CIS587 - Artificial Intelligence

Uncertainty

KB for medical diagnosis. Example.

We want to build a KB system for the **diagnosis of pneumonia**.

Problem description:

- **Disease:** pneumonia
- **Patient symptoms (findings, lab tests)**:
  - Fever, Cough, Paleness, WBC (white blood cells) count, Chest pain, etc.

Representation of a patient case:

- Statements that hold (are true) for that patient.

  E.g: Fever =*True*
  Cough =*False*
  WBCcount=*High*

Diagnostic task: we want to infer whether the patient suffers from the pneumonia or not given the symptoms

Uncertainty

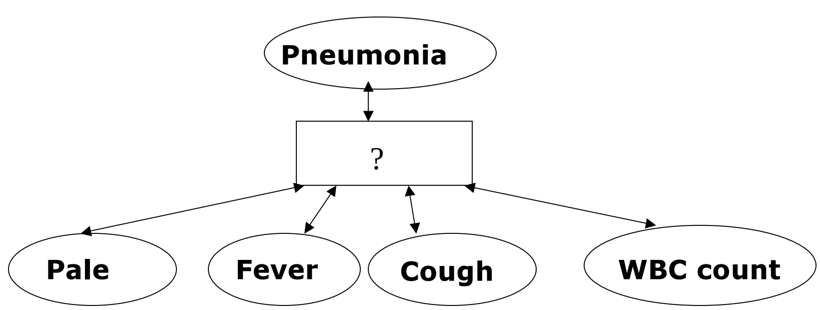
To make diagnostic inference possible we need to represent rules or axioms that relate symptoms and diagnosis

Problem: disease/symptoms relation is not deterministic (things may vary from patient to patient) – it is **uncertain**

- **Disease → Symptoms uncertainty**
  - A patient suffering from pneumonia may not have fever all the times, may or may not have a cough, white blood cell test can be in a normal range.
- **Symptoms → Disease uncertainty**
  - High fever is typical for many diseases (e.g. bacterial diseases) and does not point specifically to pneumonia
  - Fever, cough, paleness, high WBC count combined do not always point to pneumonia

Modeling the uncertainty.

- Relation between the disease and symptoms is not deterministic. **Key issues:**
- How to describe the relations in the presence of uncertainty?
- How to manipulate such knowledge to make inferences?
  - **Humans can reason with uncertainty.**



Uncertainty

- Relations at the level of detail we consider are not deterministic, they are uncertain
- Reasons for uncertainty and the need to handle it:
  - **Efficiency, capacity limits**
    - It is often impossible to enumerate and model all components of the world and their relations
  - **Observability**
    - It is impossible to observe all relevant components of the world.
- **Humans can reason with uncertainty!!!**
  - Can computer systems do the same? We need formalisms to model and manipulate uncertainty.

Methods for representing uncertainty

Default or non-monotonic logic

- Statements build on assumptions that can be retracted.

  Examples:
  - Assume that the car does not have a flat tire
  - Assume that car component works unless there is an evidence of the contrary.
    - Statements considered to be true, unless new information against them is presented. Statements are retracted or overridden
- **Problem:** exception handling, the need to enumerate all exceptions in which assumptions do not hold

Methods for representing uncertainty

Extend formalisms based on propositional and first-order logic to reflect uncertain, imprecise statements (relations)

- Typically rules with various fudge factors
- Popular in 70-80s in knowledge-based systems (e.g.,MYCIN)

If	1. The stain of the organism is gram-positive, and 2. The morphology of the organism is coccus, and 3. The growth conformation of the organism is chains
Then	**with certainty 0.7** the identity of the organism is streptococcus

Problems:

- Chaining of multiple inference rules (propagation of uncertainty)
- Combinations of rules with the same conclusions
- After some number of combinations results not intuitive

Representing uncertainty with certainty factors

- Facts (propositional statements) are assigned some certainty number reflecting the belief in that the statement is satisfied:

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

- Rules incorporate tests on the certainty values

$$(A \text{ in } [0.5,1]) \wedge (B \text{ in } [0.7,1]) \rightarrow C \text{ with CF} = 0.8$$

- Combination of multiple rules

$$(A \text{ in } [0.5,1]) \wedge (B \text{ in } [0.7,1]) \rightarrow C \text{ with CF} = 0.8$$

$$(E \text{ in } [0.8,1]) \wedge (D \text{ in } [0.9,1]) \rightarrow C \text{ with CF} = 0.9$$

$$CF(C) = \max[0.9;0.8] = 0.9$$

$$CF(C) = 0.9 * 0.8 = 0.72$$

$$CF(C) = 0.9 + 0.8 - 0.9 * 0.8 = 0.98$$

?

Methods for representing uncertainty

Probability theory

Proposition statements – represented by random variables and the assignment of (two or more) values to variables

Each value can be achieved with some probability:

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

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

Can model the effect of findings:

$$P(Pneumonia = True \mid Fever = True) = 0.02$$

$$P(Pneumonia = True \mid Fever = True, WBCcount = high, Cough = True) = 0.4$$

Subjective (or Bayesian) probability:

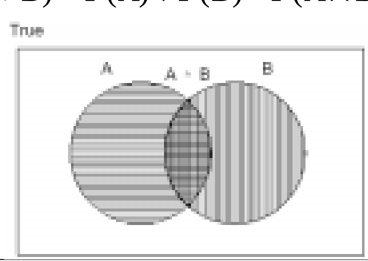
- Probabilities relate propositions to one own state of knowledge, and not assertions about the world.

Probability theory

- Well-defined theory for representing and manipulating statements with uncertainty
- **Axioms of probability:**

For any two propositions A, B.

1. $0 \le P(A) \le 1$
2. $P(True) = 1$ and $P(False) = 0$
3. $P(A \vee B) = P(A) + P(B) - P(A \wedge B)$



Modeling uncertainty with probabilities

- Assume the extension of propositional logic.

- **Propositions:**
  - statements about the world
  - assignment of values to **random variables**
- **Random variables:**
  - **Boolean**

    $Pneumonia$ is either $True, False$

  - **Multi-valued**

    $WBCcount$ is either $High, Normal, Low$

Probabilities

Unconditional probabilities (prior probabilities)

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

$P(WBCcount = high) = 0.005$

Probability distribution

- Defines probability values for all possible assignments

$$P(Pneumonia = True) = 0.001$$
$$P(Pneumonia = False) = 0.999$$

$Pneumonia$	$\mathbf{P}(Pneumonia)$
$True$	0.001
$False$	0.999

  - Probabilities sum to 1 !!!

$$P(Pneumonia = True) + P(Pneumonia = False) = 1$$

Probability distribution

Probability distribution

- Defines probability values for all possible assignments

$P(WBCcount = high) = 0.005$

$P(WBCcount = normal) = 0.993$

$P(WBCcount = high) = 0.002$

WBCcount	**P**(*WBCcount*)
high	0.005
normal	0.993
low	0.002

Joint probability distribution (for a set of variables)

- Defines probabilities for all possible assignments to values of variables in the set

$\mathbf{P}(pneumonia, WBCcount)$

Pneumonia \ *WBCcount*	*high*	*normal*	*low*
True	0.0008	0.0001	0.0001
False	0.0042	0.9929	0.0019

Joint probabilities

Joint probability distribution (for a set of variables)

- Defines probabilities for all possible assignments to values of variables in the set

$\mathbf{P}(pneumonia, WBCcount)$ 2×3 matrix

Pneumonia \ *WBCcount*	*high*	*normal*	*low*	**P**(*Pneumonia*)
True	0.0008	0.0001	0.0001	0.001
False	0.0042	0.9929	0.0019	0.999
P(*WBCcount*)	0.005	0.993	0.002	

Marginalization (summing of rows, or columns)
- summing out variables

Conditional probabilities

Conditional probability distribution

- Defines probabilities for all possible assignments, given a fixed assignment for some other variable values

$P(Pneumonia = true \mid WBCcount = high)$

$\mathbf{P}(pneumonia \mid WBCcount)$ 3 element vector of 2 elements

		WBCcount		
		high	*normal*	*low*
Pneumonia	*True*	0.08	0.0001	0.0001
	False	0.92	0.9999	0.9999
		1.0	1.0	1.0

$P(Pneumonia = true \mid WBCcount = high)$
$+ P(Pneumonia = false \mid WBCcount = high)$

CIS587 - AI

Conditional probabilities

Conditional probability distribution. Defined in terms of a joint probability

$$P(A \mid B) = \frac{P(A,B)}{P(B)} \text{ s.t. } P(B) \neq 0$$

$$P(pneumonia = true \mid WBCcount = high) = \frac{P(pneumonia = true, WBCcount = high)}{P(WBCcount = high)}$$

- **Product rule.** Join probability can be expressed in terms of conditional probabilities

$$P(A,B) = P(A \mid B)P(B)$$

- **Chain rule**. Any joint can be expressed as a product of conditionals

$$P(X_1, X_2, \ldots X_n) = P(X_n \mid X_{1,} \ldots X_{n-1})P(X_{1,} \ldots X_{n-1})$$
$$= P(X_n \mid X_{1,} \ldots X_{n-1})P(X_{n-1} \mid X_{1,} \ldots X_{n-2})P(X_{1,} \ldots X_{n-2})$$
$$= \prod_{i=1}^{n} P(X_i \mid X_{1,} \ldots X_{i-1})$$

CIS587 - AI

Bayes rule

Conditional probability.

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

Bayes rule:

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

When is it useful?

- When interested in computing the diagnostic probability, from the causal probability

$$P(cause \mid effect) = \frac{P(effect \mid cause)P(cause)}{P(effect)}$$

- **Reason:** It is often easier to assess causal probability
  - E.g. Probability of pneumonia causing fever
    vs. probability of pneumonia given fever

Bayes rule

Assume a variable A with multiple values: $a_1, a_2, \ldots a_k$

Bayes rule can be rewritten as:

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

$$= \frac{P(B = b \mid A = a_j)P(A = a_j)}{\sum_{i=1}^{k} P(B = b \mid A = a_j)P(A = a_j)}$$

Used in practice when we want to compute:

$\mathbf{P}(A \mid B = b)$ for all values of $a_1, a_2, \ldots a_k$

1. compute $P(B = b \mid A = a_j)P(A = a_j)$ for all j, and
2. obtain the result by renormalizing the probability vector with β

$$P(A = a_j \mid B = b) = \beta P(B = b \mid A = a_j)P(A = a_j)$$

$$\beta = 1 / \sum_{i=1}^{k} P(B = b \mid A = a_j)P(A = a_j)$$

Full joint distribution

- the joint distribution for all variables in the problem, **full joint probability distribution**, defines the complete probability model

  Example: pneumonia diagnosis

  Full joint defines the probability for all possible assignments of values to Pneumonia, Fever, Paleness, WBCcount, Cough

$$P(Pneumonia = T, WBCcount = High, Fever = T, Cough = T, Paleness = T)$$

$$P(Pneumonia = T, WBCcount = High, Fever = T, Cough = T, Paleness = F)$$

$$P(Pneumonia = T, WBCcount = High, Fever = T, Cough = F, Paleness = T)$$

$\cdots$ etc

- Any probabilistic query can be obtained (computed) from the full joint probability

Full joint distribution

Computation of probabilistic (inference) queries

- **Joint over smaller number of variables** is obtained through marginalization

$$P(A = a, C = c) = \sum_i \sum_j P(A = a, B = b_i, C = c, D = d_j)$$

- **Conditional probability over set of variables**, given other variables' values is obtained through marginalization and definition of conditionals

$$P(D = d \mid A = a, C = c) = \frac{P(A = a, C = c, D = d)}{P(A = a, C = c)}$$

$$= \frac{\sum_i P(A = a, B = b_i, C = c, D = d)}{\sum_i \sum_j P(A = a, B = b_i, C = c, D = d_j)}$$

Modeling uncertainty with probabilities

- Defining the full joint distribution makes it possible to represent and reason with uncertainty in a uniform way
  - We are able to handle an arbitrary inference problem

Problems:

  - **Space complexity.** To store a full joint distribution we need to remember $O(d^n)$ numbers.

    n : number of random variables, *d :* number of values
  - **Inference (time) complexity.** To compute some queries requires $O(d^n)$ steps.
  - **Acquisition problem.** Who is going to define all of the probability entries?

Medical diagnosis example

- **Space complexity**
  - Pneumonia (2 values: T,F), Fever (2: T,F), Cough (2: T,F), WBCcount (3: high, normal, low), paleness (2: T,F)
  - Number of assignments: 2*2*2*3*2=48
  - We need to define at least 47 probabilities.
- **Time complexity**
  - Assume we need to compute the marginal of Pneumonia=T from the full joint

$$P(Pneumonia = T) =$$
$$= \sum_{i \in T,F} \sum_{j \in T,F} \sum_{k=h,n,l} \sum_{u \in T,F} P(Fever = i, Cough = j, WBCcount = k, Pale = u)$$

  - Sum over: 2*2*3*2=24 combinations

Modeling uncertainty with probabilities

- Knowledge based system era (70s – early 80's)
  - Extensional non-probabilistic models
  - Space, time and acquisition bottlenecks in probability-based models froze the development and advancement of KB systems and contributed to the slow-down of AI in 80s in general

- Breakthrough (late 80s, beginning of 90s)
  - **Bayesian belief networks**
    - Give solutions to the space, acquisition bottlenecks
    - Partial solutions for time complexities