Association Based Classification

By

T.R.Lekhaa

AP/IT

Association-Based Classification

- Frequent patterns (FP) and their associations govern interesting relationships between attribute conditions (values) and class labels. → motivated classification.
- Associative rules (AR) show strong associations between attribute-value pairs (or items).
- Associative classification where association rules are generated and analyzed to classify the data.
- Idea is to search for strong associations between FPs (attr-val pair) and class labels.
- As ARs explore confident rules among multiple attributes this method may be better than DT induction which considers only one attribute at a time.

Recall…

- Association rule mining: A two step process
  - Generate FPs.
  - Rule generation.
- Example of an AR:

$$age = youth \wedge credit = OK \Rightarrow buys\_computer = yes \quad [20\%, 93\%] \quad (1)$$

The antecedents contain attributes and their values.

The consequent is the class label attribute and its value (or simply class name).

Formal way

- D := data set of tuples.
- A_i := i^{th} attribute , $1 <= i <= n$
- A_{class} = class attribute - this takes class label as its value.

So, a tuple X is denoted by n attributes and a class label:

$$X = \{ x_1, x_2, x_3 \ldots, x_n, C\}$$

Define item $p := (A_i, v)$ where v is the value of the attribute.

X satisfies p if and only if $x_i = v$

Contd…

- AR can have any number of items to form the antecedent.

$$\text{Thus, } p_1 \wedge p_2 \wedge p_3 \ldots p_l \Rightarrow A_{class} = C \ , \ l <= n$$

Confidence of a rule

For a given rule R, the percentage of tuples in D satisfying the rule antecedent that also have the class label C is called the confidence of R.

E.g. Confidence = 93% in (1) means 93% of customers in D are *young* and their credit rating is *OK* belong to the class *buys_computer = yes.*

Contd..

- ***Support of a rule***

  For a given rule R, the percentage of tuples in D satisfying the rule antecedent and have the class label C is called the support of R.

E.g. Support = 20% in (1) means 20% of customers in D are *young* and their credit rating is *OK* and belong to the class *buys_computer = yes.*


SNS COLLEGE OF ENGINEERING
COIMBATORE-107


sns
INSTITUTIONS

Thank You...