ARTIFICIAL INTELLIGENCE

L T P C

BCSE306L - 3 0 0 3

Vellore Institute of Technology
(Deemed to be University under section 3 of UGC Act, 1956)

Dr. S M SATAPATHY

Associate Professor Sr.,
School of Computer Science and Engineering,
VIT Vellore, TN, India – 632 014.

Module – 5

UNCERTAIN KNOWLEDGE & REASONING

1. Uncertainity
2. Bayes Rule Inference
3. Belief Network
4. Utility based System
5. Decision Network

UNCERTAINITY

Uncertainity :: Definition

- Uncertainty refers to epistemic situations involving imperfect or unknown information.
- It applies to predictions of future events, to physical measurements that are already made, or to the unknown.
- The world is not a well-defined place.
- There is uncertainty in the facts we know:
  - What's the temperature? Imprecise measures
  - Is Donald a good president? Imprecise definitions
  - Where is the pit? Imprecise knowledge
- There is uncertainty in our inferences
  - If I have a blistery, itchy rash and was gardening all weekend I probably have poison ivy

Sources of Uncertainty

- Uncertain data
  - missing data, unreliable, ambiguous, imprecise representation, inconsistent, subjective, noisy…
- Uncertain knowledge
- Uncertain knowledge representation
  - restricted model of the real system
  - limited expressiveness of the representation mechanism
- Inference process
  - Derived result is formally correct, but wrong in the real world
  - New conclusions are not well-founded (eg, inductive reasoning)
  - Incomplete, default reasoning methods

Reasoning under Uncertainty

- So how do we do reasoning under uncertainty and with inexact knowledge?
  - heuristics
    - ways to mimic heuristic knowledge processing
    - methods used by experts
  - empirical associations
    - experiential reasoning
    - based on limited observations
  - probabilities
    - objective (frequency counting)
    - subjective (human experience)

Using FOL for (Medical) Diagnosis

$\forall p$ *Symptom*(*p*, *Toothache*) $\Rightarrow$ *Disease*(*p*, *Cavity*)

Not correct...

$\forall p$ *Symptom(p, Toothache)* $\Rightarrow$ *Disease(p, Cavity)* $\vee$ *Disease(p, GumDisease)* $\vee$ *Disease(p, WisdomTooth)* $\vee$ *Disease(p, Abscess)*

Not complete...

$\forall p$ *Disease(p, Cavity)* $\Rightarrow$ *Symptom(p, Toothache)*

Not correct...

Uncertainty :: Example

Let action A_t = leave for airport ***t*** minutes before flight

Will A_t get me there on time?

Problems:

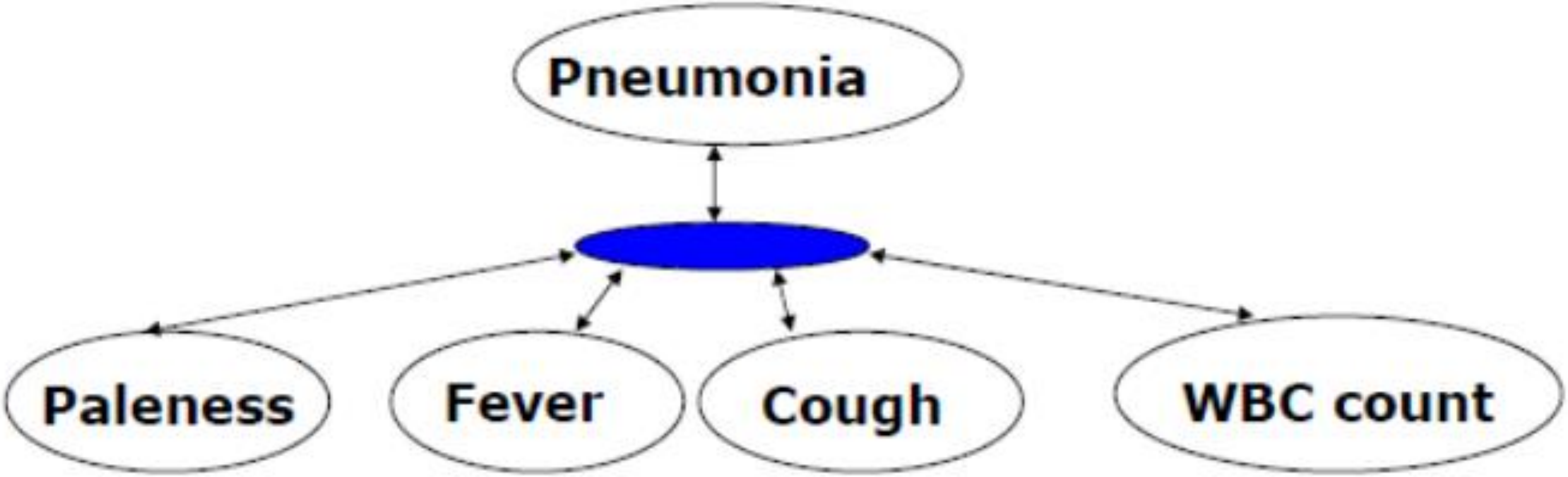
1. partial observability (road state, other drivers' plans, etc.)
2. noisy sensors (traffic reports)
3. uncertainty in action outcomes (flat tire, etc.)
4. immense complexity of modeling and predicting traffic

"A_{25} will get me there on time if there's no accident on the bridge and it doesn't rain and my tires remain intact etc."

(A_{1440} might reasonably be said to get me there on time but I'd have to stay overnight in the airport …)

Uncertainty

To make diagnostic inference possible we need to represent knowledge (axioms) that relate symptoms and diagnosis



Problem: disease/symptoms relations are not deterministic

- **They are uncertain (or stochastic) and** vary from patient to patient

Uncertainty

Why are relations uncertain?

- **Observability**
  - It is impossible to observe all relevant components of the world
  - Observable components behave stochastically even if the underlying world is deterministic
- **Efficiency, capacity limits**
  - It is often impossible to enumerate and model all components of the world and their relations
  - abstractions can become stochastic

Humans can reason with uncertainty !!!

- Can computer systems do the same?

Methods for representing Uncertainty

Extensions of the propositional and first-order logic

– Use, uncertain, imprecise statements (relations)

Example: Propositional logic with certainty factors

Very popular in 70-80s in knowledge-based systems (MYCIN)

- **Facts (propositional statements)** are assigned a **certainty value** reflecting the belief in that the statement is satisfied:

$$CF(Pneumonia = True) = 0.7$$

- **Knowledge:** typically in terms of **modular rules**

If	1. The patient has cough, and
	2. The patient has a high WBC count, and
	3. The patient has fever
Then	**with certainty 0.7**
	the patient has pneumonia

CF-Cystic fibrosis

Methods for representing Uncertainty

Probability theory

- A well defined theory for modeling and reasoning in the presence of uncertainty
- A natural choice to replace certainty factors

Facts (propositional statements)

- Are represented via **random variables** with two or more values

  Example: *Pneumonia* is a random variable

  values: ***True*** **and** ***False***

- Each value can be achieved **with some probability:**

$$P(Pneumonia = True) = 0.001$$

$$P(WBCcount = high) = 0.005$$

Handling Uncertain Knowledge

- Problems using first-order logic for diagnosis of Pneumonia:
  - Laziness:
    - Too much work to make complete rules.
    - Too much work to use them
  - Theoretical ignorance:
    - Complete theories are rare
  - Practical ignorance:
    - We can’t run all tests anyway

- Probability can be used to summarize the laziness and ignorance !

Handling Uncertain Knowledge

- Default or nonmonotonic logic:
  - Assume my car does not have a flat tire
  - Assume A_{25} works unless contradicted by evidence
- Issues: What assumptions are reasonable? How to handle contradiction?
- Rules with fudge factors:
  - $A_{25} \mapsto_{0.3}$ get there on time
  - $Sprinkler \mapsto_{0.99} WetGrass$
  - $WetGrass \mapsto_{0.7} Rain$
- Issues: Problems with combination, e.g., *Sprinkler* causes *Rain*??
- Probability
  - Model agent's degree of belief
  - Given the available evidence,
  - A_{25} will get me there on time with probability 0.04

Decision making with uncertainty

- Rational behaviour:
  - For each possible action, identify the possible outcomes
  - Compute the probability of each outcome
  - Compute the utility of each outcome
  - Compute the probability-weighted (expected) utility over possible outcomes for each action
  - Select the action with the highest expected utility (principle of Maximum Expected Utility)

Rational Decisions

- A rational decision must consider:
  - The relative importance of the sub-goals - Utility theory
  - The degree of belief that the sub-goals will be achieved - Probability theory

- Decision theory = probability theory + utility theory :

The agent is rational if and only if it chooses the action that yields the highest expected utility, averaged over all possible outcomes of the action”

Combining Beliefs & Desires under Uncertainty

- Modern agents talks of utility rather than good and evil, but the principle is exactly the same.
- Agent's preferences between world states are captured by a **Utility Function *U(S)***
- It assign a single number to express the desirability of a state.
- Utilities are combined with outcome probabilities of actions to give an expected utility for each action
- A nondeterministic ***action A*** will have possible outcome states $\boldsymbol{Result_i(A)}$ where the index $\boldsymbol{i}$ ranges over the different outcomes.
- Prior to the execution of A, the agent assigns probability $\boldsymbol{P\ (Result_i\ (A) \mid Do\ (A), E)}$ to each outcome, where E summarizes the agent's available evidence about the world and Do (A) is the proposition that action A is executed in the current state.
- The expected utility of A is

$$EU[A] = S_{i=1,\ldots,n}\, p(Result_i(A) \mid Do\ (A) U(Result_i)$$

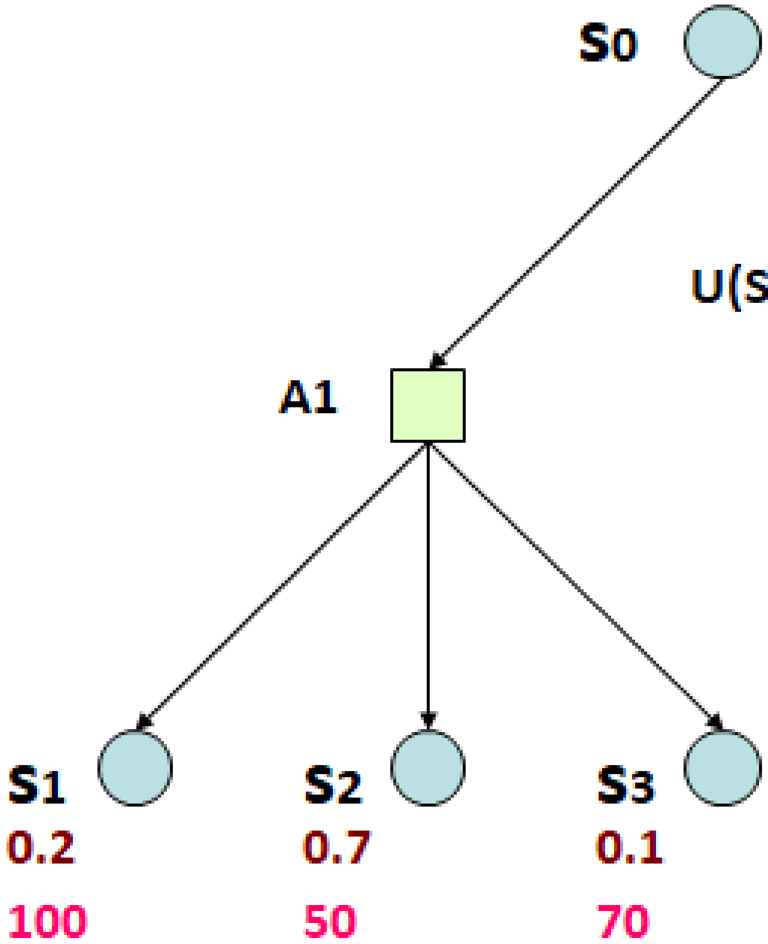
- The principle of maximum expected utility (MEU) says that a rational agent should choose an action

Expected Utility

- Random variable X with n values $x_1,\ldots,x_n$ and distribution $(p_1,\ldots,p_n)$
  E.g.: X is the state reached after doing an action A under uncertainty
- Function U of X
  E.g., U is the utility of a state
- The expected utility of A is

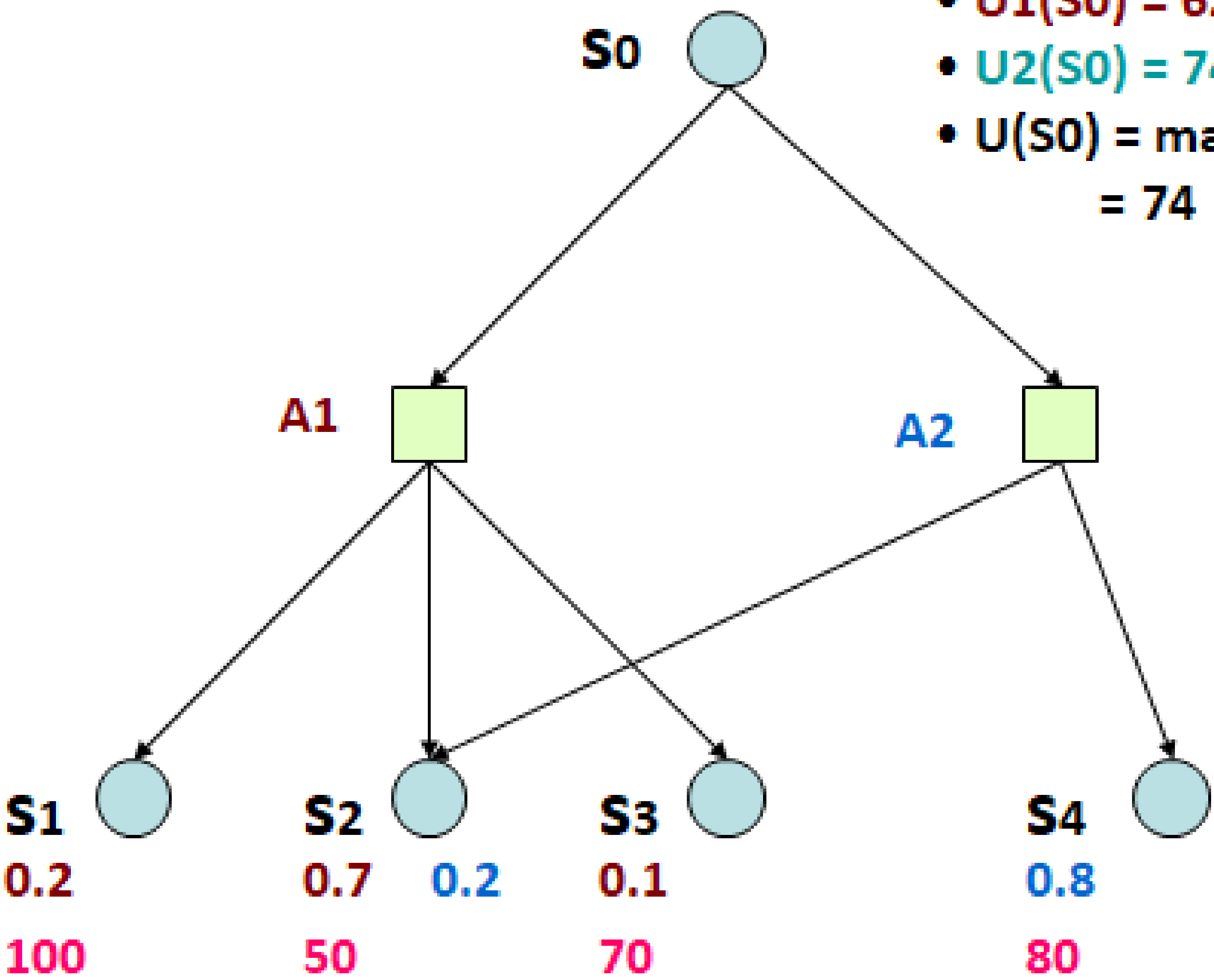
$$EU[A] = \sum_{i=1}^{n} p(x_i|A) \cup (x_i)$$

One State/One Action Example



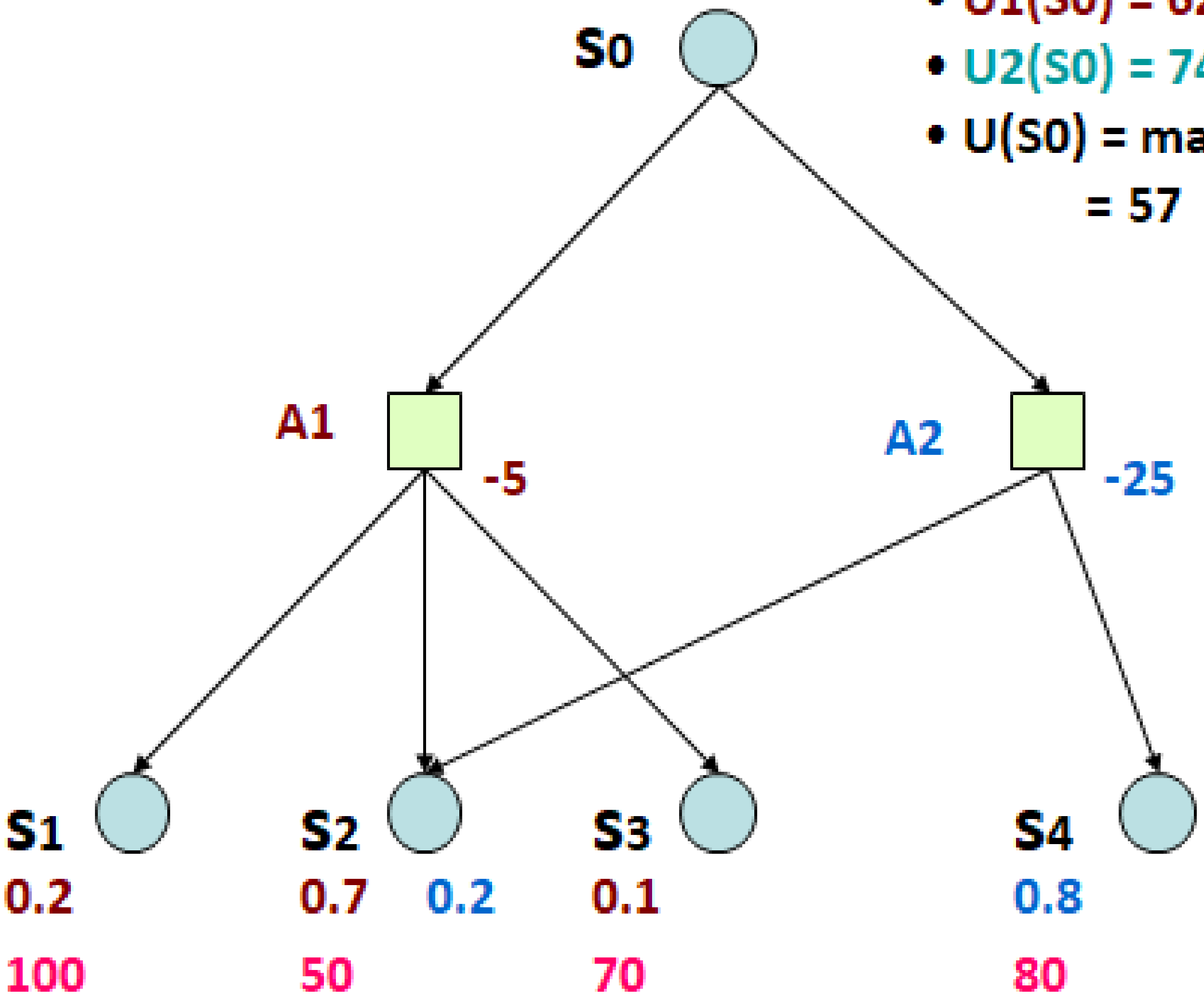
$$U(S0) = 100 \times 0.2 + 50 \times 0.7 + 70 \times 0.1$$
$$= 20 + 35 + 7$$
$$= 62$$

One State/Two Actions Example



- U1(S0) = 62
- U2(S0) = 74
- U(S0) = max{U1(S0),U2(S0)}
  = 74

Introducing Action Costs



- U1(S0) = 62 – 5 = 57
- U2(S0) = 74 – 25 = 49
- U(S0) = max{U1(S0),U2(S0)}
  = 57

Maximum Expected Utility (MEU) Principle

- A rational agent should choose the action that maximizes agent's expected utility
- This is the basis of the field of decision theory
- The MEU principle provides a normative criterion for rational choice of action

Probability

A measure of how likely it is that some event will occur; a number expressing the ratio of favorable cases to the whole number of cases possible

Subjective probability:

- Probabilities relate propositions to agent's own state of knowledge

  e.g., $P(A_{25} \mid \text{no reported accidents}) = 0.06$

These are not assertions about the world

Probabilities of propositions change with new evidence:

e.g., $P(A_{25} \mid \text{no reported accidents, 5 a.m.}) = 0.15$

Probability

Probabilities are either:

- **Prior probability** (unconditional , "obetingad")

  Before any evidence is obtained

- **Posterior probability** (conditional , "betingad")

  After evidence is obtained

The **axioms** of probability constrain the possible assignments of probabilities to propositions. An **agent that violates** the axioms will **behave irrationally** in some circumstances

Syntax

- Basic element: random variable
- Similar to propositional logic: possible worlds defined by assignment of values to random variables.
- Boolean random variables
  - e.g., Cavity (do I have a cavity?)
- Discrete random variables
  - e.g., Weather is one of <sunny,rainy,cloudy,snow>
- Elementary proposition constructed by assignment of a value to a random variable: e.g., Weather = sunny, Cavity = false (abbreviated as $\neg$ cavity)
- Complex propositions formed from elementary propositions and standard logical connectives e.g., Weather = sunny $\vee$ Cavity = false

Syntax

- Atomic event: A complete specification of the state of the world about which the agent is uncertain
  - E.g., if the world consists of only two Boolean variables Cavity and Toothache, then there are 4 distinct atomic events:

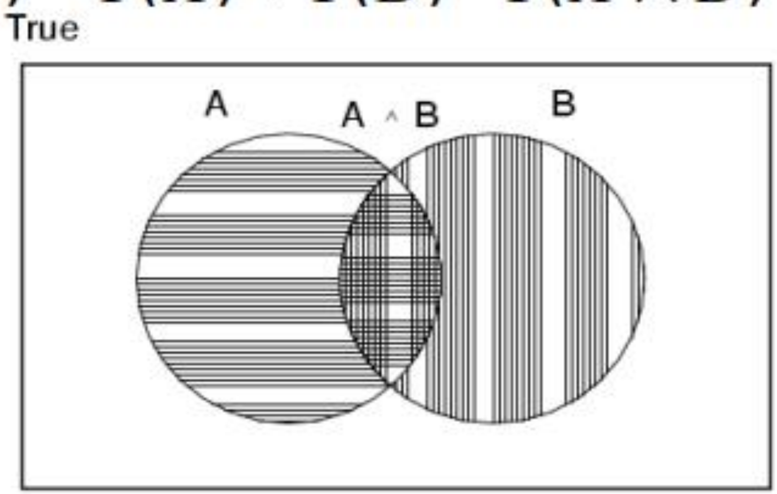
$$Cavity = false \wedge Toothache = false$$
$$Cavity = false \wedge Toothache = true$$
$$Cavity = true \wedge Toothache = false$$
$$Cavity = true \wedge Toothache = true$$

$\neg$

$\vee$

Axioms of probability

- For any propositions A, B
  - $0 \leq P(A) \leq 1$
  - $P(true) = 1$ and $P(false) = 0$
  - $P(A \vee B) = P(A) + P(B) - P(A \wedge B)$



  - These three axioms are often called ***Kolmogorov's*** axioms

Unconditional probability

- Prior or unconditional probabilities of propositions

  e.g., P(*Cavity* = true) = 0.1 and P(*Weather* = sunny) = 0.72 correspond to belief prior to arrival of any (new) evidence

- Probability distribution gives values for all possible assignments:

  P(*Weather*) = <0.72,0.1,0.08,0.1> (normalized, i.e., sums to 1)

- Joint probability distribution for a set of random variables gives the probability of every atomic event on those random variables
- **P**(*Weather*,*Cavity*) = a 4 × 2 matrix of values:

Weather =	sunny	rainy	cloudy	snow
Cavity = true	0.144	0.02	0.016	0.02
Cavity = false	0.576	0.08	0.064	0.08

- Every question about a domain can be answered by the joint distribution

Conditional probability

- Conditional or posterior probabilities

  e.g., P(cavity | toothache) = 0.8

- Indicates that if a patient is observed to have a toothache, and no other information is yet available, then the probability of the patient having a cavity will be 0.8. It is important to remember that P(A|B) can only be used when all we know is B.

- If we know more, e.g., cavity is also given, then we have

  P(cavity | toothache,catch) = 1

- New evidence may be irrelevant, allowing simplification, e.g.,

  P(cavity | toothache, sunny) = P(cavity | toothache) = 0.8

- This kind of inference, sanctioned by domain knowledge, is crucial

Conditional probability

- Definition of conditional probability:

  $P(a \mid b) = P(a \wedge b) / P(b)$ if $P(b) > 0$

- This equation can also be written as, Product rule gives an alternative formulation:

  $P(a \wedge b) = P(a \mid b) P(b) = P(b \mid a) P(a)$

- A general version holds for whole distributions, e.g.,

  $\mathbf{P}(Weather, Cavity) = \mathbf{P}(Weather \mid Cavity)\, \mathbf{P}(Cavity)$

Inference by Numeration

- Start with the joint probability distribution:

	toothache		*¬ toothache*	
	catch	*¬ catch*	*catch*	*¬ catch*
cavity	.108	.012	.072	.008
¬ cavity	.016	.064	.144	.576

- For any proposition φ, sum the atomic events where it is true: $P(\varphi) = \Sigma_{\omega:\omega \models \varphi} P(\omega)$

- Start with the joint probability distribution:

	toothache		*¬ toothache*	
	catch	*¬ catch*	*catch*	*¬ catch*
cavity	.108	.012	.072	.008
¬ cavity	.016	.064	.144	.576

- For any proposition φ, sum the atomic events where it is true: $P(\varphi) = \Sigma_{\omega:\omega \models \varphi} P(\omega)$

- $P(\textit{toothache}) = 0.108 + 0.012 + 0.016 + 0.064 = 0.2$

Inference by Numeration

- Start with the joint probability distribution:

	toothache		*¬ toothache*	
	catch	*¬ catch*	*catch*	*¬ catch*
cavity	.108	.012	.072	.008
¬ cavity	.016	.064	.144	.576

- For any proposition φ, sum the atomic events where it is true: $P(\varphi) = \Sigma_{\omega:\omega \models \varphi} P(\omega)$
- P(cavity ∨ *toothache*) = 0.108 + 0.012 + 0.072+0.008+0.016 + 0.064 = 0.28
- The above process is called **marginalization** or **summing out**

Inference by Numeration

- Start with the joint probability distribution:

	toothache		*¬ toothache*	
	catch	*¬ catch*	*catch*	*¬ catch*
cavity	.108	.012	.072	.008
¬ cavity	.016	.064	.144	.576

- Can also compute conditional probabilities:

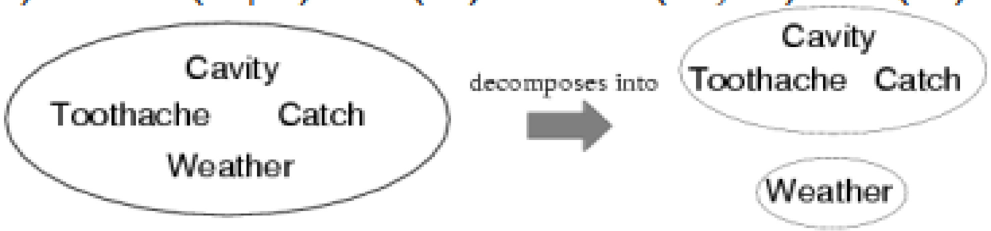
$$P(\neg cavity \mid toothache) = \frac{P(\neg cavity \wedge toothache)}{P(toothache)}$$

$$= \frac{0.016+0.064}{0.108 + 0.012 + 0.016 + 0.064}$$

$$= 0.4$$

Independence

- A and B are independent iff
  $\mathbf{P}(A|B) = \mathbf{P}(A)$ or $\mathbf{P}(B|A) = \mathbf{P}(B)$ or $\mathbf{P}(A, B) = \mathbf{P}(A)\,\mathbf{P}(B)$



$$\mathbf{P}(Toothache, Catch, Cavity, Weather)$$
$$= \mathbf{P}(Toothache, Catch, Cavity)\,\mathbf{P}(Weather)$$

- 32 entries(4x2x2x2) reduced to 12; for n independent biased coins, $O(2^n) \rightarrow O(n)$

- Absolute independence powerful but rare

- Dentistry is a large field with hundreds of variables, none of which are independent. What to do?
- $\mathbf{P}(Weather, Cavity)$ = a 4×2 matrix of values:

$Weather =$	sunny	rainy	cloudy	snow
$Cavity =$ true	0.144	0.02	0.016	0.02
$Cavity =$ false	0.576	0.08	0.064	0.08

Independence

- **P**(*Toothache, Cavity, Catch*) has $2^3 - 1 = 7$ independent entries
- If I have a cavity, the probability that the probe catches in it doesn't depend on whether I have a toothache:

  (1) **P**(*catch* | *toothache, cavity*) = **P**(*catch* | *cavity*)
- The same independence holds if I haven't got a cavity:

  (2) **P**(*catch* | *toothache,*¬*cavity*) = **P**(*catch* | ¬*cavity*)
- *Catch* is conditionally independent of *Toothache* given *Cavity*:

  P(*Catch* | *Toothache,Cavity*) = **P**(*Catch* | *Cavity*)
- In most cases, the use of conditional independence reduces the size of the representation of the joint distribution from exponential in *n* to linear in *n*.
- Conditional independence is our most basic and robust form of knowledge about uncertain environments.

Bayes' Rule

- Derive the probability of an event given another event
- Often useful for diagnosis:
  - If X are (observed) effects and Y are (hidden) causes,
  - We may have a model for how causes lead to effects (P(X | Y))
  - We may also have prior beliefs (based on experience) about the frequency of occurrence of effects (P(Y))
  - Which allows us to reason abductively from effects to causes (P(Y | X)).
- has gained importance recently due to advances in efficiency
  - more computational power available
  - better methods

Bayes' Rule

Bayes' theorem is stated mathematically as the following equation:[3]

$$\boxed{P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}}$$

where A and B are events and $P(B) \neq 0$.

- $P(A \mid B)$ is a conditional probability: the probability of event A occurring given that B is true. It is also called the posterior probability of A given B.
- $P(B \mid A)$ is also a conditional probability: the probability of event B occurring given that A is true. It can also be interpreted as the likelihood of A given a fixed B because $P(B \mid A) = L(A \mid B)$.
- $P(A)$ and $P(B)$ are the probabilities of observing A and B respectively without any given conditions; they are known as the marginal probability or prior probability.
- A and B must be different events.

Bayes' Rule ::Simple Example Problem

- A doctor knows that the disease meningitis is causes the patient to have a stiff neck, say, 50% of the time. The doctor also knows some unconditional facts: the prior probability of a patient having meningitis is 1/50,000, and the prior probability of any patient having a stiff neck is 1/20. Letting S be the proposition that the patient has a stiff neck and M be the proposition that the patient has meningitis.

$$P(S|M) = 0.5$$
$$P(M) = 1/50000$$
$$P(S) = 1/20$$
$$P(M|S) = \frac{P(S\backslash M)P(M)}{P(S)} = \frac{0.5 \text{ x } 1/50000}{1/20} = 0.0002$$

Bayes' Rule :: Example-1

- You are planning a picnic today, but the morning is cloudy
- Oh no! 50% of all rainy days start off cloudy!
- But cloudy mornings are common (about 40% of days start cloudy)
- And this is usually a dry month (only 3 of 30 days tend to be rainy, or 10%)
- **What is the chance of rain during the day?**

Bayes' Rule :: Example-1

- We will use Rain to mean rain during the day, and Cloud to mean cloudy morning.
- The chance of Rain given Cloud is written P(Rain|Cloud)
- So let's put that in the formula:
- P(Rain|Cloud) = P(Rain) P(Cloud|Rain)/P(Cloud)
- P(Rain) is Probability of Rain = 10%
- P(Cloud|Rain) is Probability of Cloud, given that Rain happens = 50%
- P(Cloud) is Probability of Cloud = 40%
- P(Rain|Cloud) = 0.1 x 0.5/0.4 = .125
- Or a 12.5% chance of rain. Not too bad, let's have a picnic!

Bayes' Rule :: Example-2

Class-labeled training tuples from the *AllElectronics* customer database.

RID	*age*	*income*	*student*	*credit_rating*	*Class: buys_computer*
1	youth	high	no	fair	no
2	youth	high	no	excellent	no
3	middle_aged	high	no	fair	yes
4	senior	medium	no	fair	yes
5	senior	low	yes	fair	yes
6	senior	low	yes	excellent	no
7	middle_aged	low	yes	excellent	yes
8	youth	medium	no	fair	no
9	youth	low	yes	fair	yes
10	senior	medium	yes	fair	yes
11	youth	medium	yes	excellent	yes
12	middle_aged	medium	no	excellent	yes
13	middle_aged	high	yes	fair	yes
14	senior	medium	no	excellent	no

The training data are in above Table. The data tuples are described by the attributes age, income, student, and credit rating. The class label attribute, buys computer, has two distinct values (namely, yes, no). Let $\mathbf{C_1}$ correspond to the class buys computer = **yes** and $\mathbf{C_2}$ correspond to buys computer = **no**. The tuple we wish to classify is **X = (age = youth, income = medium, student = yes, credit rating = fair)**

Bayes' Rule :: Example-2

- We need to maximize $P(X|C_i)P(C_i)$, for i = 1, 2. $P(C_i)$, the prior probability of each class, can be computed based on the training tuples:

  - **P(buys computer = yes) = 9/14 = 0.643**
  - **P(buys computer = no) = 5/14 = 0.357**

- To compute $P(\mathbf{X}|C_i)$, for i = 1, 2, we compute the following conditional probabilities:

Bayes' Rule :: Example-2

- P(age = youth | buys computer = yes) = ?
- P(age = youth | buys computer = no) = ?
- P(income = medium | buys computer = yes) = ?
- P(income = medium | buys computer = no) = ?
- P(student = yes | buys computer = yes) = ?
- P(student = yes | buys computer = no) = ?
- P(credit rating = fair | buys computer = yes) = ?
- P(credit rating = fair | buys computer = no) = ?

- **The above full join distribution can be written as**

$$P(Cause, Effect_1, \ldots. Effect_n) = P(Cause)\prod_i P(Effect_i \mid Cause)$$

Bayes' Rule :: Example-2

$$P(age = youth \mid buys\_computer = yes) = 2/9 = 0.222$$
$$P(age = youth \mid buys\_computer = no) = 3/5 = 0.600$$
$$P(income = medium \mid buys\_computer = yes) = 4/9 = 0.444$$
$$P(income = medium \mid buys\_computer = no) = 2/5 = 0.400$$
$$P(student = yes \mid buys\_computer = yes) = 6/9 = 0.667$$
$$P(student = yes \mid buys\_computer = no) = 1/5 = 0.200$$
$$P(credit\_rating = fair \mid buys\_computer = yes) = 6/9 = 0.667$$
$$P(credit\_rating = fair \mid buys\_computer = no) = 2/5 = 0.400$$

Bayes' Rule :: Example-2

$$P(Cause, Effect_1, \ldots. Effect_n) = P(Cause)\prod_i P(Effect_i \mid Cause)$$

$$P(X \mid buys\_computer = yes) = P(age = youth \mid buys\_computer = yes) \times$$
$$P(income = medium \mid buys\_computer = yes) \times$$
$$P(student = yes \mid buys\_computer = yes) \times$$
$$P(credit\_rating = fair \mid buys\_computer = yes)$$

= **0.222 X 0.444 X 0.667 X 0.667** = **0.044**

Similarly,

P(X | buys computer = no)

= **0.600 X 0.400 X 0.200 X 0.400** = **0.019**

To find the class, C_i, that maximizes $P(X_j C_i)P(C_i)$, we compute

P(X | buys computer = yes)P(buys computer = yes) = ?

P(X | buys computer = no)P(buys computer = no) = ?

Bayes' Rule :: Example-2

P(X | buys computer = yes)P(buys computer = yes) =0.044 X 0.643 = 0.028

P(X | buys computer = no)P(buys **computer = no)** =**0.0918 X 0.357 = 0.007**

Therefore, the naïve Bayesian classifier predicts **buys computer = yes** for tuple **X**.

Bayes' Rule :: Practice Problem - 1

Name	Give Birth	Can Fly	Live in Water	Have Legs	Class
human	yes	no	no	yes	mammals
python	no	no	no	no	non-mammals
salmon	no	no	yes	no	non-mammals
whale	yes	no	yes	no	mammals
frog	no	no	sometimes	yes	non-mammals
komodo	no	no	no	yes	non-mammals
bat	yes	yes	no	yes	mammals
pigeon	no	yes	no	yes	non-mammals
cat	yes	no	no	yes	mammals
leopard shark	yes	no	yes	no	non-mammals
turtle	no	no	sometimes	yes	non-mammals
penguin	no	no	sometimes	yes	non-mammals
porcupine	yes	no	no	yes	mammals
eel	no	no	yes	no	non-mammals
salamander	no	no	sometimes	yes	non-mammals
gila monster	no	no	no	yes	non-mammals
platypus	no	no	no	yes	mammals
owl	no	yes	no	yes	non-mammals
dolphin	yes	no	yes	no	mammals
eagle	no	yes	no	yes	non-mammals

Give Birth	Can Fly	Live in Water	Have Legs	Class
yes	no	yes	no	?

Bayes' Rule :: Practice Problem – 2

Example No.	Color	Type	Origin	Stolen?
1	Red	Sports	Domestic	Yes
2	Red	Sports	Domestic	No
3	Red	Sports	Domestic	Yes
4	Yellow	Sports	Domestic	No
5	Yellow	Sports	Imported	Yes
6	Yellow	SUV	Imported	No
7	Yellow	SUV	Imported	Yes
8	Yellow	SUV	Domestic	No
9	Red	SUV	Imported	No
10	Red	Sports	Imported	Yes

- **X = (color= red, Type = SUV, Origin = Domestic)**

Bayes' Rule :: Practice Problem - 1

A: attributes

M: mammals

N: non-mammals

$$P(A|M) = \frac{6}{7} \times \frac{6}{7} \times \frac{2}{7} \times \frac{2}{7} = 0.06$$

$$P(A|N) = \frac{1}{13} \times \frac{10}{13} \times \frac{3}{13} \times \frac{4}{13} = 0.0042$$

$$P(A|M)P(M) = 0.06 \times \frac{7}{20} = 0.021$$

$$P(A|N)P(N) = 0.004 \times \frac{13}{20} = 0.0027$$

P(A|M)P(M) >
P(A|N)P(N)

Bayes' Rule :: Practice Problem - 2

- We have P(Yes) = .5 and P(No) = .5,
- For v = Y es, we have P(Yes) * P(Red | Yes) * P(SUV | Yes) * P(Domestic|Yes) = .5 * .56 * .31 * .43 = .037
- For v = No, we have P(No) * P(Red | No) * P(SUV | No) * P (Domestic | No) = .5 * .43 * .56 * .56 = .069
- Since 0.069 > 0.037, our example gets classified as ’NO
- **X = (color= red, Type = SUV, Origin = Domestic) => Stolen=NO**

Bayes' Rule :: Practice Problem - 3

Baba’s favourite subject is science and his favourite sport is cricket. 65% of the time he studies science. The probability of him playing cricket is 0.8. The probability of him, studying science given that he plays cricket is 0.5. What is the probability of playing cricket given he studied science?

$P(C|S) = (P(S|C) * P(C)) / P(S)$

$P(C|S) = (0.5 * 0.8) / 0.65 = 0.6154$

There is a probability of 61.54% of Baba’s playing cricket given he studied science.

Bayes' Rule :: Practice Problem - 4

In a standard two of St. John school, 30% of the children have grey eyes, 50% of them have blue and the other 20%'s eyes are in other colors. One day they play a game together. In the first run, 65% of the grey eye ones, 82% of the blue eyed ones and 50% of the children with other eye color were selected. Now, if a child is selected randomly from the class, and we know that he/she was not in the first game, what is the probability that the child has blue eyes?

Bayes' Rule :: Practice Problem - 4

The general form of Bayes' rule is

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_j P(B|A_j)P(A_j)}$$

Let's say B = blue, G = grey and O = "Other color" and NR = "not selected for the first run"

$$P(B \mid NR) = \frac{P(NR \mid B)P(B)}{P(G)P(NR \mid G) + P(B)P(NR \mid B) + P(O)P(NR \mid O)}$$

On substituting values

$$P(B \mid NR) = \frac{0.5 \cdot (1 - 0.82)}{(0.3 \cdot (1 - 0.65)) + (0.5 \cdot (1 - 0.82)) + (0.2 \cdot (1 - 0.5))}$$

$$P(B \mid NR) = 0.305$$

PROBABILISTIC REASONING

- ❖ Probabilistic reasoning is a method of representation of knowledge where the concept of probability is applied to indicate the uncertainty in knowledge.

Probabilistic reasoning is used in AI:

- ❖ When we are unsure of the predicates
- ❖ When the possibilities of predicates become too large to list down
- ❖ When it is known that an error occurs during an experiment

BAYESIAN BELIEF NETWORKS (BBN)

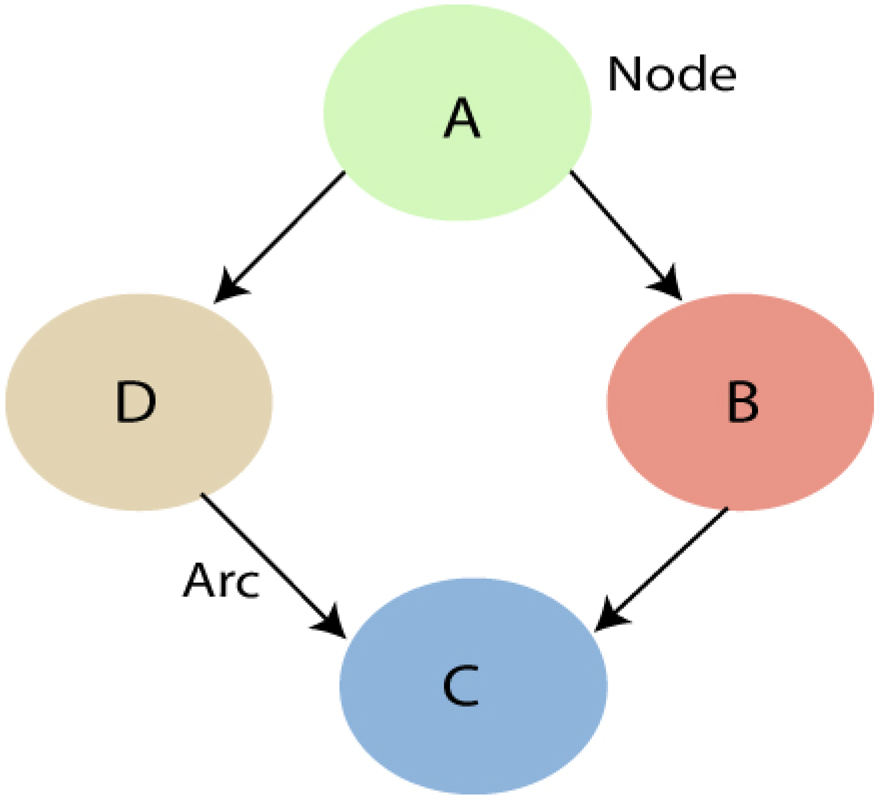
- Explicit representation of independence and conditional independence relationships in simplified probabilistic form is known as Bayesian Belief Networks
- It is a directed acyclic graph in which each node in annotated with quantitative probability information
- Also called as:
  - Belief networks
  - Probabilistic networks
  - Casual networks
  - Knowledge map
  - Bayesian Network

BAYESIAN BELIEF NETWORKS (BBN)

- We can define a Bayesian network as:

"A Bayesian network is a probabilistic graphical model which represents a set of variables and their conditional dependencies using a directed acyclic graph."

- Real world applications are probabilistic in nature, and to represent the relationship between multiple events, we need a Bayesian network. It can also be used in various tasks including prediction, anomaly detection, diagnostics, automated insight, reasoning, time series prediction, and decision making under uncertainty.
- Bayesian Network can be used for building models from data and experts opinions, and it consists of two parts:
  - Directed Acyclic Graph
  - Table of conditional probabilities.



In the above diagram, A, B, C, and D are random variables represented by the nodes of the network graph.
If we are considering node B, which is connected with node A by a directed arrow, then node A is called the parent of Node B.
Node C is independent of node A.

A Bayesian network graph is made up of nodes and Arcs (directed links)

Each **node** corresponds to the random variables, and a variable can be **continuous** or **discrete**.

Arc or directed arrows represent the causal relationship or conditional probabilities between random variables. These directed links or arrows connect the pair of nodes in the graph.
These links represent that one node directly influence the other node, and if there is no directed link that means that nodes are independent with each other

BAYESIAN BELIEF NETWORKS (BBN)

- Each node in the Bayesian network has condition probability distribution $P(X_i \mid Parent(X_i))$, which determines the effect of the parent on that node.
- Bayesian network is based on Joint probability distribution and conditional probability. So let's first understand the joint probability distribution:
- If we have variables $x_1, x_2, x_3, \ldots, x_n$, then the probabilities of a different combination of $x_1, x_2, x_3 \ldots x_n$, are known as Joint probability distribution.
- $P[x_1, x_2, x_3, \ldots, x_n]$, it can be written as the following way in terms of the joint probability distribution.

$$= P[x_1 \mid x_2, x_3, \ldots, x_n] P[x_2, x_3, \ldots, x_n]$$

$$= P[x_1 \mid x_2, x_3, \ldots, x_n] P[x_2 \mid x_3, \ldots, x_n] \ldots P[x_{n-1} \mid x_n] P[x_n].$$

BAYESIAN BELIEF NETWORKS (BBN)

- In general for each variable X_i, we can write the equation as:

$$P(X_i|Xi_{-1},........., X_1) = P(X_i\ |Parents(X_i\ ))$$

Example:

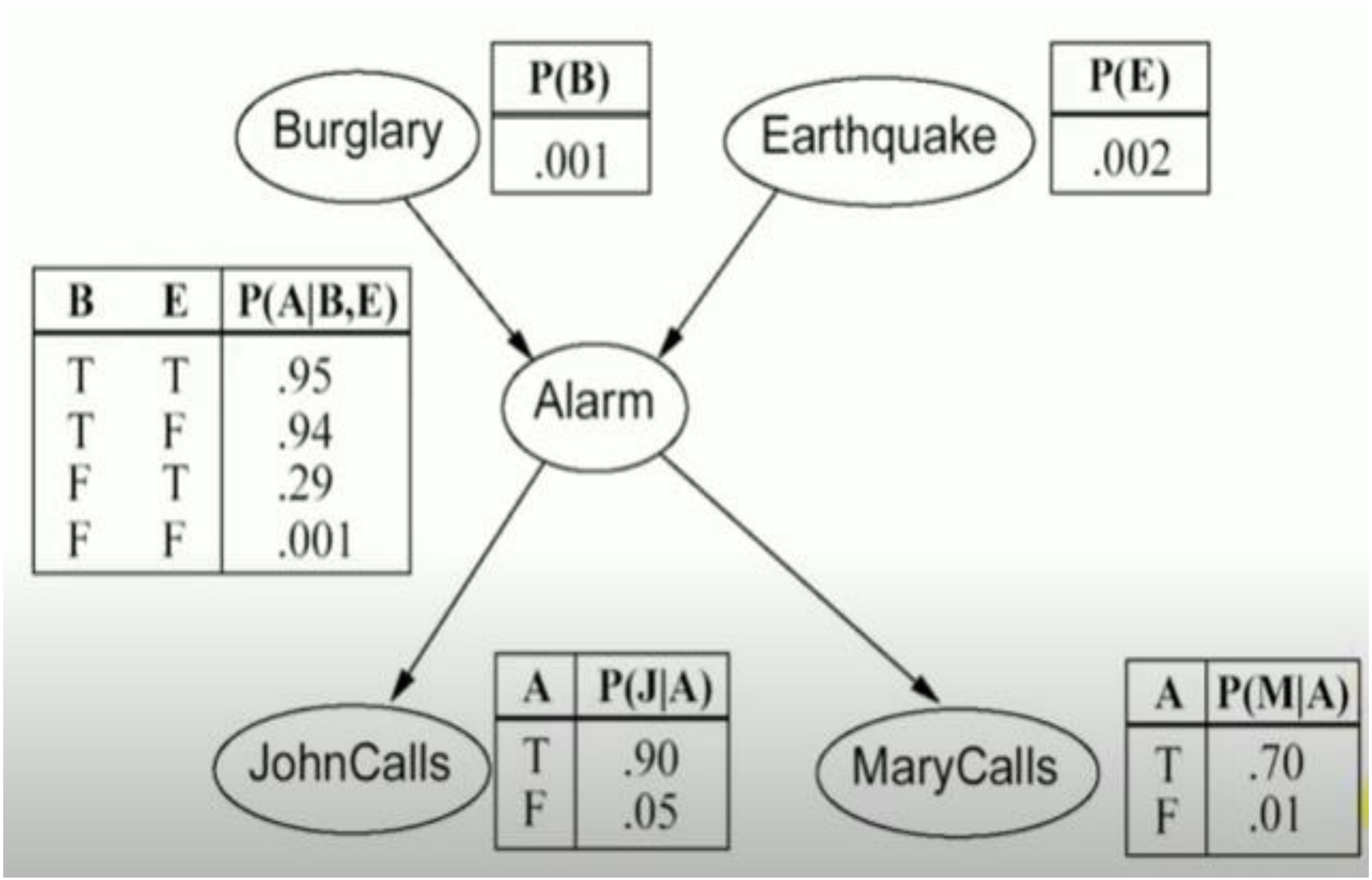
Harry installed a new burglar alarm at his home to detect burglary. The alarm reliably responds at detecting a burglary but also responds for minor earthquakes. Harry has two neighbours John and Mary, who have taken a responsibility to inform Harry at work when they hear the alarm. John always calls Harry when he hears the alarm, but sometimes he got confused with the phone ringing and calls at that time too. On the other hand, Mary likes to listen to high music, so sometimes she misses to hear the alarm. Here we would like to compute the probability of Burglary Alarm.

BAYESIAN BELIEF NETWORKS (BBN)

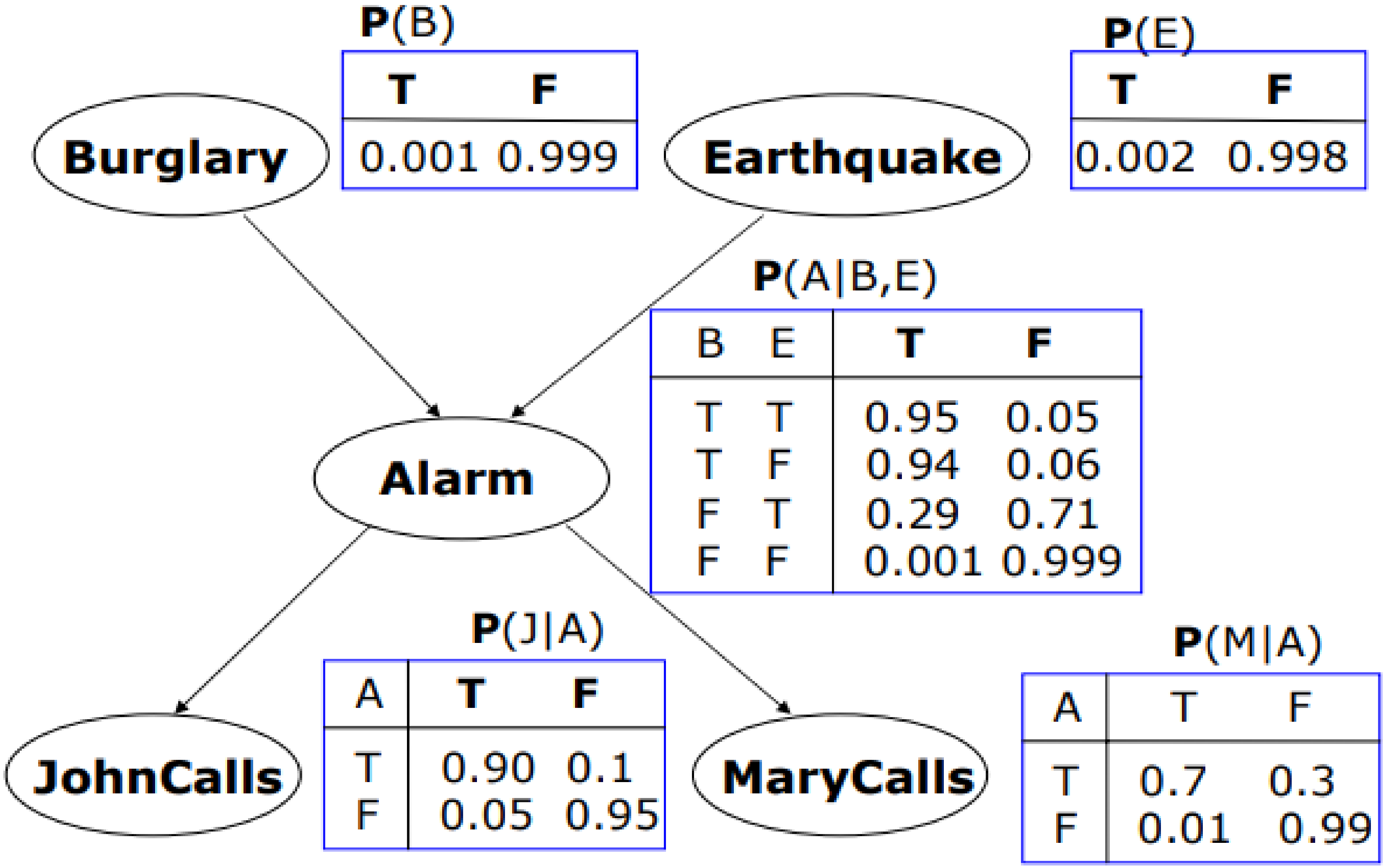
Problem:

Calculate the probability that alarm has sounded, but there is neither a burglary, nor an earthquake occurred, and John and Mary both called the Harry.

BAYESIAN BELIEF NETWORKS (BBN)



BAYESIAN BELIEF NETWORKS (BBN)



P(B)

T	F
0.001	0.999

P(E)

T	F
0.002	0.998

P(A|B,E)

B	E	T	F
T	T	0.95	0.05
T	F	0.94	0.06
F	T	0.29	0.71
F	F	0.001	0.999

P(J|A)

A	T	F
T	0.90	0.1
F	0.05	0.95

P(M|A)

A	T	F
T	0.7	0.3
F	0.01	0.99

BAYESIAN BELIEF NETWORKS (BBN)

Solution:

- ❖ The Bayesian network for the above problem is given earlier. The network structure is showing that burglary and earthquake is the parent node of the alarm and directly affecting the probability of alarm's going off, but John and Mary’s calls depend on alarm probability.
- ❖ The network is representing that our assumptions do not directly perceive the burglary and also do not notice the minor earthquake, and they also not confer before calling.
- ❖ The conditional distributions for each node are given as conditional probabilities table or CPT.

BAYESIAN BELIEF NETWORKS (BBN)

Solution:

- Each row in the CPT must be sum to 1 because all the entries in the table represent an exhaustive set of cases for the variable.
- In CPT, a boolean variable with k boolean parents contains 2K probabilities. Hence, if there are two parents, then CPT will contain 4 probability values
- List of all events occurring in this network:
  - Burglary (B)
  - Earthquake(E)
  - Alarm(A)
  - John Calls(J)
  - Mary Calls(M)

BAYESIAN BELIEF NETWORKS (BBN)

- We can write the events of problem statement in the form of probability: P[J, M, A, B, E], can rewrite the above probability statement using joint probability distribution:

$$
\begin{aligned}
P[J, M, A, B, E] &= P[J \mid M, A, B, E] . P[M, A, B, E] \\
&= P[J \mid M, A, B, E] . P[M \mid A, B, E] . P[A, B, E] \\
&= P[J \mid A] . P[M \mid A, B, E] . P[A, B, E] \\
&= P[J \mid A] . P[M \mid A] . P[A \mid B, E] . P[B, E] \\
&= P[J \mid A] . P[M \mid A] . P[A \mid B, E] . P[B] . P[E]
\end{aligned}
$$

BAYESIAN BELIEF NETWORKS (BBN)

- From the formula of joint distribution, we can write the problem statement in the form of probability distribution:

  $P(J, M, A, \neg B, \neg E)$

  $= P(J|A) * P(M|A) * P(A|\neg B \wedge \neg E) * P(\neg B) * P(\neg E).$

  $= 0.91 * 0.75 * 0.001 * 0.998 * 0.999$

  $= 0.00068045.$

- Hence, a Bayesian network can answer any query about the domain by using Joint distribution.

BAYESIAN BELIEF NETWORKS (BBN)

❖ What is the probability that John Calls?

Solution

P(J)

$= P(J|A)*P(A) + P(J|\neg A)*P(\neg A)$

$= P(J|A)*\{P(A|B,E)*P(B,E)+P(A|\neg B,E)*P(\neg B,E)+P(A|B,\neg E)*P(B,\neg E)+P(A|\neg B,\neg E)*P(\neg B,\neg E)\}$

$+ P(J|\neg A)*\{P(\neg A|B,E)*P(B,E)+P(\neg A|\neg B,E)*P(\neg B,E)+P(\neg A|B,\neg E)*P(B,\neg E)+P(\neg A|\neg B,\neg E)*P(\neg B,\neg E)\}$

$= 0.90 * 0.00252 + 0.05 * 0.9974$

$= 0.0521$

BAYESIAN BELIEF NETWORKS (BBN)

- Suppose, we are given for the evidence variables $E_1,\ldots,E_m$, their values $e_1,\ldots,e_m$, and we want to predict whether the query variable X has the value x or not.
- For this we compute and compare the following:

$$P(x \mid e_1,\ldots,e_m) = \frac{P(x, e_1,\ldots,e_m)}{P(e_1,\ldots,e_m)} = \alpha P(x, e_1,\ldots,e_m)$$

$$P(\neg x \mid e_1,\ldots,e_m) = \frac{P(\neg x, e_1,\ldots,e_m)}{P(e_1,\ldots,e_m)} = \alpha P(\neg x, e_1,\ldots,e_m)$$

$$\propto = \frac{1}{(P(x,e_1,\ldots em) + P(\neg x,e_1,\ldots em))}$$

BAYESIAN BELIEF NETWORKS (BBN)

- What is the probability that there is a burglary given that John and Mary Calls?

Solution

$$P(b \mid j,m) = \alpha P(b) \textstyle\sum_a P(j|a)\ P(m|a) \sum_e P(a|b,e)P(e)$$

$$= \alpha P(b) \textstyle\sum_a P(j|a)\ P(m|a)\ \{P(a|b,e)P(e) + P(a|b,\neg e)P(\neg e)\}$$

$$= \alpha P(b)\ [\ P(j|a)P(m|a)\ \{P(a|b,e)P(e) + P(a|b,\neg e)P(\neg e)\ \}$$

$$+ P(j|\neg a)P(m|\neg a)\ \{P(\neg a|b,e)P(e) + P(\neg a|b,\neg e)P(\neg e)\}]$$

$$= \alpha * .001*(.9*.7*(.95*.002 + .94*.998) + .05*.01*(.05*.002 + .71*.998)\ )$$

$= \alpha *$ **0.00059**

BAYESIAN BELIEF NETWORKS (BBN)

❖ What is the probability that there is a burglary given that John and Mary Calls?

Solution

$$P(\neg b \mid j,m) = \alpha\, P(\neg b) \textstyle\sum_a P(j|a)P(m|a) \textstyle\sum_e P(a|\neg b,e)P(e)$$

$$= \alpha\, P(\neg b) \textstyle\sum_a P(j|a)P(m|a) \; \{ P(a|\neg b,e)P(e) + P(a|\neg b,\neg e)P(\neg e)\}$$

$$= \alpha\, P(\neg b) \, [\, P(j|a)P(m|a) \; \{ P(a|\neg b,e)P(e) + P(a|\neg b,\neg e)P(\neg e) \}$$

$$+ P(j|\neg a)P(m|\neg a) \; \{ P(\neg a|\neg b,e)P(e) + P(\neg a|\neg b,\neg e)P(\neg e) \}]$$

$$= \alpha * .999*(.9*.7*(.29*.002 + .001*.998) + .05*.01*(.71*.002 + .999*.998))$$

$$= \alpha * .0015$$

BAYESIAN BELIEF NETWORKS (BBN)

- What is the probability that there is a burglary given that John and Mary Calls?

Solution

$$\alpha = \frac{1}{(P(b,j,m) + P(\neg b,j,m))}$$

$$\alpha = \frac{1}{(.00059 + .0015)}$$

$$= 478.5$$

$$P(b \mid j, m) = \propto * P(b, j, m)$$

$$= 478.5 * .00059$$

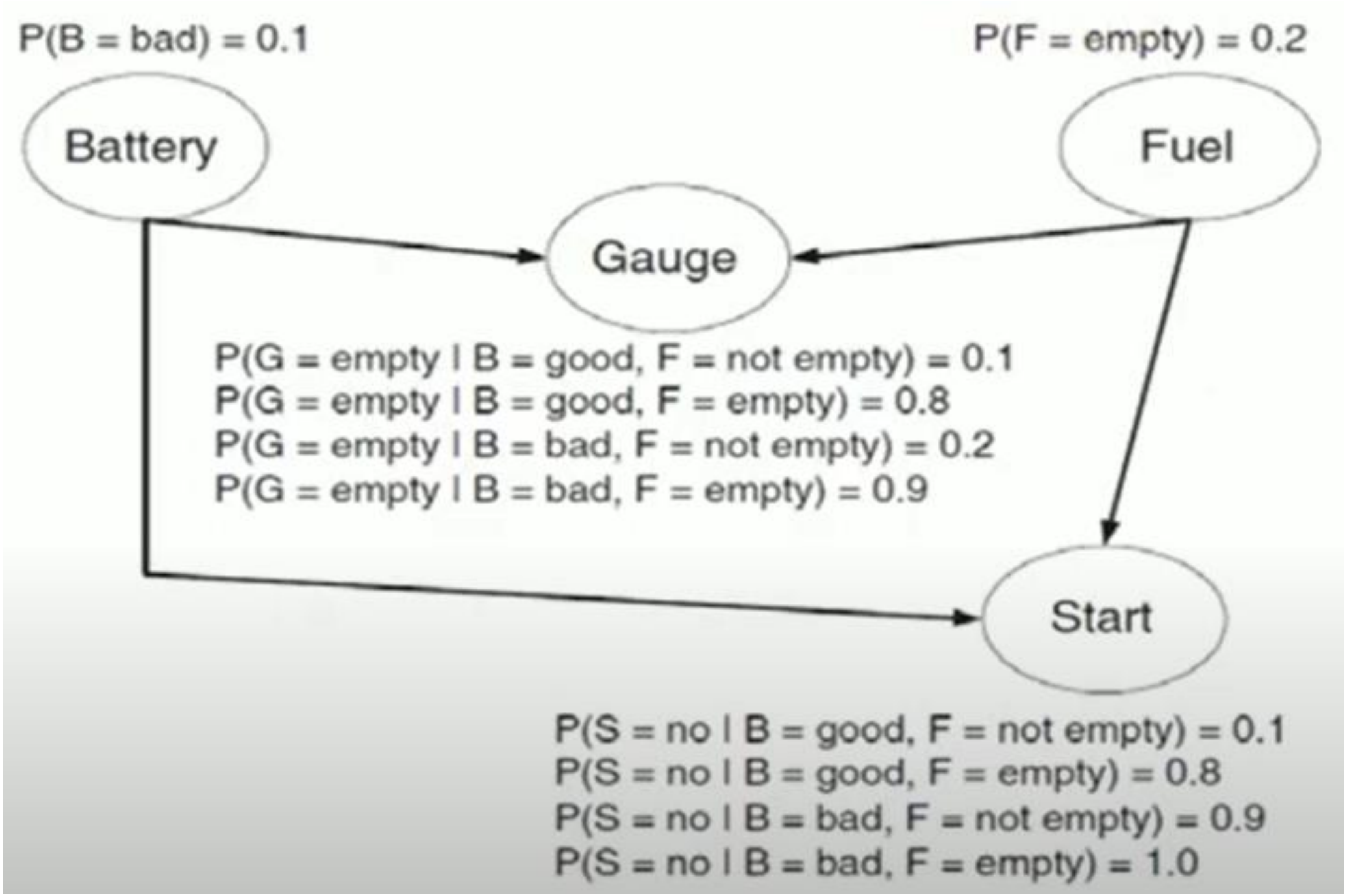
$$= 0.28$$

$$P(\neg b \mid j,m) = \propto * P(\neg b, j, m)$$

$$= 478.5 * .0015$$

$$= 0.72$$

BAYESIAN BELIEF NETWORKS (BBN)



BAYESIAN BELIEF NETWORKS (BBN)

1. P(B=Good, F=Empty, G=Empty, S=Yes)

$$P(B = good, F = empty, G = empty, S = yes)$$

$$= P(B = good) \times P(F = empty) \times P(G = empty | B = good, F = empty)$$

$$\times P(S = yes | B = good, F = empty)$$

$$= 0.9 \times 0.2 \times 0.8 \times 0.2 = 0.0288.$$

BAYESIAN BELIEF NETWORKS (BBN)

2. P(B=Bad, F=Empty, G=Not Empty, S=No)

$$
\begin{aligned}
& P(B = bad, F = empty, G = not\ empty, S = no) \\
= \ & P(B = bad) \times P(F = empty) \times P(G = not\ empty \mid B = bad, F = empty) \\
& \times P(S = no \mid B = bad, F = empty) \\
= \ & 0.1 \times 0.2 \times 0.1 \times 1.0 = 0.002.
\end{aligned}
$$

BAYESIAN BELIEF NETWORKS (BBN)

- Given the battery is bad, compute the probability that the car will start?

Solution

P(S =Yes|B=Bad)

$$= P(S = Yes|B = bad, F = Empty) * P(F = Empty) +$$
$$P(S = Yes|B = bad, F = Non - Empty) * P(F = No - Empty)$$
$$= (0 * 0.2) + (0.1 * 0.8)$$
$$= 0.08$$

Decision Theory

- Decision theory is concerned with the reasoning underlying an agent's choices, whether this is a mundane choice between taking the bus or getting a taxi, or a more far-reaching choice about whether to pursue a demanding political career.

General Tasks	Formal Tasks	Expert Tasks
Perception -	**Games**	**Engineering**
Vision and Speech	Chess, Backgammon, Checkers	Design, Fault finding, Manufacturing Planning
Natural Language -	**Mathematics**	**Scientific Analysis**
Understanding, Generation, Translation	Geometry, Logic, Integral Calculus, Proving Properties of Program	**Medical Diagnosis**
Common Sense Reasoning and Robot Control		**Financial Analysis**

Utility Theory

- ❖ The core idea behind utility theory is that every possible action or state within a given model can be described with a single, uniform value.
- ❖ This value, usually referred to as utility, describes the usefulness of that action within the given context.
- ❖ Utility theory has been used in game theory, economics, and numerous other fields.
- ❖ Value is a measurable quantity (such as the prices above).
- ❖ Utility measures how much we desire something. This can change based on personality or the context of the situation.

Utility Theory

- The main idea of utility theory is really simple: an agent's preferences over possible outcomes can be captured by a function that maps these outcomes to a real number; the higher the number the more that agent likes that outcome. The function is called a utility function.
- For example, we could say that my utility for owning various items is:

  u(apple) = 10

  u(orange) = 12

  u(basketball) = 4

  u(macbookpro) = 45

Decision Network

- Decision networks are generalized Bayes Belief Networks with chance nodes, decision nodes, and a utility node. These networks are used to make decisions under uncertainty.
- In decision networks, utility of a decision, maximum expected utility, value of information and perfect value of information are important computations that affect the decision making process.
  - Utility
  - Expected utility
  - Maximum expected utility (MEU)
  - Value of information
  - Value of perfect information

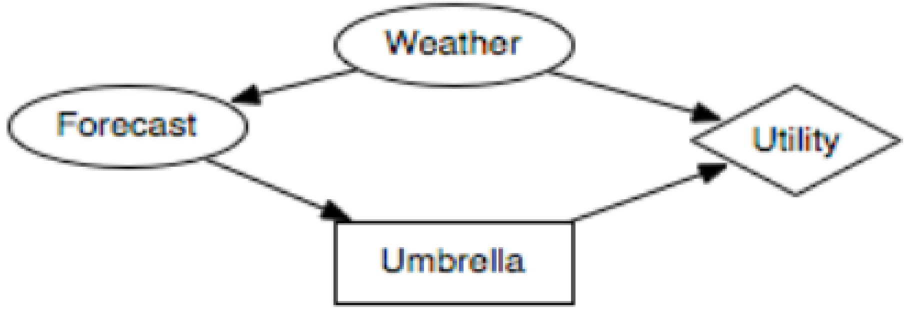
Decision Network

- Utility of a decision is a function of random variables of the decision problem.
- Expected utility of each decision is calculated as a sum of probabilities of the decision multiplied by the utility value of that decision.
- Maximum expected utility (MEU) is a decision with high expected utility from the available decisions to take.
- Value of information is the price paid to obtain a piece of information. The information obtained may or may not be useful in arriving at a correct decision.
- Value of perfect information is the price paid to obtain a piece of information that helps in making the right decision.

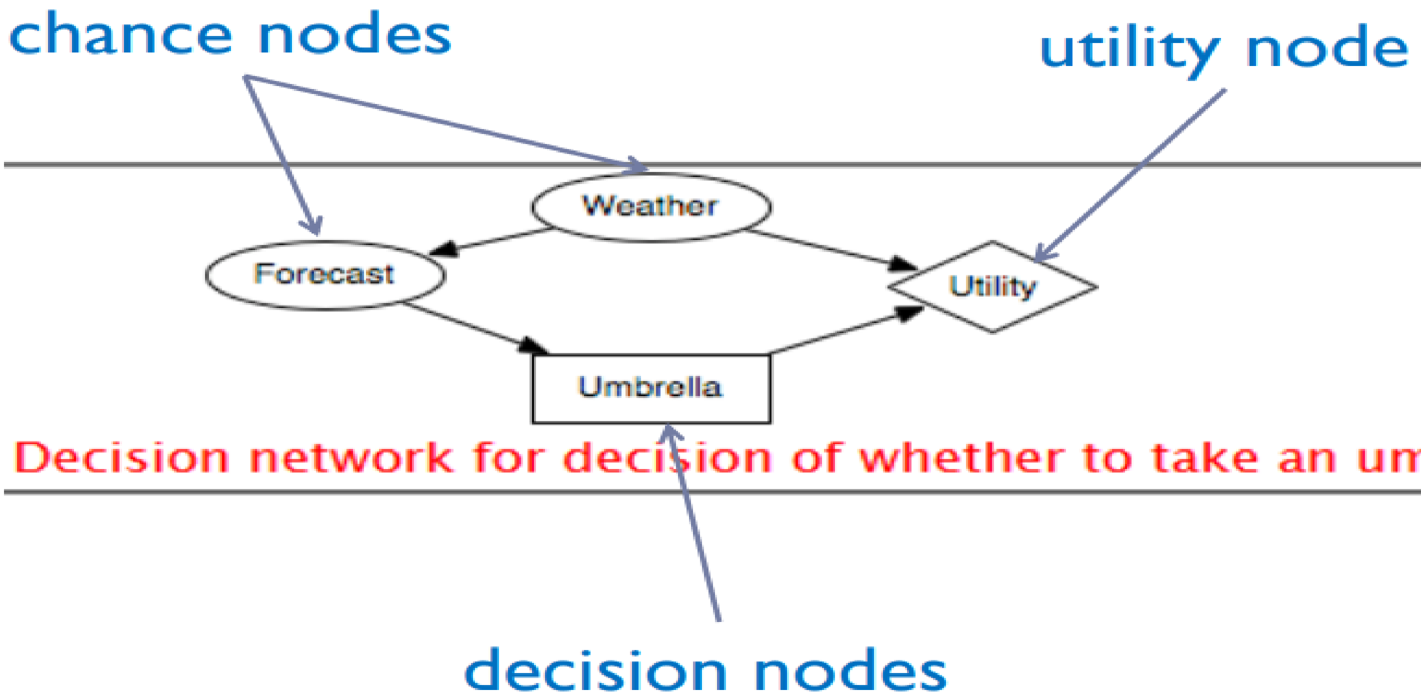
Decision Network

- Both value of information and value of perfect information may be used in a sequential decision problem where a decision impacts subsequent decisions.
- An AI agent uses decision networks to make decisions under uncertainty by:
  - Assigning conditional probabilities to each random variable (chance nodes).
  - Assigning utility to a decision.
  - Rating the utility of each possible decision.
  - Computing expected utility of each possible decision and arriving at a decision with maximum expected utility (MEU).
  - Computing value of information and value of perfect information in case of sequential decisions.

Decision Network :: Example



Decision network for decision of whether to take an umbrella

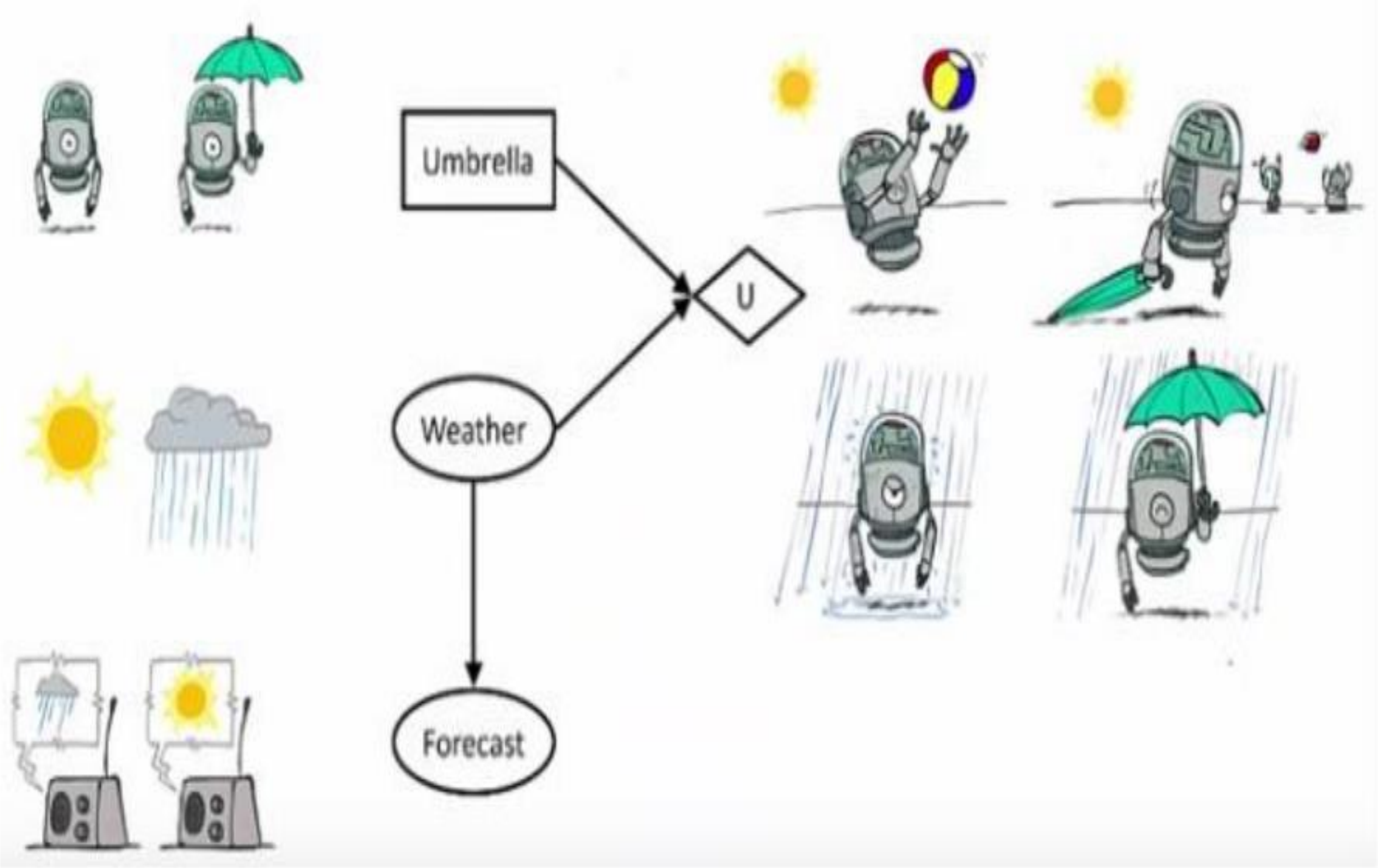


Decision network for decision of whether to take an umbrella

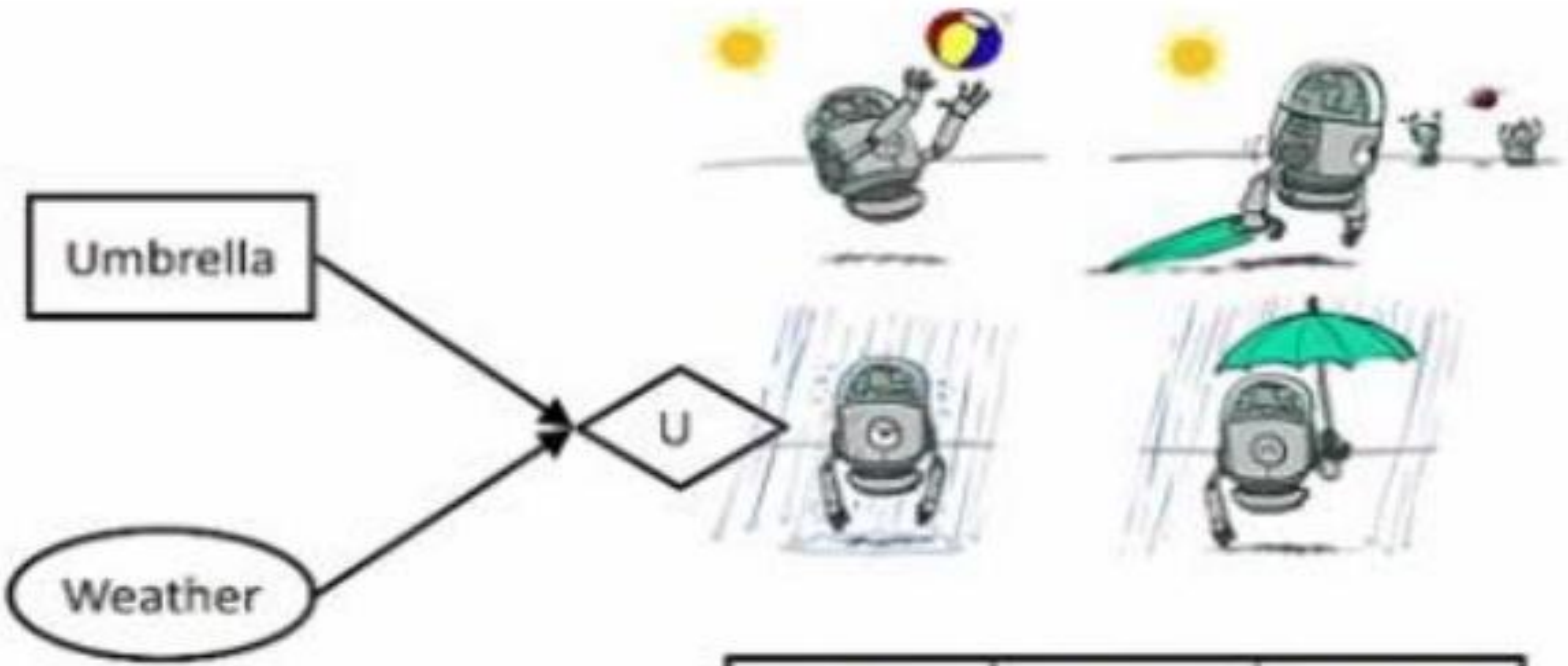
Decision Network

- Arcs coming into decision nodes (or action nodes) represent the information that will be available when the decision is made.
- Arcs coming into chance nodes represents probabilistic dependence.
- Arcs coming into the utility node represent what the utility depends on.

Decision Network :: Example



Decision Network :: Example



W	P(W)
sun	0.7
rain	0.3

A	W	U(A,W)
leave	sun	100
leave	rain	0
take	sun	20
take	rain	70

Decision Network :: Example

Umbrella = leave

$$\mathrm{EU}(\mathrm{leave}) = \sum_{w} P(w)U(\mathrm{leave}, w)$$

$$= 0.7 \cdot 100 + 0.3 \cdot 0 = 70$$

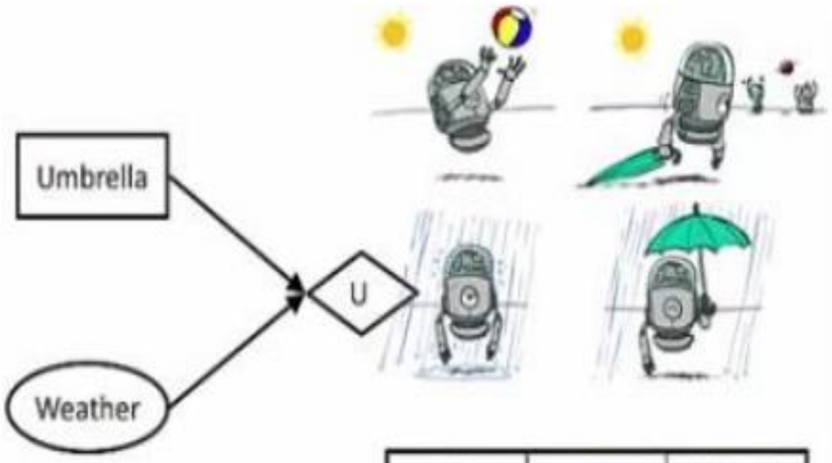
Umbrella = take

$$\mathrm{EU}(\mathrm{take}) = \sum_{w} P(w)U(\mathrm{take}, w)$$

$$= 0.7 \cdot 20 + 0.3 \cdot 70 = 35$$

Optimal decision = leave

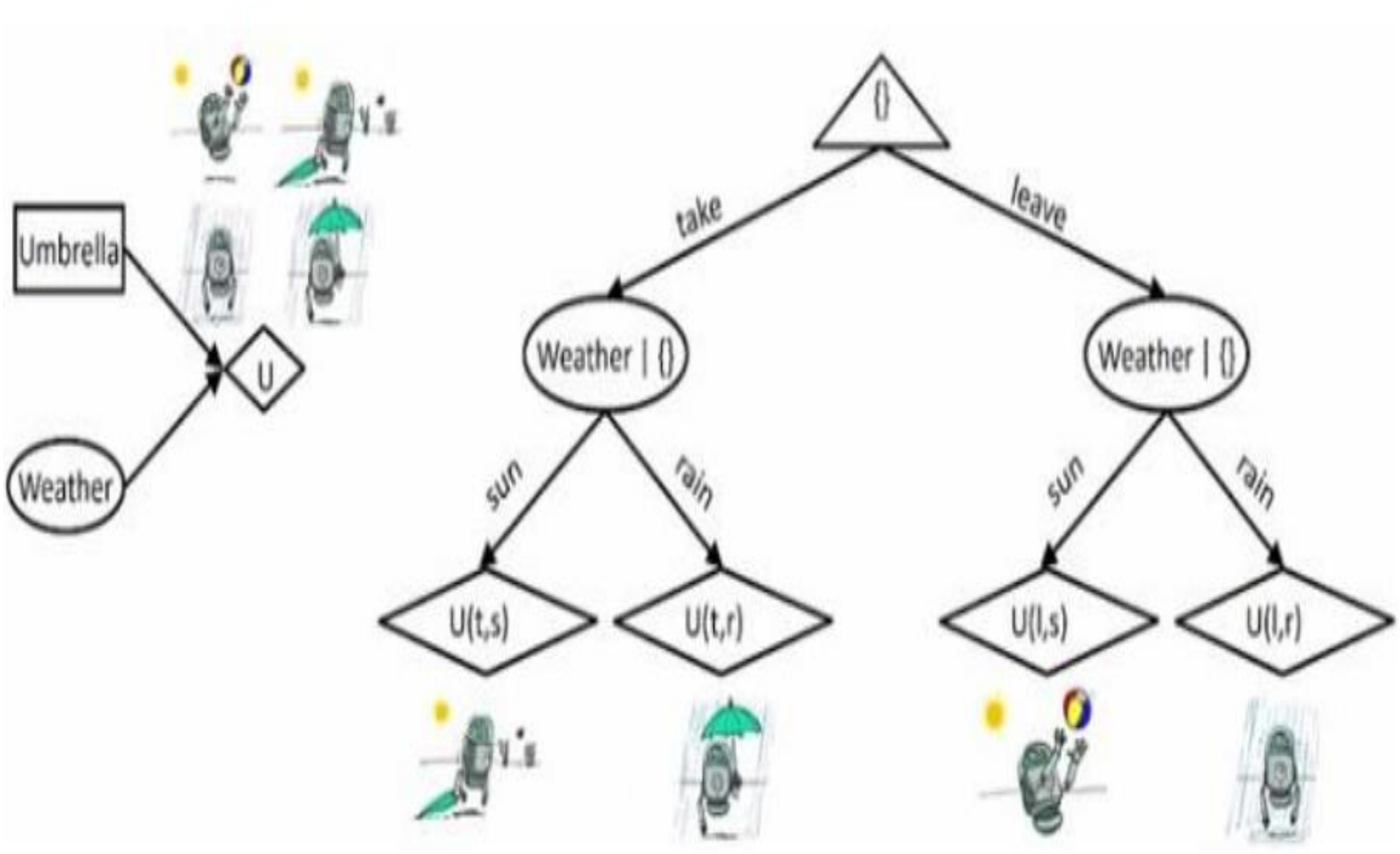
$$\mathrm{MEU}(\phi) = \max_{a} \mathrm{EU}(a) = 70$$



W	P(W)
sun	0.7
rain	0.3

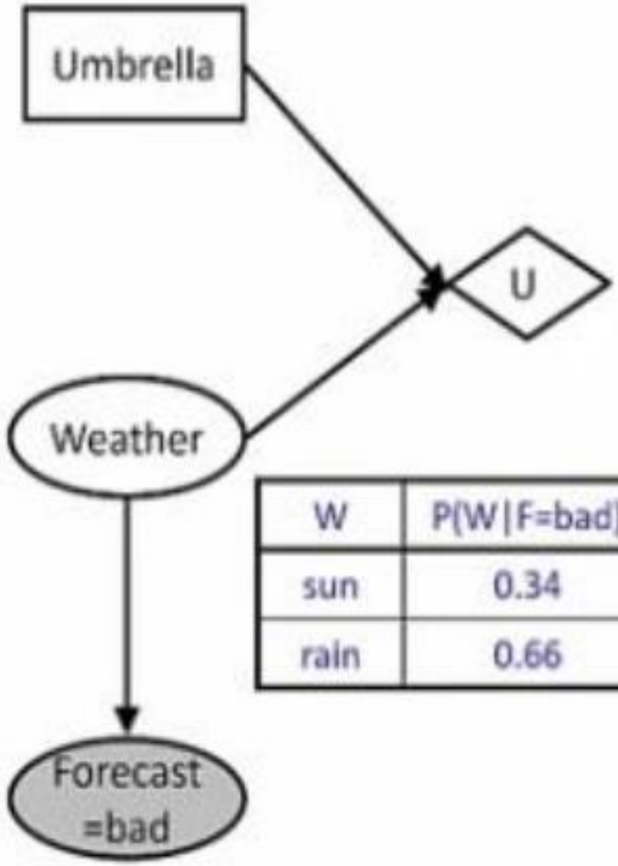
A	W	U(A,W)
leave	sun	100
leave	rain	0
take	sun	20
take	rain	70

Decision Network :: Example



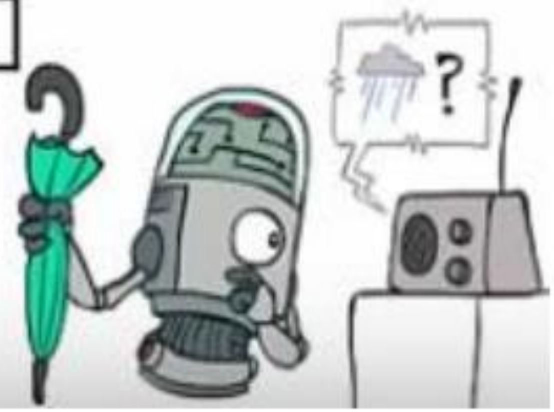
Decision Network :: Example

Complex Example



A	W	U(A,W)
leave	sun	100
leave	rain	0
take	sun	20
take	rain	70

W	P(W|F=bad)
sun	0.34
rain	0.66



Decision Network :: Example

Umbrella = leave

$$\mathrm{EU}(\text{leave}|\text{bad}) = \sum_{w} P(w|\text{bad})U(\text{leave}, w)$$

$$= 0.34 \cdot 100 + 0.66 \cdot 0 = 34$$

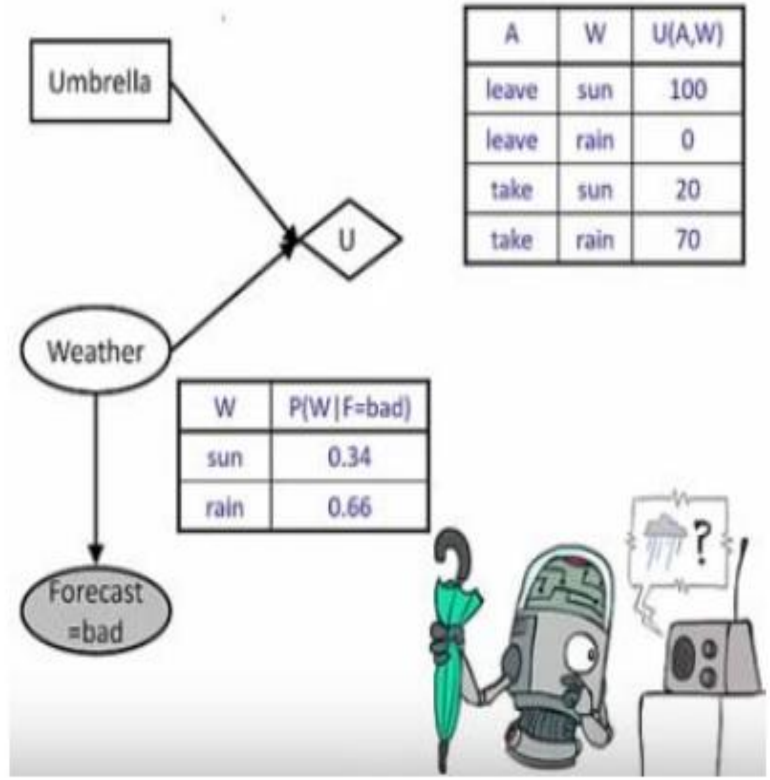
Umbrella = take

$$\mathrm{EU}(\text{take}|\text{bad}) = \sum_{w} P(w|\text{bad})U(\text{take}, w)$$

$$= 0.34 \cdot 20 + 0.66 \cdot 70 = 53$$

Optimal decision = take

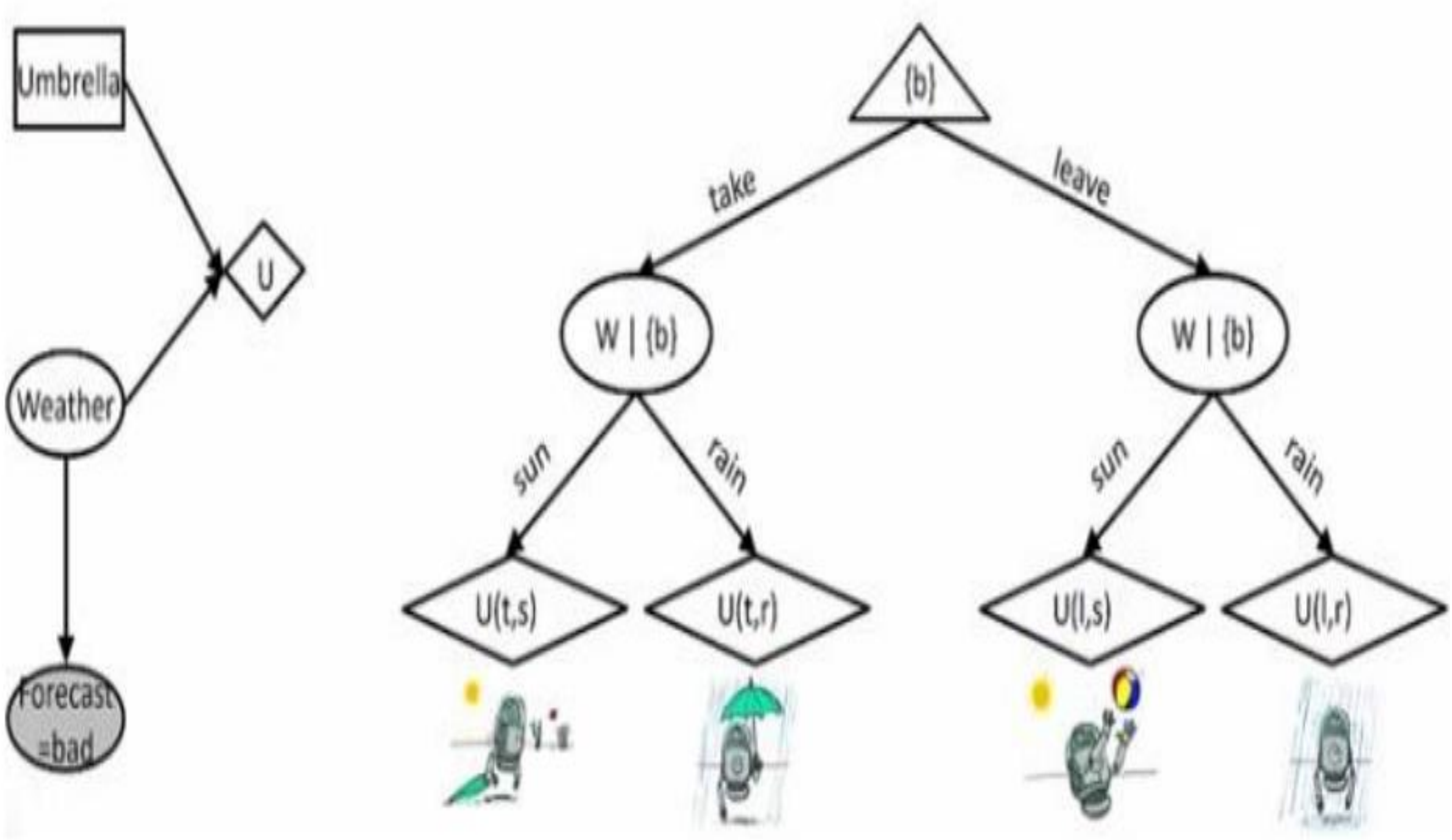
$$\mathrm{MEU}(F = \text{bad}) = \max_{a} \mathrm{EU}(a|\text{bad}) = 53$$



A	W	U(A,W)
leave	sun	100
leave	rain	0
take	sun	20
take	rain	70

W	P(W|F=bad)
sun	0.34
rain	0.66

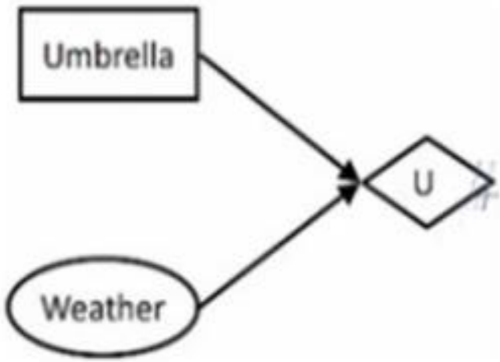
Decision Network :: Example



Decision Network :: Example – Value of Information

Optimal decision = leave

$$\text{MEU}(\emptyset) = \max_a \text{EU}(a) = 70$$

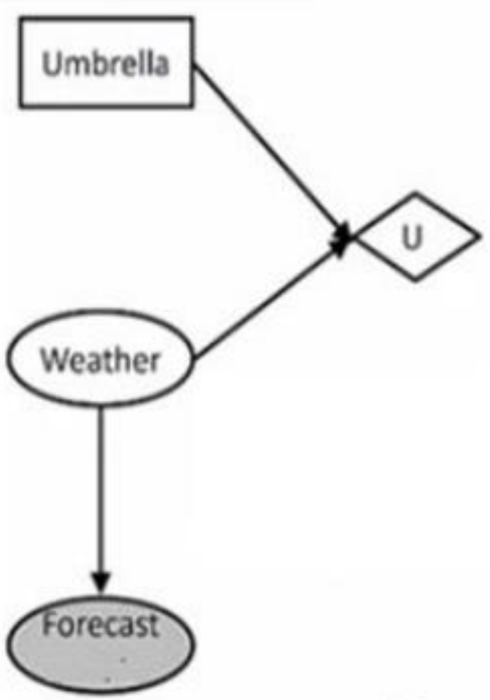


Optimal decision = take

$$\text{MEU}(F = \text{bad}) = \max_a \text{EU}(a|\text{bad}) = 53$$

MEU if forecast is good

$$\text{MEU}(F = \text{good}) = \max_a \text{EU}(a|\text{good}) = 95$$



Decision Network :: Example – Value of Information

Optimal decision = leave

$$\text{MEU}(\emptyset) = \max_a \text{EU}(a) = 70$$

Forecast distribution

F	P(F)
good	0.59
bad	0.41

Optimal decision = take

$$\text{MEU}(F = \text{bad}) = \max_a \text{EU}(a|\text{bad}) = 53$$

Value of Perfect Information (VPI)

VPI (Forecast) =

$$0.59 \cdot (95) + 0.41 \cdot (53) - 70$$

= 7.8

MEU if forecast is good

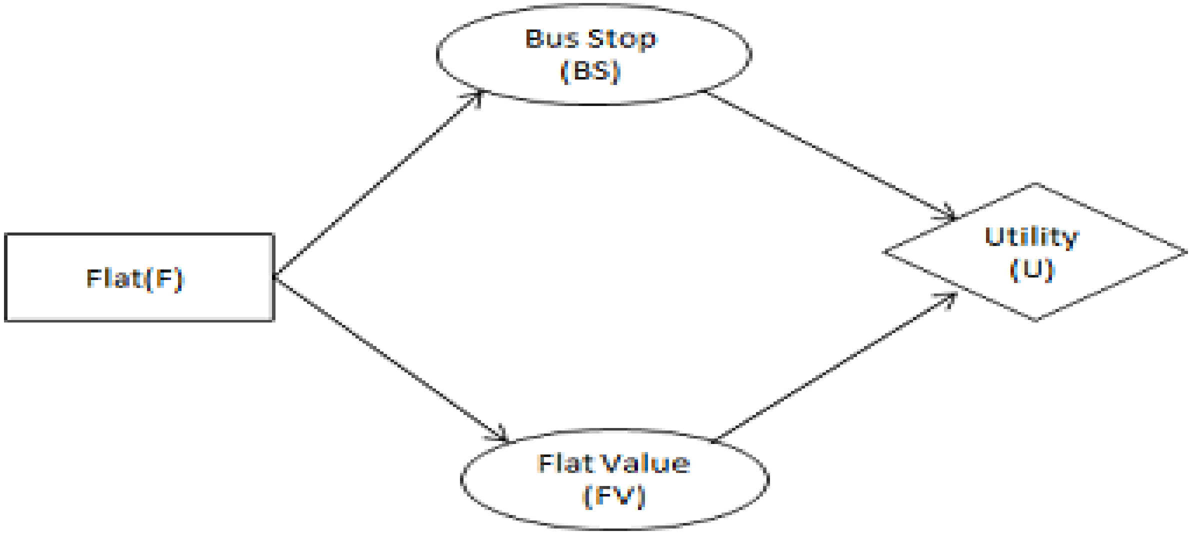
$$\text{MEU}(F = \text{good}) = \max_a \text{EU}(a|\text{good}) = 95$$

Decision Network :: Example 2

John wants to purchase a flat. He has two options: Flat A or Flat B. He wants to resell the flat in a few years. So, he is interested in how the flat's value may increase. Depending on the place he chooses, there is some probability of increase in the value of the property. John uses a bus to go to work. So, he is also interested in the distance from the bus stop to the flat.

Decision Network :: Example 2

- In the above problem, there is uncertainty about the increase of the flat's value and about the distance to the bus stop; they both depend on which flat is chosen.
- A rational decision in the above circumstances would be buying a flat that is nearest to the bus stop and also has maximum value increase amongst the two flats.
- Here is a decision network diagram for the given problem:



Decision Network :: Example 2

Bus Stop (BS):

Flat	P(BS)
A	0.5
B	0.5

Flat	BS	FV	Utility(U)
A	T	T	0.8
A	T	F	0.4
A	F	T	0.6
A	F	F	0.1
B	T	T	0.8
B	T	F	0.3
B	F	T	0.5
B	F	F	0.05

Flat Value (FV):

Flat	P(FV)
A	0.6
B	0.4

Decision Network :: Example 2

Decision	P(BS)	P(FV)	U	Expected Utility(EU)=P(BS)*U+P(FV)*U
EU(A|BS=T,FV=T)	0.5	0.6	0.8	0.88
EU(A|BS=T,FV=F)	0.5	0.6	0.4	0.44
EU(A|BS=F,FV=T)	0.5	0.6	0.6	0.66
EU(A|BS=F,FV=F)	0.5	0.6	0.1	0.11
EU(B|BS=T,FV=T)	0.5	0.4	0.8	0.72
EU(B|BS=T,FV=F)	0.5	0.4	0.3	0.27
EU(B|BS=F,FV=T)	0.5	0.4	0.5	0.45
EU(B|BS=T,FV=T)	0.5	0.4	0.05	0.045

Disclaimer

The material for the presentation has been compiled from various sources such as prescribed text books by Russell and Norvig and other tutorials and lecture notes. The information contained in this lecture/ presentation is for educational purpose only.

Thank You *for* Your Attention !