

# 【论文笔记】词向量性能评价任务

---

词向量的优劣可以从三个任务类型的八个任务进行综合评定。

## 语义属性 ( Sementic Properties)

---

### 词语相似度 ( Word Similarity )

测试过程用的是WordSimilarity-353测试集，他包含353对英文词汇和人工对这些词对之间的语义相关度的评测值。词向量的效果比较的是两个词向量的余弦距离的Pearson相关性与人工打分的平均分数的相似程度。

### 同义词选择题(Synonym Question)

测试过程用的是托福考试中的80个同义词选择题，每道选择题有四个选项，选择问题与选项中余弦距离最近的选项，并评估最终整体的准确性。

### 语义类比问题 ( Semantic Analogy Question )

完成大约9000个语义类比问题，问题类似于“ man is to (woman) as king is to queen”。通过计算 ( queen-king+man ) 的最近词向量作为问题的答案，并评估整体的准确性。

### 语法类比问题 ( Syntactic Analogy Question )

完成大约10500个语法类比问题，问题类似于“ predict is to (predicting) as dance is to dancing”，通过计算 ( dancing - dance + predict ) 的最近词向量作为问题的答案，并评估整体的准确性。

# 将词向量作为特征 (Embedding as Features)

---

## 文本分类 (Text Classification)

使用词向量的加权平均值作为文本的表示，然后应用逻辑回归来执行文本分类。每个单词的权重是其出现频率。使用的数据集是IMDB数据集。

## 命名实体识别 (Named entity recognition)

## 用词向量初始化神经网络 (Embedding as the Initialization of Neural Networks)

---

在最近的NLP任务的神经网络方法中，词向量被用来初始化第一层。

## 句子情感分析 (sentence-level sentiment classification)

使用卷积神经网络(CNN)在斯坦福情绪树库数据集上进行句子情感分类，重复实验五次，并展示这些实验的平均准确性。

## 词性标注 (Part-of-speech Tagging)

使用Ronan Collobert及其同事提出的神经网络对华尔街日报数据进行词性标注，并评估准确性。