CHAPTER 8: Sequence Labeling for Parts of Speech and Named Entities

Instructor: PhD. Nguyen Thi Quy

Group 5: 8.3 - 8.4.3

November 20, 2023

Presentation Overview

Presentation Overview

Presentation Overview

Introducing HMM

The Hidden Markov Model - HMM

- A statistical model with unknown parameters that must be determined from known parameters.
- Extends from the mathematical model: **Markov Chains**.

Applications

- Sequence labeling: NER, POS tagging
- Speech recognition
- Optical Character Recognition (OCR)
- Bioinformatics

Markov chains

Markov chains

A model that tells us something about the probabilities of sequences of random variables, states

- Sequence of states with a temporal order
- States can take values from any discrete set of values.
- **Markov assumption**: When predicting the future, the past doesn't matter

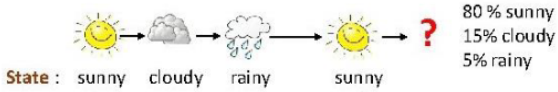
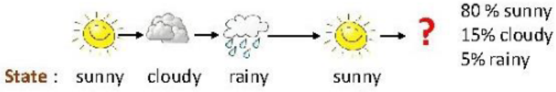


Figure: AA. Markov

Markov assumption

When predicting the future, the past doesn't matter, only the present



Markov assumption: $P(q_i = a|q_1...q_{i-1}) = P(q_1 = a|q_{i-1})$

Markov chains

Components of the Markov chains

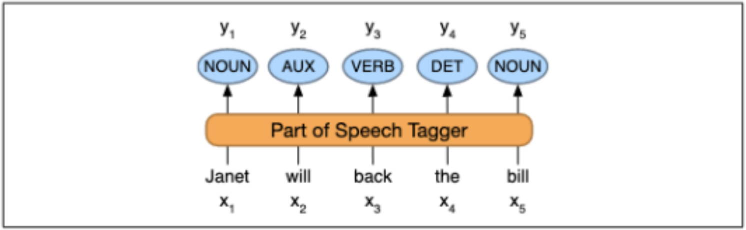
- $Q = q_1 q_2 ... q_n$: a set of N **states**
- $A = a_{11} a_{12} ... a_{N1} ... a_{NN}$: a **transition probability matrix** A, each a_{ij} representing the probability of moving from state i to state j
- $\pi = \pi_1, \pi_2, ..., \pi_n$: an **initial probability distribution** over states. π_i is the probability that the Markov chain will start in state i.

The Hidden Markov Model

A hidden Markov model (HMM) allows us to talk about both observed events and hidden events.

Unobservable Events:

- Part-of-speech
- Entity type



The Hidden Markov Model

$Q = q_1 q_2 \ldots q_N$	a set of N **states**
$A = a_{11} \ldots a_{ij} \ldots a_{NN}$	a **transition probability matrix** A, each a_{ij} representing the probability of moving from state i to state j, s.t. $\sum_{j=1}^{N} a_{ij} = 1 \quad \forall i$
$O = o_1 o_2 \ldots o_T$	a sequence of T **observations**, each one drawn from a vocabulary $V = v_1, v_2, \ldots, v_V$
$B = b_i(o_t)$	a sequence of **observation likelihoods**, also called **emission probabilities**, each expressing the probability of an observation o_t being generated from a state q_i
$\pi = \pi_1, \pi_2, \ldots, \pi_N$	an **initial probability distribution** over states. π_i is the probability that the Markov chain will start in state i. Some states j may have $\pi_j = 0$, meaning that they cannot be initial states. Also, $\sum_{i=1}^{n} \pi_i = 1$

Figure: Components of Hidden Markov Model

First order Hidden Markov Model

A first-order HMM instantiates two simplifying assumptions

1. The probability of a particular state depends only on the previous state

   Markov Assumption: $P(q_i|q_1, ...q_{i-1}) = P(q_i|q_{i-1})$

2. The probability of an output observation depends only on the state that produced it and not on any other states.

   Independence: $P(o_i|q_1, ...q_i, ..., q_T, o_1, ...o_i, ..., o_T) = P(o_i|q_i)$

HMM Tagger

A model in Natural Language Processing based on HMM, used for labeling elements in a sequence.

HMM Tagger consists of 2 components:

1. A: The probability of a tag occurring given the previous tag
2. B: The probability, given a tag, that it will be associated with a given word

HMM Tagger

The probability of a tag occurring given the previous tag

$$P(t_i|t_{i-1}) = \frac{C(t_{i-1}, t_i)}{C(t_{i-1})}$$

Example - In the WSJ corpus:

- MD occurs **13124** times
- MD is followed by VB **10471** times

Tag transition probability MD - VB:

$$P(VB|MD) = \frac{C(MD, VB)}{C(MD)} = \frac{10471}{13124} = 0.8$$

HMM Tagger

The probability of a word occurring associated with a tag

$$P(w_i|t_i) = \frac{C(t_i, w_i)}{C(t_i)}$$

Example - In the WSJ corpus:

- MD occurs **13124** times
- MD is associated with *will* **4046** times

Tag transition probability MD - VB:

$$P(will|MD) = \frac{C(MD, will)}{C(MD)} = \frac{4046}{13124} = 0.31$$

Reference

Speech and Language Processing (3rd ed. draft)
Dan Jurafsky and James H. Martin
Part I: Fundamental Algorithms, *Chapter 8: Sequence Labeling for Parts of Speech and Named Entities*

Thanks for listening!

Q&A section