

Автоматическое распознавание речи. Языковые модели

П. А. Холявин

p.kholyavin@spbu.ru

05.03.2025





Задача распознавания речи

Если $O = o_1, o_2, \dots, o_n$ – звуковая последовательность,
 $W = w_1, w_2, \dots, w_n$ – последовательность слов, то

$$\hat{W} = \underset{W \in L}{\operatorname{argmax}} P(W|O)$$

$$\hat{W} = \underset{W \in L}{\operatorname{argmax}} \frac{P(O|W) P(W)}{P(O)} = \underset{W \in L}{\operatorname{argmax}} P(O|W) P(W)$$



Языковые модели

Отвечают за $P(W)$

1. Статистические
2. Формальные



Статистические ЯМ

$$\begin{aligned} P(W) &= P(w_1, w_2, \dots, w_n) = \\ &= P(w_1)P(w_2|w_1)P(w_3|w_1, w_2) \dots P(w_n|w_1, w_2, \dots, w_{n-1}) = \\ &= \prod_{i=1}^n P(w_i|w_1, w_2, \dots, w_{i-1}) \end{aligned}$$

$P(w_i)$ – униграммы

$P(w_i|w_{i-1})$ – биграммы

три-, тетра-, ...



Вычисление N-грамм

$$P(w_n|w_{n-1}) = \frac{C(w_{n-1}w_n)}{\sum_w C(w_{n-1}w_n)} = \frac{C(w_{n-1}w_n)}{C(w_{n-1})}$$

<s> John read a book </s>

<s> I read a different book </s>

<s> John read a book by Mulan </s>

P(John read a book) =



Вычисление N-грамм

$$P(w_n|w_{n-1}) = \frac{C(w_{n-1}w_n)}{\sum_w C(w_{n-1}w_n)} = \frac{C(w_{n-1}w_n)}{C(w_{n-1})}$$

<s> John read a book </s>

<s> I read a different book </s>

<s> John read a book by Mulan </s>

$P(\text{John read a book}) = P(\text{John}|\text{<s>}) * P(\text{read}|\text{John})$
 $* P(\text{a}|\text{read}) * P(\text{book}|\text{a}) * P(\text{</s>}|\text{book})$

$P(\text{Mulan read a book}) =$



Вычисление N-грамм

$$P(w_n|w_{n-1}) = \frac{C(w_{n-1}w_n)}{\sum_w C(w_{n-1}w_n)} = \frac{C(w_{n-1}w_n)}{C(w_{n-1})}$$

<s> John read a book </s>

<s> I read a different book </s>

<s> John read a book by Mulan </s>

$P(\text{John read a book}) = P(\text{John}|\text{<s>}) * P(\text{read}|\text{John})$
 $* P(\text{a}|\text{read}) * P(\text{book}|\text{a}) * P(\text{</s>}|\text{book})$

$P(\text{Mulan read a book}) = P(\text{Mulan}|\text{<s>}) * \dots = 0$



Перплексия модели

(коэффициент неопределённости)

Вычисляется на тестовой последовательности длиной N . Чем ниже перплексия, тем лучше модель.

$$\begin{aligned}\text{perplexity}(W) &= P(w_1 w_2 \dots w_N)^{-\frac{1}{N}} \\ &= \sqrt[N]{\frac{1}{P(w_1 w_2 \dots w_N)}}\end{aligned}$$

$$\text{perplexity}(W) = \sqrt[N]{\prod_{i=1}^N \frac{1}{P(w_i | w_1 \dots w_{i-1})}}$$

Перплексия связана с кросс-энтропией:

$$H(\mathbf{W}) = -\frac{1}{N_{\mathbf{W}}} \log_2 P(\mathbf{W}) \quad PP(\mathbf{W}) = 2^{H(\mathbf{W})}$$



Сглаживание N-грамм

1. Сглаживание Лапласа (Laplace smoothing, add-1 smoothing)

$$P_{\text{Laplace}}(w_i) = \frac{c_i + 1}{N + V}$$

$$c_i^* = (c_i + 1) \frac{N}{N + V}$$

$$d_c = \frac{c^*}{c}$$



Сглаживание N-грамм

$$P_{\text{Laplace}}(w_n|w_{n-1}) = \frac{C(w_{n-1}w_n) + 1}{\sum_w (C(w_{n-1}w) + 1)} = \frac{C(w_{n-1}w_n) + 1}{C(w_{n-1}) + V}$$

<s> John read a book </s>

<s> I read a different book </s>

<s> John read a book by Mulan </s>



Сглаживание N-грамм

2. Плюс-k сглаживание

$$P_{\text{Add-k}}^*(w_n | w_{n-1}) = \frac{C(w_{n-1}w_n) + k}{C(w_{n-1}) + kV}$$

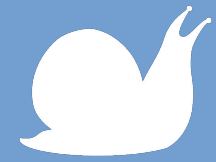


Сглаживание N-грамм

3. Откат и интерполяция

$$P_{smooth}(w_i | w_{i-n+1} \dots w_{i-1}) = \begin{cases} \alpha(w_i | w_{i-n+1} \dots w_{i-1}) & \text{if } C(w_{i-n+1} \dots w_i) > 0 \\ \gamma(w_{i-n+1} \dots w_{i-1}) P_{smooth}(w_i | w_{i-n+2} \dots w_{i-1}) & \text{if } C(w_{i-n+1} \dots w_i) = 0 \end{cases}$$

$$\begin{aligned} \hat{P}(w_n | w_{n-2} w_{n-1}) &= \lambda_1 P(w_n) \\ &\quad + \lambda_2 P(w_n | w_{n-1}) \\ &\quad + \lambda_3 P(w_n | w_{n-2} w_{n-1}) \end{aligned}$$



Сглаживание N-грамм

а) Откат Катца

$$P_{\text{BO}}(w_n | w_{n-N+1:n-1}) = \begin{cases} P^*(w_n | w_{n-N+1:n-1}), & \text{if } C(w_{n-N+1:n}) > 0 \\ \alpha(w_{n-N+1:n-1}) P_{\text{BO}}(w_n | w_{n-N+2:n-1}), & \text{otherwise.} \end{cases}$$



Сглаживание N-грамм

б) Сглаживание Гуда-Тьюринга

$$r^* = (r + 1) \frac{n_{r+1}}{n_r}$$

где n_r – количество n-грамм, встретившихся ровно r раз



Сглаживание N-грамм

в) Сглаживание Кнезера-Нея

ALGORITHM 11.3 KNESER-NEY BIGRAM SMOOTHING

$$P_{KN}(w_i | w_{i-1}) = \begin{cases} \frac{\max\{C(w_{i-1}w_i) - D, 0\}}{C(w_{i-1})} & \text{if } C(w_{i-1}w_i) > 0 \\ \alpha(w_{i-1})P_{KN}(w_i) & \text{otherwise} \end{cases}$$

where $P_{KN}(w_i) = \mathbb{C}(\bullet w_i) / \sum_{w_j} \mathbb{C}(\bullet w_j)$, $\mathbb{C}(\bullet w_i)$ is the number of unique words preceding w_i .

$\alpha(w_{i-1})$ is chosen to make the distribution sum to 1 so that we have:

$$\alpha(w_{i-1}) = \frac{1 - \sum_{w_j: C(w_{i-1}w_j) > 0} \frac{\max\{C(w_{i-1}w_j) - D, 0\}}{C(w_{i-1})}}{1 - \sum_{w_j: C(w_{i-1}w_j) > 0} P_{KN}(w_j)}$$



Сглаживание N-грамм

г) “Глупый” откат (stupid backoff)

$$S(w_i | w_{i-N+1:i-1}) = \begin{cases} \frac{\text{count}(w_{i-N+1:i})}{\text{count}(w_{i-N+1:i-1})} & \text{if } \text{count}(w_{i-N+1:i}) > 0 \\ \lambda S(w_i | w_{i-N+2:i-1}) & \text{otherwise} \end{cases}$$



Классовые модели

$$P(w_i | c_{i-n+1} \dots c_{i-1}) = P(w_i | c_i) P(c_i | c_{i-n+1} \dots c_{i-1})$$

Членами этих моделей являются не конкретные слова, а классы слов.

Классы могут быть:

1. Построены вручную
2. Частями речи
3. Результатом автоматической кластеризации



Морфемные модели

Модель основа (s)/флексия (e):

$$P(s_i | \dots w_i) = P(s_i | \dots s_{i-1})$$

$$P(e_i | \dots w_i) = P(e_i | s_i e_{i-1})$$



Адаптивные модели

Интерполяция статической и локальной динамической (кэш) моделей

$$P_{cache}(w_i | w_{i-n+1} \dots w_{i-1}) \\ = \lambda_c P_s(w_i | w_{i-n+1} \dots w_{i-1}) + (1 - \lambda_c) P_{cache}(w_i | w_{i-2} w_{i-1})$$



Адаптивные модели

TF-IDF модель

$$TF(t, d) = \frac{\text{number of times } t \text{ appears in } d}{\text{total number of terms in } d}$$

$$IDF(t) = \log \frac{N}{1 + df}$$

$$TF - IDF(t, d) = TF(t, d) * IDF(t)$$

$$Similarity(D_i, D_j) = \frac{\sum_k tfidf_{ik} * tfidf_{jk}}{\sqrt{\sum_k (tfidf_{ik})^2 * \sum_k (tfidf_{jk})^2}}$$

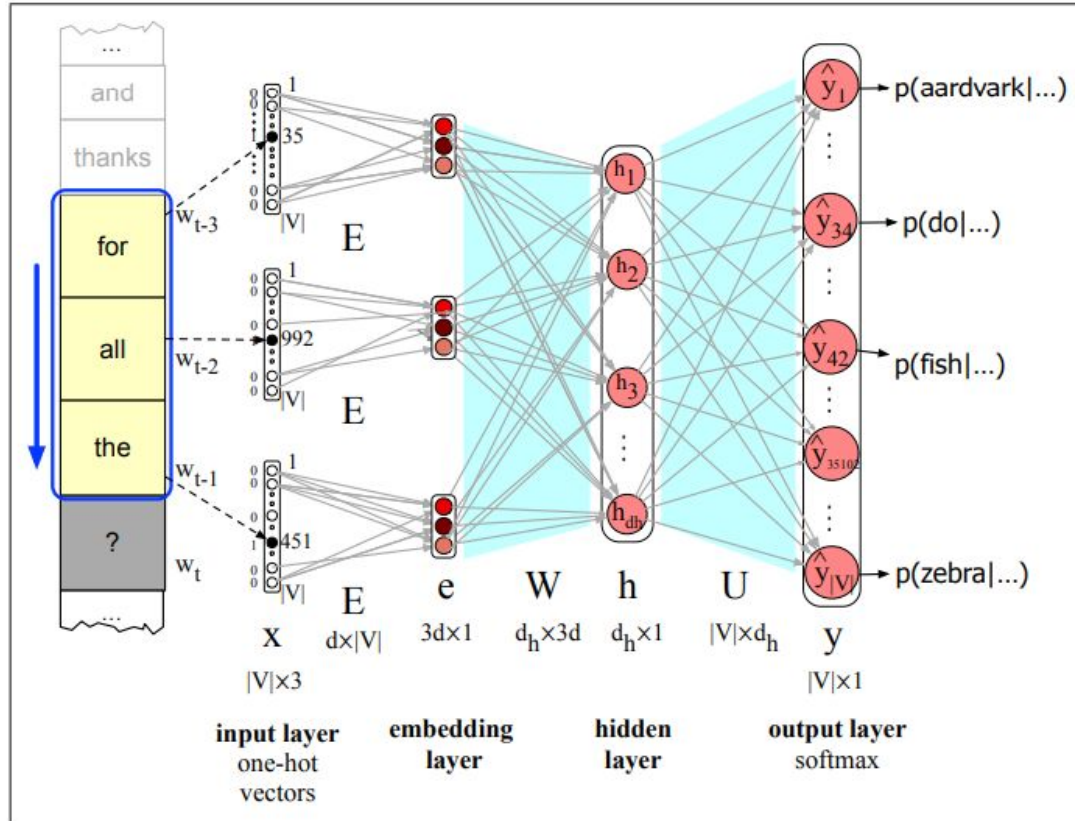


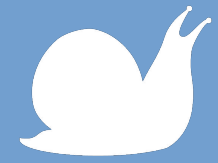
Сравнение моделей

Models	Perplexity	Word Error Rate
Unigram Katz	1196.45	14.85%
Unigram Kneser-Ney	1199.59	14.86%
Bigram Katz	176.31	11.38%
Bigram Kneser-Ney	176.11	11.34%
Trigram Katz	95.19	9.69%
Trigram Kneser-Ney	91.47	9.60%

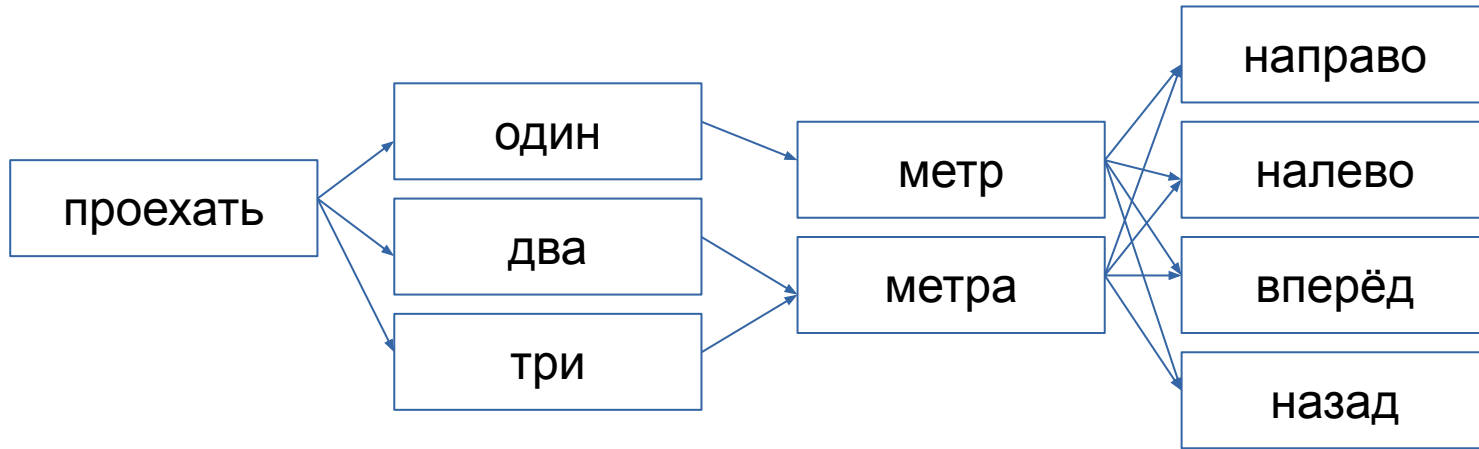


Нейронные модели



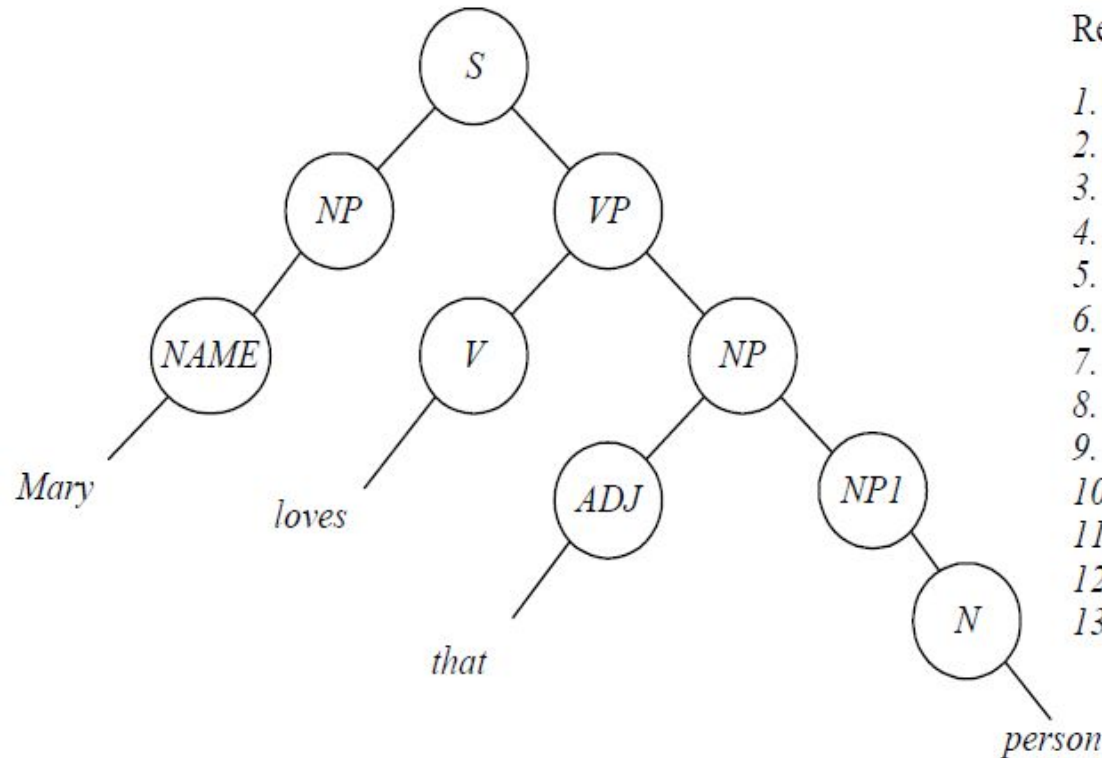


Формальные модели





Формальные модели



Rewrite Rules:

1. $S \rightarrow NP VP$
2. $VP \rightarrow V NP$
3. $VP \rightarrow AUX VP$
4. $NP \rightarrow ART NP1$
5. $NP \rightarrow ADJ NP1$
6. $NP1 \rightarrow ADJ NP1$
7. $NP1 \rightarrow N$
8. $NP \rightarrow NAME$
9. $NP \rightarrow PRON$
10. $NAME \rightarrow Mary$
11. $V \rightarrow loves$
12. $ADJ \rightarrow that$
13. $N \rightarrow person$

Спасибо за внимание!

