Statistical Reasoning

Unit 4

Brandon Syiem

Bayes Theorem

- The probability of some hypothesis H given some evidence E is

$$P(H|E) = \frac{P(E|H)\,P(H)}{P(E)}$$

- How do we get this?
    - P(A|B) = P(A,B)/P(B)
    - P(B|A) = P(A,B)/P(A)
    - Therefore,
        - P(A,B) = P(A|B)P(B) = P(B|A)P(A)
    - And, from the above

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}$$

Cont ..

- $P(A,B,C)$??
- Let us take, $D = B \cap C$
- $P(A,D) = P(A|D)\,P(D) = P(D|A)\,P(A)$
- $P(A|B,C)P(B,C) = P(A|B,C)P(B|C)P(C)$
- We could have also taken, $D = A \cap C$ or $D = A \cap B$
- So there are many combinations that need to be considered
- Also, what happens when the number of variables is large
- $P(A,B,C,D,E,F,G \ldots\ldots)$

Certainty Factors and Rule based systems

- Rule based systems represent most knowledge as **rules.**
- Rules are in the form of
  - **IF** (precondition)
  - **THEN** (conclusion)
- These rule systems can be augmented by using a measure of certainty of the conclusion.
  - Measure of belief/disbelief in the **hypothesis** given some **evidence**
- These measures of certainty are called **Certainty factors.**

Cont ..

- Certainty factor for a certain Hypothesis given some Evidence can be written as **CF[h,e].**
- Is made up of two components
  - A measure of **belief** in **h** given some **e** - range [0,1]
    - **MB[h,e]**
  - A measure of disbelief in **h** given some **e** - range [0,1]
    - **MD[h,e]**

$$\boxed{CF[h,e] = MB[h,e] - MD[h,e]}$$

Cont ..

- An Evidence can either support a hypothesis or deny it but **not both.**
- Therefore only MB or MD can be prevalent at one instance.
- Note that CF is similar to conditional probability but not the same.
  - An example will be that CF can be negative.

Combination of certainty rules

IF has-spots(X)
AND has-fever(X)
THEN has-measles(X) CF 0.5

- Let's say that we have the evidence
  - has-spots(john) with CF = 0.4 -------------> (CF1)
  - has-fever(john) with CF = 0.8 -------------> (CF2)
- To work out the CF of has-measles, we have to consider
  - The CF of each evidence.
  - The CF attached to the rule.

Cont ..

- The CF of the Conjunction(AND) of two evidence can be calculated as
  - **CF = min(CF1, CF2)**
- The CF of the Disjunction(OR) of two evidence can be calculated as
  - **CF = max(CF1, CF2)**
- The CF for the conclusion has-measles(john) can be calculated as
  - **Total CF = rule CF * total certainty of the evidence**
  - CF = 0.5 * 0.4

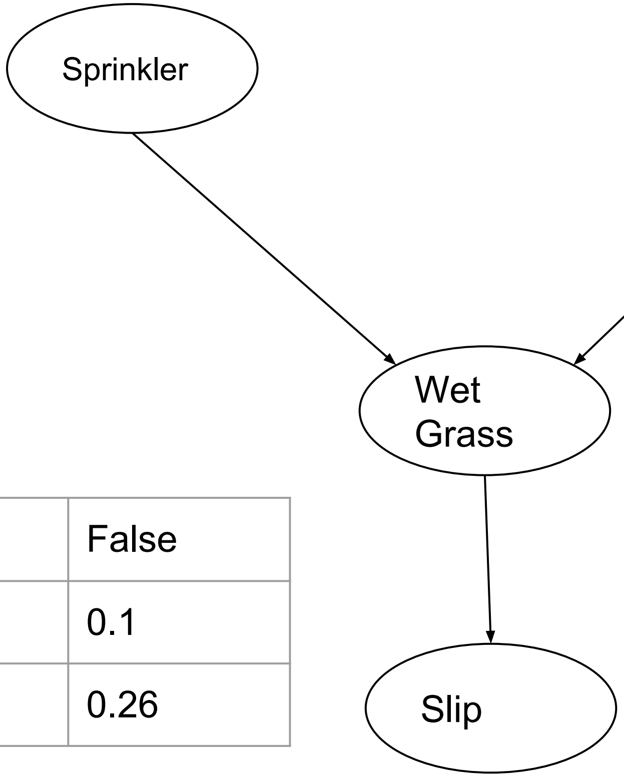
IF has-spots(X)	CF 0.4
AND has-fever(X)	CF 0.8
THEN has-measles(X)	CF 0.5

Cont ..

- If we have another rule that concludes has-measles(john), the new CF for has-measles(john) is
    - $\boxed{\textbf{CF = total CF1 + total CF2 - (total CF1 * total CF2)}}$

Bayesian Networks

- https://www.youtube.com/watch?v=TuGDMj43ehw
- https://www.bayesserver.com/docs/introduction/bayesian-networks
- It is a probabilistic graphical model
- Is a Directed Acyclic Graph (DAG)



Each entry in the ovals is a variable that can either be a proposition {True,False} or a continuous value

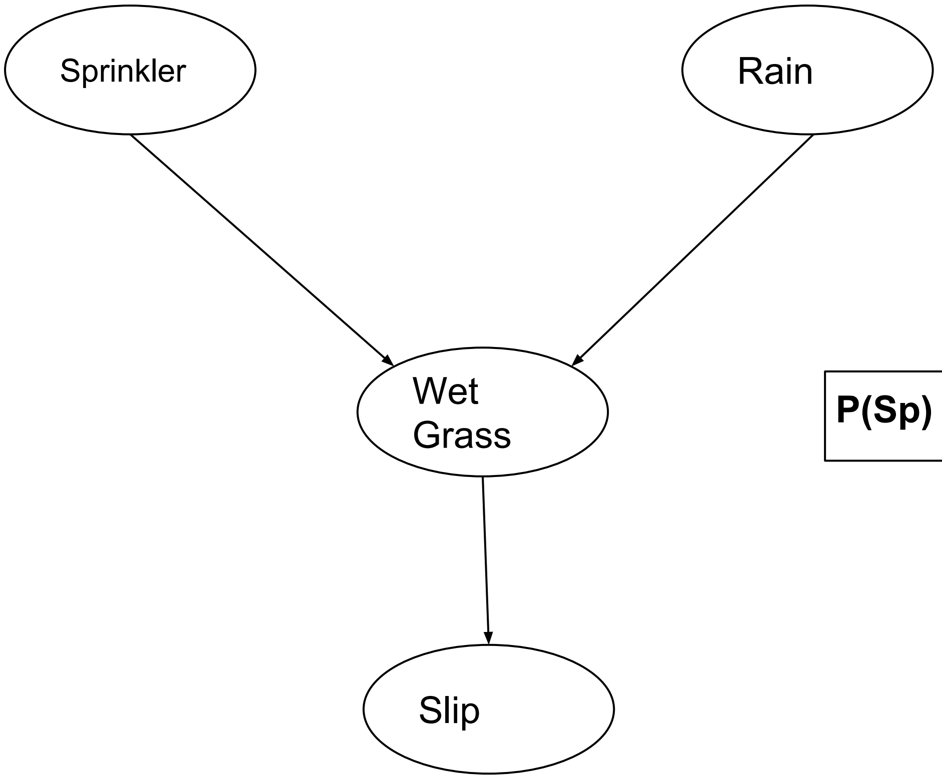
Each Variable is associated with a Probability Table

P(Slip)

	True	False
WG = TRUE	0.6	0.1
WG = FALSE	0.04	0.26

P(Rain)

True	False
0.6	0.4



- This bayesian Structure describes:
- P(Sprinkler,Rain,Wet Grass, Slip) OR
- P(Sp,R,WG,Sl)
- And is equivalent to

$$\boxed{\mathbf{P(Sp)\ P(R)\ P(WG|Sp,R)\ P(Sl|WG)}}$$

- How can we find specific probabilities?
    - Say P(R|Sl)
- Use the concept of **Marginalization**

Dempster-Shafer Theory

- Dempster-Shafer theory is an approach to combining evidence
- Dempster (1967) developed means for combining degrees of belief derived from independent items of evidence.
- His student, Glenn Shafer (1976), developed method for obtaining degrees of belief for one question from subjective probabilities for a related question.

http://arantxa.ii.uam.es/~modonnel/IC/04_3_DempsterShaferI.pdf

Cont ..

- Bayesian Networks concentrates on finding the probability of a single hypothesis given a set of evidence.

$$P(H|E_1,E_2,....)$$

- The Dempster-shafer theory deals with evidence that supports a set of hypothesis (or proposition).
- It also assigns an interval to every subset **A** of **H,** given as

$$[Bel(A),Plausibility(A)]$$

Cont ..

- Let's, say we have four different Hypothesis:

$$H = \{H1,H2,H3,H4\}$$

- We define the **Frame of Discernment** as the Power set of **H** as:

P = {∅ , {H1} , {H2} , {H3} , {H1,H2} , {H2,H3} , {H1,H3} , {H1,H2,H3}}

- Mass Function(A) :- A function that returns a value that signifies the proportion of relevant and available evidence that support A. has a range of [0,1]

  - $m(\varnothing) = 0$
  - $\sum m(A) = 1$ where, A E P

Cont ..

- Belif Function(A) :- the Belief function of **A** is equal to the sum of the **masses (mass function)** of all the subsets of **A** including **A** itself.
  - Example

$$Bel(\{H1VH2\}) = m(H1)+m(H2)+m(H1,H2)$$

- Plausibility(A) :- The plausibility of some element **A** in P is the sum of the masses of all the elements that intersect with **A**
  - **Example**

$$Pl(\{H1VH2\}) = m(H1)+m(H2)+m(H1,H2)+m(H1,H3)+m(H2,H3)+m(H1,H2,H3)$$

- Pl can also be written as : $\boxed{Pl(A) = 1 - Bel(\neg A)}$

Additional

- Fuzzy Logic

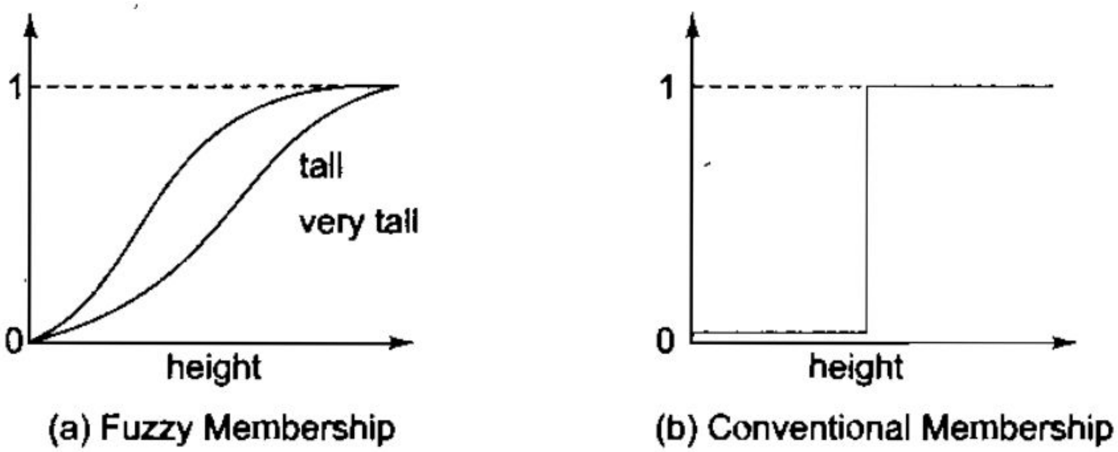


Fig. 8.4 *Fuzzy versus Conventional Set Membership*

Natural Language Processing

Introduction

- Language allows us to communicate
- Largest part of communication happens through speech
- Written language is also a large part of communication but is secondary to speech
- To create a program that understands speech, we need to have it first be able to understand language in written format and the add upon it.
- The language processing problem is divided into two parts:

- Processing written text, using lexical, syntactic, and semantic knowledge of the language as well as the required real world information
- Processing spoken language, using all the information needed above plus additional knowledge about phonology as well as enough added informal ion to handle the further ambiguities that arise in speech

Overview

- Useful to think of language as a pair (Source Language, Target Representation)
- Steps involved in natural language understanding process:
  - **Morphological Analysis**: Individual words are analysed and non-word tokens such as punctuations are separated from words.
  - **Syntactic Analysis**: Words from the morphological analyzer are transformed into structures that show how the words relate to each other. (“The ant the cloud sun” will be rejected)
  - **Semantic Analysis**: The structures created in Syntactic analysis are given meaning. (“Colorless green liquid” will be rejected)
  - **Discourse Integration:** The meaning of a sentence may depend on a previous sentence and the meaning of a following sentence may depend on the current sentence.(“We were playing with the ball. John has **it** now. **He** went home”)

Cont ..

  - **Pragmatic Analysis**: The structure representing what was said is reinterpreted to determine what was meant. (“Do you know what time it is?” to “What is the time”)
- All the above processes can either be done in sequence or all at once.

More information on the different steps can be found in pg. 288 - 290 E.Rich and K.Knight

Example

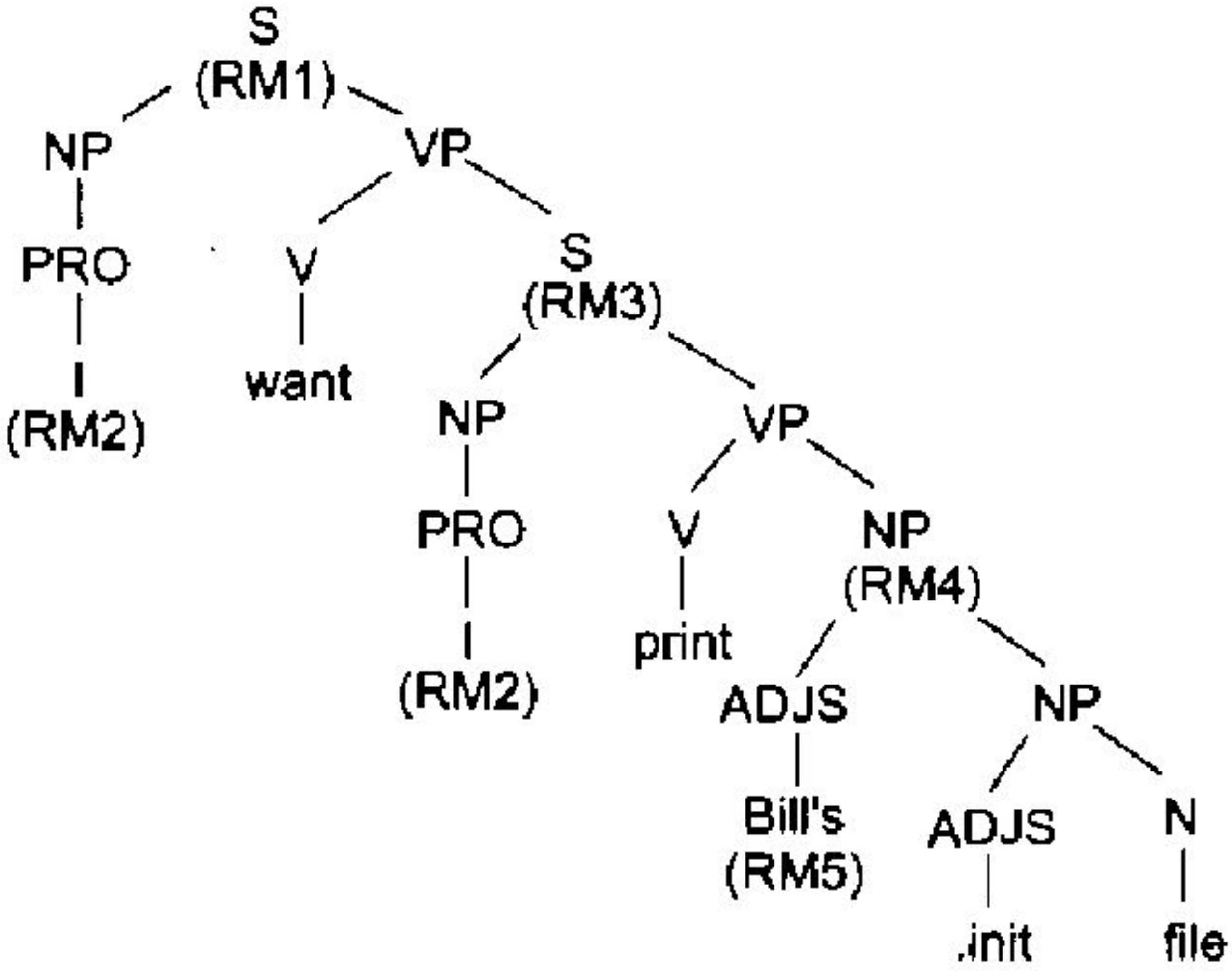
Let’s take an Example:

“I want to print Bill’s .init file”

- **Morphological Analysis has the following tasks:**
    - Break up the string into distinct words
    - Pull apart the word “Bill’s” into the proper noun Bill and the possessive suffix “s”
        - “Prints” can be pulled apart as “print” and “s” and it must also decide is the “s” represents plurality, third person singular, etc.
    - Recognize the word “.init” as a file extension that is functioning as a adjective
    - Also needs to assign ***syntactic*** categories to each word (**“Bill”** is a **proper noun**)

Cont ..

- **Syntactic Analysis**
    - Main function is to use the output of the morphological analyzer and to build a structural description of the sentence
    - Called **parsing**
    - Flat sentence structure is converted to a hierarchical structure.
    - Some Syntactic systems create ***reference markers***


S
(RM1)
NP
VP
PRO
V
S
(RM3)
(RM2)
want
NP
VP
PRO
V
NP
(RM4)
(RM2)
print
ADJS
NP
Bill's
(RM5)
ADJS
N
.init
file

Cont ..

- **Semantic Analysis**
  - It must map individual words into appropriate objects in the knowledge base
  - It must create correct structures to correspond to the way the meanings of the individual words combine with each other

User	
isa :	*Person*
* *login-name :*	must be <string>
User068	
instance :	*User*
login-name :	*Susan-Black*
User073	
instance :	*User*
login-name :	*Bill-Smith*
F1	
instance :	*File-Struct*
name :	stuff
extension :	.init
owner :	*User073*
in-directory :	/wsmith/
File-Struct	
isa :	*Information-Object*

Printing	
isa :	*Physical-Event*
* *agent :*	must be <animate or program>
* *object :*	must be <information-object>
Wanting	
isa :	*Mental-Event*
* *agent :*	must be <animate>
* *object :*	must be <state or event>
Commanding	
isa :	*Mental-Event*
* *agent :*	must be <animate>
* *performer :*	must be <animate or program>
* *object :*	must be <event>
This-System	
instance :	*Program*

RM1		{the whole sentence}
instance :	*Wanting*	
agent :	*RM2*	{I}
object :	*RM3*	{a printing event}
RM2		{I}
RM3		{a printing event}
instance :	*Printing*	
agent :	*RM2*	{I}
object :	*RM4*	{Bill's .init file}
RM4		{Bill's .init file}
instance :	*File-Struct*	
extension :	.init	
owner :	*RM5*	{Bill}
RM5		{Bill}
instance :	*Person*	
first-name :	Bill	

Cont ..

Discourse Integration

At this point, we have figured out what kinds of things this sentence is about. But we do not yet know which specific individuals are being referred to. Specifically, we do not know to whom the pronoun "I" or the proper noun "Bill" refers. To pin down these references requires an appeal to a model of the current discourse context, from which we can learn that the current user (who typed the word "I") is *User068* and that the only person named "Bill" about whom we could be talking is *User073*. Once the correct referent for Bill is known, we can also determine exactly which file is being referred to: *F1* is the only file with the extension ".init" that is owned by Bill.

Cont ..

- Pragmatic Analysis
  - If the sentence is a fact then done.
  - If intended meaning is different from what was said then apply required set of rule*

Meaning	
instance :	*Commanding*
agent :	*User068*
performer :	*This-System*
object :	*P27*
P27	
instance :	*Printing*
agent :	*This-System*
object :	*F1*

- **Book pg: - 290 (E.Rich and K.Knight)**

Syntactic Processing

- flat input from the morphological analysis is converted to a hierarchical structure that corresponds to the units of meaning in the sentence.
- Called ***parsing***.
- Semantic analysis operates on **sentence constituents** (VERBS, NOUNS, ADJECTIVES, etc)
- Two main components:
    - **Grammar** - declarative facts about the syntactic features of a language
    - **Parser** - compares grammar to input to produce parse structures.

Grammar

- Easily represented by a set of production rules.
- A simple context-free, phrase structured set of rules can be written as seen on the right.
- Symbols that can further be expanded are called *non-terminal symbols.*
- Symbols that cannot be further expanded are called *terminal symbols.*

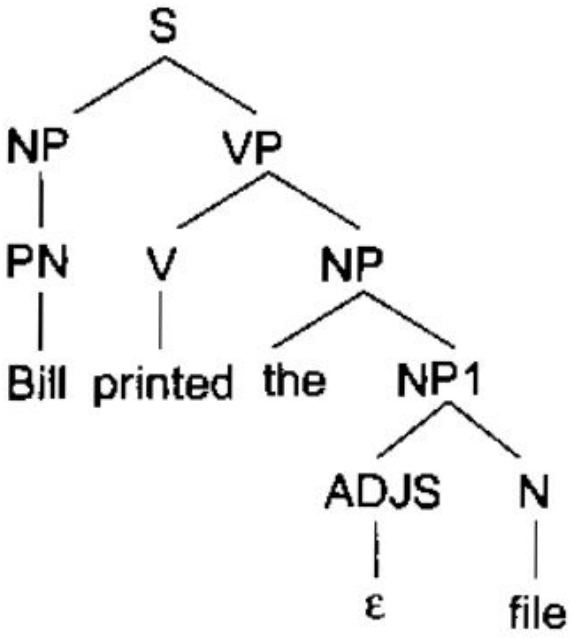
S → NP VP
NP → the NP1
NP → PRO
NP → PN
NP → NP1
NP1 → ADJS N
ADJS → ε | ADJ ADJS
VP → V
VP → V NP
N → file | printer
PN → Bill
PRO → I
ADJ → short | long | fast
V → printed | created | want

Parser

- Takes the rules of the grammar and compares them to our input sentence
- The simplest structure that can be built is a ***parse tree*** which simply records the rules and how they are matched.
- ***Parse Tree***
  - Each leaf node corresponds to a terminal node or an input word
  - Each level corresponds to the application of one grammar rule

Example Parser

“Bill printed the file”



Types of parsing

- *Top-Down Parsing*—Begin with the start symbol and apply the grammar rules forward until the symbols at the terminals of the tree correspond to the components of the sentence being parsed.
- *Bottom-Up Parsing*—Begin with the sentence to be parsed and apply the grammar rules backward until a single tree whose terminals are the words of the sentence and whose top node is the start symbol has been produced.

- Picking which is better is similar to the reasoning we applied to picking forward or backward reasoning
    - Branching factor is one of the main factors

Augmented Transition Networks

pg :- 295 -297 (E.Rich & K.Knight)