CS224/Ling 237 Homework #4 Solutions

Due Wed, 29 May 2002, 5pm

(20 points total)

1. (2 points) Information Extraction. Suppose that one wanted to learn acronyms as extracted facts from documents. So, the learned relation would have two fields, acronym and expansion, as in:

U.S.A.	United States of America
IR	Information Retrieval

Present two (significantly different) extraction rules that would be useful for learning acronyms (if you're having trouble, look at a few research papers, and see how and where acronyms are introduced). These extraction rules could either be something that would be handcoded (perhaps as a regular expression), or they could be instances of a general extraction pattern that you would hope to learn using a machine learning method. For the purposes of this problem, it is sufficient just to state them as examples. You will probably find it helpful to look at some of the examples of extraction rules in the Muslea paper 'Extraction Patterns for Information Extraction Tasks: A Survey', and you might want to borrow the notations that one or more of those systems use in presenting your answer – though do tell us which system's notation you are using.

Solutions:
There were a number of different acceptable answers for this question. Most people used hand-coded rules rather than learning algorithms.

2. (6 points) Consider a simplified version of an ambiguous sentence example that we used at the beginning of the quarter:

Fed raises interest rates greatly

A possible grammar that will be able to parse this sentence is:

S	$\rightarrow$	NP VP	NP	$\rightarrow$	*Fed*
VP	$\rightarrow$	V NP	NP	$\rightarrow$	*raises*
VP	$\rightarrow$	VP Adv	NP	$\rightarrow$	*interest*
VP	$\rightarrow$	V	NP	$\rightarrow$	*rates*
NP	$\rightarrow$	NP NP	Adv	$\rightarrow$	*greatly*
V	$\rightarrow$	*raises*	V	$\rightarrow$	*interest*
V	$\rightarrow$	*rates*			

We would now like to parse this sentence

(a) The grammar as shown isn't quite in Chomsky normal form. Turn it into a CNF grammar.
Solutions:
The only "problem" item is VP → V. We can eliminate this rule and add a new rule for S and VP, as follows:

S → NP V
VP → V Adv

Alternatively, we can add new rules for the verbs:
VP → raises
VP → rates
VP → interest

Note that we cannot *replace* the verb rules with the rules above, as this will allow verb phrases with the structure VP Adv NP, which were not permitted by the original rules.

(b) Make a well-formed substring table (parse triangle) for this sentence, and then parse it using the CKY algorithm, filling in all the cells in the triangle with the correct constituents (it isn't necessary to include further annotation as to how they were built up).
Solutions:

	NP_1	S_9 NP_{10}	S_{17} NP_{18}	S_{24} NP_{25}	S_{28}
Fed		V_2 NP_3	S_{11} NP_{12} VP_{12}	S_{19} NP_{20} VP_{21}	S_{26} VP_{27}
	raises		V_4 NP_5	S_{13} NP_{14} VP_{15}	S_{22} VP_{23}
		interest		V_6 NP_7	VP_{16}
			rates		Adv_8
				greatly	

(c) The above grammar does not capture that some of the nouns are plural and some singular, and that some of the verbs are forms (like *interest*) which would be used with plural subjects, and others like *rates* would be used with a singular subject. Furthermore, the grammar also doesn't capture that all three of the verbs shown are normally only used transitively (with a required NP object after the verb). Change the grammar so it captures all this information (but is still a CNF grammar).
Solutions:
We need to introduce separate rules for plural NPs and singular NPs, and for plural VPs

and singular VPs. To capture the fact that these verbs are used transitively, we will not allow them to be full VPs, but will eliminate the rules that we just added allowing S to be rewritten as NP V and VP to be rewritten as V Adv. If there were non-transitive verbs in our grammar, we would have to keep these rules, but add extra annotation to the verbs to distinguish between them.

S	→	NPsing VPsing	NPsing	→	*Fed*
S	→	NPplur VPplur	NPplur	→	*rates*
VPsing	→	Vsing NPsing	NPplur	→	*raises*
VPsing	→	Vsing NPplur	NPsing	→	*interest*
VPplur	→	Vplur NPsing	Vplur	→	*interest*
VPplur	→	Vplur NPplur	Vsing	→	*raises*
VPsing	→	VPsing Adv	Vsing	→	*rates*
VPplur	→	VPplur Adv	Adv	→	*greatly*
NPsing	→	NPsing NPsing			
NPsing	→	NPplur NPsing			
NPplur	→	NPsing NPplur			
NPplur	→	NPplur NPplur			

(d) Parse the sentence again using the CKY algorithm and the new grammar, and again show the completed parse triangle (without further annotation being necessary).

Solutions:

	NPsing$_1$	NPplur$_9$	S$_{14}$ NPsing$_{15}$	S$_{20}$ NPplur$_{21}$	S$_{24}$
Fed		Vsing$_2$ NPplur$_3$	VPsing$_{10}$ NPsing$_{11}$	S$_{16}$ VPsing$_{17}$ NPplur$_{18}$	S$_{22}$ VPsing$_{23}$
	raises		Vplur$_4$ NPsing$_5$	VPplur$_{12}$ NPplur$_{13}$	VPplur$_{19}$
		interest		Vsing$_6$ NPplur$_7$	
			rates		Adv$_8$
				greatly	

3. (4 points) Recall the phrase structure grammar we used in class:

S	→	NP VP	N	→	*cats*
VP	→	V NP	N	→	*claws*
VP	→	V NP PP	N	→	*people*
NP	→	NP PP	N	→	*scratch*
NP	→	N	V	→	*scratch*
NP	→	e	P	→	*with*
PP	→	P NP			

This grammar includes all the nasties: a left-recursive rule, a unary rule, a more-than-binary rule, and an empty category. We will explore using it in a chart parser to parse sentences. We would suggest drawing the chart in the style of a parse triangle, except this time numbering starting from 0 instead of 1 and treating the numbers as fenceposts. That is, for the sentence *I like brie*, the cell (0,2) will hold chart edges for *I like*, and the cell (2,3) will hold chart edges for *brie*. The parse triangle cell entries will include both passive edges (complete constituents) and active edges (incomplete dotted rules).

(a) Consider parsing this grammar with a bottom up chart parser on the sentence *Cats with claws scratch*. Show the chart that results.

Solutions:

We added the epsilon rules at the "fenceposts", treating empty categories like words by adding them during the initialization.

NP → e · NP → · N S → · NP VP NP → · NP PP S → NP · VP NP → NP · PP				
N → cats · NP → N · S → NP · VP NP → NP · PP	NP → e · PP → · P NP S → · NP VP NP → · NP PP S → NP · VP NP → NP · PP			
NP → NP PP · S → NP · VP NP → NP · PP	P → with · PP → P · NP PP → P NP · NP → NP PP · S → NP · VP NP → NP · PP	NP → e · NP → · N NP → · NP PP S → · NP VP NP → NP · PP S → NP · VP		
NP → NP PP · S → NP · VP NP → NP · PP	PP → P NP · NP → NP PP · S → NP · VP NP → NP · PP	N → claws · NP → N · NP → NP · PP S → NP · VP	NP → e · NP → · N VP → · V NP VP → · V NP PP S → NP VP NP → · NP PP S → NP · VP NP → NP · PP	
S → NP VP ·	S → NP VP ·	S → NP VP ·	N → scratch · V → scratch · NP → N · VP → V · NP VP → V · NP PP VP → V NP · VP → V NP · PP S → NP · VP NP → NP · PP S → NP VP ·	NP → e · S → · NP VP NP → · NP PP S → NP · VP NP → NP · VP

(b) Repeat this exercise parsing top down. Show the chart that results. What differences are there in the work that was done?

Solutions:

Again, we treated empty categories like words. However, you could also have only added them when they were predicted.

NP → e · S → · NP VP NP → · NP PP NP → · N S → NP · VP NP → NP · PP VP → · V NP VP → · V NP PP PP → · P NP				
N → cats · NP → N · S → NP · VP NP → NP · PP	NP → e · PP → · P NP VP → · V NP VP → · V NP PP			
NP → NP PP · S → NP · VP NP → NP · PP	P → with · PP → P · NP PP → P NP ·	NP → e · NP → · N NP → · NP PP NP → NP · PP PP → · P NP VP → · V NP VP → · V NP PP		
S → NP · VP NP → NP PP · NP → NP · PP ·	PP → P NP ·	N → claws · NP → N · NP → NP · PP	NP → e · VP → · V NP VP → · V NP PP PP → · P NP	
S → NP VP ·			N → scratch · V → scratch · VP → V · NP VP → V · NP PP VP → V NP · VP → V NP · PP	NP → e · NP → · NP PP NP → · N PP → · P NP NP → NP · PP

4. (3 points) In chapter 11 of M&S, we suggested that PCFGs have a bad bias towards using nonterminals with few expansions. Suppose that one has as a training corpus the treebank given below, where '$n\times$' indicates how many times a certain tree appears in the training corpus. What PCFG would one get from the treebank (using MLE as discussed in the text)? What is the most likely parse of the string 'a a' using that grammar. Is this a reasonable result? Was the problem of bias stated correctly in chapter 11? Discuss.

$$\left\{ 10 \times [_{\mathrm{S}}\ [_{\mathrm{B}}\ a]\ [_{\mathrm{B}}\ a]],\quad 95 \times [_{\mathrm{S}}\ [_{\mathrm{A}}\ a]\ [_{\mathrm{A}}\ a]],\quad 325 \times [_{\mathrm{S}}\ [_{\mathrm{A}}\ f]\ [_{\mathrm{A}}\ g]],\quad 8 \times [_{\mathrm{S}}\ [_{\mathrm{A}}\ f]\ [_{\mathrm{A}}\ a]],\quad 428 \times [_{\mathrm{S}}\ [_{\mathrm{A}}\ g]\ [_{\mathrm{A}}\ f]] \right\}$$

Solutions::

We have 10 applications of the rule S $\rightarrow$ B B and 856 applications of the rule S $\rightarrow$ A A.
We have 2 applications of the rule B $\rightarrow$ a.
We have 198 applications of the rule A $\rightarrow$ a.
We have 761 applications of the rule A $\rightarrow$ f.
We have 753 applications of the rule A $\rightarrow$ g.

Therefore, our PCFG is as follows:

S $\rightarrow$ A A (0.988)
S $\rightarrow$ B B (0.012)
B $\rightarrow$ a (1.000)
A $\rightarrow$ a (0.116)
A $\rightarrow$ f (0.445)
A $\rightarrow$ g (0.440)

There are two possible parses of "a a", the one that begins with S $\rightarrow$ A A and the one that begins with S $\rightarrow$ B B. The first one has probability (0.988)*(0.116)*(0.116) = 0.013, and the second has a probability (0.012)*(1.000)*(1.000) = 0.012.

Therefore, the most likely parse begins S $\rightarrow$ A A.

The problem of bias discussed in chatper 11 says that non-terminals with fewer expansions are preferred over non-terminals with more expansions. In this case, the non-terminal with more expansions (A) is chosen over the non-terminal with fewer expansions (B); however, the probabilities are much closer than may be guessed from the number of times "a a" is found as an expansion of S $\rightarrow$ A A compared with the number of times "a a" is found as an expansion of S $\rightarrow$ B B. Therefore, we do see the bias here, but the much greater probability of S being rewritten as AA dominates.

5. (2 points) Syntax and Logical Semantics I. As well as *the overpriced car runs*, we can also say *the overpriced car does run.* Starting from the grammar fragments shown in the handout, We can introduce *does* as a word of category Aux and add an additional rule VP $\rightarrow$ Aux VP. Determine a semantic representation for *does*, semantic annotations for the new syntax rule, and show a parse tree complete with semantics showing this rule used to give a correct semantic representation for *Kathy does respect Fong.*

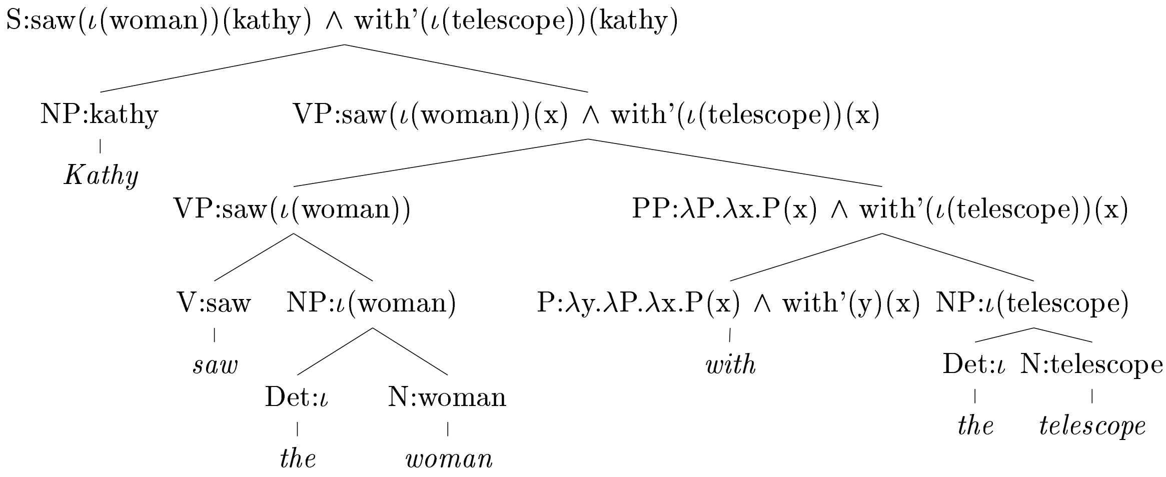
   Solutions:
   We will add the following rule:
   VP: $\alpha(\beta)$ $\rightarrow$ Aux: α VP: β
   The lexical entry for "does" is then as follows:
   does λP.P

- S:respect(fong)(kathy)
  - NP:kathy
    - *Kathy*
  - VP:respect(fong)
    - Aux λP.P
      - *does*
    - VP:respect(fong)
      - V:respect
        - *respects*
      - NP:fong
        - *Fong*

6. (3 points) Syntax and Logical Semantics II. If *woman* and *telescope* are given lexical entries like *car*, how many semantic readings are there for *Kathy saw the woman with a telescope*? Give the parse trees, complete with semantic forms. How do the generated semantic forms not match the semantic distinctions that we have generally assumed for the multiple parses of sentences such as this?

   Solutions:
   The two trees are shown below:

- S:saw(ι(λx.woman(x) $\wedge$ with'(ι(telescope))(x)))(kathy)
  - NP:kathy
    - *Kathy*
  - VP:saw(ι(λx.woman(x) $\wedge$ with'(ι(telescope))(x)))
    - V:saw
      - *saw*
    - NP:(ι(λx.woman(x) $\wedge$ with'(ι(telescope))(x))
      - Det:ι
        - *the*
      - N': λx.woman(x) $\wedge$ with'(ι(telescope))(x)
        - N:woman
          - *woman*
        - PP:λP.λx.P(x) $\wedge$ with'(ι(telescope))(x)
          - P:λy.λP.λx.P(x) $\wedge$ with'(y)(x)
            - *with*
          - NP:ι(telescope)
            - Det:ι
              - *the*
            - N:telescope
              - *telescope*



In general, we expect the meanings of the two interpretations of this sentence to be roughly "Kathy saw the woman who was holding the telescope" and "Kathy saw the woman through the telescope".

However, the two interpretations here are "Kathy saw the woman who was with the telescope" (as expected) and "Kathy saw the woman and Kathy was with the telescope" (which is not as expected).