Natural Language Processing CSCI 4152/6509 — Lecture 13 P0 Project Topics Discussion (1); HMM Model

Instructors: Vlado Keselj
Time and date: 14:35 – 15:55, 06-Nov-2025
Location: Studley LSC-Psychology P5260

Previous Lecture

- N-gram Model Smoothing (continued):
  - Witten-Bell smoothing
- **POS tagging: Introduction**
- Open word categories
- Closed word categories
- Other word categories
- **Hidden Markov Model (HMM):**
  - idea, definition, graphical representation
  - HMM assumption
- HMM POS Example

P0 Topics Discussion

- Discussion of individual projects as proposed in P0 submissions
- Projects discussed: P-03, P-07, P-08, P-09, P-11, P-12, P-20, P-21, P-22, P-23

Learning HMM (Training) with Smoothing

- Let us learn HMM from completely labeled data:
  `swat V flies N like P ants N`
  `time N flies V like P an D arrow N`
- We will use smoothing in word generation, by giving a 0.5 count to all unseen words

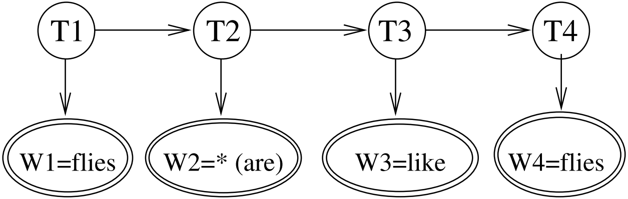
Generated Tables

T_1	$P(T_1)$
N	0.5
V	0.5

T_{i-1}	T_i	$P(T_i \| T_{i-1})$
D	N	1
N	P	0.5
N	V	0.5
P	D	0.5
P	N	0.5
V	N	0.5
V	P	0.5

T_i	W_i	$P(W_i \| T_i)$
D	an	$2/3 \approx 0.666666667$
D	*	$1/3 \approx 0.333333333$
N	ants	$2/9 \approx 0.222222222$
N	arrow	$2/9 \approx 0.222222222$
N	flies	$2/9 \approx 0.222222222$
N	time	$2/9 \approx 0.222222222$
N	*	$1/9 \approx 0.111111111$
P	like	0.8
P	*	0.2
V	flies	0.4
V	swat	0.4
V	*	0.2

Tagging Example



$$\arg\max_T \mathrm{P}(T|W=\text{sentence}) =$$

$$= \arg\max_T \frac{\mathrm{P}(T,W=\text{sentence})}{\mathrm{P}(W=\text{sentence})} = \arg\max_T \mathrm{P}(T,W=\text{sentence})$$

$$= \arg\max_T \mathrm{P}(T_1)\cdot\mathrm{P}(W_1=\text{flies}|T_1)\cdot\mathrm{P}(T_2|T_1)\cdot\mathrm{P}(W_2=*|T_2)$$

$$\cdot\mathrm{P}(T_3|T_2)\cdot\mathrm{P}(W_3=\text{like}|T_3)\cdot\mathrm{P}(T_4|T_3)\cdot\mathrm{P}(W_4=\text{flies}|T_4)$$

"Brute-Force" Approach

- Try all combinations of variable values T_1, T_2, T_3, and T_4
- Calculate the overall probability for each of them using the formula

$$\begin{array}{l} \mathrm{P}(T_1) \cdot \mathrm{P}(W_1 = \textsf{flies}|T_1) \\ \cdot \mathrm{P}(T_2|T_1) \cdot \mathrm{P}(W_2 = \textsf{*}|T_2) \\ \cdot \mathrm{P}(T_3|T_2) \cdot \mathrm{P}(W_3 = \textsf{like}|T_3) \\ \cdot \mathrm{P}(T_4|T_3) \cdot \mathrm{P}(W_4 = \textsf{flies}|T_4) \end{array}$$

- Choose the maximal probability

Brute-Force Approach (tabular view)

Efficient Tagging with HMM

- Rather than using the brute-force approach, we can incrementally optimize the product expression by partial maximization from left to right
- One way to represent this is by using a table, which leads to the dynamic programming solution, or the Viterbi algorithm
- The second way to represent this computation is using message passing, or product-sum algorithm

HMM Inference: Dynamic Programming Solution

- Brute-force approach is too inefficient
- Idea for more efficient calculation: maximize sub-products first
- Dynamic Programming approach: divide problem into sub-problems
    - with a manageable number of sub-problems
- Find maximal partial configurations up to T_1, then T_2, T_3, and T_4

Dynamic Programming Approach (graphical view)

Viterbi Algorithm Example

	T_1 $(W_1 =$ flies$)$	T_2 $(W_2 =$ *$)$	T_3 $(W_3 =$ like$)$	T_4 $(W_4 =$ flies$)$
	$\mathrm{P}(T_1)\mathrm{P}(W_1\|T_1)$	$p \cdot \mathrm{P}(T_2\|T_1)\mathrm{P}(W_2\|T_2)$	$p \cdot \mathrm{P}(T_3\|T_2)\mathrm{P}(W_3\|T_3)$	$p \cdot \mathrm{P}(T_4\|T_3)\mathrm{P}(W_4\|T_4)$
D	$0 \times 0 = 0$	DD: $0 \times 0 \times \frac{1}{3} = 0$ ND: $\frac{1}{9} \times 0 \times \frac{1}{3} = 0$ PD: 0 VD: 0 max: 0	DD: $0 \times 0 \times 0 = 0$ ND: $\frac{1}{90} \times 0 \times = 0$ PD: $\frac{1}{50} \times \frac{1}{2} \times 0 = 0$ VD: $\frac{1}{90} \times 0 \times 0 = 0$ max: 0	DD: $0 \times 0 \times 0 = 0$ ND: $0 \times 0 \times 0 = 0$ PD: $\frac{1}{225} \times 0.5 \times 0 = 0$ VD: $0 \times 0 \times 0 = 0$ max: 0
N	$0.5 \times \frac{2}{9} = \frac{1}{9}$	DN: $0 \times 1 \ldots = 0$ NN: $\frac{1}{9} \times 0 \ldots = 0$ PN: $0 \times \ldots = 0$ VN: $0.2 \times 0.5 \times \frac{1}{9} = \frac{1}{90}$ max: $\frac{1}{90}$	DN: $0 \times 1 \times 0 = 0$ NN: $\frac{1}{90} \times 0 \ldots = 0$ PN: $\frac{1}{50} \times 0.5 \times 0 = 0$ VN: $\frac{1}{90} \times 0.5 \times 0 = 0$ max: 0	DN: $0 \times 1 \times \frac{2}{9} = 0$ NN: $0 \times 0 \times \frac{2}{9} = 0$ PN: $\frac{1}{225} \times 0.5 \times \frac{2}{9} = \frac{1}{2025}$ VN: $0 \times 0.5 \times \frac{2}{9} = 0$ max: $\frac{1}{2025}$
P	$0 \times 0 = 0$	DP: $0 \times \ldots = 0$ NP: $\frac{1}{9} \times 0.5 \times 0.2 = \frac{1}{90}$ PP: $0 \times \ldots = 0$ VP: $0.2 \times 0.5 \times 0.2 = \frac{1}{50}$ max: $\frac{1}{50}$	DP: $0 \times 0 \times 0.8 = 0$ NP: $\frac{1}{90} \times 0.5 \times 0.8 = \frac{1}{225}$ PP: $\frac{1}{50} \times 0 \times 0.8 = 0$ VP: $\frac{1}{90} \times 0.5 \times 0.8 = \frac{1}{225}$ max: $\frac{1}{225}$	DP: $0 \times 0 \times 0 = 0$ NP: $0 \times 0.5 \times 0 = 0$ PP: $\frac{1}{225} \times 0 \times 0 = 0$ VP: $0 \times 0.5 \times 0 = 0$ max: 0
V	$0.5 \times 0.4 = 0.2$	DV: $0 \times \ldots = 0$ NV: $\frac{1}{9} \times 0.5 \times 0.2 = \frac{1}{90}$ PV: $0 \times \ldots = 0$ VV: $0.2 \times 0 \ldots = 0$ max: $\frac{1}{90}$	DV: $0 \times 0 \times 0 = 0$ NV: $\frac{1}{90} \times 0.5 \times 0 = 0$ PV: $\frac{1}{50} \times 0 \times 0 = 0$ VV: $\frac{1}{90} \times 0 \times 0 = 0$ max: 0	DV: $0 \times 0 \times 0.4 = 0$ NV: $0 \times 0.5 \times 0.4 = 0$ PV: $\frac{1}{225} \times 0 \times 0.4 = 0$ VV: $0 \times 0 \times 0.4 = 0$ max: 0