Graphical Models for Discovering Knowledge

Wray Buntine, Thinkbank INC.

Probabilistic Graphical Models

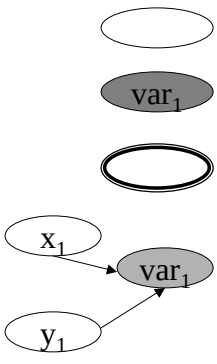
- Provide a unified qualitative and quantitative framework for representing and reasoning with probabilities and dependencies
- Provide an intuitive visual method for decomposing a problem using the notion of conditional probability

What is a “Model”?

- A model is a representation of the problem at hand showing the different variables involved, and the probabilistic and deterministic relationships between those variables.

Probabilistic Graphic Models

Graph elements



Unshaded node: unknown variable

var_1 Shaded node: known variable

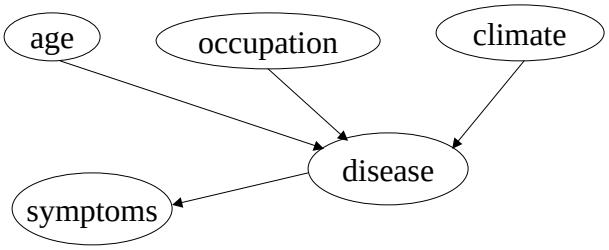
Double walled node: deterministic function

Directed arcs indicate conditional probability $p(var_1 \mid x_1, y_1)$

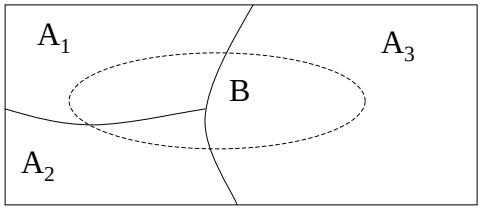
Bayesian (Causal) Network

Example: Given symptoms of some disease, what can we infer about the age, occupation, and living climate of the patient?

P(age,occ,cli,dis,sym) = P(age)P(occ)P(cli)P(dis|age,occ,cli)P(sym|disease)



Bayes Theorem



Given that B has occurred, what is the probability that A_1 occurred?

$P(A_1|B) = P(A_1 \cap B)/P(B) = P(A_1)P(B|A_1)/P(B)$

Problem Decomposition & Analysis

Graphical Models clarify problem structure without excess math.

Example: Topic Extraction from AP Newswire

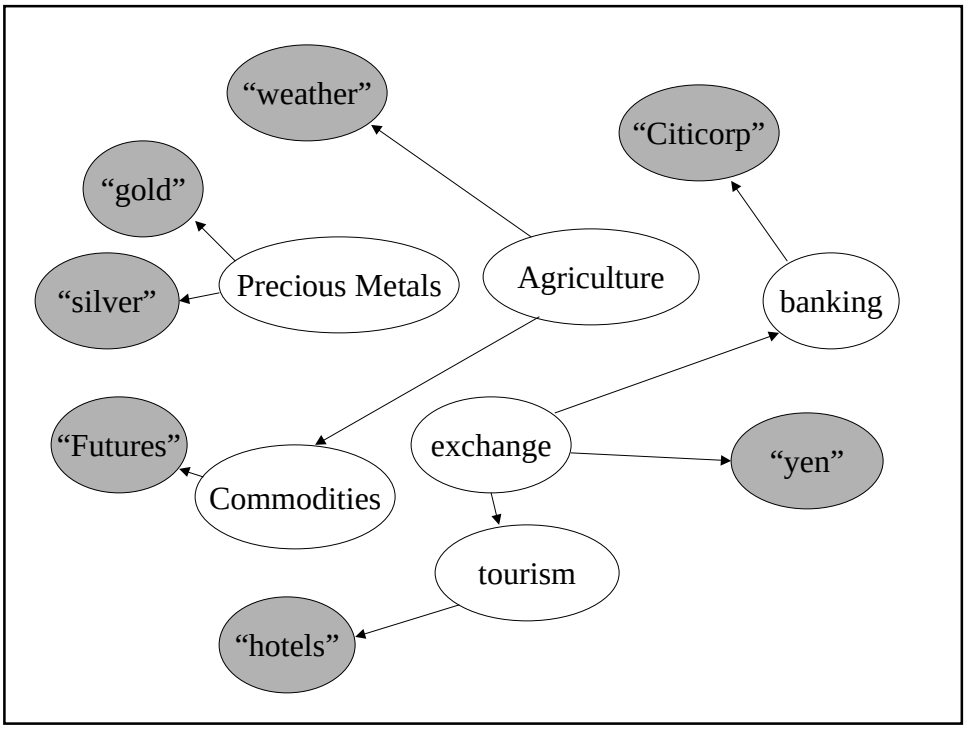
The AP Newswire provides numerous short newsclips belonging to one or more of 90 subject areas. Given that a set of 11,000 words is used in the clips, can we predict the subject of a given clip by examining the set of words used in the clip?

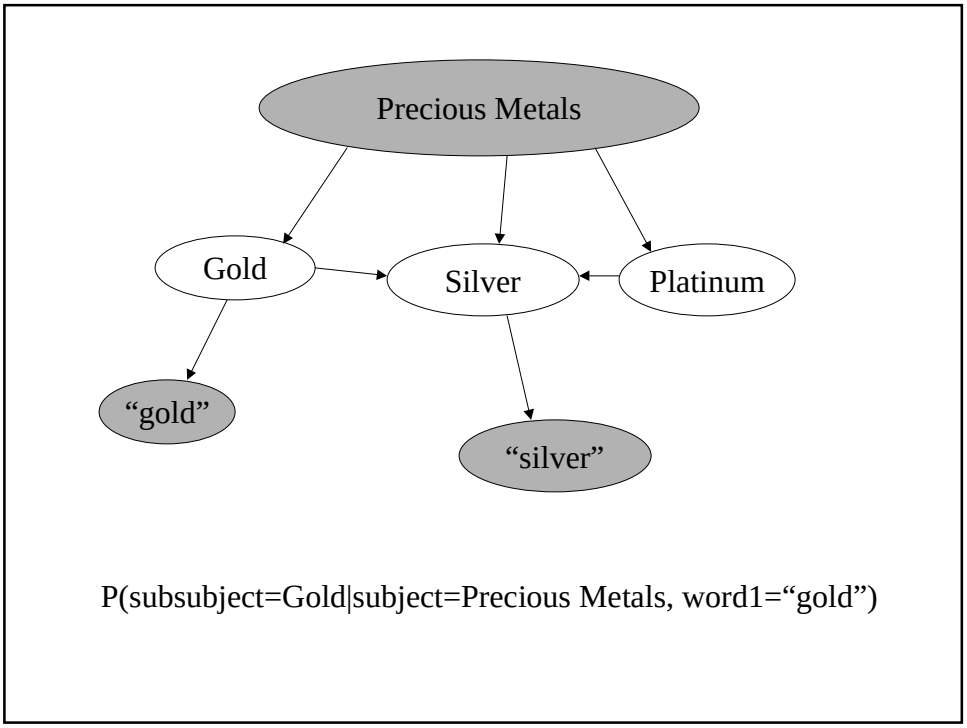
That is, can we find the Bayesian Probability:

$$P(\text{Subject A}|\text{word1, word2, word3,}\ldots)$$

Method:

(1) Establish relationships between general subjects and more specific sub-subjects
(2) Establish relationships between key words and general subjects
(3) Establish relationships between key words, general subjects and sub-subjects

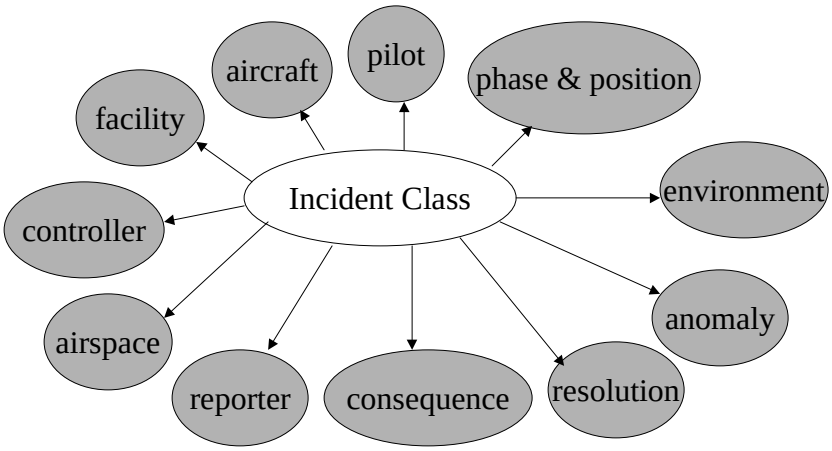




P(subsubject=Gold|subject=Precious Metals, word1=“gold”)

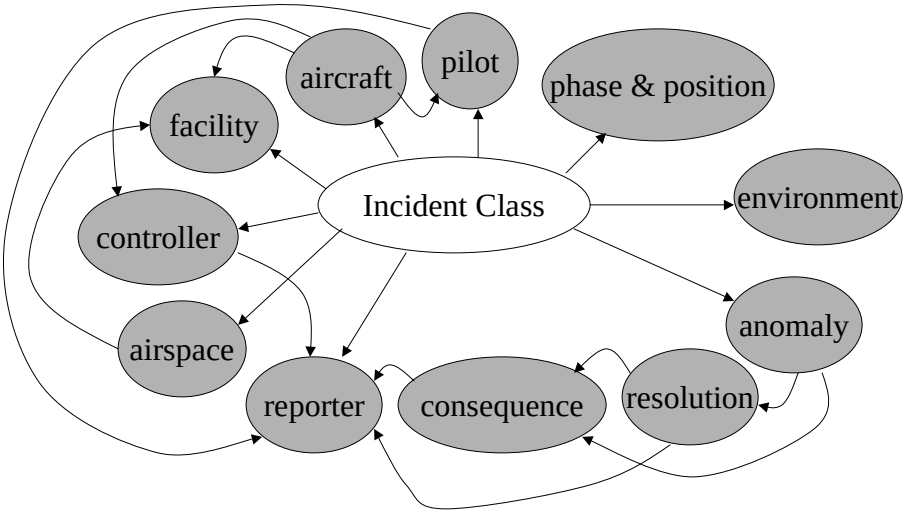
Knowledge refinement using graphical models:
Example: Classification of Aircraft accidents

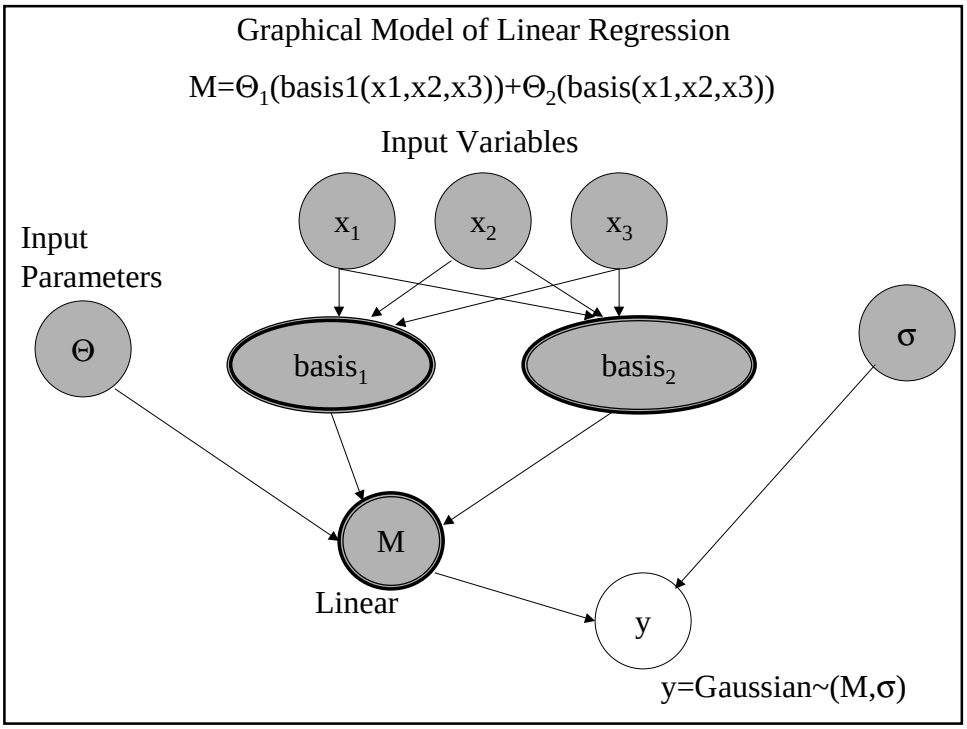
Simple case: assume all given variables are independent

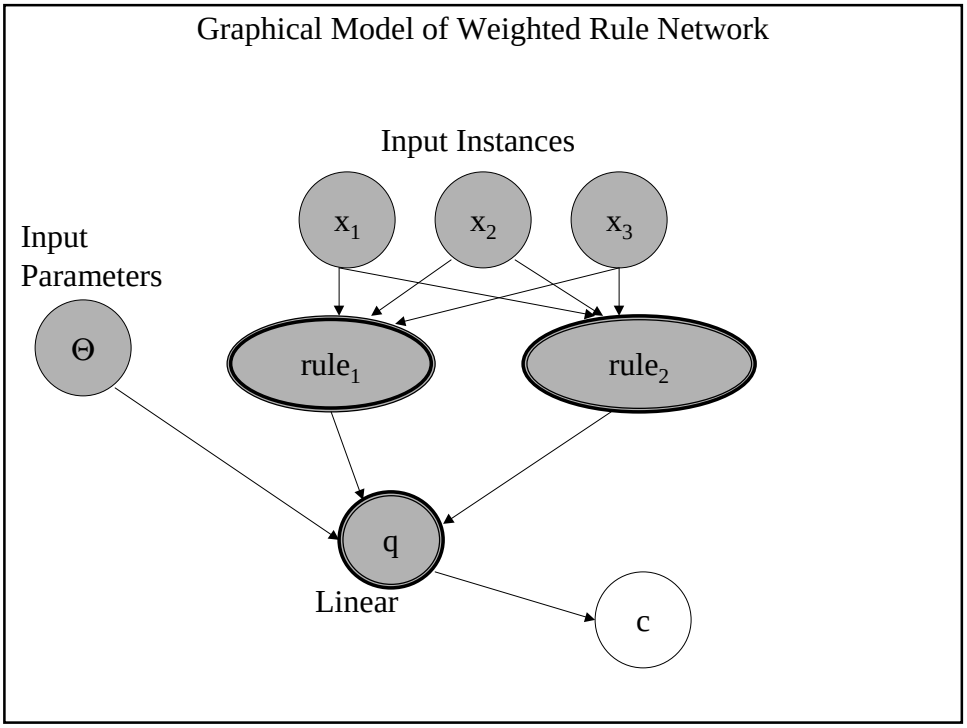


Knowledge refinement using graphical models:
Example: Classification of Aircraft accidents

Refined case: use expert knowledge to restructure dependencies

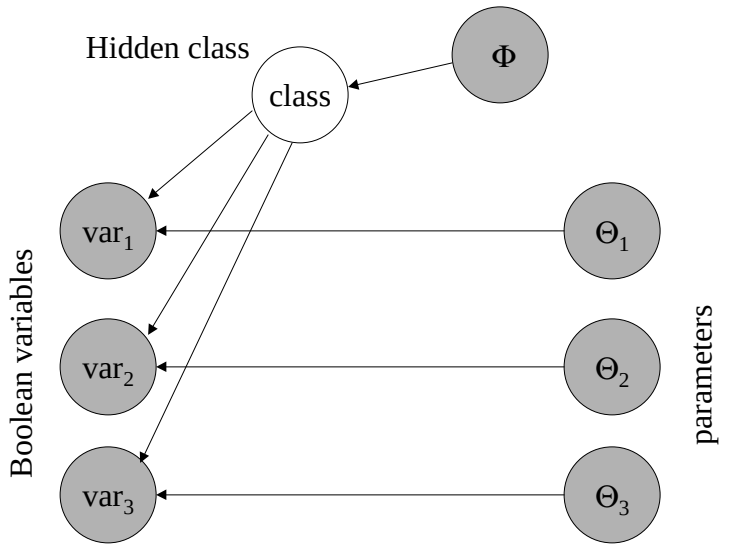



Graphical Model of Linear Regression
M=Θ1(basis1(x1,x2,x3))+Θ2(basis(x1,x2,x3))
Input Variables
x1
x2
x3
Input
Parameters
Θ
basis1
basis2
σ
M
Linear
y
y=Gaussian~(M,σ)


Graphical Model of Weighted Rule Network
Input Instances
x1
x2
x3
Input
Parameters
Θ
rule1
rule2
q
Linear
c

Autoclass fitting of a hidden class to variables & parameters

Unsupervised Learning



Conclusions

- Probabilistic Graphical Models provide a general language for problem composition
- The flexibility of Graphical Models allows representation of a wide variety of models, from linear regression to learning
- Graphical Models provide a framework for developing computational learning and discovery algorithms