Unit-2-part-2.4 Association

Dr. K . RAGHAV RAO
Professor of CSE, Dept. of MCA
KL University
krraocse@gmail.com
http://datamining.blog.com

Mining Various Kinds of Association Rules

Mining Multilevel Association

Multilevel Association Rules

- Strong association discovered at high levels of abstraction may represent commonsense knowledge.
- In next slide example show items purchased as concept hierarchy is shown in other next slide.
- Data can be generalized by replacing low-level concepts within the data by their higher-level concepts , or ancestors, from a concept hierarchy.
- Concept hierarchies for categorical attributes are often implicit within the database schema, in which case they may be automatically generated using preprocessing methods.
- Concept hierarchies for numerical attributes can be generated using discretization techniques.

Multilevel Association Rules

Association rules generated from mining data at multiple levels of abstraction are called multiple-level or multilevel association rules.

Table 5.6

TID	*items purchased*
T100	IBM-ThinkPad-T40/2373, HP-Photosmart-7660
T200	Microsoft-Office-Professional-2003, Microsoft-Plus!-Digital-Media
T300	Logitech-MX700-Cordless-Mouse, Fellowes-Wrist-Rest
T400	Dell-Dimension-XPS, Canon-PowerShot-S400
T500	IBM-ThinkPad-R40/P4M, Symantec-Norton-Antivirus-2003
...	...

Multilevel Association Rules

It is easier to find strong association rules between generalized abstractions.

Fig. Concept hierarchy for all Electronics computer items

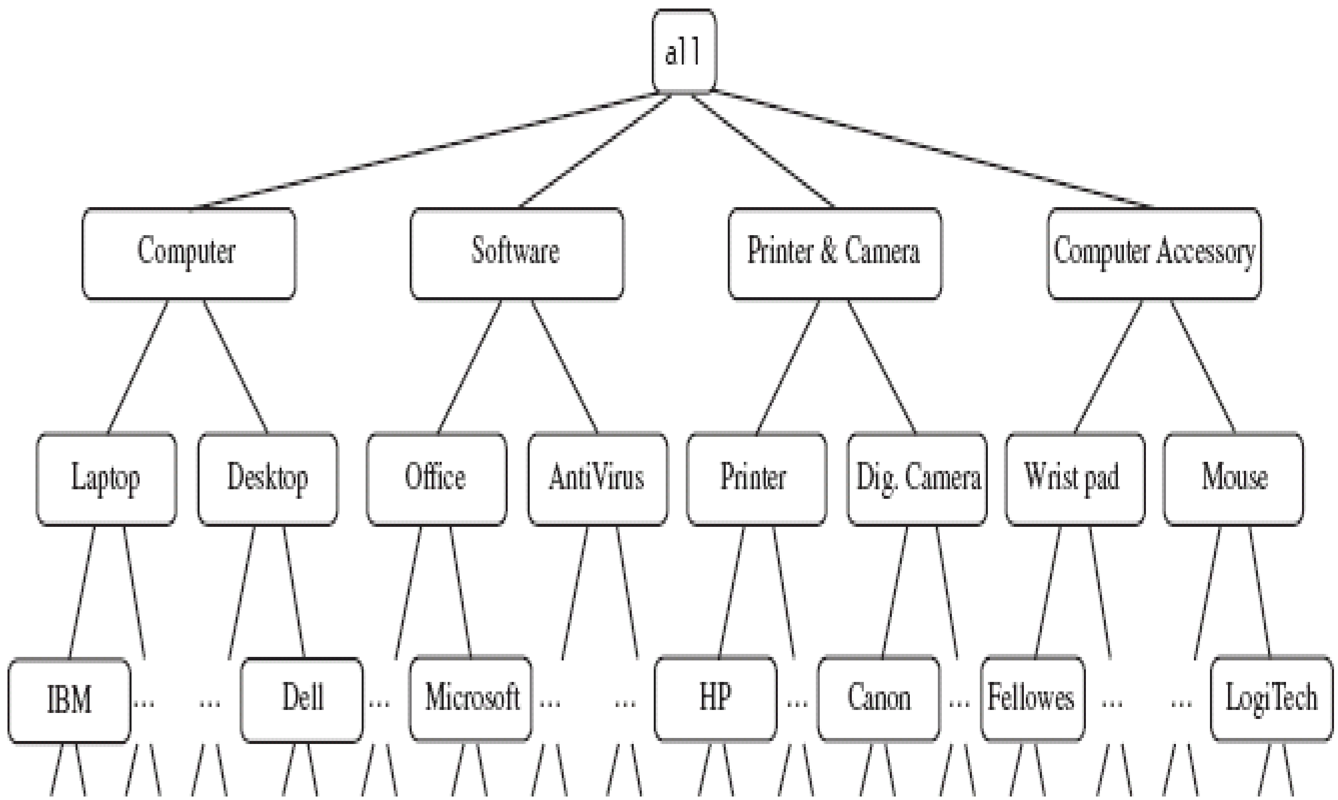


Fig 5.10

Multilevel Association: Uniform vs Reduced Support

Uniform support: same minimum support threshold is used when mining at each level of abstraction.

- Reduced support: each level of abstraction has its own minimum support threshold. The deeper the level of abstraction, the smaller the corresponding threshold is.

Multilevel Association:

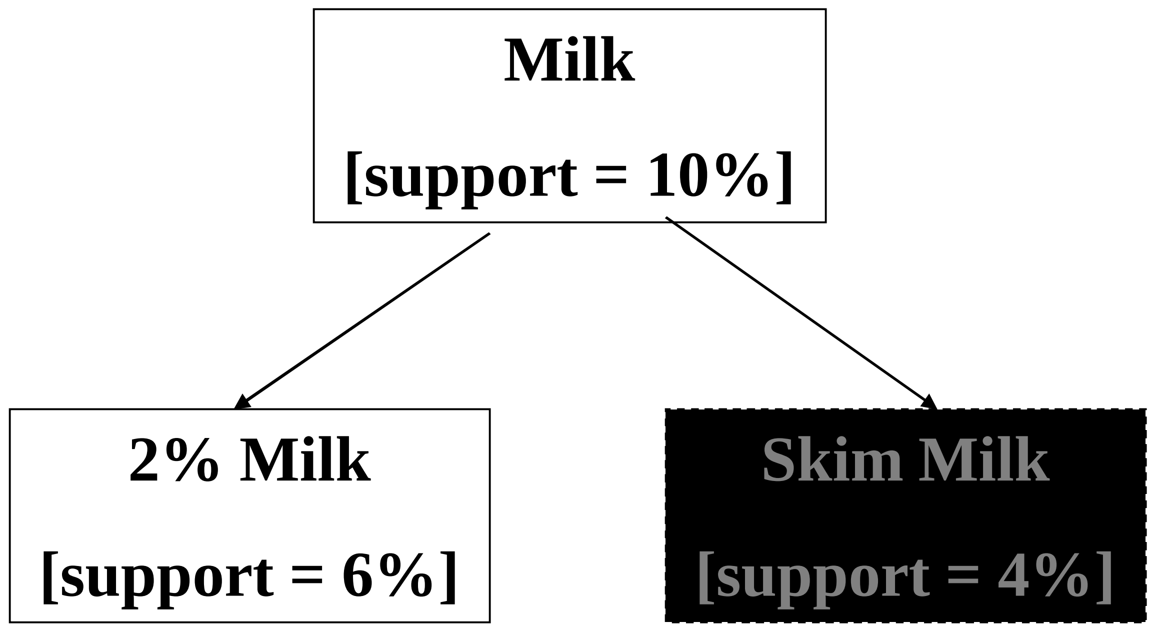
Group-based support: because users or experts often have insight as to which groups are more important than others, it is sometimes more desirable to set up user-specific ,item, or group-based mimial support thresholds when mining multilevel rules.

Uniform Support

- Same minimum support threshold for all levels

Level 1
min support = 5%

Level 2
min support = 5%



Reduced Support

- Reduced minimum support threshold at lower levels

Level 1
min support = 5%

Milk
[support = 10%]

Level 2
min support = 3%

2% Milk
[support = 6%]

Skim Milk
[support = 4%]

Mining Multilevel: Top-Down Progressive Deepening

- Find multilevel frequent itemsets
  - High-level frequent itemsets

    milk (15%), bread (10%)
  - Lower-level “weaker” frequent itemsets

    2% milk (5%), wheat bread (4%)

- Generate multilevel association rules
  - High-level strong rules

    milk → bread [8%, 70%]
  - Lower-level “weaker”rules:

    2% milk → wheat bread [2%, 72%]

Multi-level Association: Redundancy Filtering

- Some rules may be redundant due to “ancestor” relationships between items.
- Example
  - milk $\Rightarrow$ wheat bread [support = 8%, confidence = 70%]
  - 2% milk $\Rightarrow$ wheat bread [support = 2%, confidence = 72%]
- We say the first rule is an ancestor of the second rule.
- A rule is redundant if its support is close to the “expected” value, based on the rule’s ancestor.

Mining Multidimensional Association Rules

Multidimensional Association Rules

Association rules that involve two or more dimensions or predicates can be referred to as multidimensional association rules.

- Single-dimensional rules

  buys(X, "milk") → buys(X, "bread")

- Multidimensional rules(≥2 dimensions/predicates)
  - Inter-dimension assoc. rules (no repeated predicates)
  - *age(X,"19-25") ∧ occupation(X,"student") → buys(X, "coke")*
  - Hybrid-dimension assoc. rules (repeated predicates)

    age(X,"19-25") ∧ buys(X, "popcorn") → buys(X, "coke")

Multidimensional Association Rules

- Multidimensional association rules with no repeated predicates are called interdimensional association rules.

- Multidimensional association rules with repeated predicates, which contain multiple occurrences of some predicates. These rules are called hybrid-dimensional association rules.

Multi-Dimensional Association : Concepts

- Single-dimensional rules:

  buys(X, “milk”) $\Rightarrow$ buys(X, “bread”)

- Multi-dimensional rules: ❍ 2 dimensions or predicates
  - Inter-dimension association rules (*no repeated predicates*)

    age(X,”19-25”) $\wedge$ occupation(X,“student”) $\Rightarrow$ buys(X,“coke”)
  - hybrid-dimension association rules (*repeated predicates*)

    age(X,”19-25”) $\wedge$ buys(X, “popcorn”) $\Rightarrow$ buys(X, “coke”)
- Categorical Attributes
  - finite number of possible values, no ordering among values
- Quantitative Attributes
  - numeric, implicit ordering among values

Techniques for Mining MD Associations

- Search for frequent *k*-predicate set:
  - Example: {age, occupation, buys} is a 3-predicate set.
  - Techniques can be categorized by how age are treated.

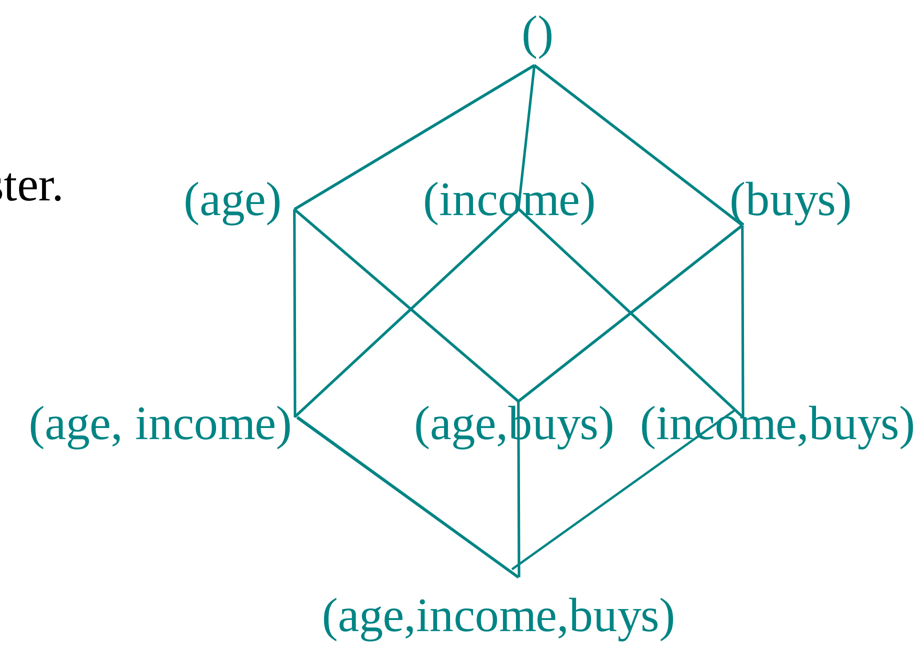
1. Using static discretization of quantitative attributes
   - Quantitative attributes are statically discretized by using predefined concept hierarchies.

2. Quantitative association rules
   - Quantitative attributes are dynamically discretized into “bins”based on the distribution of the data.

3. Distance-based association rules
   - This is a dynamic discretization process that considers the distance between data points.

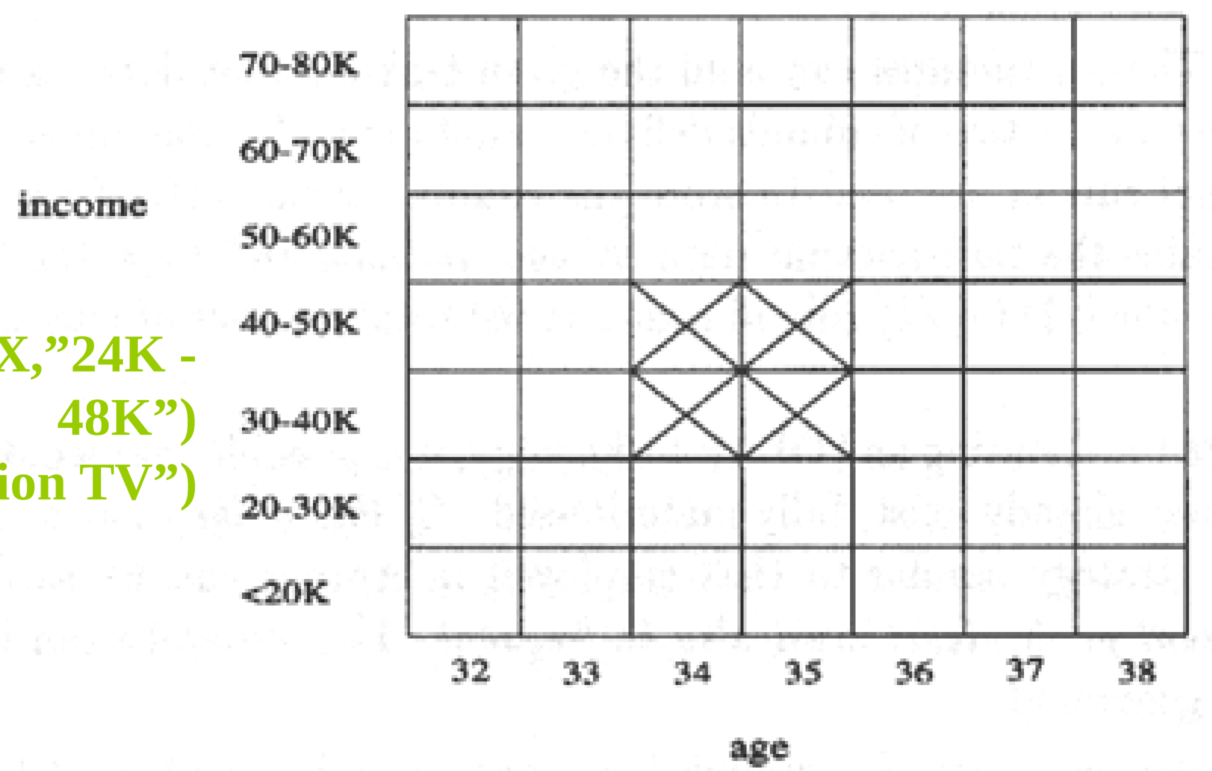
Static Discretization of Quantitative Attributes

- Discretized prior to mining using concept hierarchy.
- Numeric values are replaced by ranges.
- In relational database, finding all frequent k-predicate sets will require *k* or *k*+1 table scans.
- Data cube is well suited for mining.
- The cells of an n-dimensional
- cuboids correspond to the predicate sets.
- Mining from data cubes can be much faster.



Quantitative Association Rules

- Numeric attributes are *dynamically* discretized
  - Such that the confidence or compactness of the rules mined is maximized.
- 2-D quantitative association rules: $A_{quan1} \wedge A_{quan2} \Rightarrow A_{cat}$
- Cluster "adjacent" association rules to form general rules using a 2-D grid.
- Example:



age(X,"30-34") $\wedge$ income(X,"24K - 48K") $\Rightarrow$ buys(X,"high resolution TV")

End of Unit-II part-2