Hidden Markov Models for Information Extraction

Speaker:

Bibek Paudel

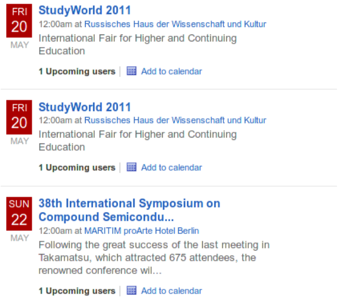
Tutor:

Maximilian Dylla

Saarland University, Germany

May 26, 2011

- *Location -> Russisches Haus der Wissenschaft und Kultur*



Motivation

- *Location -> Russisches Haus der Wissenschaft und Kultur*
- *Speaker -> Prof. Barbara Liskov*

Subject: Invitation: Prof. Barbara Liskov
From: "SWS Office Team" <office@mpi-sws.org>
Date: Mon, May 9, 2011 1:32 pm
To: sws-science@mpi-sws.org (more)
Priority: Normal
Options: View Full Header | View Printable Version | Download this as a file

As part of its Distinguished Lecture Series, the Max Planck Institute for Software Systems is pleased to announce:

Prof. Barbara Liskov, Massachusetts Institute of Technology

The Power of Abstraction

Monday, May 23, 2011 4:00 p.m. rotunda 57, TU Kaiserslautern

Simultaneous videocast to MPI-Inf room 024

Motivation

- *Location -> Russisches Haus der Wissenschaft und Kultur*
- *Speaker -> Prof. Barbara Liskov*
- Document usually contains much irrelevant text (sparse)

Motivation

- *Location -> Russisches Haus der Wissenschaft und Kultur*
- *Speaker -> Prof. Barbara Liskov*
- Document usually contains much irrelevant text (sparse)
- Find relations

Motivation

- *Location -> Russisches Haus der Wissenschaft und Kultur*
- *Speaker -> Prof. Barbara Liskov*
- Document usually contains much irrelevant text (sparse)
- Find relations
- Populate database slots with phrases from documents

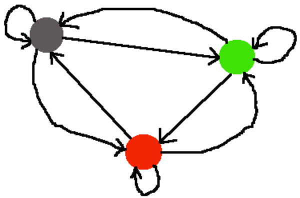
Agenda

Introduction

- Finite State Machines

Introduction

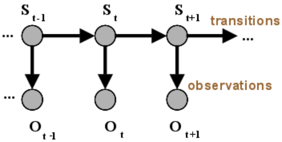
- Finite State Machines
- A generative process



Yes, Albert Einstein was born in Ulm.

Introduction

- Finite State Machines
- A generative process
- Next state depends only on current state



$$P(s,o) \propto \prod_{t=1}^{|o|} P(s_t \mid s_{t-1}) P(o_t \mid s_t)$$

Parameters: for all states $S=\{s_1, s_2, \ldots\}$

Start state probabilities: $P(s_t)$

Transition probabilities: $P(s_t|s_{t-1})$

Observation (emission) probabilities: $P(o_t|s_t)$ Usually a multinomial over atomic, fixed alphabet

Training:

Maximize probability of training observations (w/ prior)

Introduction

- Finite State Machines
- A generative process
- Next state depends only on current state
- Given some text, recover the states that generated the text

Why HMM

- Reasons for using HMM

The Task

Given:
a sequence of observations: Yes, Albert Einstein was born in Ulm.
a trained HMM

The Task

Given:
a sequence of observations: Yes, Albert Einstein was born in Ulm.
a trained HMM

- Find the most likely state sequence

The Task

Given:
a sequence of observations: Yes, Albert Einstein was born in Ulm.
a trained HMM

- Find the most likely state sequence
- Any words generated by the 'red state' are 'names'

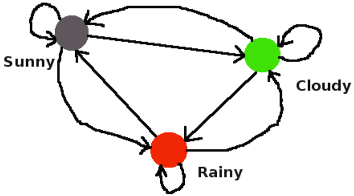
The Task

Given:
a sequence of observations: Yes, Albert Einstein was born in Ulm.
a trained HMM

- Find the most likely state sequence
- Any words generated by the 'red state' are 'names'
- Viterbi Algorithm

Agenda

An Example



S3, S3, S1, S1, S1, S3, S2, S3

An Example

- Prepare a matrix of conditional probabilities (pairwise)

$$\begin{bmatrix} a11 & a12 & a13 \\ a21 & a22 & a23 \\ a31 & a32 & a33 \end{bmatrix}$$

An Example

- Prepare a matrix of conditional probabilities (pairwise)

$$\begin{bmatrix} 0.4 & 0.3 & 0.3 \\ 0.2 & 0.6 & 0.2 \\ 0.1 & 0.1 & 0.8 \end{bmatrix}$$

An Example

- Prepare a matrix of conditional probabilities (pairwise)
- Define Observation Sequence $O = \{S_3, S_3, S_1, S_1, S_1, S_2, S_3\}$

An Example

- Prepare a matrix of conditional probabilities (pairwise)
- Define Observation Sequence $O = \{S_3, S_3, S_1, S_1, S_1, S_2, S_3\}$
- Find the probability $P(O \mid Model)$

An Example

- Prepare a matrix of conditional probabilities (pairwise)
- Define Observation Sequence $O = \{S_3, S_3, S_1, S_1, S_1, S_2, S_3\}$
- Find the probability $P(O \mid Model)$
- $P(O \mid Model) = P(S_3, S_3, S_1, S_1, S_1, S_2, S_3 \mid Model)$

An Example

- Prepare a matrix of conditional probabilities (pairwise)
- Define Observation Sequence $O = \{S_3, S_3, S_1, S_1, S_1, S_2, S_3\}$
- Find the probability $P(O \mid Model)$
- $P(O \mid Model) = P(S_3, S_3, S_1, S_1, S_1, S_2, S_3 \mid Model)$
- First order truncation by Markov property gives:

$$P(S_3) \prod P(q_t \mid q_{t-1})$$

An Example

- Prepare a matrix of conditional probabilities (pairwise)
- Define Observation Sequence $O = \{S_3, S_3, S_1, S_1, S_1, S_2, S_3\}$
- Find the probability $P(O \mid Model)$
- $P(O \mid Model) = P(S_3, S_3, S_1, S_1, S_1, S_2, S_3 \mid Model)$
- First order truncation by Markov property gives:

$$P(S_3) \prod P(q_t \mid q_{t-1})$$

$$P(S_3)P(S_3 \mid S_3)P(S_1 \mid S_3)P(S_1 \mid S_1)P(S_1 \mid S_1)P(S_2 \mid S_1)P(S_3 \mid S_2)$$

An Example

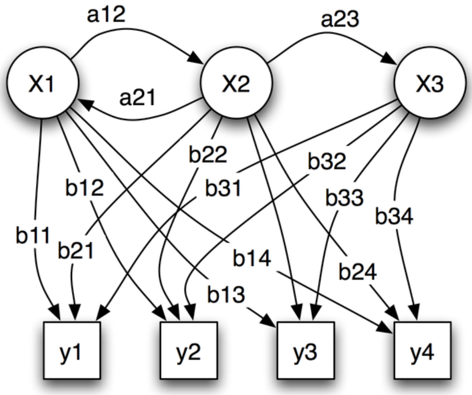


Figure: The Model

Agenda

Problems

For an observation sequence $O = O_1 ... O_T$, the three *canonical* HMM problems are:

- Evaluation Problem
  1. Given a model, find probability of Observation Sequence
  2. Forward-Backward

Problems

For an observation sequence $O = O_1 ... O_T$, the three *canonical* HMM problems are:

- Evaluation Problem
- Inference Problem
  1. Given the Observation Sequence $O = O_1 ... O_T$ and the model λ, find the most likely state sequence
  2. Choose an optimal state sequence $Q = q_1 q_2 ... q_T$

Problems

For an observation sequence $O = O_1 ... O_T$, the three *canonical* HMM problems are:

- Evaluation Problem
- Inference Problem
- Learning Problem
  1. How to model parameters in order to maximize probability of Observation Sequence ?
  2. We have to produce the actual model, the matrix

Agenda

Procedure: HowTo

Model $\lambda = (A, B, \pi)$ and observation sequence O

- Target, Prefix, Suffix and Background

Procedure: HowTo

Model $\lambda = (A, B, \pi)$ and observation sequence O

- Target, Prefix, Suffix and Background
- Lengthen, Split and Add

Procedure: HowTo

Model $\lambda = (A, B, \pi)$ and observation sequence O

- Target, Prefix, Suffix and Background
- Lengthen, Split and Add
- Shrinkage

Procedure: HowTo

Model $\lambda = (A, B, \pi)$ and observation sequence O

- Target, Prefix, Suffix and Background
- Lengthen, Split and Add
- Shrinkage
- Use dynamic programming to find the most likely state sequence

Procedure: HowTo

Model $\lambda = (A, B, \pi)$ and observation sequence O

- Target, Prefix, Suffix and Background
- Lengthen, Split and Add
- Shrinkage
- Use dynamic programming to find the most likely state sequence
- The Viterbi Algorithm

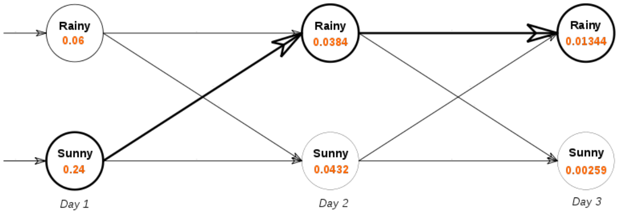


Figure: Viterbi

Procedure: The Details

- One HMM per class of information

Procedure: The Details

- One HMM per class of information
- Train on Labelled data

Procedure: The Details

- One HMM per class of information
- Train on Labelled data
- Background and Target States

Agenda

Results

	speaker	*location*	*acquired*	*dlramt*	*title*	*company*	*conf*	*deadline*	***Average***
Grown HMM	76.9	87.5	41.3	54.4	58.3	65.4	27.2	46.5	57.2
vs. SRV	+19.8	+16.0	+1.1	-1.6	—	—	—	—	+8.8
vs. Rapier	+23.9	+14.8	+12.5	+15.1	-11.7	+24.9	—	—	+13.3
vs. Simple HMM	+24.3	+5.6	+14.3	+5.6	+5.7	+11.1	+15.7	+6.7	+11.1
vs. Complex HMM	-2.1	+6.7	+7.5	-0.3	-0.3	+19.1	+0.0	-6.8	+3.0

Figure: Results

Conclusion

- Automatic generation of HMMs for IE
- Very effective and popular
- Better alternatives exist today (eg: CRF)

Sources

1. Information Extraction with HMM Structures Learned by Stochastic Optimization. Dayne Freitag and Andrew McCallum. AAAI'00.
2. A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. Lawrence R. Rabiner.