Pruning Redundant Association Rules Using Maximum Entropy Principle

Szymon Jaroszewicz and Dan A. Simovici

University of Massachusetts at Boston

email: {sj,dsim}@cs.umb.edu

Motivation

`contact-lenses` dataset

- 5 attributes
- 24 rows
- easy to analyze ‘manually’

Rules for `lenses` *data*

Manually selected rules:

- tears=reduced $\rightarrow$ no lenses
- astigmatism=no,tears=normal $\rightarrow$ soft lenses
- astigmatism=yes,tears=normal $\rightarrow$ hard lenses

- age=pre-presbyopic,prescription=hypermetrope,astigmatism=yes $\rightarrow$ none
- age=presbyopic,prescription=myope,astigmatism=no $\rightarrow$ none
- age=presbyopic,prescription=hypermetrope,astigmatism=yes $\rightarrow$ none

Use `Apriori`

minimum support: 4.2%
no minimum confidence

- 113 rules with consequent `lenses`
- 890 rules total

minimum confidence: 50%:

- 86 rules with consequent `lenses`
- 487 rules total

Need rule pruning!

Current approaches

Each subrule of a rule considered separately

Rule

$$AB \to Y$$

is not interesting if

$$\mathrm{conf}(AB \to Y) \approx \mathrm{conf}(A \to Y)$$

or

$$\mathrm{conf}(AB \to Y) \approx \mathrm{conf}(B \to Y)$$

Example I

assoc. rule	confidence
$\emptyset \to Y$	0.5
$A \to Y$	0.3
$B \to Y$	0.7

Suppose: $\mathrm{conf}(AB \to Y) = 0.3$

Is $AB \to Y$ interesting?

Example II

Apply pruning based on a single subrule

Result: $AB \to Y$ not interesting because

$$\mathrm{conf}(AB \to Y) = \mathrm{conf}(A \to Y)$$

But $AB \to Y$ tells us that:

> “If A and B are both present A’s influence is stronger”

and so it is interesting

Our approach:

take all subrules into consideration

- Rule $r : X \to Y$ introduces constraints

$$P(XY) = \mathrm{supp}(r), P(X) = \mathrm{supp}(r)/\mathrm{conf}(r)$$

- For $A \to Y$, let $\mathcal{C}$: set of constraints of all its subrules.
- Find distribution $P^{\mathcal{C}}$ over $A \cup Y$ such that
  - All constraints in $\mathcal{C}$ are satisfied
  - Entropy of $P^{\mathcal{C}}$ is maximal **(Maximum Entropy Principle)**

Our approach

- Estimate $\mathrm{conf}^{\mathcal{C}}(A \to Y)$ based on $P^{\mathcal{C}}$.
- Interestingness of $A \to Y$

  inter$(A \to Y) = |\mathrm{conf}(A \to Y) - \mathrm{conf}^{\mathcal{C}}(A \to Y)|$
- If inter$(A \to Y) < \epsilon$ prune $A \to Y$, since its explained by its subrules.

Example III

Subrules introduce constraints $\mathcal{C}$:

$\emptyset \to Y$, conf $= 0.5$

$$P(Y) = 0.5$$

$A \to Y$, supp $= 0.15$, conf $= 0.3$

$$P(A) = 0.5, P(AY) = 0.15$$

$B \to Y$, supp $= 0.35$, conf $= 0.7$

$$P(B) = 0.5, P(BY) = 0.35$$

Example IV

The MaxENT distribution is

$$P = \begin{pmatrix} 000 & 001 & 010 & 011 \\ 0.105 & 0.105 & 0.045 & 0.245 \end{pmatrix.$$

$$\begin{matrix} 100 & 101 & 110 & 111 \\ 0.245 & 0.045 & 0.105 & 0.105 \end{matrix}\Bigg),$$

So we expect

$$\mathrm{conf}^{\mathcal{C}}(AB \to Y) = 0.5$$

Example V

assoc. rule	confidence	
$\emptyset \to Y$	0.5	
$A \to Y$	0.3	
$B \to Y$	0.7	
$AB \to Y$	0.3	**interesting**
$AB \to Y$	0.5	**pruned**

Independent Attributes

- Suppose A is independent of B, Y, and jointly of BY
- Our method: $\mathrm{conf}^{\mathcal{C}}(AB \to Y) = \mathrm{conf}(B \to Y)$
  $AB \to Y$ not interesting
- Other methods:
  $\mathrm{conf}(AB \to Y) = \mathrm{conf}(B \to Y)$
  $AB \to Y$ not interesting
- Thats why those methods work well in practice

MaxENT computations

- **Iterative Proportional Scaling** - update the distribution until constraints are satisfied
- Can be slow but we are only dealing with small distributions
- Decomposition
- Special case: The only subrules present are $\emptyset \rightarrow Y$ and $X \rightarrow Y$.

`contact-lenses` *revisited*

antecedent $\rightarrow$ lenses
tears=reduced $\rightarrow$ none
astigmatism=no,tears=normal $\rightarrow$ soft
astigmatism=yes,tears=normal $\rightarrow$ hard
prescription=myope,astigmatism=yes $\rightarrow$ hard
prescription=myope,tears=normal $\rightarrow$ hard
prescription=hypermetrope,astigmatism=yes,tears=normal $\rightarrow$ none
age=pre-presb.,prescription=hypermetrope,astigmatism=yes $\rightarrow$ none
age=presbyopic,prescription=myope,astigmatism=no $\rightarrow$ none
age=presbyopic,prescription=hypermetrope,astigmatism=yes $\rightarrow$ none
... $\rightarrow$ ...

Elderly people data

antecedent → urban?	conf.[%]
$\emptyset$ → urban=no	22.4
$\emptyset$ → urban=yes	77.5
immigr=no,region=south → urban=yes	65.7
race=white → urban=yes	66.8
region=west → urban=yes	90.3
race=hisp → urban=yes	89.9
immigr=no,region=south → urban=no	34.2
alone=yes,region=south → urban=yes	66.9
immigr=before75 → urban=yes	93.4
... → ...	...

Number of rules

dataset	min. support	inter thrsh	number of rules Apriori	pruned	pruning time [s]
lenses	4%	0.3	890	40	1.3
mushroom*	16%	0.2	164125	5141	418
breast-cancer	10%	0.15	2128	74	2.8
primary-tumor*	9%	0.3	43561	67	21.8
primary-tumor*	9%	0.2	43561	432	24
car	0.5%	0.15	20669	580	30.2
splice*	9%	0.3	4847	95	5.6
elderly people	1%	0.1	247476	2056	4h:20m

* itemsets of up to 4 attributes

Further research

- Speed:
  - more special cases
  - use bounds to avoid MaxENT computations
- Methods for evaluating pruning correctness
  - domain experts
  - prediction accuracy