NNLM 理解

# 1.目的

如何判断一句话是否通顺合理？比如下面两句话：

A: 一群美国人抗议中餐馆杀鸡，老板反驳说你们平时不都吃鸡肉的吗，众人说我们吃的是鸡肉不是鸡，你杀鸡是侵犯鸡权

B: 中餐馆一群杀鸡美国人，，反驳你们平时不都吃鸡肉的吗老板说，说我们众人吃的不是鸡是鸡肉，侵犯鸡权你杀鸡是

我们如何用一个算法模型来判断上面哪句话比较通顺合理。

用 {W1, W2, … , Wt} 表达一句话的每个词汇， P(W1t)=P(W1, W2, … , Wt)表示这句话的概率即通顺合理的程度.

根据条件概率公式得到

P(W1, W2, … , Wt)

= P(Wt | Wt-1, … , W2, W1) \* P(Wt-1, … , W2, W1))

= P(Wt | Wt-1, … , W2, W1) \* P(Wt-1 | Wt-2, …, W2, W1) \* P(Wt-2, …, W2, W1)

= P(Wt | Wt-1, … , W2, W1) \* P(Wt-1 | Wt-2, …, W2, W1) \* … \* P(W2 | W1) \* P(W1)

P(Wt | Wt-1, … , W2, W1) 如何算呢？

先通过引入下面的假设降低 P(Wt | Wt-1, … , W2, W1) 的维数：

{Wt-1, … , W2, W1} 中越靠近 Wt 的越相关，Wt-1和Wt最相关，其次是Wt-2, 依次类推，因而我们进一步假设 Wt 只和靠近它的 n - 1 个词汇相关，即

P(Wt | Wt-1, … , W2, W1) = P(Wt | Wt-1, Wt-2, … , Wt-n+1)

n 一般取值都比 t 小不少，这样，求t个的条件概率就变成了求n个的条件概率;

以上推理就是

1. 用词汇的联合概率表达一个语句是否通顺；
2. 将计算联合概率转换为计算条件概率；
3. 将条件概率由不定长度的且一般较大的t维降到一般较小的n-1维;

# 2.模型架构

## 2.1 总架构

我们取前面描述的 n-1 的值为 6, 即取n=7, 整个词汇表词汇量为 50000，模型总架构表示如下，输入和词汇 Wt 之前的 6 个词汇，得到50000个Wt为不同词汇的概率，当然概率和为1，即 P(1) + P(2) + … + P(50000) = 1，这里1, 2, … , 50000 是词汇编号.



图2.1 总架构图

## 2.2 词汇编码(特征表达)

2.2.1 one-hot

2.2.2 特征降维

## 2.4 模型结构