IT3105 Project III: Recognizing Textual Entailment

Part 4

Erwin Marsi
IT-VEST 313
emarsi@idi.ntnu.no

Logic and Language Technology Group, Intelligent Systems, IDI, NTNU

Plan for today's lecture

1. Introduction
2. Machine learning for RTE
3. Features
4. Improved evaluation
5. Epilog Part III
6. Break (15 minutes)
7. Other features
8. Further improvements
9. Epilog part IV

Other features

- Lexical semantic similarity
- Named entities
- Dependency relations
- Polarity

Lexical semantic similarity (1)

(1) T: The 26-member International Energy Agency said, Friday, that member countries would **release** oil to help relieve the U.S. fuel crisis caused by Hurricane Katrina.

$\Rightarrow$

H: Foreign oil reserves will be **made available** to the U.S. in the wake of Hurricane Katrina.

- *made available* means essentially the same as *release*
- *to make available* is a **synonym** for *release*

Lexical semantic similarity (2)

(2) T: Moog's **synthesiser**, which bears his name, revolutionised music from the 1960s onwards, and was used by bands like The Beatles and The Doors.

$\Rightarrow$

H Moog's **instruments** were used by The Beatles and The Doors among others.

- *synthesizer* is a kind of *instrument*
- *synthesizer* is a **hyponym** of *instrument*
- *instrument* is a **hyperonym** of *synthesizer*

WordNet (1)

- **lexical semantics**: field of linguistics that studies the meaning of words, and relations between words as far as their meaning is concerned
- **WordNet**: a large lexical database of English nouns, verbs, adjectives and adverbs
- all words are grouped into sets of synonyms, or so-called **synsets**
- Each synset expresses a distinct **concept**
- Synsets are interlinked with conceptual-semantic and lexical relations.
- WordNet is freely and publicly available
- Frequently used in NLP, also RTE

WordNet (2)

- Example: two different senses of noun *synthesizer* (WordNet 3.1)
  1. {synthesist, synthesizer, synthesiser} (an intellectual who synthesizes or uses synthetic methods)
  2. {synthesizer, synthesiser} ((music) an electronic instrument (usually played with a keyboard) that generates and modifies sounds electronically and can imitate a variety of other musical instruments)

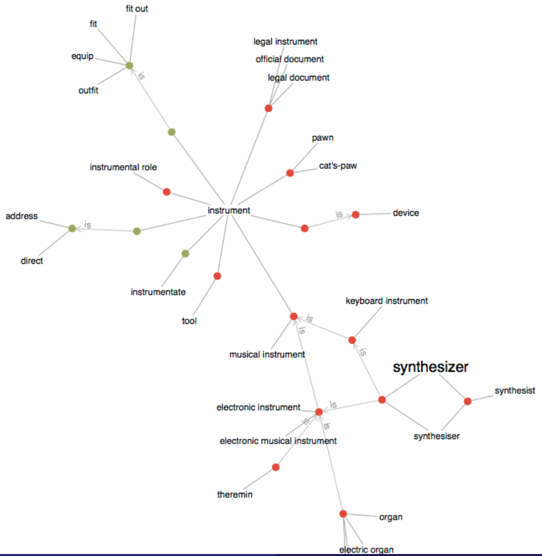
WordNet (3)

- Direct hyperonyms of the second sense of *instrument* are synsets {`keyboard instrument`} and {`electronic instrument, electronic musical instrument`}
- Both share the synset {**musical instrument**} as their hyperonym
- which in turn has *instrument* as its hyperonym.
- Therefore *instrument* is an indirect hyperonym of *synthesizer*.

WordNet (4)

- Wordnet is a large graph with synsets as nodes and lexical semantic relations as labeled edged
- some interfaces let you browse through the WordNet graph

WordNet (5)


fit out
fit
equip
outfit
is
legal instrument
official document
legal document
pawn
cat's-paw
instrumental role
instrument
address
is
direct
is
device
instrumentate
tool
keyboard instrument
is
is
musical instrument
is
synthesizer
synthesist
electronic instrument
is
is
is
electronic musical instrument
synthesiser
theremin
organ
electric organ

Semantic similarity (1)

- We can learn from WordNet that *synthesizer* and *instrument* are semantically related by following hyperonym relations
- However, by the same method we can establish that *dinosaur* and *chewing gum* are related!
- Because ultimately we end up at a very abstract concept (*entity*) that covers almost every noun
- Requires some measure of graph distance

Semantic similarity (2)

- Path length works to certain extent
- Problem: some parts of WordNet are more dense than others
  - *musical instrument* and *instrument* are "conceptually close"
  - but *security blanket* and *thing* are conceptually distant
- Quite a lot of alternative measures for semantic similarity have been proposed
- In practice, the Jiang & Conrath and the Lin measures appear good choices

Word senses (1)

- Same word can have multiple senses
- *synthesizer* can be musical instrument or someone synthesizing ideas
- The latter sense is a hyperonym of *intelectual*, which is in itself a hyperonym of *person*
- However, *synthesizer* and the *Moog* are clearly not referring to the same person!
- Problem: how do we identify the correct sense of a word?

Word senses (2)

- **Word Sense Disambiguation** (WSD): tasks of automatically determining the correct sense of word, provided a predefined set of senses
- Hard problem in NLP
- No sense tags in preprocessed RTE data
- Solutions:
  1. Apply off-the-shelve WSD software
  2. Compute similarity for all senses of words involved, and take maximum score

Lexical Semantic Similarity measure (1)

$$LSS(H, T) = \frac{\sum_{w_i \in H} \arg\max_{w_j \in T} Sim(w_i, w_j)}{\sum_{w_i \in H} Sim(w_i, w_i)}$$

- function *Sim* returns the lexical similarity between two words as a real number between zero and one
- by definition $Sim(w, w) = 1$
- *argmax* operator selects the w_j for which $Sim(w_i, w_j)$ is maximal

Lexical Semantic Similarity measure (2)

- Alternative: instead of words, use word + POS, or even word + POS + sense number
- Or weight by IDF weight of H words
- Smart feature engineering has a big impact on performance

Named entities

- **Named-Entity Recognition** is the process of tagging elements in text into predefined categories such as the names of persons, organizations, locations, expressions of times, quantities, monetary values, percentages, etc

(3) T: The 26-member [$_{ORG}$ International Energy Agency] said, Friday, that member countries would release oil to help relieve the [$_{LOC}$ U.S.] fuel crisis caused by [$_{MISC}$ Hurricane Katrina] .

$\Rightarrow$

H: Foreign oil reserves will be made available to the [$_{LOC}$ U.S.] in the wake of [$_{MISC}$ Hurricane Katrina] .

Named entities (2)

- Assumption: all named entities in H must also occur in T, otherwise entailment is unlikely
- There are obviously counter examples, but works out reasonably well as a feature
- Requires some normalized measure of shared named entities

Dependency relations (1)

- Tree edit distance for RTE has some drawbacks:
  - expensive for long sentences
  - hard to tune edit costs
- Fortunately there are other ways to include syntactic information
- Dependency analysis of sentence consists of set of $< \mathit{head}, \mathit{relation}, \mathit{dependent} >$ triples
- Therefore we can approximate the similarity between the dependency trees for T and H by the cardinality of the intersection of the sets of dependency triples for T and H respectively
- Again, normalization is required and weighting may be added

Dependency relations (2)

- Recall hat BLEU score is essentially average precision over word n-grams
- Same idea may be applied to dependency triples so that
  - single triple is analogous to a word bigram
  - two triples matching on either head or dependent (think: domino) are analogous to a word trigram
  - three matching triples to a 4-gram
- Amounts to matching paths through the dependency trees of H and T

Polarity (1)

(4) T Drew Walker, NHS Tayside's public health director, said: "It is important to stress that this is not a confirmed case of rabies."
H A case of rabies was confirmed.

- Tricky: overlap between T and H is nearly perfect, yet no entailment
- Reason: presence of *not* in T
- **polarity** difference: whereas T is a *negative* statement, H is a *positive* statement

Polarity

- Define a set of *Negatives* that toggle polarity: $Negatives = \{not, refuse, wrong, deny, no, false, ignore, cannot, never, unsuccessfully\}$
- Define a simple polarity feature POL, where n_t is the number of words in both Negatives and T and n_h is the number of words in both Negatives and H

$$POL(T, H) = \begin{cases} 0 & \text{if } n_t \text{ and } n_h \text{ have the same parity} \\ 1 & \text{otherwise} \end{cases}$$

Further improvements

- Text normalization
- Co-reference resolution
- More training data
- RTE per task

Text normalization

- Problem: "*200 dollar*" does not match "*$200*" or "*two hundred dollars*"
- **Text normalization** is transforming all variation in expressions of time, date, currency, and so on, to a standard form
- Expected to facilitate matching such expressions between H and T

Co-reference resolution

(5) T: Britain said, Friday, that it has barred cleric, Omar Bakri, from returning to the country from Lebanon, where **he** was released by police after being detained for 24 hours

H: Bakri was briefly detained, but was released.

- Problem: the anaphoric expression *he* and to its antecedent *Omar Bakri* are one and the same person
- **Co-reference Resolution**: automatically relating anaphoric expression to their antecedent
- In practice, the accuracy of Co-reference Resolution systems is rather low
- Application may require other knowledge, e.g. deriving the triple $< OmarBikiri, obj, detain >$ from the triple $< he, obj, detain >$ and the identity $he = OmarBikiri$

More training data

- "There's no data like more data" is a famous quote attributed to Fred Jelinek
- More training data from other RTE challenges is available on the web
- Please stay away from *test* data
- Notice: differences between data sets due to changes to the task and the annotation procedure over the years
- Also *preprocessed* versions may not be available

RTE per task

- Pairs in the RTE data stem from four (related) NLP task: Information Extraction (IE), Information Retrieval (IR), Question Answering (QA) and Automatic Summarization (SUM)
- `pair` element in the XML file has `task` attribute
- Task may be included as extra feature
- Can even use RTE system tuned for the task

How to go on from here

- Reread project description Part IV
- Read fourth part of lecture notes
- Choose your baseline RTE system
- Choose a strategy for improving your RTE system
- Obtain required resources (see lecture notes)
- Implement improvements
- Evaluate and optimize as long as time is available
- Download test data when released, run your system and submit output
- Write final report
- Prepare demonstration