NLP Exam Questions   2000

1)

The following context-free grammar and lexicon generates the examples below it with multiple derivations and therefore multiple associated interpretations.
Describe how a probabilistic version of the context-free grammar (PCFG) can be created, defining the constraints which must hold for the resulting PCFG to be interpretable as a stochastic language model. [8 marks]
How accurate would the resulting PCFG be at assigning the semantically appropriate derivations highest probability for the examples and other structurally similar sentences? [6 marks]
Define an improved probabilistic model for discriminating alternative derivations. What problems would arise in the implementation of this model? [6 marks]

a) S → NP VP
b) NP → Det N
c) NP → NP PP
d) N → N N
e) VP → V NP
f) VP → V NP PP
g) VP → VP PP
h) PP → P NP

N → car | park | tree | boy | toy | morning ...
V → hit | ...
P → in | with ...
Det → a | the ...

a car hit the tree in the park
the boy hit the toy car park with a toy car in the morning

Solution Notes

A) PCFG basic constraint, alternative expansions of distinct LHS symbols sum to one – leads to a lg model in which P(S) and P(D(S)) is defined, see s8 lecture notes for details. Estimating the probs. by max likelihood requires an annotated corpus of sents+derivs to be useful. Other techniques likely to be less effective. Smoothing might be needed if some rules aren't exemplified in training corpus.

B) Depending on the estimates for rules g) vs e) or f) and rules g) vs c) the PCFG will prefer argument PPs over adjuncts and adjectival PP adjuncts over adverbial PP adjuncts. This will determine the derivations assigned to the examples, but structurally identical sentences with different words may have the opposite interpretation. toy car park is sructurally ambiguous but the model

cannot distinguish the two derivations probabilistically.

C) A probabilistic model which took account of the words would do better, but this would greatly increase the number of parameters to be estimated and therefore require a good smoothing scheme to avoid overfitting the training data. The lecture notes outline a couple of possible models and briefly describe why Add1 smoothing isn't optimal.

2)

The following narrative contains several types of anaphoric relation between pronouns or definite noun phrases and discourse antecedents. Identify them. [4 marks]
Describe how they might be recovered using either statistical or knowledge-based techniques. [6 marks]
What are the relative advantages and disadvantages of each approach? [10 marks]

The house was beautiful Kim and Sandy especially liked the windows. They wanted to buy it, so they rang the estate agent.

Solution Notes

A) distinction between anaphoric definite descriptions and those than link to 'understood' discourse referents arising from the house buying scenario (estate agent); 'bridging' inferences from lexical content (house, windows); anaphoric pronouns, chains of anaphoric links

B) lecture covered discourse structure based approach (Sidner's theory) and a statistical, salience weights approach (Kennedy and Boguraev), outline of either or related approaches with link to example narrative showing what would happen with each anaphor (see s9 in lecture notes).

C) House-windows requires a kb approach; the bridging inference also supports the inference that the second clause is an elaboration of the first (hence dash) so discourse subordinate. K and S-They, K and S is highest agreeing antecedent in discourse ie. there is no possible antecedent in first clause – either approach should get this, as windows should be ruled out / less prefered in dir. obj. slot. house-it agrees uniquely and house is still overall topic, but the kb approach is more likely to have access to this info. K and S-They-they subj-subj link via intermediate They – both approaches should get this.