

CS4740 Intro to NLP

- **Today: sequence tagging applications in NLP**
 - part-of-speech tagging
 - hidden Markov model (HMM)
 - ➔ – named entity recognition (NER)
 - MEMMs

NE Identification

- **Identify all named locations, named persons, named organizations, dates, times, monetary amounts, and percentages.**

The delegation, which included the commander of the U.N. troops in Bosnia, Lt. Gen. Sir Michael Rose, went to the Serb stronghold of Palc, near Sarajevo, for talks with Bosnian Serb leader Radovan Karadzic.

Este ha sido el primer comentario publico del presidente Clinton respecto a la crisis de Oriente Medio desde que el secretario de Estado, Warren Christopher, decidiera regresar precipitadamente a Washington para impedir la ruptura del proceso de paz tras la violencia desatada en el sur de Libano.

1. Locations
2. Persons
3. Organizations

Figure 1.1 Examples. Examples of correct labels for English text and for Spanish text.

Guidelines need to be specified

- *The Wall Street Journal* : artifact or organization?
- *White House* : organization or location?
- Is a street name a location?
- Should *yesterday* and *last Tuesday* be labeled as dates?
- Is *mid-morning* a time?

Examples

1. **MATSUSHITA ELECTRIC INDUSTRIAL CO.** HAS REACHED AGREEMENT ...
2. IF ALL GOES WELL, **MATSUSHITA** AND ROBERT BOSCH WILL ...
3. **VICTOR CO. OF JAPAN (JVC)** AND SONY CORP. ...
4. IN A FACTORY OF **BLAUPUNKT WERKE**, A **ROBERT BOSCH SUBSIDIARY**, ...
5. **TOUCH PANEL SYSTEMS**, CAPITALIZED AT 50 MILLION YEN, IS OWNED ...
6. **MATSUSHITA EILL** DECIDE ON THE PRODUCTION SCALE. ...

Figure 2.1 English Examples. Finding names ranges from the easy to the challenging. Company names are in boldface. It is crucial for any name-finder to deal with the underlined text.

Training Data

- **Usually indicate NEs via SGML, XML, JSON**
 - Mark boundaries of expression
 - Label span with appropriate name class

Approaches to NE identification

- **Handcrafted finite state patterns**
 - <proper noun>+ <corporate designator> → <corporation>
 - Can't easily capture typical naming conventions
 - “Boston Power & Light” (corporation, electric utility)
 - Time-consuming to define
 - Maintenance is a problem
 - E.g. moving to NYT from WSJ
 - Not generally portable to new languages

HMM's for NE identification

- View NE identification as a word tagging task
 - e.g. part-of-speech tagging
- Local cues to identify named entities
- Goal: Train an HMM to label every word with one of the NE name classes or with a *not-a-name* class.
- Alternative: MEMMs, CRFs ...

Identifinder [Bikel et al. 1997, 1999]

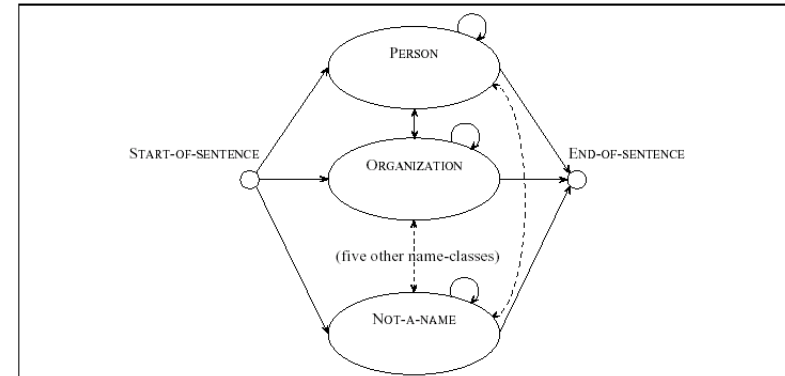
- First Hidden Markov model for recognizing and classifying named entities
- Outperformed other learning algorithms on standard data sets [MUC-6, MUC-7, MET-1]
- Competitive with approaches based on handcrafted rules on mixed case text
- Superior on text where case information isn't available

Identifinder

- **Handles 7 classes of NE' s**
 - entity
 - person
 - organization
 - location
 - time expression
 - date
 - time
 - numeric expression
 - money
 - percent

High-level view

A hidden Markov model represents the process of generating the sequence of words and labels



BBN' s Identifinder (Bikel et al. 1999)

NE Results Using HMM' s

Table 5.1 F-measure Scores. This table illustrates IdentiFinder's performance as compared to the best reported scores for each category.

	Language	Best Rules	IdentiFinder
Mixed Case	English (WSJ)	96.4	94.9
Upper Case	English (WSJ)	89	93.6
Speech Form	English (WSJ)	74	90.7
Mixed Case	Spanish	93	90

CS4740 Intro to NLP

- **Today: sequence tagging applications in NLP**
 - part-of-speech tagging
 - hidden Markov model (HMM)
 - named entity recognition (NER)
 - ➔ MEMMs

Hidden Markov Models

$Q = q_1 q_2 \dots q_N$ a set of N **states**
 $A = a_{11} a_{12} \dots a_{n1} \dots a_{nm}$ a **transition probability matrix** A , each a_{ij} representing the probability of moving from state i to state j , s.t. $\sum_{j=1}^n a_{ij} = 1 \quad \forall i$
 $O = o_1 o_2 \dots o_T$ a sequence of T **observations**, each one drawn from a vocabulary $V = v_1, v_2, \dots, v_V$
 $B = b_i(o_t)$ a sequence of **observation likelihoods**, also called **emission probabilities**, each expressing the probability of an observation o_t being generated from a state i
 q_0, q_F a special **start state** and **end (final) state** that are not associated with observations, together with transition probabilities $a_{01} a_{02} \dots a_{0n}$ out of the start state and $a_{1F} a_{2F} \dots a_{nF}$ into the end state

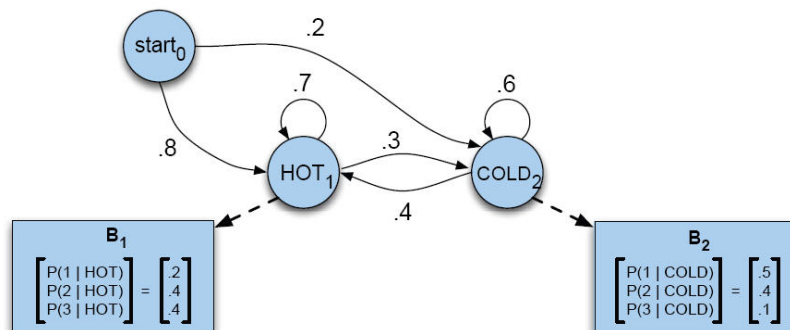
Figure, copyright J&M 2nd ed

HMMs for entity detection

American		B _{ORG}
Airlines		I _{ORG}
,		O
a		O
unit		O
of		O
AMR		B _{ORG}
Corp.		I _{ORG}
,		O
immediately		O
matched		O
the		O
move		O
,		O
spokesman		O
Tim		B _{PER}
Wagner		I _{PER}
said		O
.		O

Figure, copyright J&M 2nd ed

HMM for weather prediction

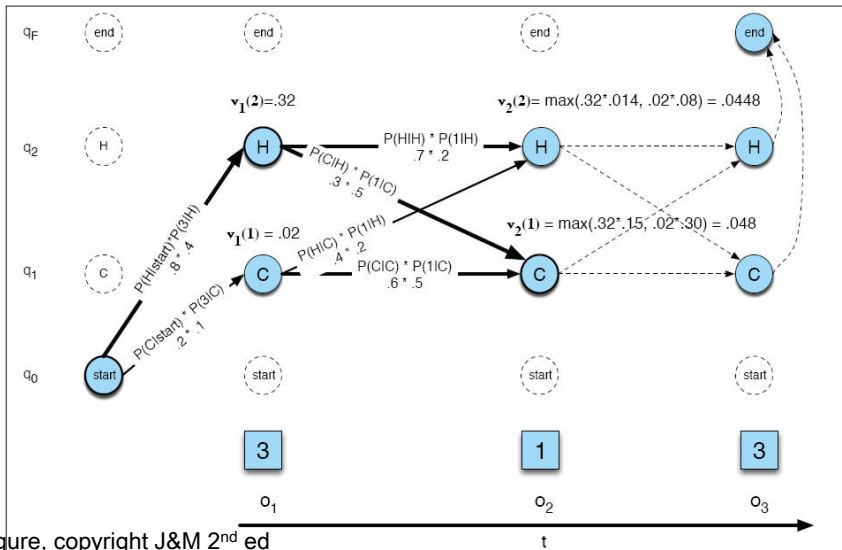


Figure, copyright J&M 2nd ed

HMM equations

- Finding the most likely tag (state) T sequence given a sequence of observations O**

Viterbi

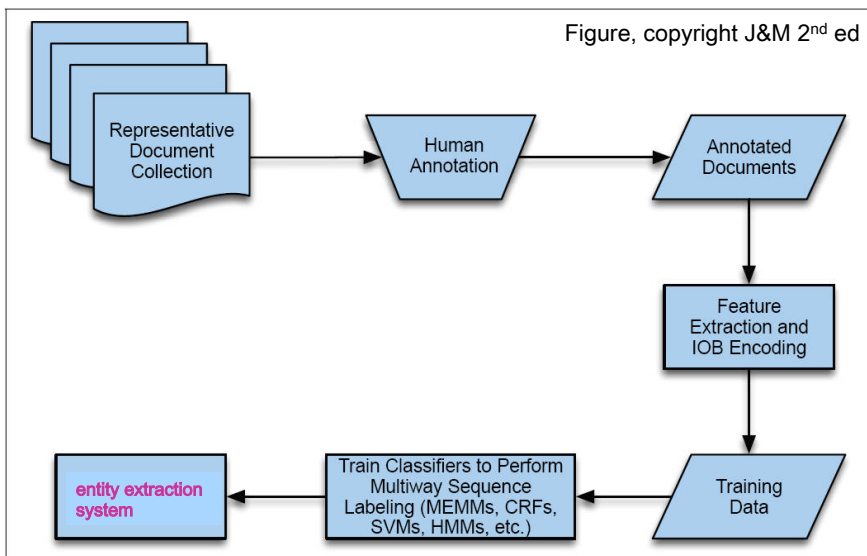


Classification approach???

Features	Label
American	B _{ORG}
Airlines	I _{ORG}
,	O
a	O
unit	O
of	O
AMR	B _{ORG}
Corp.	I _{ORG}
,	O
immediately	O
matched	O
the	O
move	O
,	O
spokesman	O
Tim	B _{PER}
Wagner	I _{PER}
said	O
.	O

Figure, copyright J&M 2nd ed

End-to-end process



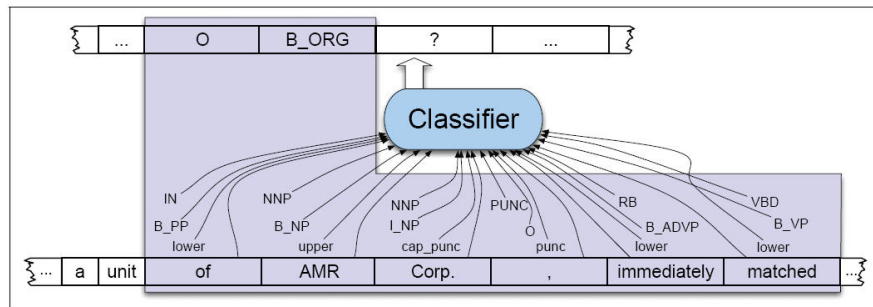
Classification approach???

Features	Label
American	B _{ORG}
Airlines	I _{ORG}
,	O
a	O
unit	O
of	O
AMR	B _{ORG}
Corp.	I _{ORG}
,	O
immediately	O
matched	O
the	O
move	O
,	O
spokesman	O
Tim	B _{PER}
Wagner	I _{PER}
said	O
.	O

Figure, copyright J&M 2nd ed

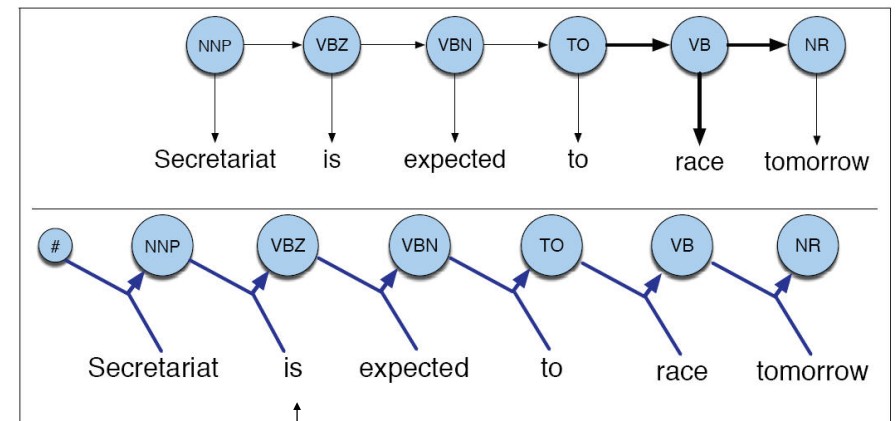
Feature extraction

- We'd like to be able to include lots of features as in classification-based approaches (e.g. SVMs, dtrees)



Figure, copyright J&M 2nd ed

Not possible with HMMs



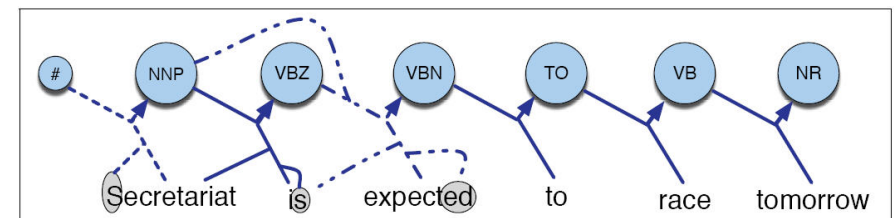
Maximum entropy Markov model (MEMM)

Figure, copyright J&M 2nd ed

MEMM equations

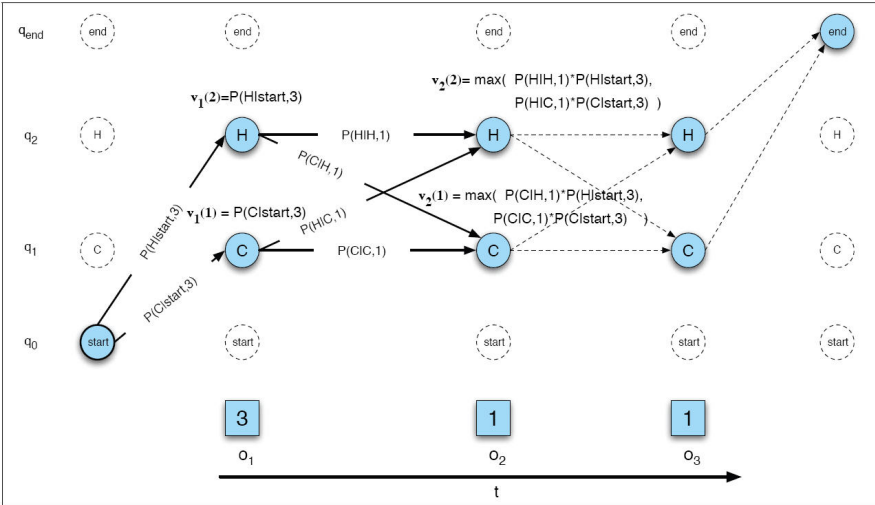
MEMM for p-o-s tagging

- Condition on many features of the input
 - Capitalization
 - Morphology
 - Earlier words
 - Earlier tags



Figure, copyright J&M 2nd ed

Decoding/inference in MEMMs



Figure, copyright J&M 2nd ed

- **Next class**
 - Sentiment/opinion analysis