

Домашнее задание №4

Анаул Андрей, ИУ7-11М

①

Найти оптимальный разбор предложения Time flies like an arrow,

| | $\langle S \rangle$ | ADJ | N | V | CONJ | DET |
|---------------------|---------------------|------|------|------|------|------|
| $\langle S \rangle$ | 0,0 | 0,2 | 0,2 | 0,1 | 0,01 | 0,4 |
| ADJ | 0,0 | 0,2 | 0,12 | 0,05 | 0,2 | 0,01 |
| N | 0,0 | 0,05 | 0,05 | 0,5 | 0,2 | 0,01 |
| V | 0,0 | 0,1 | 0,1 | 0,01 | 0,2 | 0,3 |
| CONJ | 0,0 | 0,1 | 0,1 | 0,2 | 0,0 | 0,2 |
| DET | 0,0 | 0,3 | 0,1 | 0,0 | 0,0 | 0,0 |

если матрицы A и B имеют следующий вид:

вероятность 0,01 время - это сущ.

$$p(\text{time} | N) = 0,01$$

$$p(\text{flies} | V) = 0,01$$

$$p(\text{like} | V) = 0,02$$

$$p(\text{arrow} | N) = 0,01$$

$$p(\text{time} | V) = 0,001$$

$$p(\text{flies} | N) = 0,0005$$

$$p(\text{like} | N) = 0,001$$

$$p(\text{arrow} | ADJ) = 0,01$$

$$p(\text{time} | ADJ) = 0,0005$$

$$p(\text{like} | CONJ) = 0,05$$

$$p(\text{an} | DET) = 0,1$$

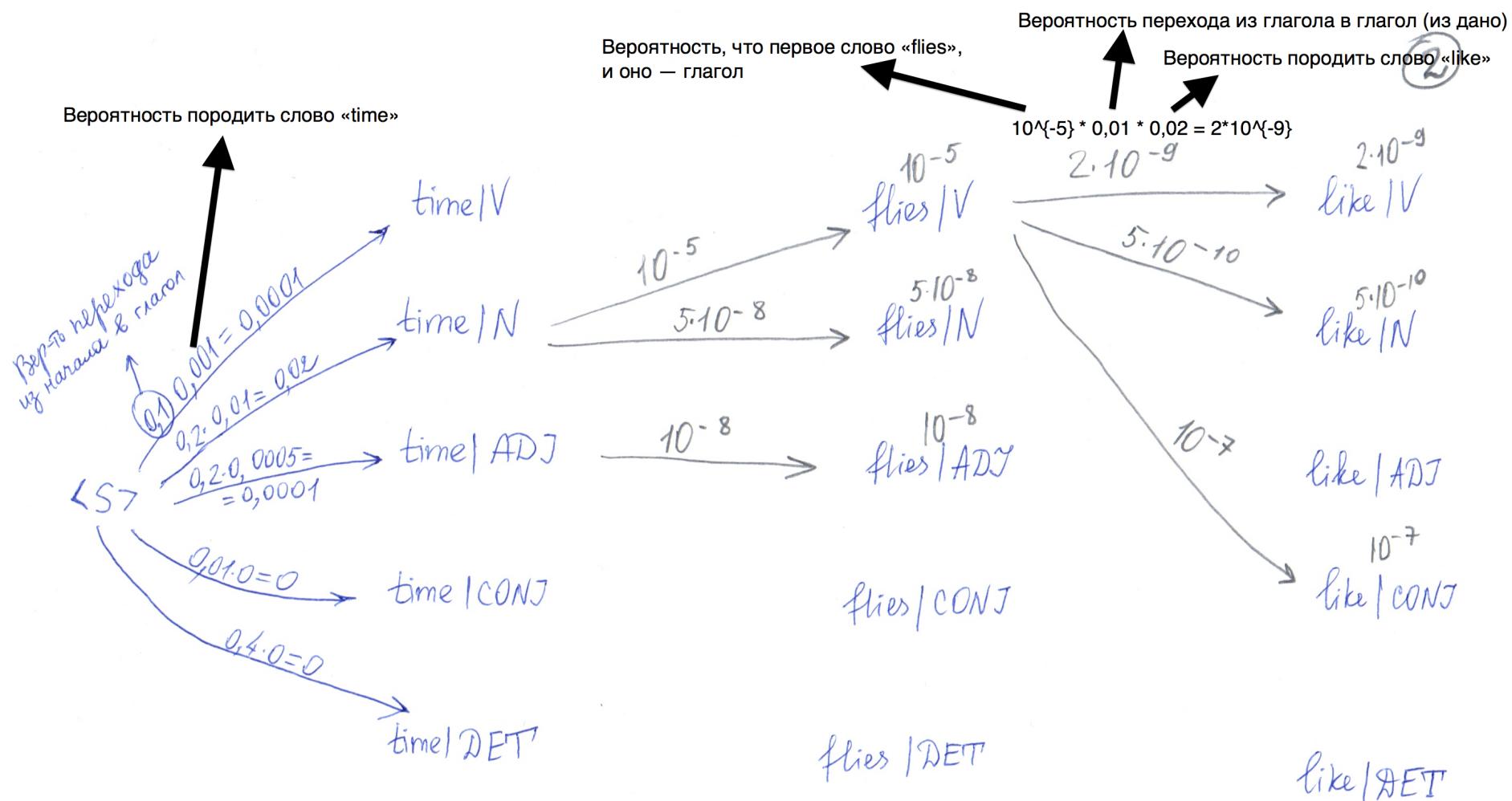
Алгоритм Виттерби позволяет сделать наиболее вероятное предположение о последовательности состояний скрытой Марковской модели на основе наблюдаемости наблюдений.

Скрытое состояния - частотное теги

Наблюдаемое состояния - слова текста

Given word sequence w_1, \dots, w_N

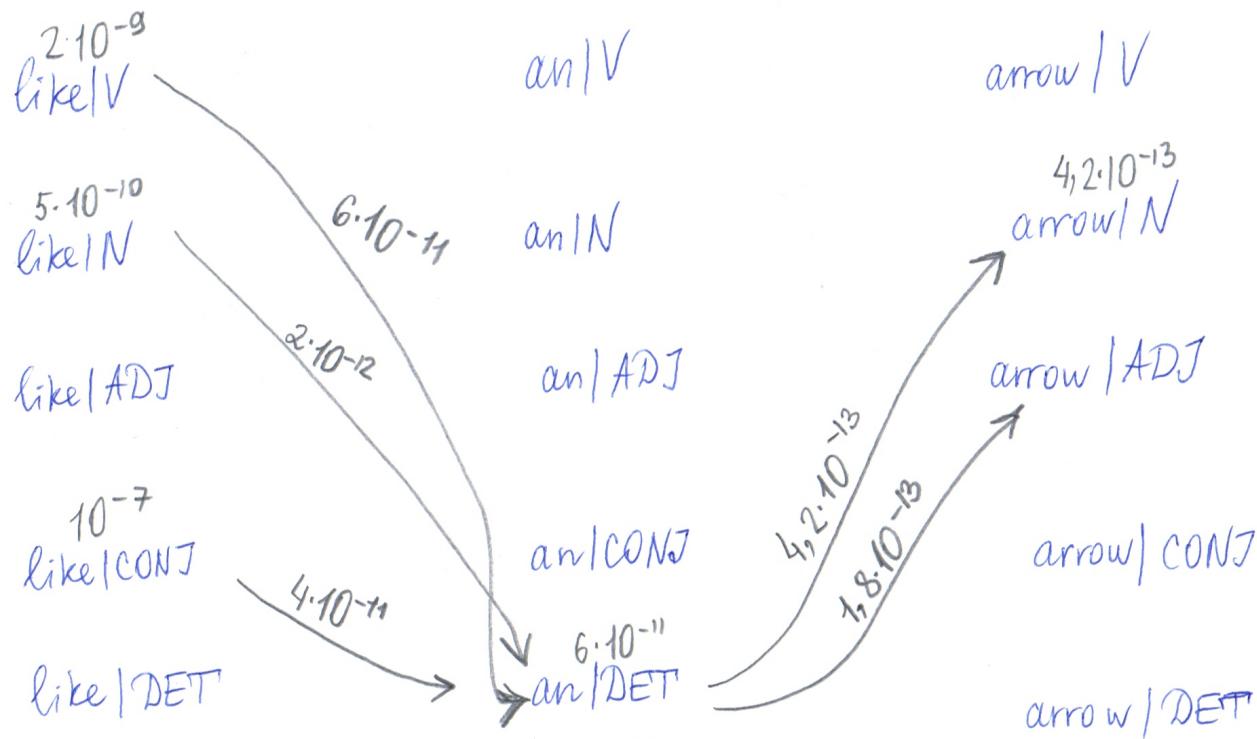
Given pos tags t_1, \dots, t_N , $t_i \in [V, N, ADJ, CONJ, DET]$



На самом деле мы рисуем стрелочки от каждого слова первого столбика к каждому слову второго, но в конце написания Марковской модели нам необходимо для каждого слова «выцепить» наиболее высокую вероятность из всех стрелочек*. И в конечном варианте Марковской модели рисуются наиболее значимые стрелочки, как и в этой схеме.

*На это указывает argmax в формуле

(3)



Bablog: Time^{cyus.} flies^{m.} like^{cot03} an^{getepnum.} arrow^{cyus.}
 N V CONJ DET N