Posts and Telecommunication Institute of Technology
Faculty of Information Technology 1

Introduction to Artificial Intelligence

Probabilistic inference

Dao Thi Thuy Quynh

Outline

- Inference with uncertain evidence
- Principle of probabilistic inference
- Some concepts of probability

Inference with uncertain evidence (1/2)

- Logic
  - Allows knowledge representation and inference/reasoning
  - Requires clear, complete, certain, non-contradictory knowledge

- Real world
  - Ambiguity, uncertainty, lack of information, and contradictions

Inference with uncertain evidence (2/2)

- Factors affecting the clarity and certainty of knowledge and information
  - Information with randomness
    - Play cards, flip a coin
  - Theory is not clear
    - For example, the mechanism of disease is not known
  - Lack of factual information
    - Insufficient patient testing
  - Factors related to the problem are too big and too complicated
    - It is impossible to represent all elements
  - Errors when getting information from the environment
    - Measuring devices have errors

Approaches

- Multivalued logic
  - Use more logical values (not only “true”, “false”)
- Fuzzy logic
  - An expression can take value "true" with a value in the range of [0,1]
- Possibility theory
  - Events or formulas are assigned a number representing the likelihood that event occurs
- Probabilistic inference
  - The inference result returns the probability that a certain event or formula is true

Outline

- Inference with uncertain evidence
- Principle of probabilistic inference
- Some concepts of probability

Principle of probabilistic inference (1/2)

- Instead of inferring about the "true" or "false" of a proposition (2 values), inferring about the "belief" that the proposition is true or false (infinite)
  - Assign a belief value to each proposition
  - Express the belief value as a probability value; using probability theory to work with this value
  - For proposition A
    - Assign a probability $P(A)$: $0 \leq P(A) \leq 1$;
    - $P(A) = 1$ if A is true, $P(A) = 0$ if A is false
  - Example:
    - $P\ (Cold = true) = 0.6$
    - the patient has a cold with a probability of 60%, "Cold" is a random variable that can receive True or False

Principle of probabilistic inference (2/2)

- The nature of probability used in inference
  - Statistical nature: based on experiments and observations
    - It is not always possible to determine
  - Probability based on subjectivity: the degree of confidence/belief that the event is true or false by experts/users
    - Used in probabilities

Outline

- Inference with uncertain evidence
- Principle of probabilistic inference
- Some concepts of probability

Axioms of probability and some basic properties

Axioms of probability

1. $0 \le P(A = a) \le 1$ for all a in the value domain of A
2. $P(True) = 1$, $P(False) = 0$
3. $P(A \vee B) = P(A) + P(B) - P(A \wedge B)$

Basic properties

1. $P(\neg A) = 1 - P(A)$
2. $P(A) = P(A \wedge B) + P(A \wedge \neg B)$
3. $\Sigma_a P(A = a) = 1$: sum for all a in the value domain of A

Joint probability (1/2)

- In the form $P(V_1 = v_1, V_2 = v_2, \ldots, V_n = v_n)$
- Full joint probability distribution: includes probabilities for all combinations of values of all random variables
- Example: given 3 Boolean variables: Chim, Non, Bay

Chim (C)	Non (N)	Bay (B)	P
T	T	T	0.0
T	T	F	0.2
T	F	T	0.04
T	F	F	0.01
F	T	T	0.01
F	T	F	0.01
F	F	T	0.23
F	F	F	0.5

Joint probability (2/2)

- If all joint probabilities are given, we can compute the probabilities for all propositions related to the problem

- Example:
  - $P(Chim = T) = P(C) = 0.0 + 0.2 + 0.04 + 0.01 = 0.25$
  - $P(Chim = T, Bay = F) = P(C, \neg B) = P(C, N, \neg B) + P(C, \neg N, \neg B) = 0.2 + 0.01 = 0.21$

Conditional probability (1/2)

- Plays an important role in probabilistic inference
  - Inferring the probability of an outcome given a set of evidences
  - Example:
    - $P(A|B) = 1$ equivalent to $B \Rightarrow A$ in logic
    - $P(A|B) = 0.9$ equivalent to $B \Rightarrow A$ with a probability of 90%
    - Given evidences (observations) $E_1, \dots, E_n$ compute $P(Q|E_1, \dots, E_n)$
- Definition of conditional probability
  - $P(A|B) = \frac{P(A \land B)}{P(B)} = \frac{P(A,B)}{P(B)}$
  - Example: Compute
    - $P(\neg Chim \mid Bay)$

Conditional probability (2/2)

- Properties of conditional probability
  - $P(A,B) = P(A|B)P(B)$
  - Chain rule: $P(A,B,C,D) = P(A|B,C,D)\,P(B|C,D)\,P(C|D)\,P(D)$
  - Conditional chain rule: $P(A,B|C) = P(A|B,C)\,P(B|C)$
  - Bayes' theorem $P(A|B) = \frac{P(A)\,P(B|A)}{P(B)}$
  - Conditional Bayes' theorem: $P(A|B,C) = \frac{P(B|A,C)\,P(A|C)}{P(B|C)}$
  - $P(A) = \Sigma_b\{P(A|B=b)\,P(B=b)\}$, sum for all values b of B
  - $P(\neg B|A) = 1 - P(B|A)$

Combining multiple evidences

- Example:
  - Compute $P\,(\neg Chim \mid Bay, \neg Non) = \frac{P\,(\neg Chim, Bay, \neg Non)}{P\,(Bay, \neg Non)}$

- The general case: given the joint probability table, we can compute
  - $P(V_1 = v_1, \dots, V_k = v_k \mid V_{k+1} = v_{k+1}, \dots, V_n = v_n)$
  - Sum of rows with $V_1 = v_1, \dots, V_n = v_n$ divided by sum of rows with $V_{k+1} = v_{k+1}, \dots, V_n = v_n$

Probabilistic independence

- An event A is said to be independent of another event B if $P(A|B) = P(A)$
  - Meaning: knowing that B occurred does not change the probability that A occurred
  - Therefore $P(A,B) = P(A)P(B)$

- A is conditionally independent of B given C if
  - $P(A|B,C) = P(A|C)$ or $P(B|A,C) = P(B|C)$
  - Meaning: B doesn't tell us anything about A if we already know C
  - Therefore $P(A,B|C) = P(A|C)P(B|C)$

Using Bayes' theorem

- Bayes' theorem plays an important role in inference
- To compute $P(A|B)$ we can compute $P(B|A)$ which is easier
  - Example: the probability of getting the flu with a headache and the probability of having a headache with the flu

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

Example (1/3)

- A person has a positive result with disease B
- Testing equipment is not completely accurate
  - The equipment gives a positive result for 98% sick people
  - The equipment gives a positive result for 3% people who are not sick
- 0.8% of the population has this disease
- Question: Is this person sick?

Example (2/3)

- Notation: event that a person has disease is B, event that a person has a positive result is A
- According to the problem data we have:
  - $P(B) = 0.008, P(\neg B) = 1 - 0.008 = 0.992$
  - $P(A|B) = 0.98, P(\neg A|B) = 1 - 0.98 = 0.02$
  - $P(A|\neg B) = 0.03, \mathrm{P}(\neg \mathrm{A}|\neg B) = 1 - 0.03 = 0.97$
- We need to compare probabilities $P(B|A)$ and $P(\neg B|A)$
- Using Bayes' theorem:
  - $P(B|A) = \frac{P(A|B)P(B)}{P(A)} = \frac{0.98*0.008}{P(A)} = \frac{0.00784}{P(A)}$
  - $P(\neg B|A) = \frac{P(A|\neg B)P(\neg B)}{P(A)} = \frac{0.03*0.992}{P(A)} = \frac{0.02976}{P(A)}$
- $P(\neg B|A) > P(B|A)$, ➔ Not sick

Example (3/3)

- To compare $P(B|A)$ and $P(\neg B|A)$ we can not compute specifically 2 probability values, instead we compute: $\frac{P(B|A)}{P(\neg B|A)}$
  - 2 expressions have the same denominator $P(A)$
  - The result of testing disease depends on the value $\frac{P(B|A)}{P(\neg B|A)}$ which is greater than or less than 1

- When we need to specifically compute this probability, we do:

$P(B|A) + P(\neg B|A) = 1$ so $\frac{P(A|B)P(B)}{P(A)} + \frac{P(A|\neg B)P(\neg B)}{P(A)} = 1$

Then, $P(A) = P(A|B)P(B) + P(A|\neg B)P(\neg B) = 0.00784 + 0.02976 = 0.0376$

So, $P(\neg B|A) = 0.79; P(B|A) = 0.21$

Combining Bayes' theorem and probabilistic independence

- Compute $P(A|B,C)$, where B and C are conditionally independent given A
  - Bayes' theorem $P(A|B,C) = \frac{P(B,C|A)*P(A)}{P(B,C)}$
  - Probabilistic independence $P(B,C|A) = P(B|A) * P(C|A)$
  - Therefore $P(A|B,C) = \frac{P(B|A)*P(C|A)*P(A)}{P(B,C)}$
- Example:
  - Give 3 binary valuables: liver disease BG, jaundice VD, anemia TM
  - Assume that VD is independent with TM
  - Know that $P(BG) = 10^{-7}$
  - Someone has VD disease
  - Know that $P(VD) = 2^{-10}$ and $P(VD|BG) = 2^{-3}$
  - a) What is the probability that a tester has disease?
  - b) Know that a person has TM disease and $P(TM) = 2^{-6}$, $P(TM|BG) = 2^{-1}$. Compute the probability that the tester has BG.