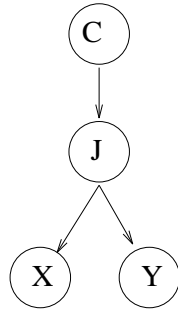


CS 530 — Principles of AI
Written Exercises
Out: November 27, 2001
Due: December 6, 2001

Problem 1. The goal of this problem is to familiarize yourself with Belief Networks as flexible specifications of probability distributions. In class, I described the following network as a clustering model:

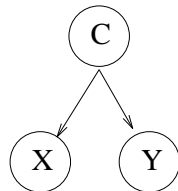


C is a variable for the underlying category of an object; X and Y are observable object features (assume they are discrete), and J is a hidden variable representing which cluster the object belongs to.

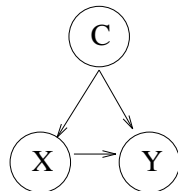
1a. What probability distributions do you have to specify to determine a joint distribution on X, Y, J and C in this network?

1b. Suppose that the distribution of objects is as follows. A quarter of all objects are in class $C = 1$ and have $X = \text{true}$ and $Y = \text{false}$. A quarter of all objects are in class $C = 1$ and have $X = \text{false}$ and $Y = \text{true}$. A quarter are in class $C = 2$ and have $X = \text{true}$ and $Y = \text{true}$. A quarter are in class $C = 2$ and have $X = \text{false}$ and $Y = \text{false}$. Specify parameters for this network (as outlined problem 1a) to capture this distribution exactly.

1c. Can the belief network shown below describe the distribution of problem 1b exactly? Why or why not?



1d. Can the belief network shown below describe the distribution of problem 1b exactly? Why or why not?

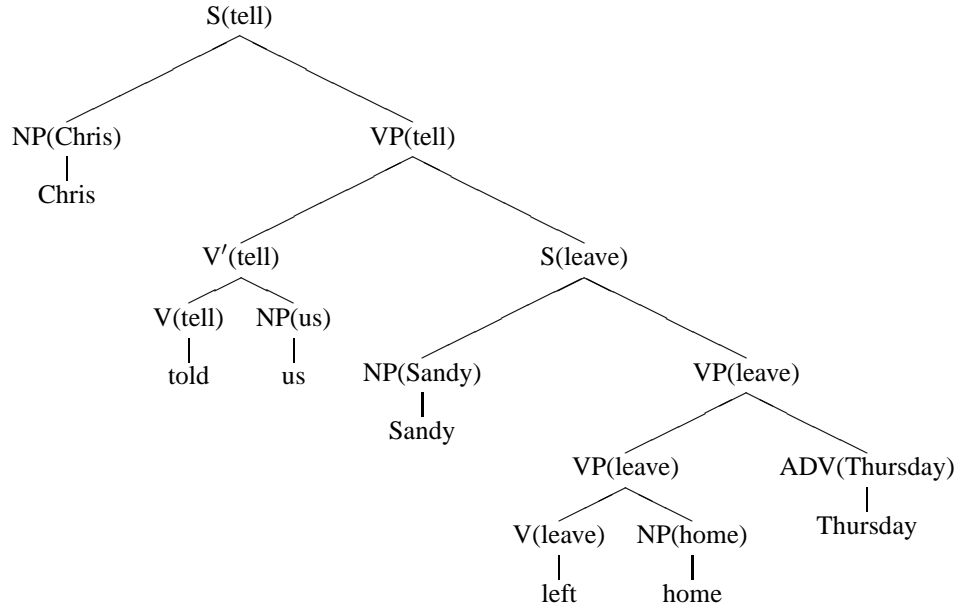


Problem 2. This problem looks at probabilistic context-free grammars as representations of linguistic forms, and some of their strengths and limitations.

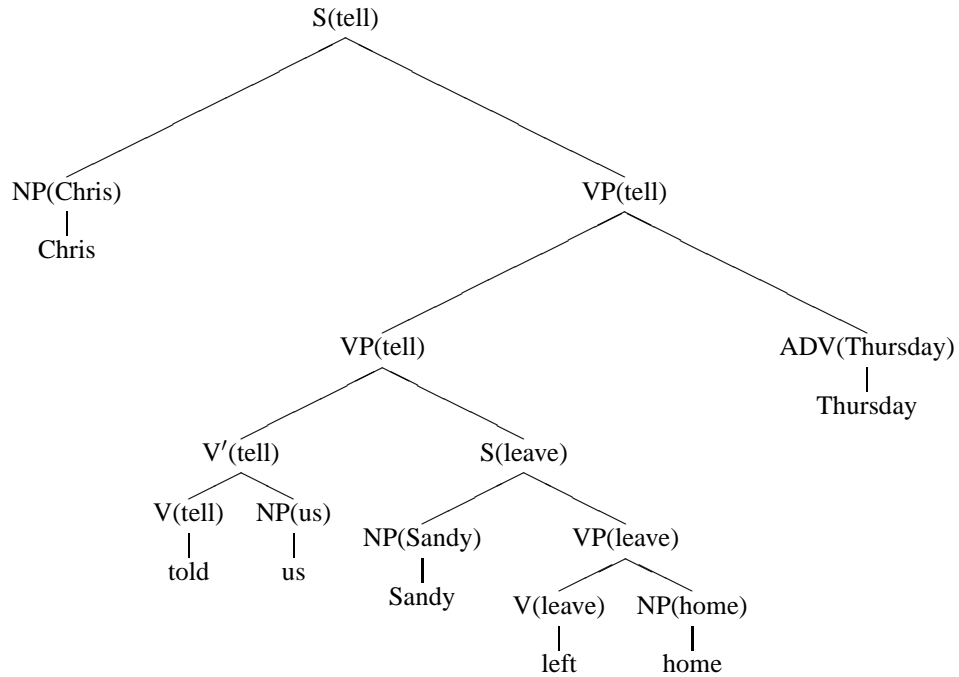
Sentences such as the following are ambiguous:

Chris told us Sandy left home Thursday.

Problem 2a. Which reading corresponds to the tree below:



Problem 2b. Which reading corresponds to the tree below:



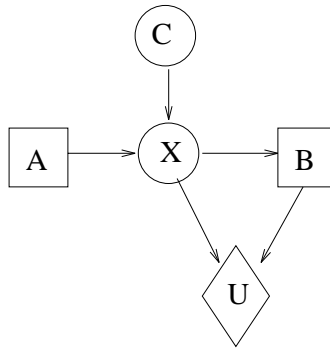
Problem 2c. What rule occurrences are required for the analysis in 2a but are not required for the analysis in 2b? And what rule occurrences are required for the analysis in 2b but not for the analysis in 2a?

Problem 2d. Using the notation $P(X \rightarrow YZ)$ for the probability parameter of a context-free rule in a PCFG, and your answer to problem 2c, write a concise mathematical rule that describes which of analysis 2a and 2b is assigned higher probability by an arbitrary PCFG.

Problem 2e. Suppose that we use a large body of text to discover what day of the week different kinds of events tend to happen. Can a PCFG model represent this generalization so as to use it to disambiguate the structures of 2a and 2b?

Problem 2f. Suppose that we find, in a large body of text, that adverbs tend to modify the closest verb. Can a PCFG model represent this preference and use it to disambiguate the structures of 2a and 2b?

Problem 3. Here is an influence diagram (the extension of Belief Networks that describes choice as well as uncertainty).



In summary, this diagram describes the following situation. The world starts out in a hidden state represented by the variable C , which is true or false. Without knowing C , the agent has to make a choice of action A , either a or a' . A and C probabilistically determine an effect X , which is true or false. The agent then observes X and makes a choice at B of b or b' . The utility of the outcome depends on X and B .

Here are parameters of the model which yield a complete specification of the agent's decision-making.

$P(C)$		0.6
$P(X A,C)$	A	C
	a	T
	a	F
	a'	T
	a'	F
		0.9
		0.4
		0.3
		0.3
$U(X,B)$	X	B
	T	b
	T	b'
	F	b
	F	b'
		0.5
		0.3
		0.4
		0.7

Problem 3a. Formulate an equivalent tree decision model.

Problem 3b. What is the optimal policy in the tree decision model? What is its utility?

Problem 3c. Explain how the independence assumptions of the influence diagram allow you to specify the optimal policy more compactly.