

ЯЗЫКОВЫЕ МОДЕЛИ

- › Общая постановка задачи
- › Частный пример постановки
- › N -граммные языковые модели
- › Возможные применения

ОБЩАЯ ПОСТАНОВКА ЗАДАЧИ

- » Оценка распределения вероятностей последовательностей слов

$$P(\omega_1, \dots, \omega_m)$$

ЧАСТНЫЙ ПРИМЕР

› Вероятность при условии предыдущих слов

$$P(\omega_i | \omega_1, \dots, \omega_{i-1})$$

$$P(\omega_1, \dots, \omega_m) = \prod_{i=1}^m P(\omega_i | \omega_1, \dots, \omega_{i-1}) \approx$$

$$\begin{aligned} P(\omega_1, \dots, \omega_m) &= \prod_{i=1}^m P(\omega_i | \omega_1, \dots, \omega_{i-1}) \approx \\ &\approx \prod_{i=1}^m P(\omega_i | \omega_{i-(n-1)}, \dots, \omega_{i-1}) \end{aligned}$$

$$P(\omega_1, \dots, \omega_m) = \prod_{i=1}^m P(\omega_i | \omega_1, \dots, \omega_{i-1}) \approx$$
$$\approx \prod_{i=1}^m P(\omega_i | \omega_{i-(n-1)}, \dots, \omega_{i-1})$$

$$P(\omega_i | \omega_{i-(n-1)}, \dots, \omega_{i-1}) =$$
$$= \frac{\text{count}(\omega_{i-(n-1)}, \dots, \omega_{i-1}, \omega_i)}{\text{count}(\omega_{i-(n-1)}, \dots, \omega_{i-1})}$$

➤ Обычно:

- ▶ $N = 2$ (биграммы)
- ▶ $N = 3$ (триграммы)

» Проблемы

- ▶ Для больших N мало статистики
- ▶ Малые N недостаточно хорошо моделируют осмысленный текст
- ▶ Плохие оценки вероятностей, если мало статистики

ПРИМЕНЕНИЯ ЯЗЫКОВЫХ МОДЕЛЕЙ

- » Оценивать вероятность появления текста (например, можно использовать её в классификаторах)
- » Генерировать тексты

Вся Солнечная система заполнена микроскопическими пылинками и астероидами, и в советское время такие опыты? Воды выступала дальняя половина пояса астероидов, в этой фазе человек испытывает так называемую сонную амнезию, в том числе предназначенные для управления маркетинговыми кампаниями в социальных сетях, могут предотвращать скапливание тромбоцитов, что в одной из самых высокоэффективных вариантов горючего для космических ракет?

- › Общая постановка задачи
- › Частный пример постановки
- › N -граммные языковые модели
- › Возможные применения