Question Answering and Information Extraction

CMSC 473/673
UMBC
December 11th, 2017
Course Announcement 1: Project

Due Wednesday 12/20, 11:59 AM

Late days cannot be used

Any questions?
Course Announcement 2: Final Exam

No mandatory final exam

December 20th, 1pm-3pm: optional second midterm/final

Averaged into first midterm score

No practice questions

Register by Monday 12/11:
https://goo.gl/forms/aXflKkP0BIrRxhOS83
Course Announcement 3: Evaluations

Please fill them out! (We do pay attention to them)

Links from StudentCourseEvaluations@umbc.edu
Recap from last time...
Pat and Chandler agreed on a plan.

He said Pat would try the same tactic again.

is “he” the same person as “Chandler?”
What are Named Entities?

Named entity recognition (NER)

Identify proper names in texts, and classification into a set of predefined categories of interest

Person names
Organizations (companies, government organisations, committees, etc)
Locations (cities, countries, rivers, etc)
Date and time expressions
Measures (percent, money, weight etc), email addresses, Web addresses, street addresses, etc.
Domain-specific: names of drugs, medical conditions, names of ships, bibliographic references etc.

Cunningham and Bontcheva (2003, RANLP Tutorial)
## Two kinds of NE approaches

<table>
<thead>
<tr>
<th>Knowledge Engineering</th>
<th>Learning Systems</th>
</tr>
</thead>
<tbody>
<tr>
<td>rule based</td>
<td>requires some (large?) amounts of annotated training data</td>
</tr>
<tr>
<td>developed by experienced language engineers</td>
<td>some changes may require re-annotation of the entire training corpus</td>
</tr>
<tr>
<td>make use of human intuition</td>
<td>annotators can be cheap</td>
</tr>
<tr>
<td>requires only small amount of training data</td>
<td></td>
</tr>
<tr>
<td>development could be very time consuming</td>
<td></td>
</tr>
<tr>
<td>some changes may be hard to accommodate</td>
<td></td>
</tr>
</tbody>
</table>

Cunningham and Bontcheva (2003, RANLP Tutorial)
Baseline: list lookup approach

System that recognises only entities stored in its lists (gazetteers).

Advantages - Simple, fast, language independent, easy to retarget (just create lists)

Disadvantages – impossible to enumerate all names, collection and maintenance of lists, cannot deal with name variants, cannot resolve ambiguity

Cunningham and Bontcheva (2003, RANLP Tutorial)
Shallow Parsing Approach
(internal structure)

Internal evidence – names often have internal structure. These components can be either stored or guessed, e.g. location:

- Cap. Word + {City, Forest, Center, River}
  Sherwood Forest

- Cap. Word + {Street, Boulevard, Avenue, Crescent, Road}
  Portobello Street

Cunningham and Bontcheva (2003, RANLP Tutorial)
NER and Shallow Parsing: Machine Learning

Sequence models (HMM, CRF) often effective

BIO encoding

Pat and Chandler Smith agreed on a plan.

B-NP O B-NP I-NP B-VP O B-NP I-NP
Pat and Chandler agreed on a plan.

He said Pat would try the same tactic again.
Model Attempt 1: Binary Classification

Pat and Chandler agreed on a plan.

He said Pat would try the same tactic again.

naïve approach (take all non-positive pairs): highly imbalanced!

possible problem: not transitive

for a mention $m$, select the closest preceding coreferent mention

otherwise, no antecedent is found for $m$

Soon et al. (2001): heuristic for more balanced selection
Anaphora

does a mention have an antecedent?

Chris told Pat he aced the test.
Pat and Chandler agreed on a plan. He said Pat would try the same tactic again.

**advantage**: featurize based on all (or some or none) of the *clustered mentions*

**disadvantage**: clustering doesn’t address anaphora
Pat and Chandler agreed on a plan.

He said Pat would try the same tactic again.

Model 3: Cluster-Ranking Model (Rahman and Ng, 2009)

learn to rank the clusters and items in them.
Stanford Coref (Lee et al., 2011)

- Mention Detection
  - Sieve 1: Speaker Identification
  - Sieve 2: String Match
  - Sieve 3: Relaxed String Match
  - Sieve 4: Precise Constructs
  - Sieve 5: Strict Head Match A
  - Sieve 6: Strict Head Match B
  - Sieve 7: Strict Head Match C
  - Sieve 8: Proper Head Noun Match
  - Sieve 9: Relaxed Head Match
  - Sieve 10: Pronoun Match

- More global decisions
- Recall increases

Post Processing
Applications

- Question answering
- Information extraction
- Machine translation
- Text summarization
- Information retrieval
Applications

Question answering
Information extraction
Machine translation
Text summarization
Information retrieval
IBM Watson

https://www.youtube.com/watch?v=C5Xnxjq63Zg

https://youtu.be/WFR3lOm_xhE?t=34s
What happened with Watson?

(let’s ask Google)
What Happened with Watson?

David Ferrucci, the manager of the Watson project at IBM Research, explained during a viewing of the show on Monday morning that several of things probably confused Watson. First, the category names on Jeopardy! are tricky. The answers often do not exactly fit the category. Watson, in his training phase, learned that categories only weakly suggest the kind of answer that is expected, and, therefore, the machine downgrades their significance. The way the language was parsed provided an advantage for the humans and a disadvantage for Watson, as well. “What US city” wasn’t in the question. If it had been, Watson would have given US cities much more weight as it searched for the answer. Adding to the confusion for Watson, there are cities named Toronto in the United States and the Toronto in Canada has an American League baseball team. It probably picked up those facts from the written material it has digested. Also, the machine didn’t find much evidence to connect either city’s airport to World War II. (Chicago was a very close second on Watson’s list of possible answers.) So this is just one of those situations that’s a snap for a reasonably knowledgeable human but a true brain teaser for the machine.

https://www.huffingtonpost.com/2011/02/15/watson-final-jeopardy_n_823795.html
How many children does the Queen have?
How many children does the Queen have?

Elizabeth II

Children: Charles, Prince of Wales, Anne, Princess Royal, Prince Andrew, Duke of York, Prince Edward, Earl of Wessex

Born: April 21, 1926 (age 91), Mayfair, London, United Kingdom

Spouse: Prince Philip, Duke of Edinburgh (m. 1947)

Grandchildren: Prince William, Duke of Cambridge, Prince Harry, MORE

People also ask

- How many grandchildren does the Queen have and what are their names?
- Who is the Queen's youngest son?
- How many times has the Queen been married?
- Who is next in line to the throne?
There are still errors (but some questions are harder than others)
Question Answering Motivation

- Question answering
- Information extraction
- Machine translation
- Text summarization
- Information retrieval
Question Answering Motivation

- Question answering
- Information extraction
- Machine translation
- Text summarization
- Information retrieval
Three people have been fatally shot, and five people, including a mayor, were seriously wounded as a result of a Shining Path attack today.

**ATTACK template**
- **Type**: Gun attack
- **Perp**: Shining Path
- **# killed**: 3

**BUSINESS NEGOTIATION template**
- ....
Remember Our Logical Forms?

\[ \exists e, x, y \text{ Eating}(e) \land \text{Agent}(e, x) \land \text{Theme}(e, y) \]

Papa ate the caviar
Two Types of QA

Closed domain

Often tied to structured database

Open domain

Often tied to unstructured data
Remember Our Logical Forms?

\[ \exists e, x, y \; Eating(e) \land Agent(e, x) \land Theme(e, y) \]

Papa ate the caviar

Corpus → KB
Open Domain:
START (1993-Present; Katz, 1997)

START, the world's first Web-based question answering system, has been on-line and continuously operating since December, 1993. It has been developed by Boris Katz and his associates of the InfoLab Group at the MIT Computer Science and Artificial Intelligence Laboratory. Unlike information retrieval systems (e.g., search engines), START aims to supply users with "just the right information," instead of merely providing a list of hits. Currently, the system can answer millions of English questions about places (e.g., cities, countries, lakes, coordinates, weather, maps, demographics, political and economic systems), movies (e.g., titles, actors, directors), people (e.g., birth dates, biographies), dictionary definitions, and much, much more. Below is a list of some of the things START knows about, with example questions. You can type your question above or select from the following examples. less...

SynTactic Analysis using Reversible Transformations

http://start.csail.mit.edu

“START, the world’s first Web-based question answering system, has been on-line and continuously operating since December, 1993. It has been developed by Boris Katz and his associates of the InfoLab Group at the MIT Computer Science and Artificial Intelligence Laboratory. Unlike information retrieval systems (e.g., search engines), START aims to supply users with "just the right information," instead of merely providing a list of hits. Currently, the system can answer millions of English questions about places (e.g., cities, countries, lakes, coordinates, weather, maps, demographics, political and economic systems), movies (e.g., titles, actors, directors), people (e.g., birth dates, biographies), dictionary definitions, and much, much more. Below is a list of some of the things START knows about, with example questions. You can type your question above or select from the following examples. less...

“START Server is built on two foundations the sentence level Natural Language processing capability provided by the START Natural Language system (Katz, 1990) and the idea of natural language annotations for multimedia information segments. This paper starts with an overview of sentence level processing in the START system and then explains how annotating information segments with collections of English sentences makes it possible to use the power of sentence level natural language processing in the service of multimedia information access. The paper ends with a proposal to annotate the World Wide Web.” (Katz, 1997)

SynTactic Analysis using Reversible Transformations
http://start.csail.mit.edu

“The START Server is built on two foundations the sentence level Natural Language processing capability provided by the START Natural Language system (Katz, 1990) and the idea of natural language annotations for multimedia information segments. This paper starts with an overview of sentence level processing in the START system and then explains how annotating information segments with collections of English sentences makes it possible to use the power of sentence level natural language processing in the service of multimedia information access. The paper ends with a proposal to annotate the World Wide Web.” (Katz, 1997)

Decompose sentences into (subject, verb, object) triples

“De-questionify” the input

How many children does the queen have →

The queen has how many children

Apply any needed inference rules

Query against knowledge base

SynTactic Analysis using Reversible Transformations
http://start.csail.mit.edu
how many children does the queen have

**Queen (Snow White)**

Children: Snow White (daughter in the original version, stepdaughter since the 1819 revision)

**Source:** Wikipedia

**Queen Fabiola of Belgium**

Children: none

**Source:** Wikipedia

**Evil Queen (Disney)**

Children: Snow White (stepdaughter)
Evie (daughter; in *Descendants* only)

**Source:** Wikipedia

**Queen of Hearts (Alice’s Adventures in Wonderland)**

Children: Ten Hearts

**Source:** Wikipedia
Basic System

Input Question → Question Analysis → Question Classification → Query Construction

Corpus → Document Retrieval → Sentence Retrieval → KB

Sentence NLP → Answer Extraction → Answer Validation → Answer

To Learn More:
NLP
Information Retrieval (IR)
Information Extraction (IE)
Aspects of NLP

POS tagging
Stemming
Shallow Parsing (chunking)
Predicate argument representation
  verb predicates and nominalization
Entity Annotation
  Stand alone NERs with a variable number of classes
Dates, times and numeric value normalization
Identification of semantic relations
  complex nominals, genitives, adjectival phrases, and adjectival clauses
Event identification
Semantic Parsing
Basic System

Input Question ➔ Question Analysis ➔ Question Classification ➔ Query Construction

Corpus ➔ Document Retrieval ➔ Sentence Retrieval ➔ KB

Sentence NLP ➔ Answer Extraction ➔ Answer Validation ➔ Answer

To Learn More:
NLP
Information Retrieval (IR)
Information Extraction (IE)
Question Classification

Albert Einstein was born in 14 March 1879.

Albert Einstein was born in Germany.

Albert Einstein was born in a Jewish family.
Question Classification Taxonomy

LOC: other Where do hyenas live?

NUM: date When was Ozzy Osbourne born?

LOC: other Where do the adventures of "The Swiss Family Robinson" take place?

LOC: other Where is Procter & Gamble based in the U.S.?

HUM: ind What barroom judge called himself The Law West of the Pecos?

HUM: gr What Polynesian people inhabit New Zealand?

SLP3: Figure 28.4

http://cogcomp.org/Data/QA/QC/train_1000.label
Basic System

Input Question → Question Analysis → Question Classification → Query Construction

Corpus → Document Retrieval

Sentence NLP → Answer Extraction

Sentence Retrieval → Answer Validation

KB → Answer

To Learn More:
NLP
Information Retrieval (IR)
Information Extraction (IE)

Neves --- https://hpi.de/fileadmin/user_upload/fachgebiete/plattner/teaching/NaturalLanguageProcessing/NLP09_QuestionAnswering.pdf
Document & Sentence Retrieval

**NLP Techniques:**
- Vector Space Model
- Probabilistic Model
- Language Model

**Software:**
- Lucene
- sklearn
- nltk

**tf-idf**
- $\text{tf-idf}(d, w)$
  - **Term Frequency, Inverse Document Frequency**
  - $\text{tf: frequency of word } w \text{ in document } d$
  - $\text{idf: inverse frequency of documents containing } w$

  $$\frac{\text{count}(w \in d)}{\# \text{ tokens in } d} \times \log\left(\frac{\# \text{ documents}}{\# \text{ documents containing } w}\right)$$
Current NLP QA Tasks

TREC (Text Retrieval Conference)
http://trec.nist.gov/
Started in 1992

Freebase Question Answering
e.g., https://nlp.stanford.edu/software/sempre/
Yao et al. (2014)

WikiQA
Visual Question Answering

http://www.visualqa.org/
Course Goals

• Be introduced to some of the core problems and solutions of NLP (big picture)
• Learn different ways that success and progress can be measured in NLP
• Relate to statistics, machine learning, and linguistics
• Implement NLP programs
• Read and analyze research papers
• Practice your (written) communication skills
Course Recap

Basics of Probability
- Requirements to be a distribution ("proportional to", $\propto$)
- Definitions of conditional probability, joint probability, and independence
  - Bayes rule, (probability) chain rule

Basics of language modeling
- Goal: model (be able to predict) and give a score to language (whole sequences of characters or words)
  - Simple count-based model
  - Smoothing (and why we need it): Laplace (add-$\lambda$), interpolation, backoff
  - Evaluation: perplexity

Tasks and Classification (use Bayes rule!)
- Posterior decoding vs. noisy channel model
- Evaluations: accuracy, precision, recall, and $F_\beta$ ($F_1$) scores
  - Naïve Bayes (given the label, generate/explain each feature independently) and connection to language modeling

Maximum Entropy Models
- Meanings of feature functions and weights
  - Use for language modeling or conditional classification ("posterior in one go")
  - How to learn the weights: gradient descent

Distributed Representations & Neural Language Models
- What embeddings are and what their motivation is
  - A common way to evaluate: cosine similarity

Latent Models
- What is meant by "latent"
- Expectation Maximization
  - Basic Example: Unigram Mixture Model (3 coins)

Machine Translation Alignment
- Family of methods for learning word-to-word translations
  - IBM Model 1
  - Can be used beyond MT (e.g., semantics, paraphrasing)

Hidden Markov Model
- Basic Definition: generative bigram model of latent tags
  - 3 Tasks: Likelihood, Most-Likely Sequence, Parameter Estimation
  - 3 Basic Algorithms: Forward (Backward), Viterbi, Baum-Welch
  - 2 Types of Decoding: Viterbi & Posterior

Semi-Supervised Learning
- Labeled data (small amount) + unlabeled data (large amount)
  - Apply EM to get fractional counts to re-estimate parameters

Word Modeling

Latent Sequences
Basics of Probability
Requirements to be a distribution ("proportional to", ⋆)
Definitions of conditional probability, joint probability, and independence
Bayes rule, (probability) chain rule

Basics of language modeling
Goal: model (be able to predict) and give a score to language (whole sequences of characters or words)
Simple count-based model
Smoothing (and why we need it): Laplace (add-λ), interpolation, backoff
Evaluation: perplexity

Tasks and Classification (use Bayes rule!)
Posterior decoding vs. noisy channel model
Evaluations: accuracy, precision, recall, and Fp (F1) scores
Naïve Bayes (given the label, generate/explain each feature independently) and connection to language modeling

Maximum Entropy Models
Meanings of feature functions and weights
Use for language modeling or conditional classification ("posterior in one go")
How to learn the weights: gradient descent

Distributed Representations & Neural Language Models
What embeddings are and what their motivation is
A common way to evaluate: cosine similarity

Latent Models
What is meant by “latent”
Expectation Maximization
Basic Example: Unigram Mixture Model (3 coins)

Machine Translation Alignment
Family of methods for learning word-to-word translations
IBM Model 1
Can be used beyond MT (e.g., semantics, paraphrasing)

Hidden Markov Model
Basic Definition: generative bigram model of latent tags
3 Tasks: Likelihood, Most-Like Sequence, Parameter Estimation
3 Basic Algorithms: Forward (Backward), Viterbi, Baum-Welch
2 Types of Decoding: Viterbi & Posterior

Semi-Supervised Learning
Labeled data (small amount) + unlabeled data (large amount)
Apply EM to get fractional counts to re-estimate parameters

Syntactic Parsing
Basic linguistic intuitions
Capturing of some ambiguities and light semantics

Constituency Parsing
Basic Definition: generative tree
3 Tasks: Likelihood, Most-Like Sequence, Parameter Estimation
3 Basic Algorithms: Inside, Viterbi, Outside
2 Types of Decoding: Viterbi & Posterior

Semi-Supervised Learning
Labeled data (small amount) + unlabeled data (large amount)
Apply EM to get fractional counts to re-estimate parameters

Dependency Parsing
Word-to-word relations
Shift-reduce parsing
Greedy vs. beam search

Semantic Forms
Roles, Frames, and Labeling
Ways to get human judgments (methodology and infrastructure)
Lexical & knowledge resources

Latent Structures
Deep Learning
Natural Language Processing

What society thinks I do
What my friends think I do
What other computer scientists think I do

What mathematicians think I do
What I think I do
What I actually do
Conditional vs. Sequence

Naive Bayes

SEQUENCE

HMMs

GENERAL GRAPHS

Generative directed models

Logistic Regression

SEQUENCE

Linear-chain CRFs

GENERAL GRAPHS

General CRFs

We’ll cover these in 678
Gradient Ascent

$$\arg \max_{\theta} \ F (\theta)$$
Pick Your Toolkit

PyTorch
Deeplearning4j
TensorFlow
DyNet
Caffe

Keras
MxNet
Gluon
CNTK
...

Comparisons:
https://deeplearning4j.org/compare-dl4j-tensorflow-pytorch
https://github.com/zer0n/deepframeworks (older---2015)
A major area of computational linguistics is that of "ambiguity resolution". It turns out that many things people say in a language – English, for example – can have more than one meaning!

Consider the phrase "fruit flies like a banana". Is it describing the taste of fruit flies, or rather flying fruit? How can a computer hope to figure this out?

Many have focused on statistical modelling of language, but this approach is approximate. I agree!

What do YOU know about computational linguistics? Ever read a little paper called "Non-Statistical Models for Unsupervised Prepositional Phrase Attachment"?

That was me! It was some of my earliest work on head word tuples!

Shit man, you know more about this than I do! You know what? You should be the one doing the talking here!
Thank you for a great semester!

ITE 358
ferraro@umbc.edu

Natural language processing

Semantics

Vision & language processing

Learning with low-to-no supervision