Christopher Robin is alive and well. He lives in England. He is the same person that you read about in the book, Winnie the Pooh. As a boy, Chris lived in a pretty home called Cotchfield Farm. When Chris was three years old, his father wrote a poem about him. The poem was printed in a magazine for others to read. Mr. Robin then wrote a book. He made up a fairy tale land where Chris lived. His friends were animals. There was a bear called Winnie the Pooh. There was also an owl and a young pig, called a piglet. All the animals were stuffed toys that Chris owned. Mr. Robin made them come to life with his words. The places in the story were all near Cotchfield Farm. Winnie the Pooh was written in 1925. Children still love to read about Christopher Robin and his animal friends. Most people don’t know he is a real person who is grown now. He has written two books of his own. They tell what it is like to be famous.

1. Christopher Robin was born in England.
2. Winnie the Pooh is a title of a book.
3. Christopher Robin’s dad was a magician.
4. Christopher Robin must be at least 65 now.

This is an Inference Problem
Illinois' bored of education

...Nissan Car and truck plant is ...
...divide life into plant and animal kingdom

(The Art) (can N) (will MD) (rust V) V,N,N

The dog bit the kid. He was taken to a veterinarian
a hospital

Tiger was in Washington for the PGA Tour
Christopher Robin is alive and well. He lives in England. He is the same person that you read about in the book, Winnie the Pooh. As a boy, Chris lived in a pretty home called Cotchfield Farm. When Chris was three years old, his father wrote a poem about him. The poem was printed in a magazine for others to read. Mr. Robin then wrote a book. He made up a fairy tale land where Chris lived. His friends were animals. There was a bear called Winnie the Pooh. There was also an owl and a young pig, called a piglet. All the animals were stuffed toys that Chris owned. Mr. Robin made them come to life with his words. The places in the story were all near Cotchfield Farm. Winnie the Pooh was written in 1925. Children still love to read about Christopher Robin and his animal friends. Most people don’t know he is a real person who is grown now. He has written two books of his own. They tell what it is like to be famous.

1. Who is Christopher Robin?  
2. When was Winnie the Pooh written?  
3. What did Mr. Robin do when Chris was three years old?  
4. Where did young Chris live?  
5. Why did Chris write two books of his own?
An Owed to the Spelling Checker

I have a spelling checker, it came with my PC
It plane lee marks four my revue
Miss steaks aye can knot sea.
Eye ran this poem threw it, your sure reel glad two no.
Its vary polished in it’s weigh
My checker tolled me sew.
A checker is a bless sing, it freeze yew lodes of thyme.
It helps me right awl stiles two reed
And aides me when aye rime.
Each frays come posed up on my screen
Eye trussed to bee a joule...
Intelligent Access to Information

- Access free form text
  - News articles; reports: maintenance, projects,…..
  - E-mail; web data
  - Mixed form information: layout intensive (lists, tables, databases…)
- As if it was a data base
  
  Ability to identify the semantics of the text

- Specific Tasks:
  
  Basic recognition, categorization & tagging tasks; semantic analysis; semantic integration; textual entailment,…. Done within and across documents

- Techniques: Machine Learning and Inference
Intelligent Access to Information: Tasks

- Named Entity (Semantic Categorization):
  - Identifying names of entities and concepts in the text

  [JFK was busy; the parking lots were full] LOC

- Semantic Relations:
  - Identifying relations between entities in documents

  [Dr. ABC joined Microsoft, Redmond and will lead the SearchIt project.]

- Cross-Document Entities Identification and Tracing
  - Robust Reading of Text; overcome variability in writing

  [The JFK problem; The Michael Jordan Problem]

- Temporal integration of Information
  - Tracking of entities along time; information integration; change.

  [Dr. ABC joined Google to save the AnswerIt project.]

- Question Answering
When searching for Wilson, we found 5 names you might be interested in:

Woodrow Wilson

Woodrow Wilson
In his low-key way, though, Wilson sometimes is.

Bill Clinton
San Francisco 49ers
Pepsi-Cola
Romney

Related Entities »

White House
San Francisco 49ers
Pepsi-Cola

Related news articles (5):

Luke Wilson

(a Luke Wilson who currently can be seen working the same sweat shop quality in "Blue Streak"), is a recently dumped young bachelor who misses his ditzy girlfriend, Cheryl (Kathleen Robertson), perhaps a little less than he does his peep, Memaw, who is now a child of joint custody.

Related Entities »

Doo Park
Doo Park
Bruce McCullough
Doo Park

Related news articles (2):

Owen Wilson

In his low-key way, though, Wilson sometimes is.

Related Entities »

Doo Park
Boston

Related news articles (2):

Pete Wilson

Screen shot from a CCG demo
http://L2R.cs.uiuc.edu/~cogcomp

More work on this problem:
Scaling up
Integration with DBs
Temporal Integration/Inference

......
Understanding Questions

- What is the question asking? (different from Googling)
- Beyond finding candidate passages; choose the right one.

**Q:** What is the fastest automobile in the world?

**A1:** ...will stretch Volkswagen’s lead in the world’s fastest growing vehicle market. Demand for cars is expected to soar

**A2:** ...the Jaguar XJ 220 is the dearest (415,000 pounds), fastest (217mph) and most sought after car in the world.

- Context: News articles (SJM, LAT, ...)
- And, what if the answers require aggregation,...
Not So Easy

WAP WAP WAP

CALVIN! WHAT ARE YOU DOING TO THE COFFEE TABLE?!!

*

IS THIS SOME SORT OF TRICK QUESTION, OR WHAT?
Eyeing the huge market potential, currently led by Google, Yahoo took over search company Overture Services Inc. last year.

Is it true that...

(Textual Entailment)

Yahoo acquired Overture
Overture is a search company
Google is a search company
Google owns Overture

------
Why Textual Entailment?

- A fundamental task that can be used as a building block in multiple NLP and information extraction applications
- Has multiple direct applications
A key problem in natural language understanding is to abstract over the inherent syntactic and semantic variability in natural language. Multiple tasks attempt to do just that.

- Relation Extraction:
  Dole’s wife, Elizabeth, is a native of Salisbury, N.C. \( \subseteq \)
  Elizabeth Dole was born in Salisbury, N.C

- Information Integration (Data Bases)
  Different database schemas represent the same information under different titles.

- Information retrieval:
  Multiple issues, from variability in the query and target text, to relations

- Summarization; Paraphrasing
  Multiple techniques can be applied; all are entailment problems.
Direct Application: Semantic Verification

- **Given:**
  
  A long contract that you need to **ACCEPT**

- **Determine:**
  
  Does it satisfy the 3 conditions that you really care about?

---

**Cognitive Computation Group**

**University of Illinois at Urbana-Champaign**
What’s Needed?

- A collection of tools that are essential for any intelligent use of text...
- Robust text analysis tools
  - Tokenization; POS tagging; Shallow parsing;
  - Parsing, Semantic Parsing,…
- Name Entity Classifiers
  - people; locations; organizations; transport; materials…
- Information Extraction
  - functional phrases (e.g., job descriptions; acquisitions)
- Relations/Event recognizers
  - born_in(A,B); capital_of(C,D); killed(A,B)
  - implicit relations in sentences…
  - A lot more...(knowledge; better semantic inference)
Classification: Ambiguity Resolution

Illinois' bored of education [board]
Nissan Car and truck plant; plant and animal kingdom
(This Art) (can N) (will MD) (rust V) V,N,N
The dog bit the kid. He was taken to a veterinarian; a hospital
Tiger was in Washington for the PGA Tour

→ Finance; Banking; World News; Sports

Important or not important; love or hate
Classification

- The goal is to learn a function $f: X \rightarrow Y$ that maps observations in a domain to one of several categories.

- **Task:** Decide which of \{board, bored\} is more likely in the given context:
  - $X$: some representation of:
    - The Illinois’ ______ of education met yesterday…
  - $Y$: \{board, bored\}

- **Typical learning protocol:**
  - Observe a collection of labeled examples $(x,y) \in X \times Y$
  - Use it to learn a function $f:X \rightarrow Y$ that is **consistent** with the observed examples, and (hopefully) performs well on new, previously unobserved examples.
Theoretically: generalization bounds
- How many example does one need to see in order to guarantee good behavior on previously unobserved examples.

Algorithmically: good learning algorithms for linear representations.
- Can deal with very high dimensionality ($10^6$ features)
- Very efficient in terms of computation and # of examples. On-line.

Key issues remaining:
- Learning protocols: how to minimize interaction (supervision); how to map domain/task information to supervision; semi-supervised learning; active learning; ranking.
- What are the features? No good theoretical understanding here.

Classification is Well Understood

Attributes (node labels):
- person
  - name("Mohammed Atta")
  - gender(male)
- location
  - city name("Prague")
- country
  - name("Czech Republic")
- meeting
- date
  - month(April)
  - year(2001)
- participant
- nationality
  - country name("Iraq")
- affiliation
- organization

Roles (edge labels):
- meeting (participant)
- before
  - word(an) tag(DT)
  - word(intelligence) tag(NN)
  - word(Iraqi) tag(JJ)
- before
  - after
Semantic Parse (Semantic Role Labeling)

Semantic Role Labeling Output

Input Text:
A car bomb that exploded outside the U.S. military base in Beniji killed 11 Iraqi citizens.

Result: Complete!

- General Explanation of Argument Labels

<table>
<thead>
<tr>
<th>Argument</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>bomb [A1]</td>
</tr>
<tr>
<td>car</td>
<td>killer [A0]</td>
</tr>
<tr>
<td>bomb</td>
<td>bomb (Reference) [R-A1]</td>
</tr>
<tr>
<td>that</td>
<td>V: explode</td>
</tr>
<tr>
<td>exploded</td>
<td>outside</td>
</tr>
<tr>
<td></td>
<td>location [AM-LOC]</td>
</tr>
<tr>
<td>the</td>
<td>military</td>
</tr>
<tr>
<td>U.S.</td>
<td>base</td>
</tr>
<tr>
<td></td>
<td>temporal [AM-TMP]</td>
</tr>
<tr>
<td>in</td>
<td>Beniji</td>
</tr>
<tr>
<td></td>
<td>location [AM-LOC]</td>
</tr>
<tr>
<td>killed</td>
<td>V: kill</td>
</tr>
<tr>
<td>11</td>
<td>corpse [A1]</td>
</tr>
<tr>
<td>Iraqi</td>
<td>citizens</td>
</tr>
</tbody>
</table>
Unsupervised Text Models, Topic Modeling

- Use co-occurrence information to discover ‘topics’
- Used in:
  - Visualization of Text Collections
  - Text Segmentation
  - Summarization (summarize along topics)
  - Sentiment Prediction
Role of Learning

- “Solving” a natural language problem requires addressing a wide variety of questions.

- Learning is at the core of any attempt to make progress on these questions.

- Learning has multiple purposes:
  - Knowledge Acquisition
  - Classification/Context sensitive disambiguation
  - Integration of various knowledge sources to ensure robust behavior
Why Natural Language?

- Main channel of Communication
- Knowledge Acquisition
- Important for Cognition and Engineering perspectives
- A grand application: Human computer interaction.
  - Language understanding and generation
  - Knowledge acquisition
  - NL interface to complex systems
  - Querying NL databases
  - Reducing information overload
Why Learning in Natural Language?

- Challenging from a Machine Learning perspective
  - There is no significant aspect of natural language that can be studied without giving learning a principle role.

- Language Comprehension: a large scale phenomenon in terms of both knowledge and computation

- Requires an integrated approach: need to solve problems in learning, inference, knowledge representation...

- There is no “cheating”: no toy problems.
This Course

- There are many topics to cover.

- A very active field of research, many ideas are floating around, most of which will not stay.

- Rather than covering problems - we will cover some of the main ideas and techniques.

- Attempt to abstract away from specific works and understand the main paradigms.

- Move towards: Beyond Classification
  - Knowledge Representation and Inference
    - Joint Inference
    - Joint Learning
  - Let me know if you have specific interests
This Course

- Representation-less Approaches
  - Statistics
- Paradigms
  - Generative and Discriminative
  - Understanding why things work
- Classification: Learning Algorithms
  - Generative and Discriminative algorithms
  - The ubiquitous Linear Representation
  - Features and Kernels
  - Latent Variable Models
- Inference
  - Generative models; Conditional Models
  - Inference with Classifiers; Joint Inference; Joint Learning
    - CRF; Structured Perceptrons and SVMs; Constraint Conditional Models
    - Hierarchical Bayesian Models

Problems

- Language Modeling (Verb Classification?)
- (Dependency) Syntactic Parsing
- (Shallow) Semantic Analysis (Semantic Role Labeling)
- Inferring Sequential Structure
The Course Plan

- Introduction
  - Why Learning; Learning vs. Statistics

- Learning Paradigms
  - Basics of Generalization Theory
  - Generative Models
  - LSQ (Probabilistic Approaches Work)

- Power of Generative Models
  - Modeling
  - HMM, K-Means, EM

- Discriminatory Algorithm
  - Classification & Inference
  - Linear Learning Algorithms

- Learning Structured Representations
  - Representation & Inference
  - Sequential and General structures

- Features
  - Feature extraction languages
  - Kernels (over structures)
  - Latent Variable Models

- Using Unlabeled/Noisy Data
  - Semi-Supervised Methods
  - Domain Adaptation

- Hierarchical Bayesian Models
  - Topic Models and Their Applications
  - Non-Param methods in NLP
More Detailed Plan (1)

1. Introduction to Natural Language Learning
   - Why is it difficult?
   - Statistics vs. Learning
   - When do we need learning?
   - Examples of problems

2. Statistics and Information Theory
   - Corpus based work: data and tasks.

3. Learning Paradigms (Basics)
   - PAC Learning
   - Bayesian Learning
   - Examples
More Detailed Plan (II)

3. Learning Algorithms
   - Examples
   - General Paradigm: feature based representation
   - Linear functions
   - On line algorithms
   - Support-Vector Machines

4. Probabilistic Classifiers
   - Naïve Bayes
   - HMMs (Predictions and model learning)
   - Max Entropy
   - LSQ: why do probabilistic algorithms work?

5. Sequential Learning (aka Sequence Labeling)
   - HMMs (with Classifiers), PMMs
   - Discriminative Learning Methods: CRFs, SVMStruct, Max-Margin Markov Networks,...
   - Constraint Conditional Models
More Detailed Plan (III)

5. **Inference: Complex Models**
   - Parsing: Generative and Condit. Models, Reranking
   - Inference as constrained optimization
   - Generative vs Discriminative

6. **Features**
   - Kernels for Structured Data (Convolution Kernels,...)
   - Inducing Features from Probabilistic Models
   - Latent Variable Models

7. **Unsupervised Models**
   - Topic Modeling
   - Hierarchical Bayesian Models for NLP (Dirichlet Processes,...)

8. **Relaxing Supervision**
   - Semi supervised learning
   - Domain-Adaptation
Who Are You?

- Undergrads? Ph.D students? Post Ph.D?

- Background:
  - Natural Language
  - Learning
  - Algorithms/Theory of Computation

- Survey
Expectations

- Interaction is important! Please, ask questions and make comments.
- Read; Present; Work on projects.
- Independence: (do, look for, read) more than surface level
- Rigor: advanced papers will require more Math than you know...
- Critical Thinking: don’t simply believe what’s written; criticize and offer better alternatives
Structure

- Lectures (60/75 min)
- You lecture (15/75 min) (once or twice)
- Critical discussions of papers (4)

Assignments:
- One small experimental assignment
- A final project, divided into two parts
  - Dependency parsing + Semantic Role Labeling
Next Time

- Examine some of the philosophical themes and leading ideas that motivate statistical approaches to linguistics and natural language and to

- Begin exploring what can be learned by looking at statistics of texts.