Data mining, machine learning, and uncertainty reasoning

林偉川

Inverse resolution

- intra-construction

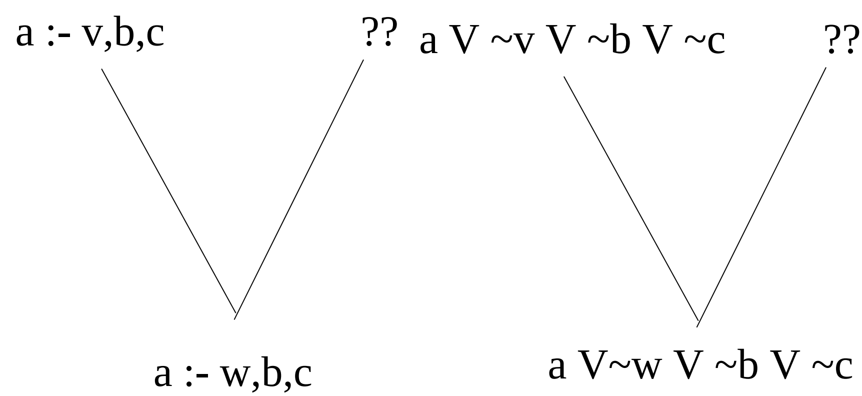
$$\left\{\begin{array}{l} a :- v, b, c \\ a :- w, b, c \end{array}\right\} \Rightarrow \left\{\begin{array}{l} a :- v, u \\ a :- w, u \\ u :- b, c \end{array}\right\}$$

- inter-construction

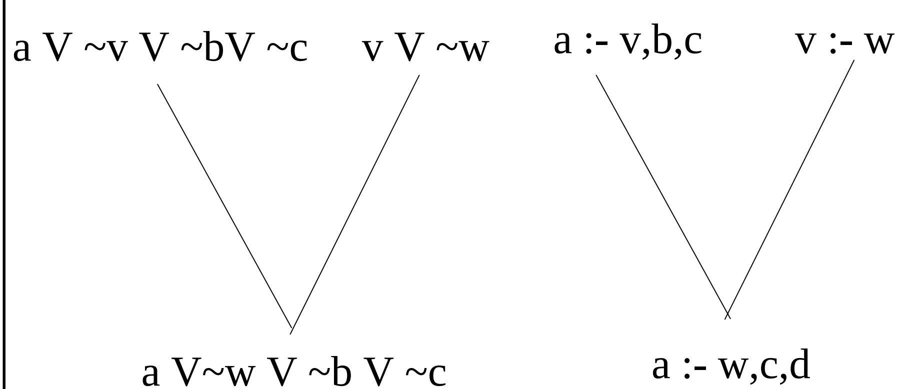
$$\left\{\begin{array}{l} a :- v, b, c \\ a :- w, b, c \end{array}\right\} \Rightarrow \left\{\begin{array}{l} a :- u, b, c \\ u :- v \\ u :- w \end{array}\right\}$$

- All of these operators can be derived from the resolution principle that is very popular in AI.

Inverse resolution -- inter-construction



Inverse resolution -- inter-construction



Inverse resolution -- intra-construction

a V ~v V~u   u V ~b V ~c

a :- v,u   u :- b,c

a V ~v V ~b V ~c

a :- v,b,c

Inverse resolution -- intra-construction

a V ~w V~u   u V ~b V ~c

a :- w,u   u :- b,c

a V ~w V ~b V ~c

a :- w,c,d

Theory Revision

- Suppose background knowledge contains partial information about the family relation
- Additional information about the sex of the individual persons is provided in terms of the predicate male and female
- The learner is expected to derive the definition of the predicate father, previously not present in the background knowledge

Theory Revision

- The learner starts with some strongly restricted language such as initial constraint demands that each literal in the clause body is allowed to contain as arguments only those constants and variables that also appear in the head P(x,y) :- q(x,y), r(x)
- Father(jack, bill) means that the system searches for all literals containing no other arguments except for jack and bill
- If found such literals, the system connects them with conjunctions

Theory Revision

- Suppose that the system's background knowledge contains the following predicates:
  …
  parent(jack,bill). parent(tom, jack). parent(tom, eve). parent(eve, bill). male(tom). male(jack). male(bill). female(eve). painter(bill). singer(jack).

Theory Revision

- If there are no other predicates containing either of the arguments jack or bill, the attempt to construct the concept is described as following:
  father(jack, bill) :- parent(jack, bill),male(jack),
  male(bill), painter(bill),singer(jack)
- The learner turns the constants into variables and obtains the initial clause as followed:
  father(x,y) :- parent(x,y), male(x), male(y),
  painter(y), singer(x)

Theory Revision

- The fact jack is bill's father has nothing to do with bill's profession
- The authors of CLINT provided the learner with the ability to refine the initial description of the concept by way of a simple dialog with the user
- During the dialog, the learner examines each predicate and checks it necessity by creating new examples and asking the user to classify them.

Theory Revision

- Is father(tom, jack) true? is positively answered by the user ➔ literal painter(y) is unnecessary (singer(x) is also unnecessary because Tom is still his father)
- Whether it is necessary that y be male. Knowing that eve is female, the learner finds in the background knowledge the literal parent(tom, eve) and asks the user the following question:
  Is father(tom, eve) true?
- A positive answer indicates that male(y) is also unnecessary

Theory Revision

- On the other hand, the question:
  Is father(eve, bill) true?
- A negative answer indicates that male(x) cannot be discarded from the clause. Therefore, the result is listed followed:
  father(x,y) :- parent(x,y), male(x)
- In the course of this verification, the original clause can totally alter or, all literals will be deleted from the body
- The system proceeds by alleviating some of the constraints

Theory Revision

- If the one imposed on the predicates arguments, the body predicates will be allowed to contain one and only one argument that does not appear in the head, as is the case of the clause:
  grandparent(x, y) :- parent(x, z), parent(z, y)
- In this way, the system generalize the concept description with the objective to cover those positive instances that have not been covered by the previous description

Theory Revision

- The description can become too general that it also cover negative examples ➔ respective measures must be taken to rectify this inconvenience
- The solution implemented in CLINT consists of building the explanation tree for negative example, identifying the culprit clause c responsible for the coverage of the negative example, deleting c from the knowledge base, and re-generalizing the resulting knowledge structure so that all positive examples that have previously been covered by c become covered again.

Another case study of MYCIN

Problem domain:

http://www.computing.surrey.ac.uk/research/ai/PROFILE/mycin.html#Certainity%20Factors

- Selection of antibiotics for patients with serious infections.
- Medical decision making, particularly in clinical medicine is regarded as an "art form" rather than a "scientific discipline": this knowledge must be systemized for practical day-to-day use and for teaching and learning clinical medicine.

Another case study of MYCIN

- **<u>Target Users:</u>** Physicians and possibly medical students and paramedics.
- **<u>Originators</u>** Edward Shortliffe, (1972)
  Department of Medicine and Computer Science, Heuristic Programming Project, Stanford University School of Medicine
- <u>References:</u> **Shortliffe**, Edward H.(1981), "CONSULTATION SYSTEMS FOR PHYSICIANS: The Role of Artificial Intelligence Techniques". In **Webber**, Bonnie L.and **Nilsson**, Nils J. (Eds). <u>Readings in Artificial Intelligence</u>. pp323-333. Tioga Publishing Company. Palo Alto, California.

Another case study of MYCIN

<u>MYCIN -- A typical consultation</u>

1. Patient's name: (first - last)
   ** **FRED BRAUN** (Questions 1-3 request background patient data).
2. Sex:
   ** **M**
3. Age:
   ** **55**
4. Are there any cultures for Fred Braun which may be related to the present illness, and from which organisms have been grown successfully in the microbiology laboratory?
   ** **Y** - - - CULTURE - 1 - - -
5. From what site was the specimen for CULTURE - 1 taken?
   ** **BLOD**
   = **BLOOD** (System corrects typing error).

Another case study of MYCIN

6. Please give the date and time when CULTURE-1 was obtained. (mo/da/yr time)
   ** **JUN 20, 1977** The first organism isolated from the blood culture of 20-JUN-77 (CULTURE-1) will be referred to as: - - ORGANISM - 1 - -
7. Enter the laboratory - reported identity of ORGANISM-1:
   ** **UNKNOWN** (Typically identity is not known yet).
8. The strain (gram or Ziehl-Neelsen acid-fast) of ORGANISM-1:
   ** **NEG** (Preliminary lab results give some clues).
9. Is ORGANISM - 1 rod or coccus (etc.):
   ** **ROD**
10. What is the form of the individual organisms (e.g. lancet-shaped for cocci, fusiform for rods, etc)?
    ** **FUSIFORM** (… more questions follow in order to gather sufficient information to infer the identity and significance of the infecting organisms ...)

Another case study of MYCIN

23. What is the suspected portal of entry of the gramneg rod from the blood culture of 20-JUN-77 (ORGANISM-1) into this sterile site?
    ** **GI (6)** (User indicates the uncertainty of his response).
24. Is Fred Braun a burn patient?
    ** **N**
25. Is Fred Braun a compromised host (e.g. alcoholic, sickle-cell-disease, im-munosuppressed)?
    ** **N**
26. Is Fred Braun febrile?
    ** **Y**

- (... after about 50-60 questions, MYCIN prints the diagnostic hypotheses on which therapy will be based ...)

Homework

- Try to find the related work about the system that has a simple dialog with the user. In this system, this system can obtain some intelligent information that is not existed in the predefined predicate!!

Constructive induction

- The problem of determining the appropriate representation space for learning!!
- The learner analyze the examples, describing them in terms of some predefined set of attributes or predicate, and produces the expected concept description using the operators of the description language such as conjunctions, disjunctions, and negations of the attribute value or predicates ➔ induction empirical
- TDIDT and AQ algorithms are used to described the concepts by a subset of attributes of the learning examples

Classic methods of learning

2 essential learning principles

- Divide-and-Conquer
  - The entire set of examples is split into subsets that are more easy to handle (TDIDT)
- AQ-philosophy
  - Based on the idea of progressive coverage of the training data by consecutively generated decision rules

TDIDT Algorithm

S … the set of examples

1. Find the “best” attribute *at* (***if this can be found!!!***)
2. Split the set S into the subset S_1, S_2, …, so that all examples in the subset S_i have at=v_i. Each subset constitutes a node in the decision tree
3. For each S_i : if all examples in S_i belong to the same class (⊗ or ⊕), then create a leaf of the decision tree and label it with this class label (such as x or z) . Otherwise, perform the same procedure (go to step 1) with S=Si (such as *at3*)

TDIDT Algorithm

- The entire set of examples is split into subsets that are more easy to handle
- With a properly defined evaluation function, the TDIDT algorithm will derive the proper decision tree
- This Divide-and-conquer algorithm terminates when all subsets are labeled or when no further attributes splitting the unlabelled sets are available
- Complete the full decision tree can cover the table example but they include some negative examples
- The most difficult is how to find the first best attribute in step 1!!

AQ algorithm

1. Divide all examples into the subsets PE ⊕'s and NE ⊗'s
2. Choose randomly or by design one example from PE and call it the ***seed***
3. Find a set of maximally general rules characterizing the seed. The limit of the generalization is defined by the set NE: a generalized description of the seed is not allowed to cover any object from NE. the set of rules obtained is called the ***star***

AQ algorithm

4. According to some preference criterion, select the best rule in the star
5. If this rule, jointly with all previously generated rules, covers all objects from PE, then stop. Otherwise, find another seed among the uncovered examples in PE and go to step 3.

Step 3 is done by a special star generation procedure

It constructs a set of decision rules with different relationship among the individual rules

Criterion for applying AQ learning

- Maximize the number of ⊕'s covered by the rule
- Minimize the number of attributes involved
- Maximize the estimate of generality
- Minimize the costs of attribute-value measurement
- Use attribute selection criteria used in decision tree learning such as entropy, gain ratio, …etc.
- Decision rules may be logical intersect, disjoint, or linearly order (requires their evaluation in a sequential order)

Constructive induction

- Inverse resolution is to construct new predicate that are to facilitate the learning process. Often, this method necessitates an interactive learning procedure
- The user possessing the knowledge about which predicates make sense is asked to acknowledge the new predicate and assign it a name
- Meat-eating animals with claws are accepted and given the name predators; big animals with yellow skin will probably not make a useful concept ➔ will be rejected by the learner

AQ algorithm

- Constructive induction in analogy-based learning is based on the idea of second-order schemata as implemented in CIA system.
- Learning system based on AQ algorithm can easily incorporate background knowledge because such knowledge is often represented by decision rules
- The background knowledge is more typical for predicate-logic-based learning system

Constructive induction

- The principle of constructive induction is combined with the deeper studies of the nature of various representation language
- The essence of the system consists in storing typical schemata of predicate expressions, such as:
  p(x,y) :- q(x,xw), q(y,yw), r(xw,yw)
  where [predicates] p, q, r,
  [arguments] x, y, xw, yw represent variables

Constructive induction

- This schema can, by suitable substitutions, be instantiated into the following clauses:
  lighter(x,y) :- weight(x,xw), weight(y,yw), less(xw,yw)
  where Θ={p/lighter, q/weight, r/less }
  same-color(x,y) :- color(x,xc), color(y,yc), eq(xc,yc)
  where Θ={p/same-color, q/color, r/eq }
  brothers(x,y) :- son(x,xp), son(y,yp), eq(xp,yp)
  where Θ={p/brothers, q/son, r/eq }
- Second-order schemata lend themselves quite straightforwardly to constructive induction.

Constructive induction

- In CIA's setting, after proper substitutions, the body of a schema becomes a subset of the body of some clause whose head is unknown.
- For illustration, the schema:
  p(x,y) :- q(x,xw), q(y,yw), r(xw,yw)
  can become a subset of the clause
  :- male(f), male(c), parent(f,m1), parent(m2,c), eq(m1,m2)
  after the substitutions:
  Θ={ q/parent, r/eq }, ρ={x/f, y/c, xw/m1, yw/m2 }
- The instantiated schema is : (grandparent ➔ p)
  p(f,c) :- parent(f,m1), parent(m2,c), r(m1,m2)

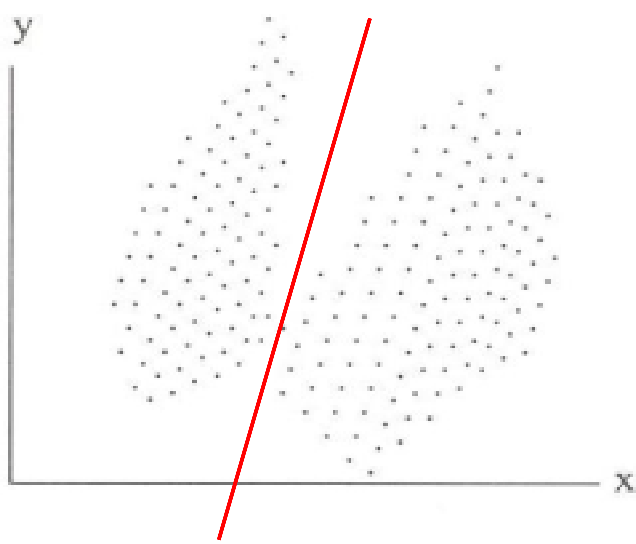
Artificial discovery

- Artificial discovery is an instructive illustration of the general way of thinking in machine learning
- Supervised learning where the learner seeks to develop a concept description from examples that have been pre-classified by the teacher
- Unsupervised learning whose task is to generate conceptual taxonomies from non-classified objects

Artificial discovery

- An object that has been recognized as a member of a certain category inherits the general properties of the categories
- A related task is carried out by some traditional statistical techniques such as cluster analysis.

Traditional task for cluster analysis



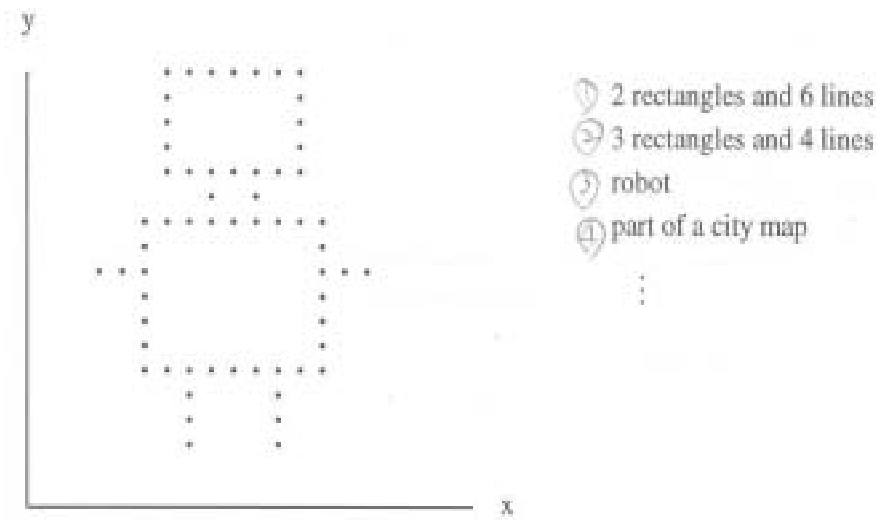
Artificial discovery

- The picture represent objects described by 2 numeric attributes, x and y. Apparently, the objects can be partitioned into 2 groups which are easy to discover by relatively simple algorithms exploiting the notion of similarity, as measured by numeric distance between objects
- Unfortunately, not every kind of similarity can be assessed numerically

Artificial discovery

- Is the distance between cat and giraffe greater than the distance between dog and elephant?
- Even though these distances can be transformed into numbers, any such a transformation would be difficult and subjective

Concept-discovery task of machine learning



Concept-discovery task of machine learning

- The objects are already pre-ordered in a way that can be described conceptually, and several interpretations can be offered depending on the particular context
- Traditional distance-based cluster analysis will hardly produce reasonable outcomes on these data ➔ this task appears to be trivial for human

Concept-discovery task of machine learning

- In machine learning, the search for concept hidden in a set of objects is studied by the discipline known under the name of concept formation
- The further research is to discover not only concepts but also laws defining the relations among them, with the ambition of creating a computer-based system to assist human researchers in such disciplines as chemistry or biology

Concept formation

- Focus on the unsupervised concept learning and divides this field into 2 different subfields: concept discovery ➔deriving concepts from a batch, and incremental concept formation ➔ gradually forms the concepts from a stream of examples
- Concept discovery by conceptual clustering

Concept formation

- Conceptual clustering is a novel form of clustering in which clusters are not just collections of entities processing numerical similarity. Rather, the clusters are understood as groups of objects that together represent a concept
- Conceptual clustering produces not only clusters, but also descriptions of related concepts