Introduction

What is Natural Language Understanding?

Natural language understanding:
- often refers to full comprehension/semantic processing of language;
- here, natural language understanding is used to contrast with natural language generation.

Understanding:

- Text $\Rightarrow$ Analyses (parse trees, logical forms, discourse segmentation, etc.)

Generation:

- Non-linguistic input (logical forms, database entries, etc.) or text $\Rightarrow$ Text

Course Mechanics

Reading: Jurafsky and Martin (2009: Ch. 6.8).

Relation to other Courses

NLP courses:
- Advanced Natural Language Processing OR Foundations of Natural Language Processing;
- Machine Translation; Natural Language Understanding; Natural Language Generation; Topics in Natural Language Processing.

Additional prerequisites:
- IAML or PMR;
- CPSLP or equivalent programming experience.

Related course: Text Technologies.
Course Content

We will cover topics in syntax, semantics, and discourse:

- few topics in detail, rather than a broad overview.

Key concepts and methods:

- probabilistic (e.g., Bayesian) versus statistical approaches;
- discriminative versus generative models;
- unsupervised methods;
- human language processing.

Assignments will involve practical work with NLTK.

Probabilistic vs. Statistical Approaches

Most current NLP systems use statistical methods:

<table>
<thead>
<tr>
<th>Decision trees</th>
<th>Naive Bayes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Support vector machines</td>
<td>PCFGs, HMMs</td>
</tr>
<tr>
<td>Nearest neighbor</td>
<td>Maximum entropy models</td>
</tr>
<tr>
<td>Heuristic approaches</td>
<td>Conditional random fields</td>
</tr>
</tbody>
</table>

Statistical: involves function(s) computed from data:

- counts, means, probabilities, thresholds set using development set, etc.

Discriminative vs. Generative Models

Only some of these are probabilistic:

- observed variables \((x_1, \ldots, x_n)\); e.g., word string;
- hidden variables \((y_1, \ldots, y_n)\); e.g., tag sequence;
- model parameters \(\theta\), e.g., \(P(x_i = \text{cat}| y_i = \text{NN})\).

Most probabilistic models in NLP are generative, trained to maximize the joint distribution over the training data:

\[
\hat{\theta} = \arg\max_{\theta} \prod_{i=1}^{n} P_{\theta}(y_i, x_i)
\]
Discriminative vs. Generative Models

An alternative is to use discriminative models, trained to maximize the following conditional distribution:

$$\hat{\theta} = \arg\max_{\theta} \prod_{i=1}^{n} P_\theta(y_i|x_i)$$

This means we use maximum conditional likelihood estimation (MCLE) instead of standard maximum likelihood estimation (MLE) for generative models (Johnson 2001).

Figure from Johnson (2001).

Supervised vs. Unsupervised Methods

Standard statistical NLP systems use a supervised paradigm:

**Training:**

- Labeled training data
- Statistics
- Machine learning system
- Prediction procedure

**Evaluation/deployment:**

- New unlabeled data
- Statistics
- Prediction procedure (from training)
- Labeled output
**Supervised vs. Unsupervised Methods**

Recent work aims to improve unsupervised NLP systems, to reduce annotation costs and broaden applicability:

Some unsupervised tasks we’ll cover:

Part of speech induction:

```
walk
runners
keyboard
desalinated
```

```
walk.VVB
runners.NNS
keyboard.NN
desalinate.VVD
```

**Modeling Human Language Processing**

We will also upon some issues from cognitive science in this course:

- humans understand text and speech effortlessly in real time, robust to ambiguity and noise;
- principles of human language processing (from linguistics, psychology) can inform engineering solutions:
  - humans learn language from exposure to linguistic data;
  - human learning is mostly unsupervised;
  - human language processing is often heuristic and approximative;
    - it operates with limited memory and input data.
- but: important not to overstate analogies (planes and birds both fly, but using very different mechanisms).
Examples of how human language processing can inform NLP models:

- human parsing is incremental (structures are built word by word): constrains possible parsing algorithms;
- human lexical processing is subject to priming (words with similar meanings are processed faster): constrains possible semantic representations;
- human language learning uses limited input (about 350M words during lifetime): constraints training set size.

Assessment will consist of:

- two assessed assignments, worth 15% each (30% total);
- a final exam (120 minutes), worth 70%.

Due dates:

- Assignment 1 (semantic parsing): issued Jan 30, due Feb 13;

Assignment deadlines will be preceded by feedforward sessions in which you can ask questions about the assignments.
Plagiarism

**Definition:** Plagiarism is the act of copying or including in one’s own work, without adequate acknowledgment, intentionally or unintentionally, the work of another. It is academically fraudulent and an offence against University discipline.

**Details:**
http://www.inf.ed.ac.uk/teaching/plagiarism.html

---

Examples of plagiarism:

- including extracts from another person’s work without using quotation marks and acknowledgment of source;
- summarizing others’ work without acknowledgment;
- using others’ ideas or help without acknowledgment;
- copying another student’s work, with or without their knowledge or agreement;
- collaborating with students or others on work that should be completed individually;
- cutting and pasting text, illustrations, diagrams, etc. from electronic sources without acknowledging the URL.

---

**References**

