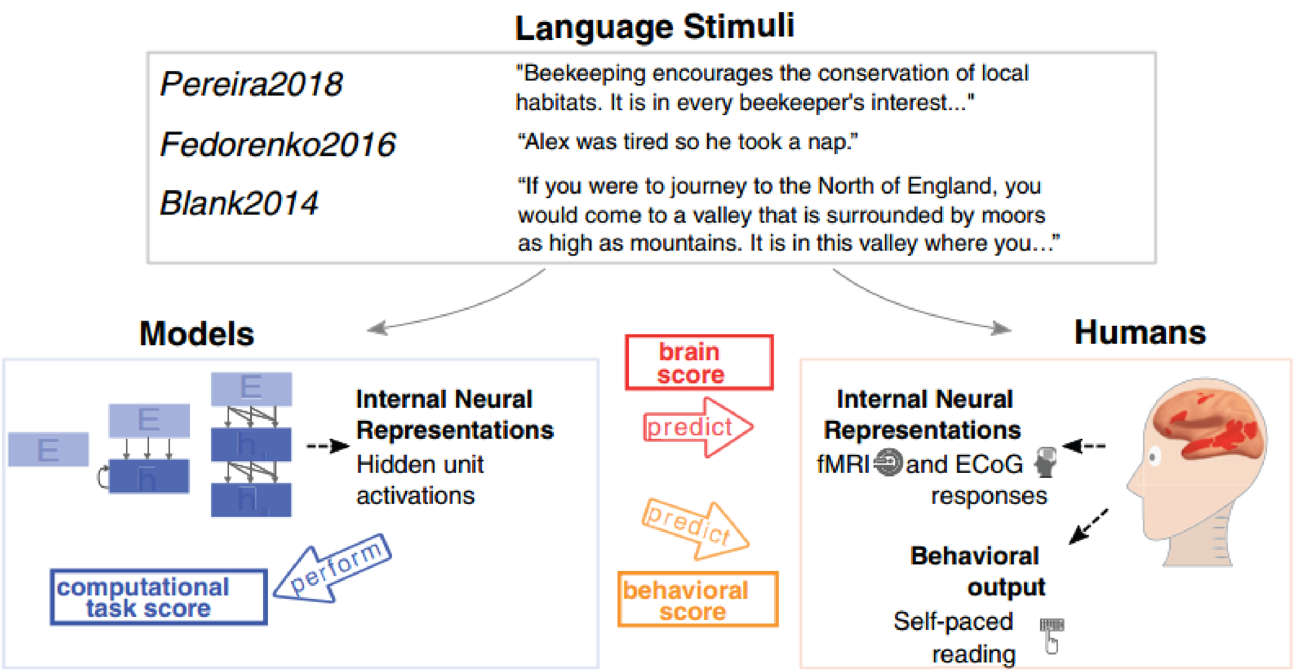
Broderick, M. et al. *Current Biology*, 2018.

语义特征　研究案例

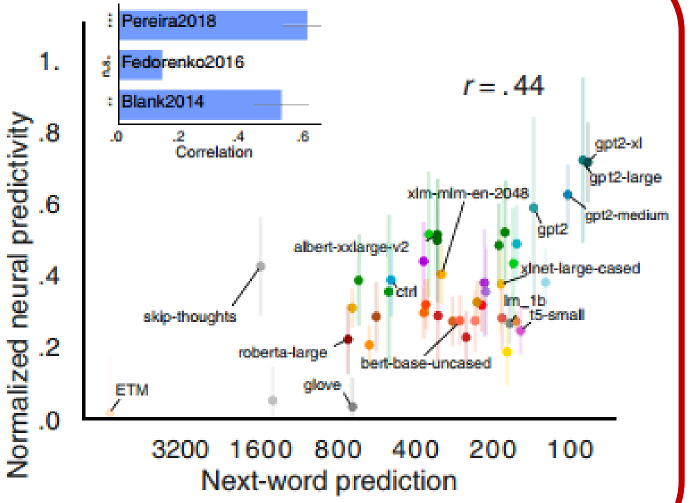
■ 共现矩阵　■ 词向量　■ 上下文有关

案例3：DNN语义表征建模行为与神经响应 -> 预测下一个字的计算机制

如何应用



研究结果



Schrimpf, M. et al. *PNAS*, 2021.

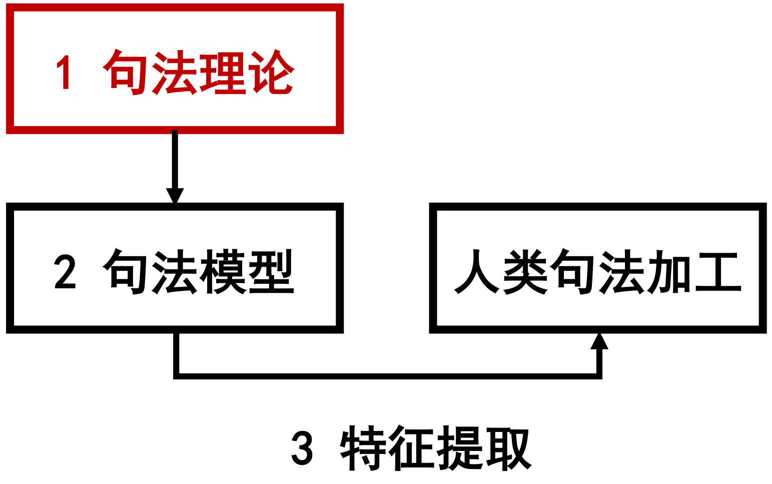
句法特征

1. 概述
   - 代码平台
   - 模型结构
   - 语言任务
2. 特征计算
   - 词汇特征
   - 语义特征
   - 句法特征
3. 其他案例
   - 可控文本生成
   - 群体认知

句法特征　句法解析

句法理论

- **成分句法**（Constituency grammar）
- **依存句法**（Dependency grammar）
- 其他句法
  - 组合范畴句法（Combinational category grammar）
  - Minimalist grammar



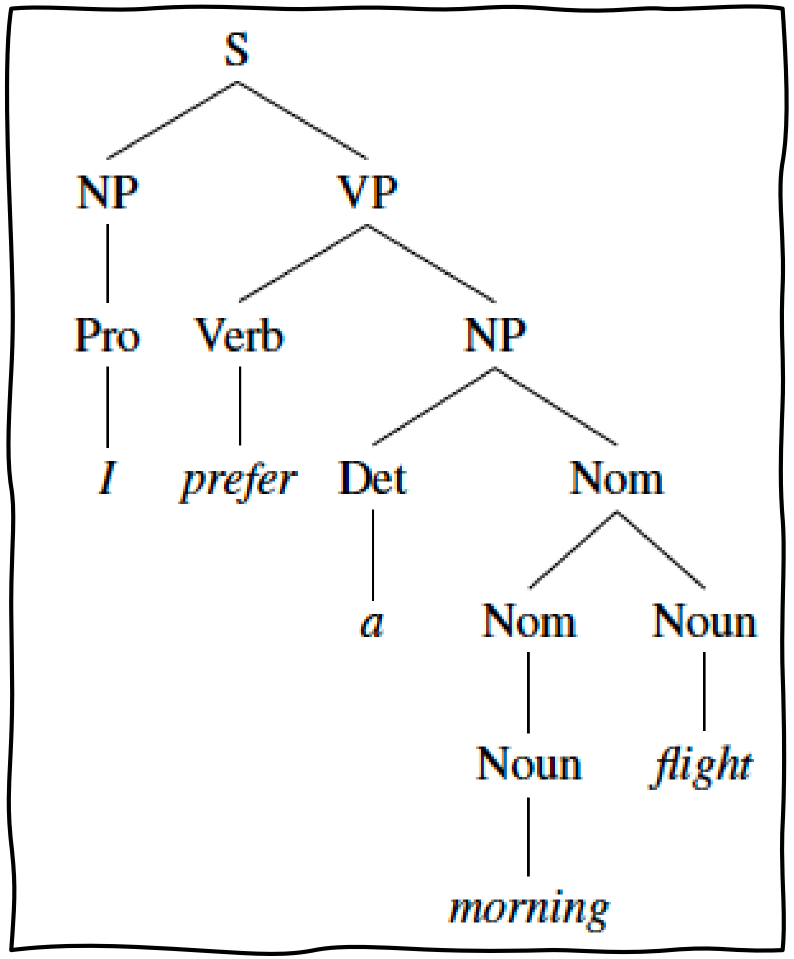
[Speech and Language Processing (3rd ed.)](), [Universal Dependencies]()

句法特征　句法解析

句法理论

- **成分句法**（Constituency grammar）
  - 成分（Constituency）
  - 上下文无关文法（Context-free grammars，CFGs）

Grammar Rules		
S	→	*NP VP*
NP	→	*Pronoun*
	|	*Proper-Noun*
	|	*Det Nominal*
Nominal	→	*Nominal Noun*
	|	*Noun*
VP	→	*Verb*
	|	*Verb NP*
	|	*Verb NP PP*
	|	*Verb PP*
PP	→	*Preposition NP*



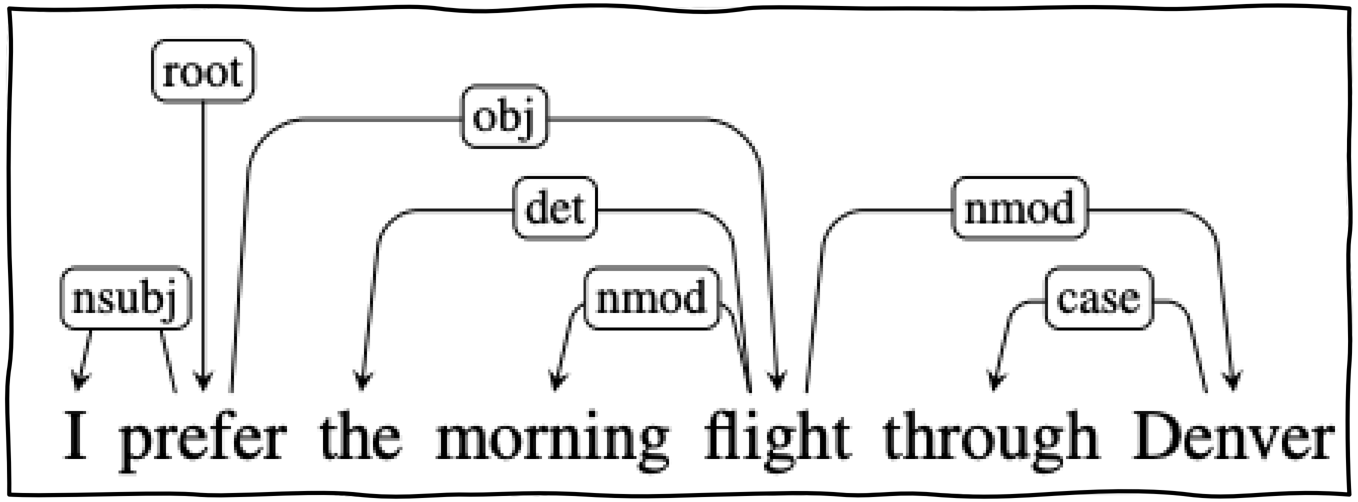
括号表示方法: $[_S\ [_{NP}\ [_{Pro}\ \textbf{I}]]\ [_{VP}\ [_V\ \textbf{prefer}]\ [_{NP}\ [_{Det}\ \textbf{a}]\ [_{Nom}\ [_{Nom}\ [_N\ \textbf{morning}]]\ [_N\ \textbf{flight}]\ ]]]]$

Speech and Language Processing (3rd ed.), Universal Dependencies

句法特征　句法解析

句法理论

- **依存句法**（Dependency grammar）
  - 核心词（Head）、依存词（Dependent）、依存关系（Dependency Relation）
  - 投射性（Projectivity）：交叉依存



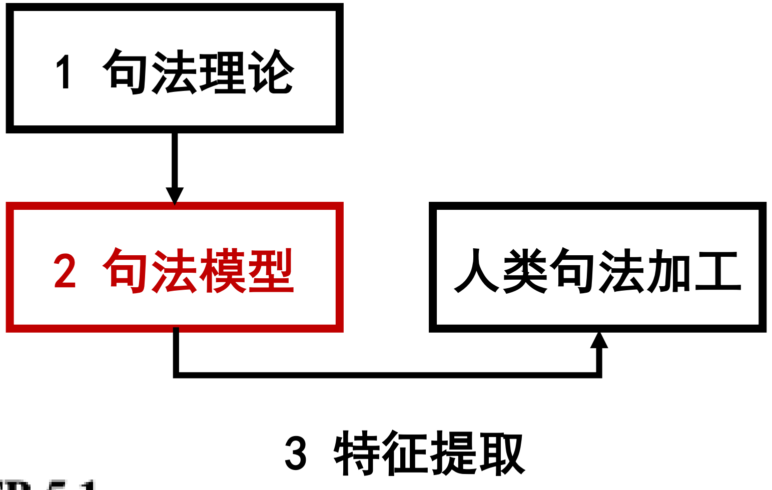
Speech and Language Processing (3rd ed.), Universal Dependencies

句法特征　句法解析

句法解析模型

- 数据: 句法树库（Treebanks）

- **模型表现**
  - 英文：接近 95% 准确率
  - 中文：接近 90% 准确率



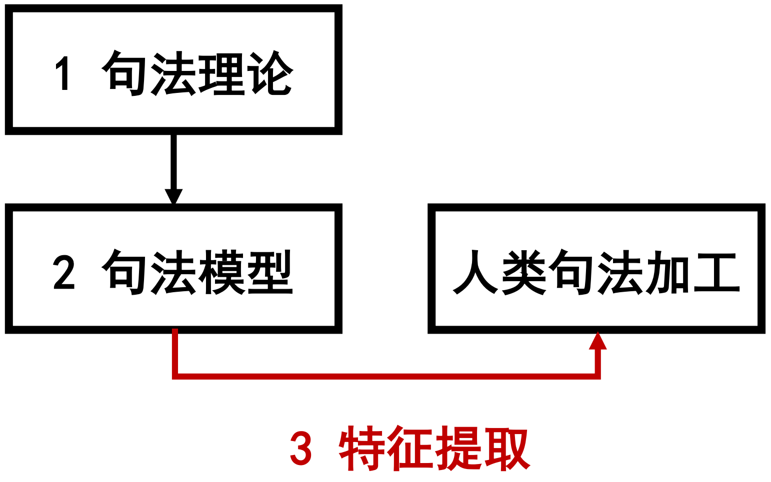
Type	Model	English PTB-SD 3.3.0 UAS	English PTB-SD 3.3.0 LAS	Chinese PTB 5.1 UAS	Chinese PTB 5.1 LAS
Transition	Ballesteros et al. (2016)	93.56	91.42	87.65	86.21
	Andor et al. (2016)	94.61	92.79	–	–
	Kuncoro et al. (2016)	**95.8**	**94.6**	–	–
Graph	Kiperwasser & Goldberg (2016)	93.9	91.9	87.6	86.1
	Cheng et al. (2016)	94.10	91.49	88.1	85.7
	Hashimoto et al. (2016)	94.67	92.90	–	–
	Deep Biaffine	95.74	94.08	**89.30**	**88.23**

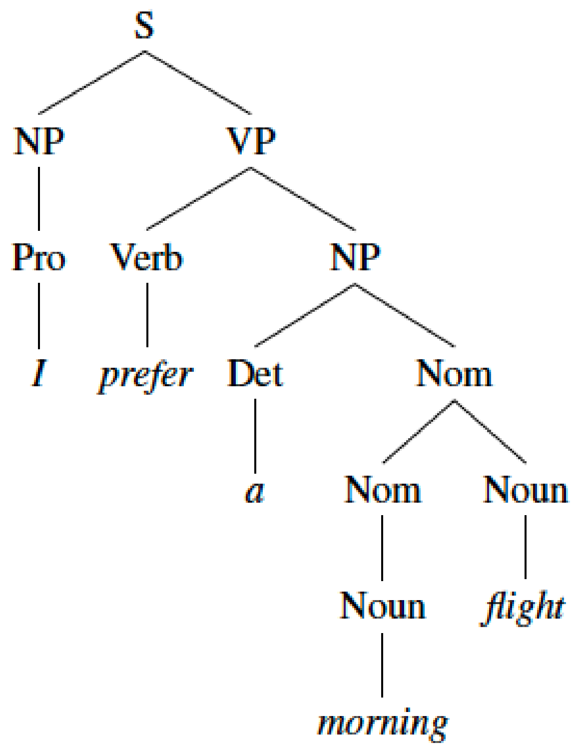
Speech and Language Processing (3rd ed.), Universal Dependencies

句法特征　句法解析

定义句法特征：结构复杂性

- Tree depth
- Distance of dependency (Wilson, 2018)
- **Structure building** (Brennan, 2012, Nelson, 2017)
- Parsing steps (Top-down/bottom-up) (Brennan, 2023)
- …



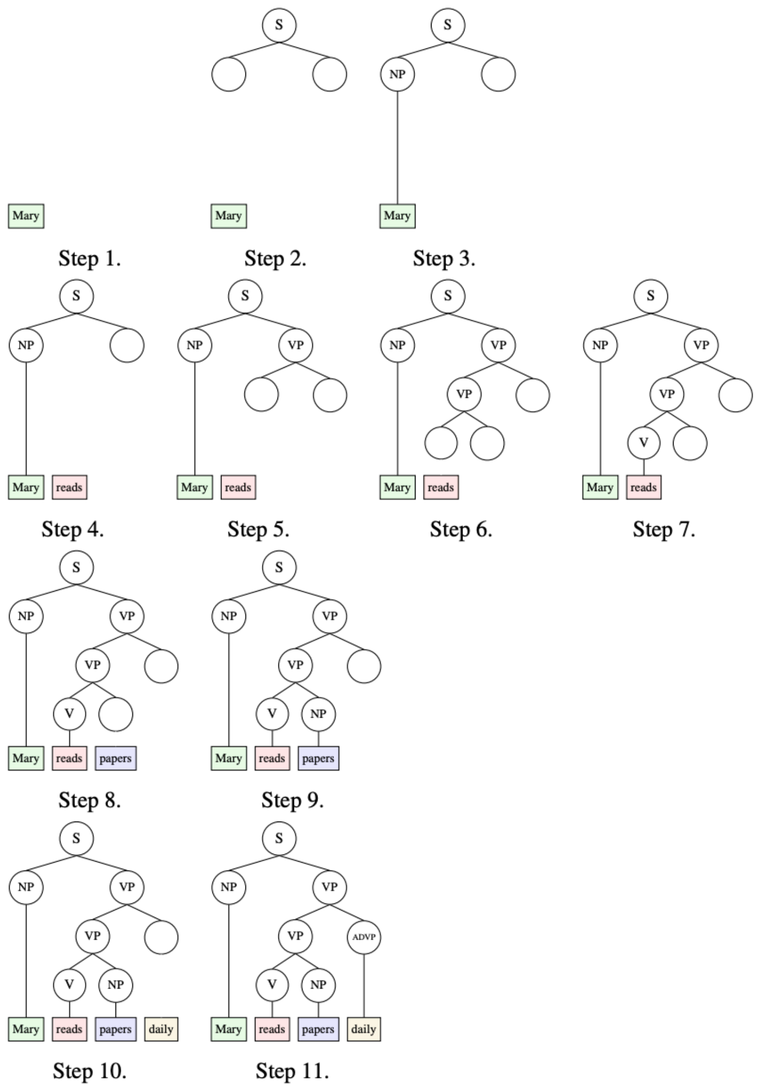


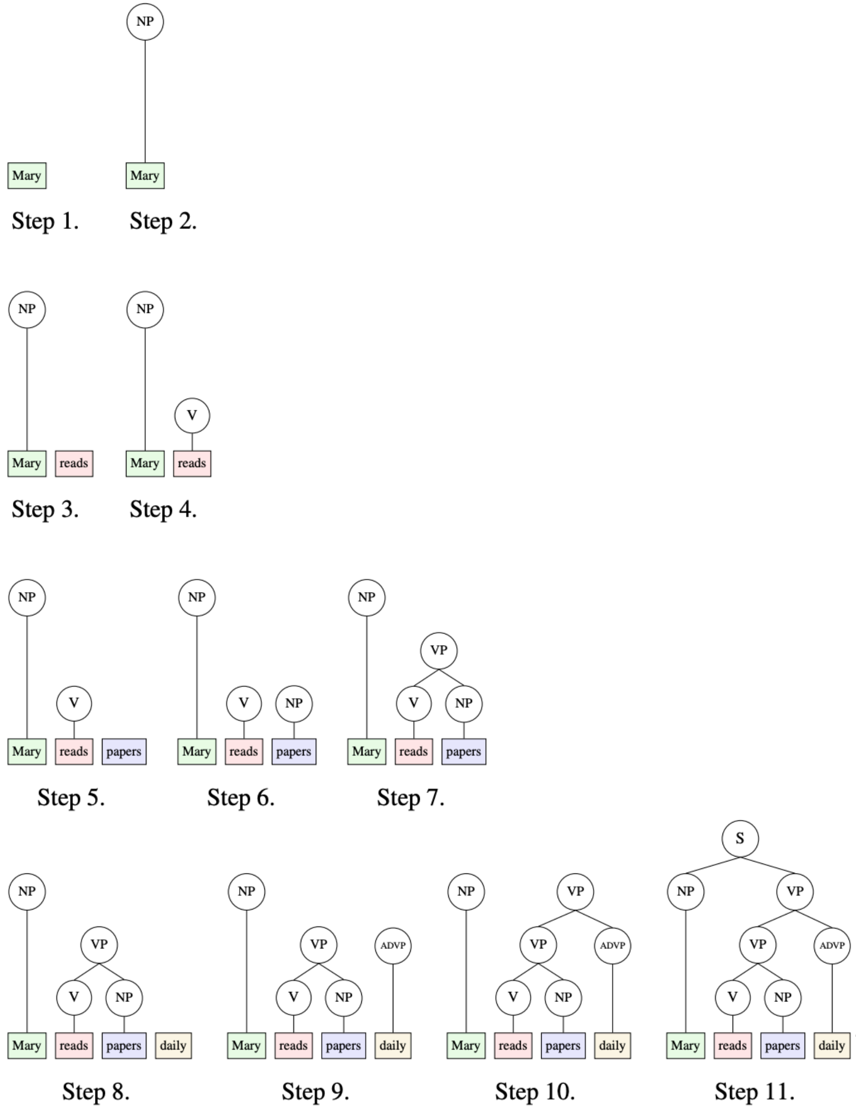
$[_S\ [_{NP}\ [_{Pro}\ \textbf{I}]]\ [_{VP}\ [_V\ \textbf{prefer}]\ [_{NP}\ [_{Det}\ \textbf{a}]\ [_{Nom}\ [_{Nom}\ [_N\ \textbf{morning}]]\ [_N\ \textbf{flight}]]]]]$

e.g. Number of node closing -> number of right brackets

句法特征　句法解析

- **Parsing steps** (Top-down/bottom-up) (Brennan, 2023)





句法特征　句法解析

句法特征抽取实践：基于Hanlp

```python
Hanlp = hanlp.load(hanlp.pretrained.mtl.CLOSE_TOK_POS_NER_SRL_DEP_SDP_CON_ELECTRA_SMALL_ZH)
doc = Hanlp('欢迎大家参加工作坊！', tasks=['dep', 'con'])
doc.pretty_print()
```

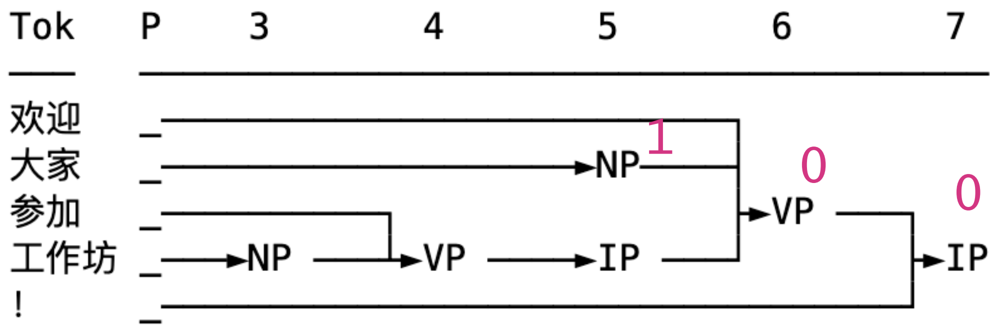
✓ 0.1s

```
Dep Tre Tok    Relat Tok    P    3        4        5        6        7
─────── ───    ───── ───    ──────────────────────────────────────────
┌┬────── 欢迎   root  欢迎    _ ───────────────────────────────┐
││ └───► 大家   dobj  大家    _ ──────────────────────►NP ───┤
│└─►┌─── 参加   dep   参加    _ ──────────────┐               ├►VP ───┐
│   └──► 工作坊 dobj  工作坊  _ ───►NP ───────┴►VP ────►IP ───┘       ├►IP
└──────► ！     punct ！      _ ──────────────────────────────────────┘
```

句法特征 句法解析

句法特征抽取实践：基于Hanlp

- **成分句法**：
  - 数据结构：嵌套列表
  - 标注方式：[Chinese Treebank]()




```python
Hanlp = hanlp.load(hanlp.pretrained.mtl.CLOSE_TOK_POS_NER_SRL_DEP_SDP_CON_ELECTRA_SMALL_ZH)
doc = Hanlp('欢迎大家参加工作坊！')
tree = doc['con']
tree[0, 0, 1]
```

✓ 0.5s Python

```
['NP', [['PN', ['大家']]]]
```

句法特征　句法解析

句法特征抽取实践：基于Hanlp

- **成分句法：**
  - 数据结构：嵌套列表
  - 标注方式：Chinese Treebank

输入

“欢迎大家参加工作坊！”

tree

`tree = doc[‘con’]`

括号表示法: `tree.pformat()`

(TOP(IP(VP(VV欢迎)(NP(PN大家))(IP(VP(VV参加)(NP(NN工作坊)))))(PU！)))

Chomsky范式: `tree.chomsky_normal_form()`

(TOP(IP(VP(VV欢迎)(<u>VP|<NP-IP></u>(NP(PN大家))(IP(VP(VV参加)(NP(NN工作坊))))))(PU！)))

句法树深度: `tree.height()`

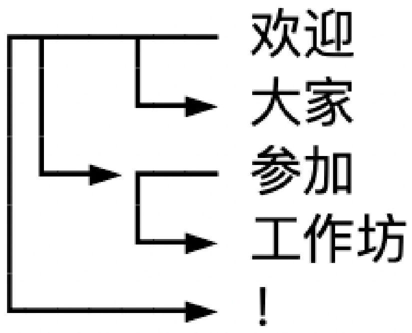
括号结构

word	left_bracket	right_bracket
欢迎	3	1
大家	2	2
参加	2	1
工作坊	2	4
！	1	2

句法特征　句法解析

句法特征抽取实践： 基于Hanlp

- **依存句法：**
  - 数据结构：[(head, relation), ...]
  - 标注方式：Universal dependency（UD）




```
doc = Hanlp('欢迎大家参加工作坊！')
Hanlp = hanlp.load(hanlp.pretrained.mtl.CLOSE_TOK_POS_NER_SRL_DEP_SDP_CON_ELECTRA_SMALL_ZH)
doc['tok/coarse'], doc['dep']
```

✓ 0.5s

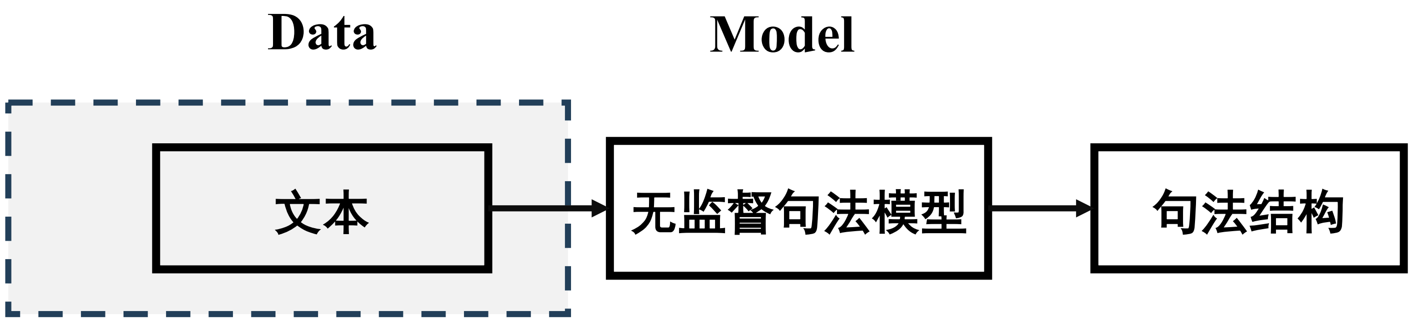
```
(['欢迎', '大家', '参加', '工作坊', '！'],
 [(0, 'root'), (1, 'dobj'), (1, 'dep'), (3, 'dobj'), (1, 'punct')])
```

	word	head	rel
0	欢迎	0	root
1	大家	1	dobj
2	参加	1	dep
3	工作坊	3	dobj
4	！	1	punct

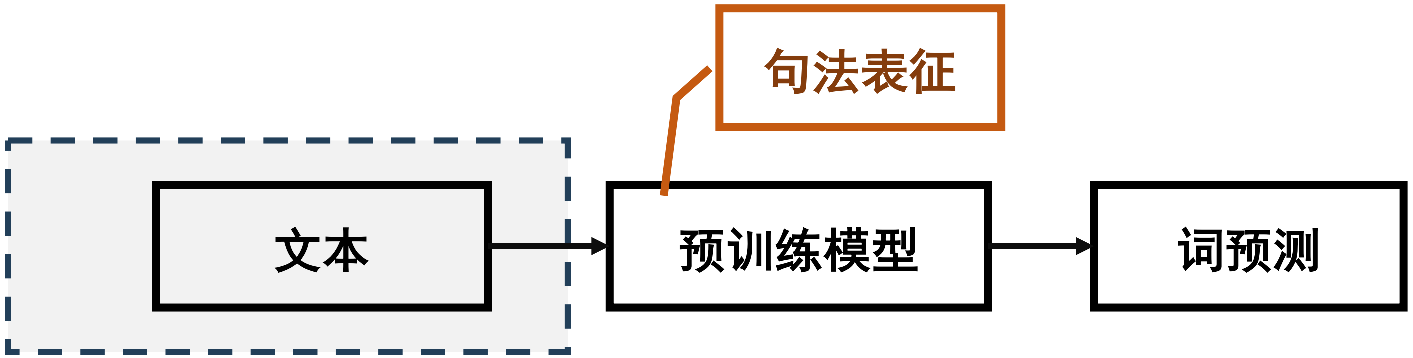
句法特征 句法解析

讨论

- 模型是否能够增量地进行句法解析？(Kitaev, 2022)
- 模型是否能够无监督地构建句法结构? (Cao, 2020)



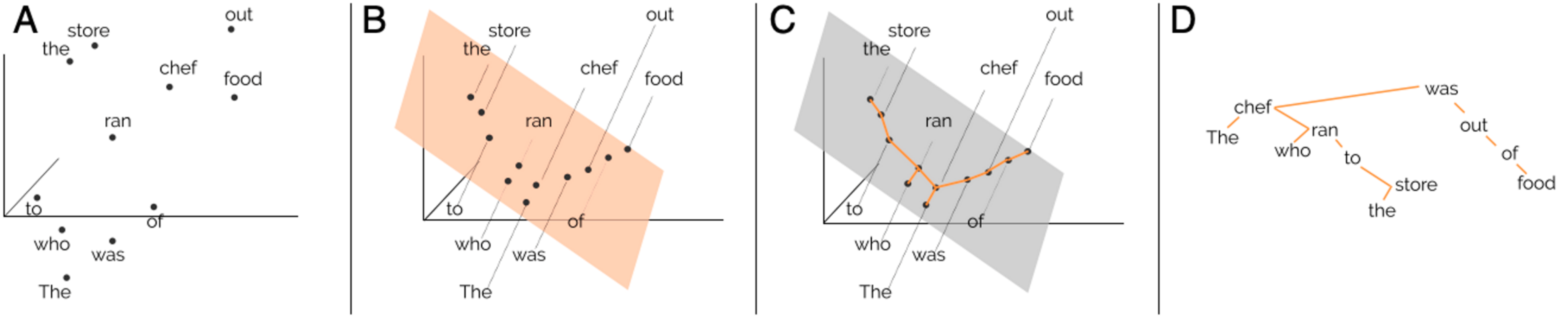
- 我们需要显式的句法结构或特征吗?



句法特征　句法表征

模型中的句法表征

- 注意力 & **词嵌入** (Manning, 2019; Coenen & Reif, 2019; Rogers, 2020)
  - 词嵌入可以线性投射到一个潜空间中，再通过向量的距离还原句法结构。



Each of the words of the sentence *The chef who ran to the store was out of food* is internally represented in context as a vector.

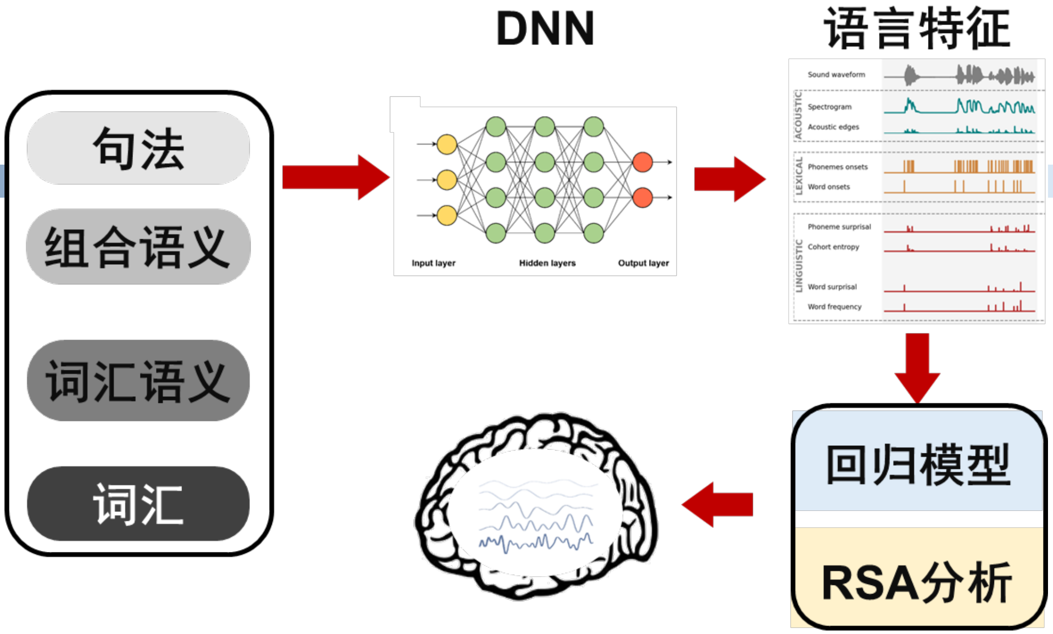
A structural probe finds a linear transform of that space under which squared L_2 distance between vectors best reconstructs tree path distance between words.

Once in this latent space, the structure of the tree is globally represented by the geometry of the vector space, meaning words that are close in the space are close in the tree.

In fact, the tree can be approximately recovered by taking a minimum spanning tree in the latent syntax space.

小结

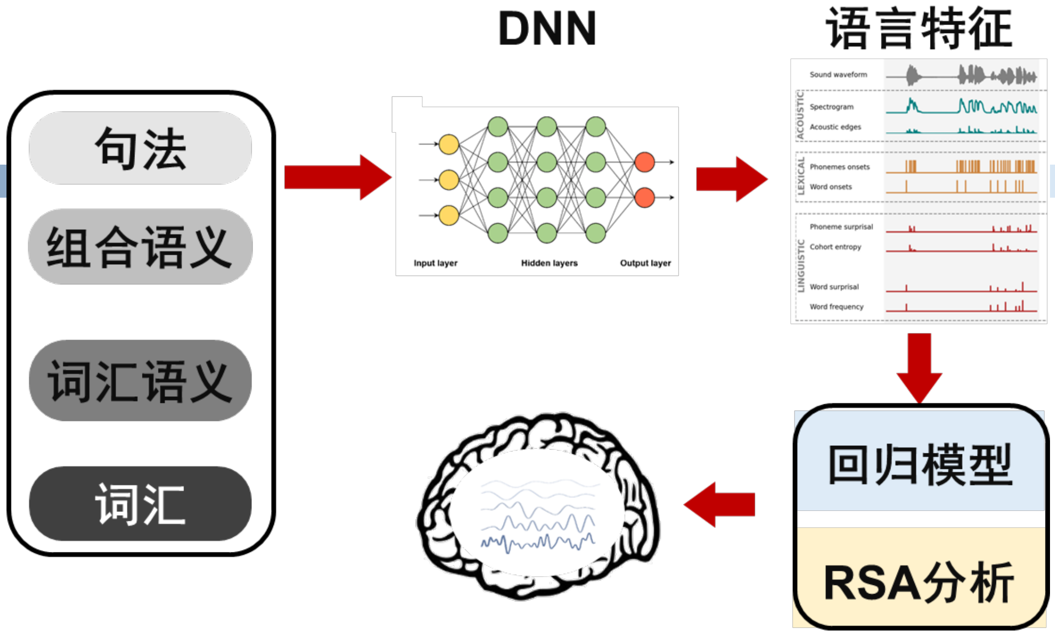
AI帮助解决了什么



- 人类标注 -> AI自动标注
- 严格的实验设计、非自然语料、考察单一变量 -> 对自然语音展开分析、同时考察多个变量
- 繁琐精细的语料设计 -> AI进行语料初筛
- 帮助回答大脑的语言加工机制：如预测下一个字等

小结

AI没有帮助解决什么



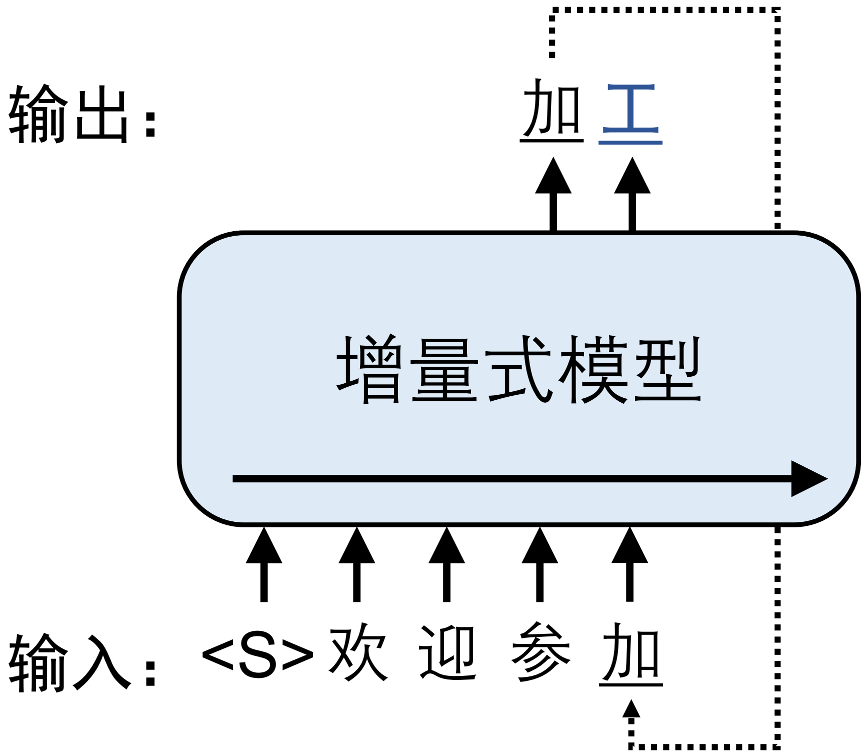
- 多种特征相互关联，较难分离不同特征的神经响应
- 未提供基础的科学理论上的突破（新方法研究旧问题验证旧理论）

可控文本生成

1. 概述
   - 代码平台
   - 模型结构
   - 语言任务
2. 特征计算
   - 词汇特征
   - 语义特征
   - 句法特征
3. 其他案例
   - 可控文本生成
   - 群体认知

文本生成

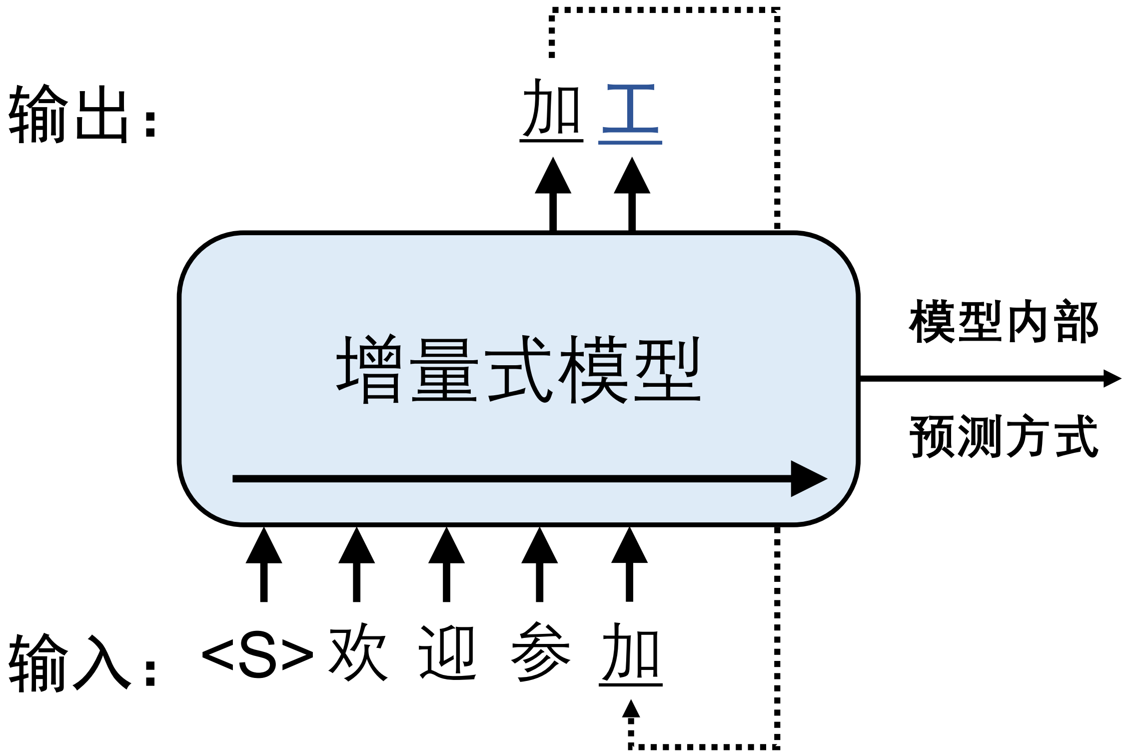
逐步将输出的词作为输入继续预测



语料：欢迎参加工作坊！

可控文本生成　文本生成

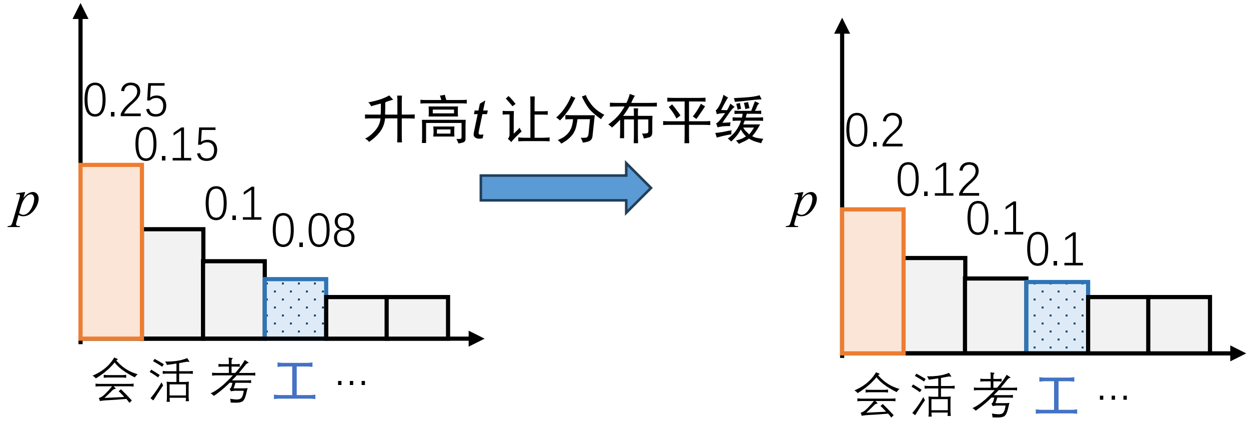
逐步将输出的词作为输入继续预测



语料：欢迎参加工作坊！

两种不同的预测方式

1. 贪婪搜索：每次选择最高概率的词
2. 随机采样：根据概率分布随机选择词
   可以使用超参数改变概率分布
   - 温度t：控制概率分布的陡峭程度



输入：欢迎参加 ?

可控文本生成　文本生成

输入: Welcome to

1. 贪婪搜索：每次选择最高概率的词

```
==================== 贪婪搜索 ============
Iter 0: Welcome to the e-commerce world!
Iter 1: Welcome to the e-commerce world!
Iter 2: Welcome to the e-commerce world!
Iter 3: Welcome to the e-commerce world!
Iter 4: Welcome to the e-commerce world!
```

2. 随机采样：根据概率分布随机选择词

温度 t = 0.1

```
==================== 随机搜索，温度参数=0.1 ======
Iter 0: Welcome to the official website of the
Iter 1: Welcome to the e-commerce world!
Iter 2: Welcome to the e-commerce world!
Iter 3: Welcome to the official website of the
Iter 4: Welcome to the official website of the
```

温度 t = 1.0

```
==================== 随机搜索，温度参数=1.0 ======
Iter 0: Welcome to the Tsugonia website
Iter 1: Welcome home! Check out our new site for
Iter 2: Welcome to La Casa Veronique, which
Iter 3: Welcome to the swedish-
Iter 4: Visit us at our new home. You'
```

文本生成

推荐使用的文本生成参数：

- 如果需要固定答案，选择贪婪搜索
- 如果需要生成多样的语料，采用随机采样：
  - 温度 t = 0.7~1.0，温度越高，文本越多样

可控文本生成

要求模型根据生成属性 (和输入), 生成受控文本

生成属性:
<积极情绪> → **输出:**
我很开心

生成属性:
<消极情绪>
输入:
今天我去考试 → **输出:**
结果考得很差

可控文本生成

生成属性可以相互组合

[Negative] The potato is a pretty bad idea. It can make you fat, it can cause you to have a terrible immune system, and it can even kill you.. . .
[Positive] The potato chip recipe you asked for! We love making these, and I've been doing so for years. I've always had a hard time keeping a recipe secret. I think it's the way our kids love to eat them – so many little ones.
[Politics] [Positive] To conclude this series of articles, I will present three of the most popular and influential works on this topic. The first article deals with the role of women's political participation in building a political system that is representative of the will of the people.

Dathathri S, et al. ICLR, 2020.

可控文本生成

生成属性可以规定**语言特征**

input (Semantic Content)	food : Japanese
output text	Browns Cambridge is good for Japanese food and also children friendly near The Sorrento .
input (Parts-of-speech)	PROPN AUX DET ADJ NOUN NOUN VERB ADP DET NOUN ADP DET NOUN PUNCT
output text	Zizzi is a local coffee shop located on the outskirts of the city .
input (Syntax Tree)	(TOP (S (NP (*) (*) (*)) (VP (*) (NP (NP (*) (*))))))
output text	The Twenty Two has great food
input (Syntax Spans)	(7, 10, VP)
output text	Wildwood pub serves multicultural dishes and is ranked 3 stars
input (Length)	14
output text	Browns Cambridge offers Japanese food located near The Sorrento in the city centre .

Li LX, et al. NIPS, 2022.

可控文本生成　生成属性

可控文本生成

1. 部署微调模型实现可控文本生成

2. 利用chatGPT等大语言模型 + 上下文学习范式实现可控文本生成

群体认知

1. 概述
   - 代码平台
   - 模型结构
   - 语言任务
2. 特征计算
   - 词汇特征
   - 语义特征
   - 句法特征
3. 其他案例
   - 可控文本生成
   - 群体认知

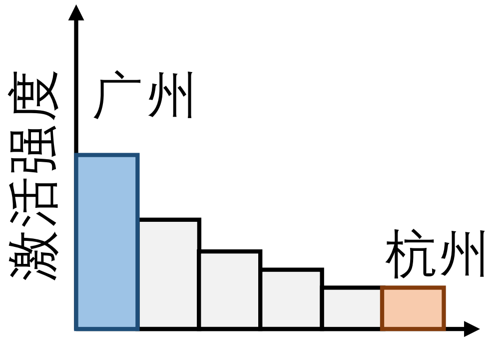
群体认知　　概述

语言 vs. 群体认知

人们在日常生活中产生的文本能够反映群体认知系统的激活强度

P(“广州美食”) > P(“杭州美食”)

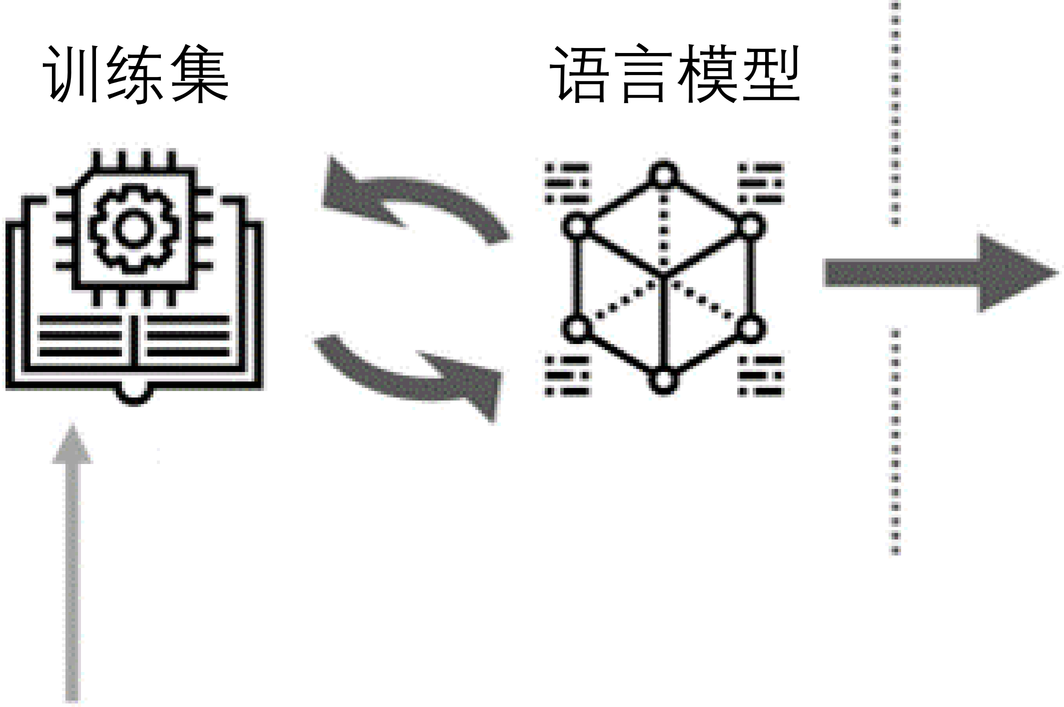
关于“美食”的群体认知



Language is a mirror of mind in a deep and significant sense.

----Noam Chomsky

性别歧视



P('**She** is a doctor') < P('**He** is a doctor')

P('She is a **doctor**') < P('She is a **nurse**')

包含偏见的文本

Sun T, et al. ACL, 2019.

群体认知 性别歧视

性别歧视

Association 强度与社会劳工统计结果高度相关

- 女性工作者越多的职业，其词向量与female的association越强

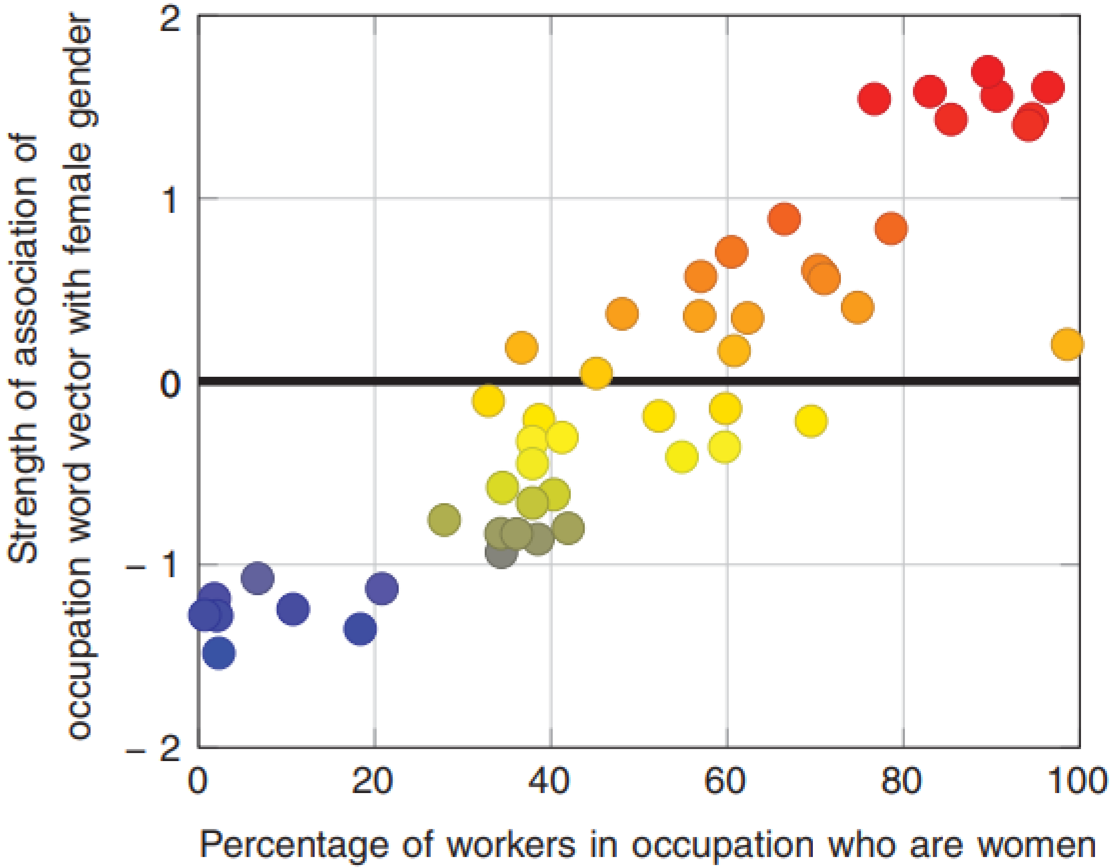


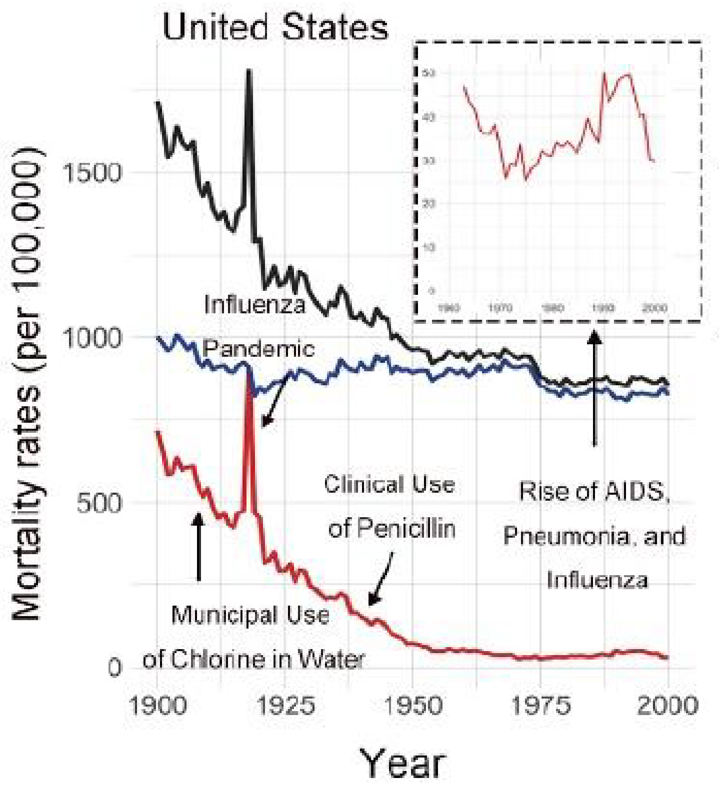
Fig. 1. Occupation-gender association. Pearson's correlation coefficient ρ = 0.90 with $P < 10^{-18}$.

Caliskan A, et al. Science, 2017.

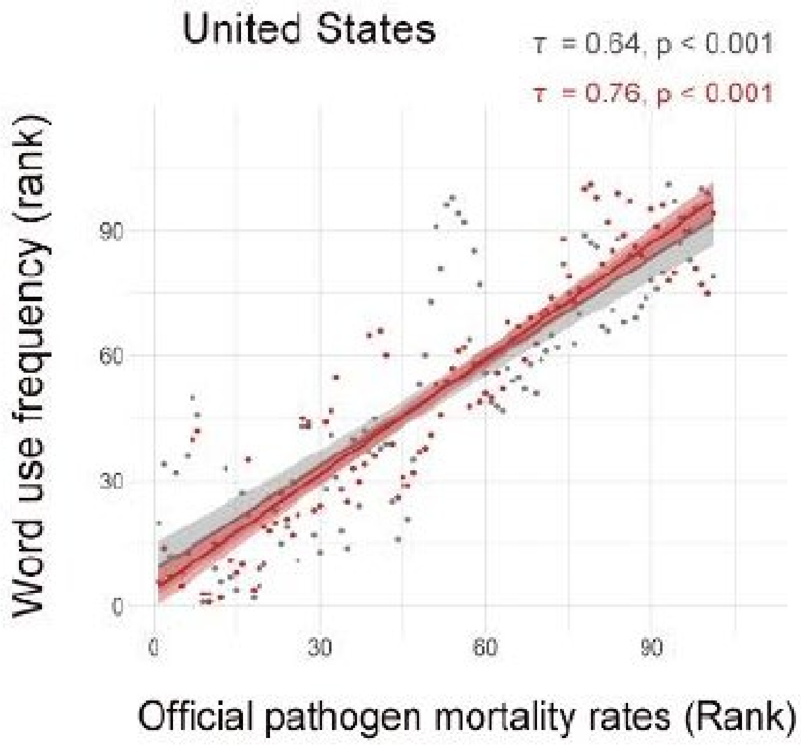
群体认知　　语义空间

语义空间能够反映群体认知变化

流行病死亡率



死亡率 vs. 疾病相关词的词频



流行病越严重的年份，疾病相关词的使用频率越高。

Fu Z, et al. preprint, 2023.

小结

1. 语言模型能够为心理语言学提供特定结构的语言材料

2. 语言模型能够反映群体认知

联系我们

丁鼐，研究员，浙江大学
ding_nai@zju.edu.cn

邹家杰，博士后，浙江大学
jiajiezou@zju.edu.cn

刘威，博士研究生，浙江大学
liuweizju@zju.edu.cn

陈泓桦，博士研究生，浙江大学
honghuachen@zju.edu.cn

意见反馈、代码资源

演示代码: https://y1ny.github.io/assets/AI_for_psycholinguistics.ipynb