Machine Learning in Natural Language Applications

Daniel Marcu
marcu@isi.edu

Lectures

Mar 27. Supervised Learning Methods in NLP: Bayes; Decision Trees.
April 3 Dealing with little or no data. Bootstrapping Methods in NLP.
April 5 Unsupervised Learning in NLP and Experiment Design.

NLP tasks used as underlying themes

- Word Sense Disambiguation.
- Tagging
  - Parts of Speech;
  - Named Entities (Names, Organizations, Dates, Locations, etc.)
  - Discourse units (clauses).

What this section of the course is about

- Learning frameworks/models/algorithms/tools that NLP researchers and engineers use.
- Learning methods that are not taught in the CS562 class (Knight and Marcu).
Word Sense Disambiguation

Motivation: words have multiple meanings

Words are ambiguous: (LDOCE, 3rd edition)

Equal spaces. Plant the seeds at regular intervals.

Every week/day, etc. He phones us every Sunday at six, regular as a clockwork.

Often. Old John is one of the bar’s regular customers.

Usual. Driving the van was a change from his regular duties.

Shape. A regular hexagon; regular features.

Normal. A regular coke.

Ordinary. Regular teachers just don’t have the training to deal with problem children.

Friendly. A regular guy.

Keep regular. Regular bowel movement.

Grammar. The verb “dance” is regular.

Emphasis. The child’s so bossy – a regular little dictator.

Permanency. A regular soldier.

However, dealing with ambiguity is not straightforward. Some of the potential problems are explained by:

1. dictionary inconsistencies (different dictionaries list different senses – see handout).
2. the fuzzyness/vagueness of meanings.

Criteria for classifying meanings:

1. Mental representations (slippery definition).
2. Clustering contexts.
3. Translation.
4. Dictionaries (vague definitions; true ambiguities).

Word sense disambiguation is important in Information Retrieval.
regular (regularity) [n] 1. [regular] a regular series of things has the same amount of time or space between each thing and the next. All breathing rate steps or equal and regular intervals (steady) not the beats between each time, every day, every hour.

2. EVERY DAY/EVERY WEEK = (usually before noun) happening or doing something many times and often at the same time every day, every week, etc.

3. Oft (usually before noun) happening or doing something very often, usually (customer) visitor of Old Joe's is one of be's regular customers.

4. USUAL (usually before noun) normal or usual: having the usual rate of change from one regular state.

5. SHAPE (event) shaped, matched, similar, or equal size: a regular triangle, a regular polygon, a regular shape with equal angles.

6. NORMAL (only before noun) especially AmE of a normal or standard size a regular cake.

7. ORDINARY (especially AmE) ordinary, without any special feature or qualities. Regular teachers just don't have the training to deal with problem children.

8. FRIENDLY (informal) friendly, very friendly, a regular guy.

9. KEEP REGULAR INFORMED: (a) to get rid of waste (b) not to waste too much time doing the same thing every day. REGULAR INFORMED: (b) not to waste too much time doing the same thing every day.

10. GIANT (a) to make large, huge, massive, or enormous. (b) very big, large, or great. (c) huge, enormous, or immense. (d) very big, large, or great.}

11. EMPHASIZING (only before noun) to make the impression you think someone is like. (a) a regular little dictator! (b) a regular soldier/army man.

12. REGULAR PROPHET/REGULARS: making your own words, nouns, or adjectives in the following way.

regular (regard) [v] 1. REGULAR PROPHETS/REGULARS: making your own words, nouns, or adjectives in the following way.

regular (regard) [regard] 1. REGULAR PROPHETS/REGULARS: making your own words, nouns, or adjectives in the following way.

regular (regard) [regard] 1. REGULAR PROPHETS/REGULARS: making your own words, nouns, or adjectives in the following way.
1 Tagging

- The/DT Canadian/JJ government/NN plans/VBZ to/TO auction/VB on/IN Tuesday/NNP 750/CD million/CD Canadian/JJ dollars/NN of/IN bonds/NNS.

- [PERSON: Hussein], grandson of [PERSON: Prophet Mohammad], killed at [LOCATION: Karbala] by [ORGANIZATION: army of [PERSON: Caliph Yazid]] in [DATE: 680].

Tagging is used pervasively in many NLP applications (Machine Translation; Question Answering; Parsing; etc.)