CBA: Classification Based on Association

- Basic Concept
  - Let X = {a set of items}, Y = class_label
  - Ruleitem:
    - Notation: X-> Y
  - condsupCount:
    - The number of cases in the database that contain X
  - rulesupCount:
    - The number of cases in the database that contain the ruleitem (X -> Y)
  - Support
    - The frequency of (X -> Y) in the database
      - (rulesupCount / total number of cases in the database) * 100%
  - Confidence
    - The frequency of (X -> Y) in the cases that contain X
      - (rulesupCount / condsupCount) * 100%

CBA: Classification Based on Association

- Basic Concept

  Example:

A	B	C	D	E	Class
0	0	1	1	0	Y
0	0	0	1	1	N
0	1	1	1	0	Y
1	1	1	1	0	Y
0	1	0	0	1	N

- Ruleitem:
  - (C -> Y)
- Support
  - (C -> Y): 3 / 5
- Confidence
  - (C -> Y): 3 / 3

CBA: Classification Based on Association

- Basic Concept

  Example:

A	B	C	D	E	Class
0	0	1	1	0	Y
0	0	0	1	1	N
0	1	1	1	0	Y
1	1	1	1	0	Y
0	1	0	0	1	N

- Ruleitem:
  - (D -> Y)
  - (D -> N)
- Support
  - (D -> Y): 3 / 5
  - (D -> N): 1 / 5
- Confidence
  - (D -> Y): 3 / 4
  - (D -> N): 1 / 4

CBA: Classification Based on Association

- Rule Generation (CARs)
  - Frequent: the support is above *minsup*
  - Accurate: the confidence is above *minconf*
  - Rules are sorted based on
    1. Confidence
    2. Support
    3. The order of generation
- Build a classifier

CBA: Classification Based on Association

- Example

A	B	C	D	E	Class
0	0	1	1	0	Y
0	0	0	1	1	N
0	1	1	1	0	Y
1	1	1	1	0	Y
0	1	0	0	1	N

Min support = 40%;
Min conf = 50%

1- candidates

RuleItem	Confidence	Support
A -> Y	1/1 = 100%	1/5=20 %
B -> Y	2/3 = 67%	2/5 = 40%
B -> N	1/3 = 33%	1/5 = 20%
C -> Y	3/3 = 100%	3/5 = 60%
D -> Y	3/4 = 75%	3/5 = 60%
D -> N	1/4 = 25%	1/5 = 20%
E -> N	2/2 = 100%	2/5 = 40%

1- frequent & accurate

RuleItem	Confidence	Support
B -> Y	2/3 = 67%	2/5 = 40%
C -> Y	3/3 = 100%	3/5 = 60%
D -> Y	3/4 = 75%	3/5 = 60%
E -> N	2/2 = 100%	2/5 = 40%

CBA: Classification Based on Association

- Example

A	B	C	D	E	Class
0	0	1	1	0	Y
0	0	0	1	1	N
0	1	1	1	0	Y
1	1	1	1	0	Y
0	1	0	0	1	N

Min support = 40%;
Min conf = 50%

2- candidates

RuleItem	Confidence	Support
BC -> Y	2/2 = 100%	2/5 = 40%
BD -> Y	2/2 = 100%	2/5 = 40%
CD -> Y	3/3 = 100%	3/5 = 60%

2- frequent & accurate

RuleItem	Confidence	Support
BC -> Y	2/2 = 100%	2/5 = 40%
BD -> Y	2/2 = 100%	2/5 = 40%
CD -> Y	3/3 = 100%	3/5 = 60%

CBA: Classification Based on Association

- Example

A	B	C	D	E	Class
0	0	1	1	0	Y
0	0	0	1	1	N
0	1	1	1	0	Y
1	1	1	1	0	Y
0	1	0	0	1	N

Min support = 40%;
Min conf = 50%

3- candidates

RuleItem	Confidence	Support
BCD -> Y	2/2 = 100%	2/5 = 40%

3- frequent & accurate

RuleItem	Confidence	Support
BCD -> Y	2/2 = 100%	2/5 = 40%

CBA: Classification Based on Association

- Example frequent & accurate rules (CARs):

A	B	C	D	E	Class
0	0	1	1	0	Y
0	0	0	1	1	N
0	1	1	1	0	Y
1	1	1	1	0	Y
0	1	0	0	1	N

Min support = 40%;
Min conf = 50%

Order By
•Confidence
•Support
•Order of generation

RuleItem	Confidence	Support
B -> Y	2/3 = 67%	2/5 = 40%
C -> Y	3/3 = 100%	3/5 = 60%
D -> Y	3/4 = 75%	3/5 = 60%
E -> N	2/2 = 100%	2/5 = 40%
BC -> Y	2/2 = 100%	2/5 = 40%
BD -> Y	2/2 = 100%	2/5 = 40%
CD -> Y	3/3 = 100%	3/5 = 60%
BCD -> Y	2/2 = 100%	2/5 = 40%

RuleItem	Confidence	Support
C -> Y	3/3 = 100%	3/5 = 60%
CD -> Y	3/3 = 100%	3/5 = 60%
E -> N	2/2 = 100%	2/5 = 40%
BC -> Y	2/2 = 100%	2/5 = 40%
BD -> Y	2/2 = 100%	2/5 = 40%
BCD -> Y	2/2 = 100%	2/5 = 40%
D -> Y	3/4 = 75%	3/5 = 60%
B -> Y	2/3 = 67%	2/5 = 40%

CBA: Classification Based on Association

- Example

A	B	C	D	E	Class
0	0	1	1	0	Y
0	0	0	1	1	N
0	1	1	1	0	Y
1	1	1	1	0	Y
0	1	0	0	1	N

Min support = 40%;
Min conf = 50%

CARs:

RuleItem	Confidence	Support
C -> Y	3/3 = 100%	3/5 = 60%
CD -> Y	3/3 = 100%	3/5 = 60%
E -> N	2/2 = 100%	2/5 = 40%
BC -> Y	2/2 = 100%	2/5 = 40%
BD -> Y	2/2 = 100%	2/5 = 40%
BCD -> Y	2/2 = 100%	2/5 = 40%
D -> Y	3/4 = 75%	3/5 = 60%
B -> Y	2/3 = 67%	2/5 = 40%

Classifiers:

	Default Class	Overall Accuracy
✓	N	100%

Classifiers:
C -> Y
Default: N