开始：Noisy Channel Model

Noisy Channel Model

$$p(\text{text}|\text{source}) \propto p(\text{source}|\text{text})\, p(\text{text})$$

应用场景：

语音识别，机器翻译，拼写纠错，OCR，密码破解

$$p(\text{text}|\text{source}) \propto p(\text{source}|\text{text})\, p(\text{text})$$

机器翻译

拼写纠错

$$p(\text{text}|\text{source}) \propto p(\text{source}|\text{text})\, p(\text{text})$$

语音识别

密码破解

结束：Noisy Channel Model

开始：语言模型介绍

Language Model

语言模型用来判断：是否一句话从语法上通顺

比较：

今天是周日 VS 今天周日是

全民AI是趋势 VS 趋势全民AI是

Language Model

语言模型用来判断：是否一句话从语法上通顺

今天是周日　　今天周日是

全民AI是趋势　　趋势全民AI是

Language Model

目标

Compute the probability of a sentence or sequence of words. $p(s) = p(w_1, w_2, w_3, w_4, w_5 \ldots w_n)$

P(全民AI是趋势)

结束：语言模型介绍

开始：Chain Rule和Markov Assumption

Recap: Chain Rule

- $p(A, B, C, D)$

- $p(w_1, w_2, w_3, w_4, w_5 \ldots w_n)$

Chain Rule for Language Model

- $p(w_1, w_2, w_3, w_4, w_5 \ldots w_n)$

- p(今天，是，春节，我们，都，休息)

Chain Rule for Language Model

- p(休息 | 今天，是，春节，我们，都)

Markov Assumption

- p(休息 | 今天，是，春节，我们，都)
- p(休息 | 今天，是，春节，我们，都)
- p(休息 | 今天，是，春节，我们，都)

Markov Assumption

- $p(w_1, w_2, w_3, w_4, w_5 \ldots w_n)$
- $p(w_1, w_2, w_3, w_4, w_5 \ldots w_n)$
- $p(w_1, w_2, w_3, w_4, w_5 \ldots w_n)$

Language Model (Use 2nd Order)

$p(是|今天) = 0.01$

$p(今天) = 0.002$

$p(周日|是) = 0.001$

$p(周日|今天) = 0.0001$

$p(周日) = 0.02,$

$p(是|周日) = 0.0002$

比较: 今天是周日 VS 今天周日是

结束：Chain Rule和Markov Assumption

开始：Unigram, Bigram, N-gram

Language Model : Unigram

- $p(w_1, w_2, w_3, w_4, w_5 \ldots w_n)$
- p(今天，是，春节，我们，都，休息)
- p(今天，春节，是，都，我们，休息)

Language Model : Bigram

- $p(w_1, w_2, w_3, w_4, w_5 \ldots w_n)$
- p(今天，是，春节，我们，都，休息)
- p(今天，春节，是，都，我们，休息)

Language Model : N-gram

- $p(w_1, w_2, w_3, w_4, w_5 \ldots w_n)$
- p(今天，是，春节，我们，都，休息)
- p(今天，春节，是，都，我们，休息)

结束：Unigram, Bigram, N-gram

开始：估计语言模型的概率

Unigram：Estimating Probability

- $p(w_1, w_2, w_3, w_4, w_5 \ldots w_n)$

Unigram：Estimating Probability

语料库

今天 的 天气 很好 啊

我 很 想 出去 运动

但 今天 上午 有 课程

训练营 明天 才 开始

今天 开始 训练营 课程

今天 没有 训练营 课程

Bigram：Estimating Probability

- $p(w_1, w_2, w_3, w_4, w_5 \ldots w_n)$

Bigram：Estimating Probability

语料库

今天 的 天气 很好 啊

我 很 想 出去 运动

但 今天 上午 想 上课

训练营 明天 才 开始

今天 上午 想 出去 运动

今天 上午 的 天气 很好 呢

N-gram：Estimating Probability

语料库

今天 上午 的 天气 很好

我 很 想 出去 运动

但 今天 上午 有 课程

训练营 明天 才 开始

今天 上午 有 课程

今天 没有 训练营 课程

N-gram：Estimating Probability

- $p(w_1, w_2, w_3, w_4, w_5 \ldots w_n)$

结束：估计语言模型的概率

开始：评估语言模型：Perplexity

Evaluation of Language Model

Q: 训练出来的语言模型效果好还是坏?

- 理想情况下
  1. 假设有两个语言模型 A,B
  2. 选定一个特定的任务比如拼写纠错
  3. 把两个模型A,B都应用在此任务中
  4. 最后比较准确率，从而判断A,B的表现

Evaluation of Language Model

- 理想情况下
  1. 假设有两个语言模型 A,B
  2. 选定一个特定的任务比如拼写纠错
  3. 把两个模型A,B都应用在此任务中
  4. 最后比较准确率，从而判断A,B的表现

Q: 有没有更简单的评估方法？ 比如不需要放在特定的任务中验证?

Evaluation of Language Model

核心思路

今天____

今天天气____，

今天天气很好，____

今天天气很好，适合____

今天天气很好，适合出去____

Perplexity

$$\text{Perplexity} = 2^{-(x)}$$

x: average log likelihood

Perplexity

$$\text{Perplexity} = 2^{-(x)} \qquad x\text{: average log likelihood}$$

训练好的Bigram

$p(\text{天气}|\text{今天}) = 0.01$

$p(\text{今天}) = 0.002$

$p(\text{很好}|\text{天气}) = 0.1$

$p(\text{适合}|\text{很好}) = 0.01$

$p(\text{出去}|\text{适合}) = 0.02,$

$p(\text{运动}|\text{出去}) = 0.1$

今天

今天天气

今天天气很好，

今天天气很好，适合

今天天气很好，适合出去

今天天气很好，适合出去运动

Perplexity

$$\text{Perplexity} = 2^{-(x)} \qquad x\text{: average log likelihood}$$

Training 38 million words, test 1.5 million words, WSJ

N-gram Order	Unigram	Bigram	Trigram
Perplexity	962	170	109

Slide Credit: Dan Jurafsky

Recap: Estimating Probability

语料库

今天 上午 的 天气 很好

我 很 想 出去 运动

但 今天 上午 有 课程

训练营 明天 才 开始

今天 训练营 没有

今天 没有 训练营 课程

结束：评估语言模型：Perplexity

开始：Add-one Smoothing

Smoothing

- Add-one Smoothing
- Add-K Smoothing
- Interpolation
- Good-Turning Smoothing

Add-one Smoothing (Laplace Smoothing)

$$P_{MLE}(w_i|w_{i-1}) = \frac{c(w_{i-1}, w_i)}{c(w_i)}$$

$$P_{Add-1}(w_i|w_{i-1}) = \frac{c(w_{i-1}, w_i) + \mathbf{1}}{c(w_i) + \mathbf{V}}$$

Add-one Smoothing (Laplace Smoothing)

$$P_{Add-1}(w_i|w_{i-1}) = \frac{c(w_{i-1}, w_i) + \mathbf{1}}{c(w_i) + \mathbf{V}}$$

语料库

今天 上午 的 天气 很好

我 很 想 出去 运动

但 今天 上午 有 课程

训练营 明天 才 开始

结束：Add-one Smoothing

开始：Add-K Smoothing

Add-K Smoothing (Laplace Smoothing)

$$P_{Add-k}(w_i|w_{i-1}) = \frac{c(w_{i-1}, w_i) + \mathbf{k}}{c(w_i) + \boldsymbol{kV}}$$

语料库

今天 上午 的 天气 很好

我 很 想 出去 运动

但 今天 上午 有 课程

训练营 明天 才 开始

Add-K Smoothing (Laplace Smoothing)

$$P_{Add-k}(w_i|w_{i-1}) = \frac{c(w_{i-1}, w_i) + \mathbf{k}}{c(w_i) + \boldsymbol{kV}}$$

Add-K Smoothing (Laplace Smoothing)

$$P_{Add-k}(w_i|w_{i-1}) = \frac{c(w_{i-1}, w_i) + \mathbf{k}}{c(w_i) + \boldsymbol{kV}}$$ → 怎么选择?

训练集语料库

今天 上午 的 天气 很好

我 很 想 出去 运动

但 今天 上午 有 课程

训练营 明天 才 开始

验证集语料库

今天 上午 想 出去 运动

明天 才 开始 训练营

结束：Add-K Smoothing

开始：Interpolation

Interpolation

C(in the kitchen) = 0

C(the kitchen) = 3

C(kitchen) = 4

C(arboretum) = 0

p(kitchen | in the) =

p(arboretum | in the)=

Interpolation

C(in the kitchen) = 0

C(the kitchen) = 3

C(kitchen) = 4

C(arboretum) = 0

p(kitchen | in the) =

p(arboretum | in the)=

Interpolation

C(in the kitchen) = 0

C(the kitchen) = 3

C(kitchen) = 4

C(arboretum) = 0

p(kitchen | in the) =

p(arboretum | in the)=

核心思路

在计算Trigram概率时同时考虑Unigram, Bigram, Trigram出现的频次

Interpolation

$$p(w_n|w_{n-1},w_{n-2}) = \boldsymbol{\lambda_1} p(w_n|w_{n-1},w_{n-2}) + \boldsymbol{\lambda_2} p(w_n|w_{n-1}) + \boldsymbol{\lambda_3} p(w_n)$$

$$\boldsymbol{\lambda_1 + \lambda_2 + \lambda_3 = 1}$$

结束：Interpolation