Modelling Semantic Plausibility in Human Sentence Processing

IGK Meeting
13.7.2005
Ulrike Padó

An Example

The hunter shot by the teenager was only 30 years old.

... is difficult for most people when reading „by the teenager"

An Example II

The deer shot by the hunter was only used as a trophy.

... is easy to read for most people.

- Structure is identical for both sentences
- Initial interpretation (up to the by-phrase) is a main clause both times

- But: Plausibility of that interpretation differs!
  ⇨Implausible main clause is easier to abandon

Goal

- Model this effect (and others) that depend on *semantic plausibility* (thematic fit)
- Model plausibility effects *probabilistically*, using (semantically annotated) corpora

Overview

- Modelling Human Sentence Processing
- Our Model of (Shallow) Semantics
  - Overcoming Data Sparseness
  - Comparison to Selectional Preference Methods
  - Comparison to a Role Labeller
- Dealing with Multiple Arguments
- Conclusions

Properties of Human Sentence Processing

- *Incrementality*: Interpret incomplete input
- *Shaped by experience*: Frequent words/structures are processed more quickly and easily.
- *Modularity*: ERP results indicate distinct loci and time courses for syntactic and semantic processing

⇨ Should be reflected in a model

PCFG-Based Models

- A structure‘s goodness is its probability given by the probabilistic grammar
  (Jurafsky, 1996; Crocker & Brants, 2000)
- Models predict difficulty if best (most likely) structure changes
  - Predict difficulty at by-PP in “NP V by-PP …” because best structure changes from main clause to reduced relative interpretation

PCFG-Based Models II

- Models are
  - Incremental ✓
  - Experience-based ✓
  (depending on grammar)
- But: Only syntactic!
  - „One module‟
  - Make the same predictions for both example sentences

Adding semantics

- Rate syntactic structure by its plausibility (based on its verb-argument-role triples)
- Compute ratings: Probabilistically assign *thematic roles* to verb-argument pairs
  - [The deer Ag] shot ... MC ✗
  - [The deer Pat] shot ... RR ✓
- Then create a modular architecture:
  - Base final ranking of structures on predictions from both syntax and semantics

Overview

- Modelling Human Sentence Processing
- Our Model of (Shallow) Semantics
  - Overcoming Data Sparseness
  - Comparison to Selectional Preference Methods
  - Comparison to a Role Labeller
- Dealing with Multiple Arguments
- Conclusions

A Model of (Shallow) Semantics

- Use thematic roles to link to semantics of verb-argument relations
- Estimate plausibility of the verb-argument relation as its probability:

$$Plausibility_{v,r,a} = P(r,a,v,f,gf) = P(v) * P(f|v) * P(gf|v,f) * P(r|v,f,gf) * P(a|v,f,gf,r)$$

Evaluation: Tasks

- Predict human plausibility ratings for verb-role-argument triples
  - Correlate predictions and ratings
  - Reporting coverage and correlation strength/significance
- Predict the correct role
  - Correct: Role with the highest human rating
  - Predicted: Most probable role
  - Reporting coverage and F score

Test and Training Data

- Training Data: FrameNet corpus

  Killing: [The deer *Victim*] shot [by the hunter *Killer*] ...

  Verbs introduce frames (situation descriptions), which define a set of possible participants.

- Test Data: Human plausibility ratings McRae et al. (1998), Trueswell et al. (1994)

shoot	deer	agent	1.0
shoot	deer	patient	6.4
shoot	hunter	agent	6.9
shoot	hunter	patient	2.8

Overcoming Data Sparseness

- Without smoothing, we can predict only 6% of test items!
- Combine two complementary smoothing methods:
  - Good-Turing/Linear Interpolation: Assign probabilities to unseen counts
  - Class-based Smoothing of P(a|r,v,f,gf): Use verb clusters from training data, WordNet noun synsets

Smoothing Results

Smoothing	Coverage	Correlation (Spearman's ρ)
No Smoothing	6%	–
Class-based	19%	0.494, *
GT/LI	90.6%	ns
Class-based + GT/LI	90.6%	0.302, *

How does this compare to… Selectional Preference Models?

- Selectional preference models estimate the goodness of an argument in a grammatical relation to a verb
- Do they predict human data?
- Compare against standard models:
  - Resnik 1993
  - Clark & Weir 2001
  - Li & Abe 1998
- Different approaches to class-based smoothing using WordNet

Comparison to Selectional Preference Models

Test Set	Model	Coverage	ρ
McRae	Module	90.6%	0.302, *
	Resnik	93.5%	ns
	Clark&Weir	70.3%	ns
	Li&Abe	90.6%	ns
Trueswell	Module	80.8%	0.422, **
	Resnik	69.2%	0.440, **
	Clark&Weir	61.5%	ns
	Li&Abe	71.2%	ns

How does this compare to... a standard Role Labeller?

- Giuglea&Moschitti (2004): Role labeller for FrameNet roles.
  - Use only standard features to build a vanilla labeller
  - F=80.5 on FN test data (gold boundaries)
  - Use re-normalised confidence values for prediction
- Both tasks: Correlation and Labelling

Comparison to standard Role Labeller

Model	Correlation		Labelling	
	Coverage	ρ	Coverage	F score
Baseline	–	–	100%	37.5
Module	90.6%	0.271, *	100%	59.4
Labeller	100%	ns	100%	43.8

Labeller performs poorly: It doesn't pick up on the semantic cues in the data.

Overview

- Modelling Human Sentence Processing
- Our Model of (Shallow) Semantics
  - Overcoming Data Sparseness
  - Comparison to Selectional Preference Methods
  - Comparison to a Role Labeller
- Dealing with Multiple Arguments
- Conclusions

Multiple Arguments

- We need to be able to process multiple arguments, each with its own array of role predictions
- Optimal role assignment may change:

He packed the bag into the van.

Treating Multiple Arguments

- Find best overall role combination: Viterbi-style
- Only seen role combinations are allowed
  - Ensure that all roles exist in the frame
  - Ensure that role combination makes sense
- Model predictions:
  - He packed [the bag *Goal*]
  - He packed [the bag *Theme*] [into the van *Goal*]

Conclusions

- A corpus-based way of modelling (shallow) semantics
- Good performance in comparison to selectional preference approaches
- Good performance in comparison to role labeller
- Ability to process multiple arguments per verb and output optimal role set

Future Work

- Combine semantic module with a syntactic module (incremental probabilistic parser)
- Define linking of model predictions to observations (reading times)
- Model reading time data