

N-граммы

- ставят модели, к-е предсказывают след слово после N-грамм на основе вероятности их сочетания

N-грамма - посл-ть n слов

Во всех случаях предсказываем вер-ть след-го слова / или посл-ти слов  
языковые модели

$$P(A, B) = P(B|A)P(A)$$

$$P(A|B) = \frac{P(AB)}{P(B)}$$

ex. 3 б. ш. 2 ч. ш.

{B} - появл. белого при 1<sup>ой</sup> выш.

{A} - появл. бел. при 2<sup>ой</sup> выш.

$$P(A|B) = \frac{\frac{3}{5} \cdot \frac{2}{4}}{\frac{3}{5}} = \frac{1}{2}$$

Мы не можем рассчитать точную вероятность слова при условии длинной последовательности пред-  
шествующих слов

Мы можем аппроксимировать  
вероятность, упростив её

Оценка вероятностей  
n-грамм

$$P(w_i | w_{i-1}) = \frac{\text{count}(w_{i-1}, w_i)}{\text{count}(w_{i-1})}$$

→ Метод максимального правдоподобия. MLE  
Maximum Likelihood Estimate.

Качество модели n-грамм:

Перлексия: насколько хорошо модель  
предсказывает выборку  
(чем ниже значение перлексии,  
тем лучше)

▼ Которая неопределённость

→ внутр. оценка  
ред-т от модели  
фактическая нагрузка  
растёт линейно

Перлексия как обратная вероят-  
ность тестового набора,

нормализованная по кон-ву  
слов <sup>тестовый набор</sup>

1

$$P(W) = \frac{1}{\prod_{i=1}^N P(w_i, w_2, \dots, w_N)}$$

Тестовый набор  $\rightarrow$  локальная вероят-  
ность, первый уровень пермис-  
сии

$$P(W) = P(w_1 \dots w_N) = P(w_1) P(w_2) \dots P(w_N) = \prod_{i=1}^N P(w_i)$$

Как нормализовать вер-ть?

$$\ln(P(W)) = \ln\left(\prod_{i=1}^N P(w_i)\right) = \sum_{i=1}^N \ln P(w_i)$$

$$\frac{\ln P(W)}{N} = \frac{\sum_{i=1}^N \ln P(w_i)}{N}$$

$$\left( e^{\ln P(W)} \right)^{\frac{1}{N}} = \left( e^{\sum_{i=1}^N \ln P(w_i)} \right)^{\frac{1}{N}}$$

$$P(W)^{\frac{1}{N}} = \left( \prod_{i=1}^N P(w_i) \right)^{\frac{1}{N}} \rightarrow 1$$