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

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 WILL DECIDE ON THE PRODUCTION SCALE ... 

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

Guidelines need to be specified
- The delegation, which included the commander of the [soldiers] in *Srebrenica*, Lt. Gen. Sir Michael Rose, went to the Serb stronghold of [Prijedor], near Sarajevo, for talks with Bosnian Serb leader Radovan Karadzic.

Guidelines need to be specified
- Este ha sido el primer comentario publico del presidente *Clinton* respecto a la crisis de Orienete Medio desde que el secretario de Estado, *Warren Christopher*, decidió regresar precipitadamente a *Washington* para impedir la ruptura del proceso de paz tras la violencia desatada en el sur de *Libano*.

Guidelines need to be specified

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

• Usually indicate NE’s via SGML or XML
  – 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

Identifinder [Bikel et al. 1997, 1999]

• Hidden Markov model that learns to recognize and classify named entities.
• Outperforms 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
HMM’s for NE identification

• View NE identification as a sequence of word classification tasks
• Successful for other “word tagging” tasks, 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 ...

High-level view

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

Using the HMM

• Goal: find the most likely sequence of name classes, NC, given a sequence of words W
  – W: Banks filed bankruptcy papers
  – Compare the probability of
    <person, not-a-name, not-a-name, not-a-name>
    <not-a-name, not-a-name, not-a-name, not-a-name>
    ...
  – Viterbi algorithm is a dynamic programming algorithm that performs this computation efficiently.

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.

<table>
<thead>
<tr>
<th>Case</th>
<th>Language</th>
<th>Best Rules</th>
<th>Identifinder</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mixed Case</td>
<td>English (WSJ)</td>
<td>96.4</td>
<td>94.9</td>
</tr>
<tr>
<td>Upper Case</td>
<td>English (WSJ)</td>
<td>89</td>
<td>93.6</td>
</tr>
<tr>
<td>Speech Form</td>
<td>English (WSJ)</td>
<td>74</td>
<td>90.7</td>
</tr>
<tr>
<td>Mixed Case</td>
<td>Spanish</td>
<td>93</td>
<td>90</td>
</tr>
</tbody>
</table>
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

A Hidden Markov Model (HMM) is a statistical model that represents a sequence of observations where the underlying state sequence is not directly observable. It is widely used in natural language processing for tasks such as part-of-speech tagging, named entity recognition, and speech recognition.

### Components of an HMM
- **States** ($Q = q_1 q_2 \ldots q_N$): A set of $N$ states.
- **Transition Probabilities** ($A = a_{ij}$): A transition probability matrix $A$, each $a_{ij}$ representing the probability of moving from state $i$ to state $j$, s.t. $\sum_{i} a_{ij} = 1 \text{ for all } j$.
- **Observations** ($O = o_1 o_2 \ldots o_T$): A sequence of $T$ observations, each one drawn from a vocabulary $V = v_1, v_2, \ldots, v_V$.
- **Emission Probabilities** ($B = b_i(o_t)$): A sequence of observation likelihoods, also called emission probabilities, each expressing the probability of an observation $o_t$ being generated from a state $i$.
- **Start and End States** ($q_0, q_F$): A special start state and end (final) state that are not associated with observations, together with transition probabilities $a_{01}a_{02}\ldots a_{0N}$ out of the start state and $a_{NF}a_{N2}\ldots a_{NF}$ into the end state.

### Examples

**HMMs for entity detection**

<table>
<thead>
<tr>
<th>Entity</th>
<th>POS</th>
<th>B-ORG</th>
<th>I-ORG</th>
<th>O</th>
</tr>
</thead>
<tbody>
<tr>
<td>American</td>
<td>AMR</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Airlines</td>
<td>AMR</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.</td>
<td>.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>unit</td>
<td>.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>of</td>
<td>.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AMR Corp.</td>
<td>.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.</td>
<td>.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>immediately</td>
<td>.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>matched</td>
<td>.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>the</td>
<td>.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>move</td>
<td>.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.</td>
<td>.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>spokesman</td>
<td>.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tim Wagner</td>
<td>.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>said</td>
<td>.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.</td>
<td>.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**HMM for weather prediction**

A weather prediction HMM can be used to predict the weather conditions at different times of the day. The model takes into account the current weather conditions and the transitions between different states to predict future conditions.

Figure, copyright J&M 2nd ed
HMM equations

Viterbi

Classification approach???

<table>
<thead>
<tr>
<th>Features</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>American</td>
<td>BORG</td>
</tr>
<tr>
<td>Airlines</td>
<td>BORG</td>
</tr>
<tr>
<td>.</td>
<td>O</td>
</tr>
<tr>
<td>a</td>
<td>O</td>
</tr>
<tr>
<td>unit</td>
<td>O</td>
</tr>
<tr>
<td>of</td>
<td>O</td>
</tr>
<tr>
<td>AMR</td>
<td>BORG</td>
</tr>
<tr>
<td>Corp.</td>
<td>BORG</td>
</tr>
<tr>
<td>immediately</td>
<td>O</td>
</tr>
<tr>
<td>matched</td>
<td>O</td>
</tr>
<tr>
<td>the</td>
<td>O</td>
</tr>
<tr>
<td>move</td>
<td>O</td>
</tr>
<tr>
<td>,</td>
<td>O</td>
</tr>
<tr>
<td>spokesman</td>
<td>O</td>
</tr>
<tr>
<td>Tim</td>
<td>BPER</td>
</tr>
<tr>
<td>Wagner</td>
<td>BPER</td>
</tr>
<tr>
<td>said</td>
<td>O</td>
</tr>
<tr>
<td>.</td>
<td>O</td>
</tr>
</tbody>
</table>

End-to-end process
Feature extraction

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

MEMM equations

MEMM for p-o-s tagging

- Condition on many features of the input
  - Capitalization
  - Morphology
  - Earlier words
  - Earlier tags
Decoding/inference in MEMMs

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