Natural Language Processing: Part II Overview of Natural Language Processing (L90): ACS

Simone Teufel (Materials by Ann Copestake)

Computer Laboratory
University of Cambridge

October 2018

Outline of today's lecture

Lecture 1: Introduction

Overview of the course
Why NLP is hard
Scope of NLP
A sample application: sentiment classification
More NLP applications
NLP subtasks

Part II / ACS / MLSALT

- ▶ Part II – Paper 10 UoA
  - ▶ 12 lectures, no supervisions
  - ▶ Assessed by Practical (with demonstrators) with 2 tasks: Sentiment Classification (extended from MLRD); Text Understanding QA
- ▶ ACS L90
  - ▶ Overview of NLP: other modules go into much greater depth: L90 intended for people with no substantial background in NLP.
  - ▶ Same 12 lectures as Part II, plus Practical Sentiment Classification Plus-Plus (with demonstrators).
- ▶ MLSALT, Engineering: Milica Gasic

Also note:

- ▶ Lecture notes in batches.
- ▶ No notes for lecture 12: topic changes every year.
- ▶ Slides: on web page (in advance where possible), but possible (slight) differences to slides used in lecture.
- ▶ Glossary in lecture notes.
- ▶ Webpage with links to demos etc.
- ▶ Recommended Book: Jurafsky and Martin (2008).
- ▶ Linguistics background: Bender (2013).

NLP and linguistics

NLP: the computational modelling of human language.

1. Morphology — the structure of words: lecture 2.

2. Syntax — the way words are used to form phrases: lectures 3, 4 and 5.

3. Semantics
   - Compositional semantics — the construction of meaning based on syntax: lecture 6.
   - Lexical semantics — the meaning of individual words: lecture 7.
   - Distributional semantics — representing words by their neighbourhoods: lectures 8 and 9.

4. Pragmatics — meaning in context: lecture 10.

5. Language generation — lecture 11.

6. Some current research — lecture 12.

Querying a knowledge base

User query:

- ▶ Has my order number 4291 been shipped yet?

Database:

ORDER

Order number	Date ordered	Date shipped
4290	2/2/13	2/2/13
4291	2/2/13	2/2/13
4292	2/2/13	

USER: Has my order number 4291 been shipped yet?
DB QUERY: order(number=4291,date_shipped=?)
RESPONSE: Order number 4291 was shipped on 2/2/13

Why is this difficult?

Similar strings mean different things, different strings mean the same thing:

1. How fast is the TZ?
2. How fast will my TZ arrive?
3. Please tell me when I can expect the TZ I ordered.

Ambiguity:

▶ Do you sell Sony laptops and disk drives?

▶ Do you sell (Sony (laptops and disk drives))?

▶ Do you sell ((Sony laptops) and disk drives)?

Wouldn't it be better if . . . ?

The properties which make natural language difficult to process are essential to human communication:

- ▶ Flexible
- ▶ Learnable but compact
- ▶ Emergent, evolving systems

Synonymy and ambiguity go along with these properties.
Natural language communication can be indefinitely precise:

- ▶ Ambiguity is mostly local (for humans)
- ▶ Semi-formal additions and conventions for different genres

xkcd 191

Some NLP applications

- ► spelling and grammar checking
- ► predictive text
- ► optical character recognition (OCR)
- ► screen readers
- ► augmentative and alternative communication
- ► machine aided translation

- ► lexicographers' tools
- ► information retrieval
- ► document classification
- ► document clustering
- ► information extraction
- ► sentiment classification
- ► text mining

More NLP applications . . .

- ▶ question answering
- ▶ summarization
- ▶ text segmentation
- ▶ exam marking
- ▶ language teaching

- ▶ report generation
- ▶ machine translation
- ▶ natural language interfaces to databases
- ▶ email understanding
- ▶ dialogue systems

Opinion mining: what do they think about me?

- ▶ Task: scan documents (webpages, tweets etc) for positive and negative opinions on people, products etc.
- ▶ Find all references to entity in some document collection: list as positive, negative (possibly with strength) or neutral.
- ▶ Construct summary report plus examples (text snippets).
- ▶ Fine-grained classification: e.g., for phone, opinions about: design, performance, battery life . . .

iPhone 8 review (Guardian 29/9/2017)

> *The iPhone 8 has Apple's latest and best processor.*
> *The six-core A11 Bionic has two high-performance*
> *cores and four power-efficient cores and is apparently*
> *the most powerful so far because it can use a*
> *combination of all six at once.*
> *Performance was excellent, but I struggled to see a*
> *real difference in day-to-day speed compared to the*
> *iPhone 7. But what I'm very pleased to be able to*
> *report is that Apple has finally improved battery life for*
> *the 4.7in iPhone.*
> *We're not talking a two-day battery here, but the*
> *iPhone 8 lasted just over 26 hours . . .*

Sentiment classification: the research task

- ► Full task: information retrieval, cleaning up text structure, named entity recognition, identification of relevant parts of text. Evaluation by humans.

- ► Research task: preclassified documents, topic known, opinion in text along with some straightforwardly extractable score.

- ► Movie review corpus (Pang et al 2002): strongly positive or negative reviews from IMDb, 50:50 split, with rating score.

IMDb: An American Werewolf in London (1981)

Rating: 9/10

Ooooo. Scary.
The old adage of the simplest ideas being the best is once again demonstrated in this, one of the most entertaining films of the early 80's, and almost certainly Jon Landis' best work to date. The script is light and witty, the visuals are great and the atmosphere is top class. Plus there are some great freeze-frame moments to enjoy again and again. Not forgetting, of course, the great transformation scene which still impresses to this day.
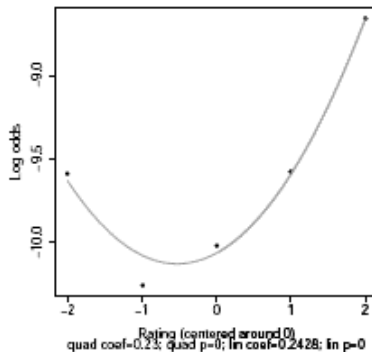*In Summary: Top banana*

Bag of words technique

- ▶ Treat the reviews as collections of individual words.
- ▶ Classify reviews according to positive or negative words.
- ▶ Could use word lists prepared by humans, but machine learning based on a portion of the corpus (training set) is preferable.
- ▶ Use human rankings for training and evaluation.
- ▶ Pang et al, 2002: Chance success is 50% (corpus artificially balanced), bag-of-words gives 80%.

Sentiment words

from Potts and Schwarz (2008)
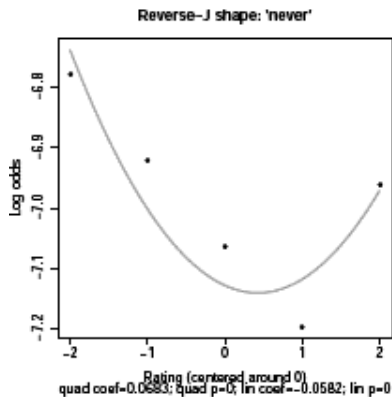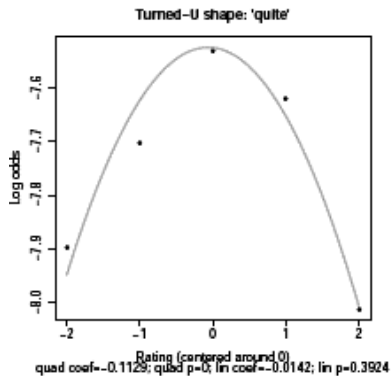
Sentiment words

never

from Potts and Schwarz (2008)

Sentiment words

quite

Turned-U shape: 'quite'
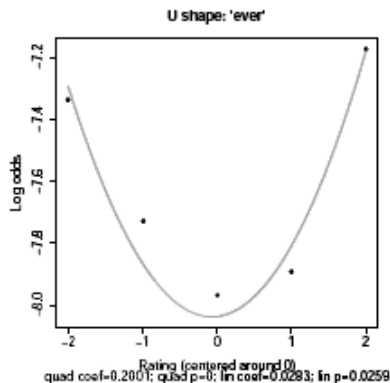
from Potts and Schwarz (2008)

Sentiment words: ever

ever

from Potts and Schwarz (2008)

Some sources of errors for bag-of-words

- Negation:

    Ridley Scott has never directed a bad film.

- Overfitting the training data:
  e.g., if training set includes a lot of films from before 2005,
  Ridley may be a strong positive indicator, but then we test
  on reviews for 'Kingdom of Heaven'?

- Comparisons and contrasts.

Contrasts in the discourse

> *This film should be brilliant. It sounds like a great plot, the actors are first grade, and the supporting cast is good as well, and Stallone is attempting to deliver a good performance. However, it can't hold up.*

More contrasts

> *AN AMERICAN WEREWOLF IN PARIS is a failed attempt . . . Julie Delpy is far too good for this movie. She imbues Serafine with spirit, spunk, and humanity. This isn't necessarily a good thing, since it prevents us from relaxing and enjoying AN AMERICAN WEREWOLF IN PARIS as a completely mindless, campy entertainment experience. Delpy's injection of class into an otherwise classless production raises the specter of what this film could have been with a better script and a better cast . . . She was radiant, charismatic, and effective . . .*

Doing sentiment classification 'properly'?

- ► Morphology, syntax and compositional semantics:
  who is talking about what, what terms are associated with
  what, tense . . .

- ► Lexical semantics:
  are words positive or negative in this context? Word
  senses (e.g., *spirit*)?

- ► Pragmatics and discourse structure:
  what is the topic of this section of text? Pronouns and
  definite references.

- ► Getting all this to work well on arbitrary text is very hard.

- ► Ultimately the problem is AI-complete, but can we do well
  enough for NLP to be useful?

IR, IE and QA

- ▶ Information retrieval: return documents in response to a user query (Internet Search is a special case)
- ▶ Information extraction: discover specific information from a set of documents (e.g. company joint ventures)
- ▶ Question answering: answer a specific user question by returning a section of a document:

  What is the capital of France?
  Paris has been the French capital for many centuries.

MT

- ▶ Earliest attempted NLP application.
- ▶ High quality only if the domain is restricted (or with very close languages: e.g., Swedish-Danish).
- ▶ Utility greatly increased in 1990s with increase in availability of electronic text.
- ▶ Spoken language translation is viable for limited domains.

Human translation?

I am not in the office at the moment. Please send any work to be translated.

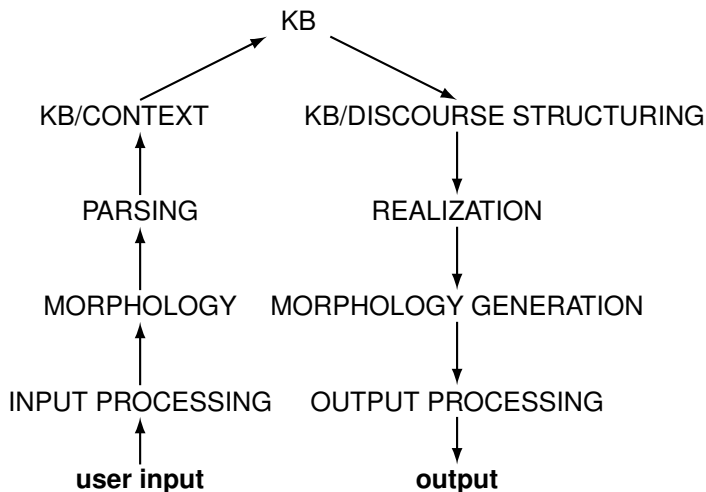# Natural language interfaces and dialogue systems

All rely on a limited domain:

- LUNAR: classic example of a natural language interface to a database (NLID): 1970–1975
- SHRDLU: (text-based) dialogue system: 1973
- Current spoken dialogue systems

Limited domain allows disambiguation: e.g., in LUNAR, *rock* had one sense.

NLP subtasks

- ▶ input preprocessing: speech recognizer, text preprocessor or gesture recognizer.
- ▶ morphological analysis (2)
- ▶ part of speech tagging (3)
- ▶ parsing: this includes syntax and compositional semantics (4, 5, 6)
- ▶ disambiguation, inference (6, 7, 8, 9)
- ▶ context processing (10)
- ▶ discourse structuring (11)
- ▶ realization (11)
- ▶ morphological generation (2)
- ▶ output processing: text-to-speech, text formatter, etc.

Subtasks in natural language interface to a knowledge base

General comments

- ▶ Even 'simple' applications might need complex knowledge sources.
- ▶ Applications cannot be 100% perfect.
- ▶ Applications that are $< 100\%$ perfect can be useful.
- ▶ Aids to humans are easier than replacements for humans.
- ▶ NLP interfaces compete with non-language approaches.
- ▶ Typically: shallow processing on arbitrary input or deep processing on narrow domains.
- ▶ Limited domain systems require expensive expertise to port or large amounts of (expensive) data.