

Study Week 7 - NLP & Computer Vision

Named Entity Recognition : Hidden Markov Model

Hannah Do

Introduction

❖ Named Entity Recognition :

Subtask of information extraction that seeks to locate and classify named entities mentioned in unstructured text into pre-defined categories such as person names, organizations, locations, medical codes, time expressions, quantities, monetary values, percentages, etc.

TimesBy Adam Goldman **ORG** and Michael S. SchmidtAug **PERSON** . 13 **CARDINAL** , 2018WASHINGTON **CARDINAL** — Peter Strzok **PERSON** , the F.B.I. **GPE** senior counterintelligence agent who disparaged President Trump **PERSON** in inflammatory text messages and helped oversee the Hillary Clinton **PERSON** email and Russia **GPE** investigations, has been fired for violating bureau policies, Mr. Strzok **PERSON** 's lawyer said Monday **DATE** .Mr. Trump and his allies seized on the texts — exchanged during the 2016 **DATE** campaign with a former F.B.I. **GPE** lawyer, Lisa Page — in **PERSON** assailing the Russia **GPE** investigation as an illegitimate “witch hunt.” Mr. Strzok **PERSON** , who rose over 20 years **DATE** at the F.B.I. **GPE** to become one of its most experienced counterintelligence agents, was a key figure in the early months **DATE** of the inquiry. Along with writing the texts, Mr. Strzok **PERSON** was accused of sending a highly sensitive search warrant to his personal email account. The

Hidden Markov Model

- Used for variety of sequence labeling tasks such as Speech Recognition, POS tagging, NER, etc.
- Generative model that uses joint probability of paired observation and label sequences.
- Assumes independence of each word from its context (Markov property)
- Quick Learning, global maximization of the joint probability over the whole observation - a complete analysis of input sentence is made before the decision of best sequence of labels.

Markov Chain

$$\mathbf{x} = (x_1 x_2 \dots x_n)$$

Observation sequence of words of length n

$$\mathbf{s} = (s_1 s_2 \dots s_n)$$

Sequence of states that provides for word sequence \mathbf{x}

$$\mathbf{t} = (t_1 t_2 \dots t_n)$$

Sequence of entity annotation

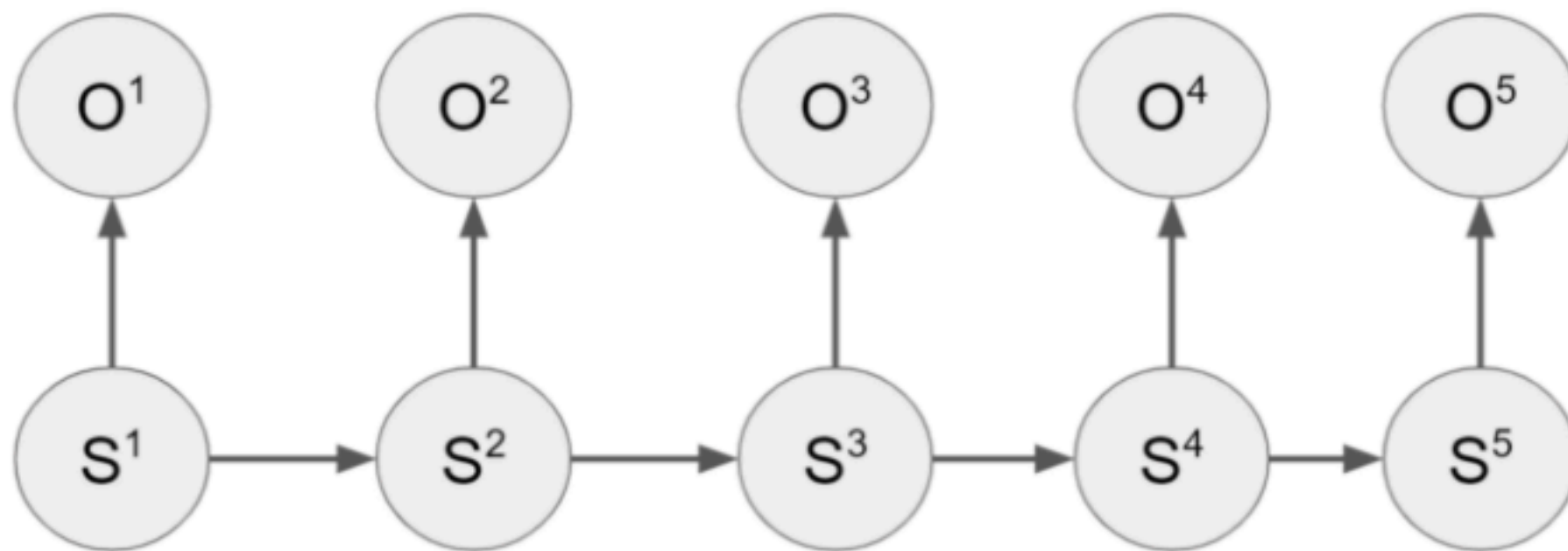
$$P(\mathbf{s}, \mathbf{x}) = P(\mathbf{x}|\mathbf{s})P(\mathbf{s})$$

HMM-based classifier belongs to Naive
Bayes classifiers

$$P(\mathbf{s}, \mathbf{x}) = \prod_{i=1}^n P(x_i | s_i) P(s_i | s_{i-1})$$

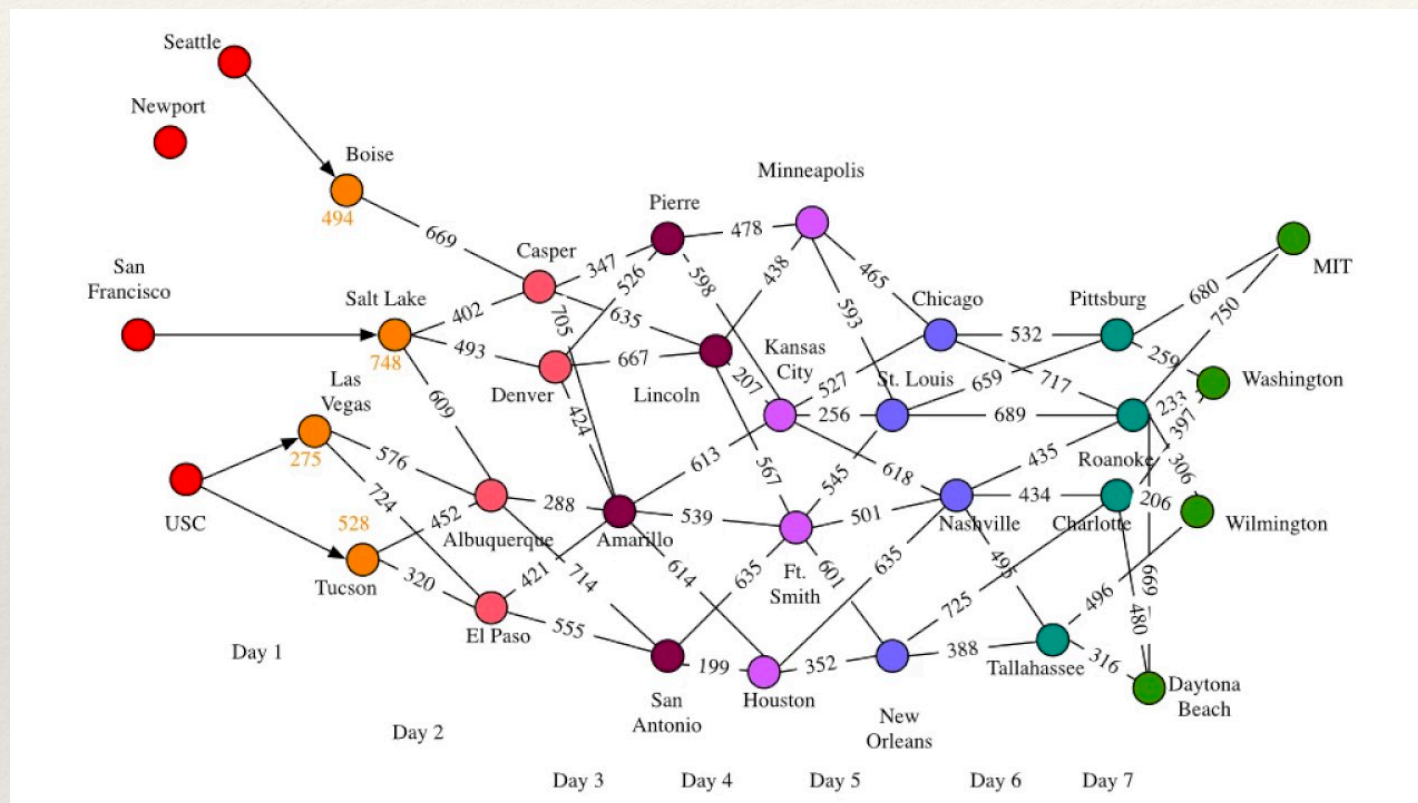
Based on Markov property, state S_i only depends on previous state S_{i-1} . The joint probability of a state sequence is represented above.

Why 'Hidden' Markov Model?



The dependency structure between states and observations in an HMM

Viterbi Algorithm

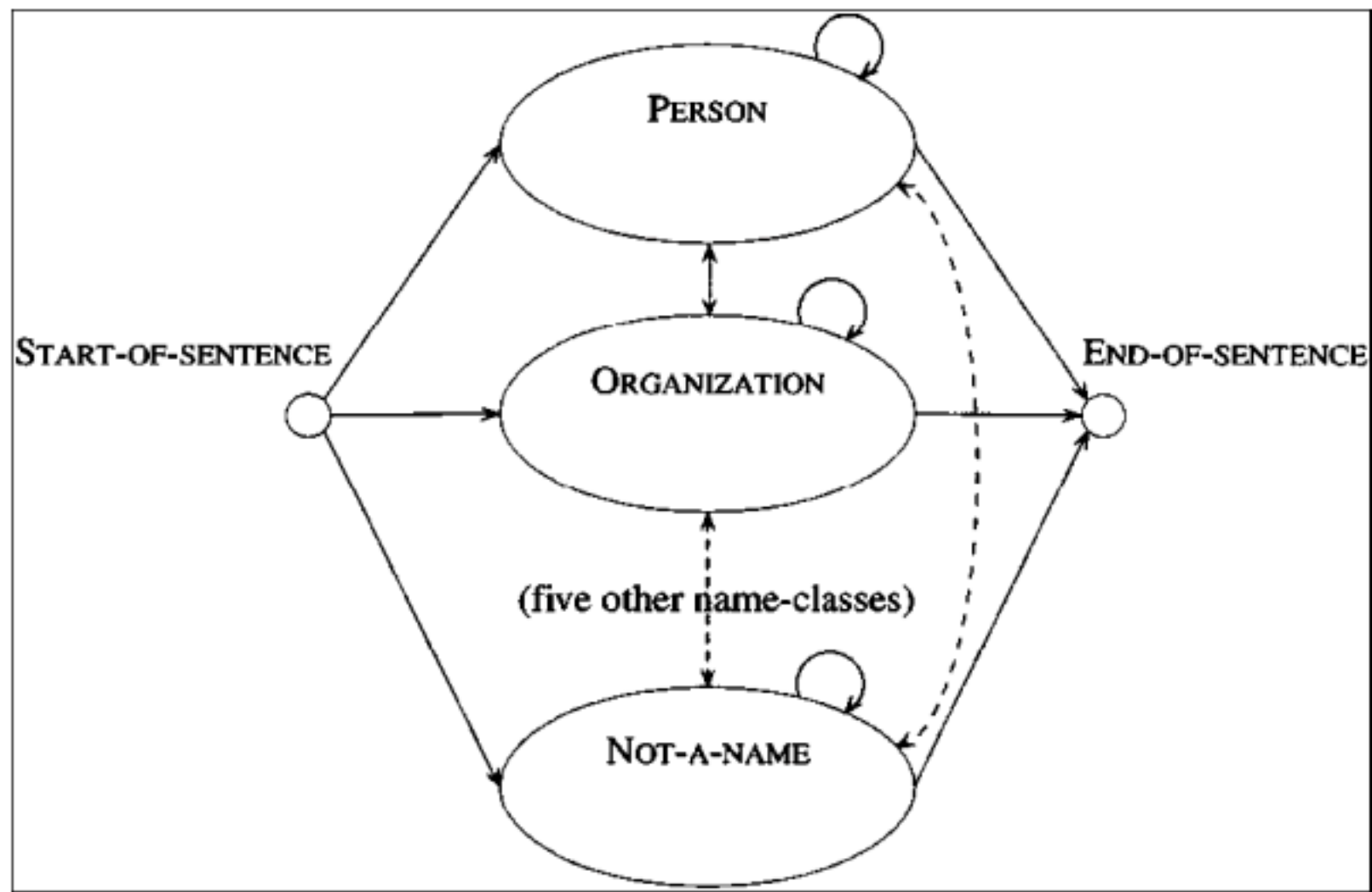


Viterbi Algorithm systematically eliminates those paths that cannot be part of the most likely path because they diverge and remerge with another path that has a smaller path metric.

After all n samples have been received, the path with the overall minimum path metric is selected as the most likely path and the input sequence associated with that path is the maximum-likelihood.

Keith Chugg, Viterbi Algorithm
(<https://www.youtube.com/watch?v=6JVqutwtzmo>)

Applications



Named Entity Recognition using HMM and MEMM models
(https://www.cse.iitb.ac.in/~cs626-460-2012/seminar_ppts/ner.pdf)

References

1. “Conditional Random Fields vs. Hidden Markov Models in a biomedical Named Entity Recognition task” Natalia Ponomareva, Paolo Rosso, Ferran Pla, Antonio Molina Universidad Politecnica de Valencia
2. Keith Chugg, Viterbi Algorithm (<https://www.youtube.com/watch?v=6JVqutwtzmo>)
3. Named Entity Recognition and Classification with scikit-learn (<https://towardsdatascience.com/named-entity-recognition-and-classification-with-scikit-learn-f05372f07ba2>)