From association rules to interpretable classification models - a tutorial

Tomas Kliegr
Department of Information and Knowledge Engineering
Faculty of informatics and statistics
University of Economics, Prague
Outline

- Association rules
- Classification based on Association rules
- CBA algorithm
- Evaluation and comparison with other algorithms
- Extensions and implementations
- Summary
Outline

• Association rules
• Classification based on Association rules
• CBA algorithm
• Evaluation and comparison with other algorithms
• Extensions and implementations
• Summary
Association rules - introduction

• Serve for discovering interesting patterns in data
• Conjunctive rules
• Exhaustive - all rules are discovered that meet user-set pattern and constraints
• Initially developed for analysis of shopping baskets and recommendation.
• The most well-known algorithm is Apriori (Agrawal, 1994)

IF milk and diapers
THEN beer
Association rules – how they can be used

When customer buys item X, then he will also buy item Y
Outline

- Association rules
- Classification based on Association rules
- CBA algorithm
- Evaluation and comparison with other algorithms
- Extensions and implementations
- Summary
Association rules – importance

The Apriori algorithm was soon after its publication in 1994 considers as a breakthrough:

„ ... Association rules are among data mining’s biggest successes.“

_Hastie et al. Elements of Statistical Learning_
Association rules – use for classification

The Apriori algorithm was soon after its publication in 1994 considers as a breakthrough:

„... Association rules are among data mining’s biggest successes.“

Hastie et al. Elements of Statistical Learning

The contribution of the algorithm lied in the ability to process large multidimensional data in short time.
Association rules – use for classification

The Apriori algorithm was soon after its publication in 1994 considered as a breakthrough:

„... Association rules are among data mining’s biggest successes.“

*Hastie et al. Elements of Statistical Learning*

The contribution of the algorithm lied in the ability to process large multidimensional data in short time.

In 1998, the algorithm was adapted for the **classification task** in:

Outline

• Association rules
• Classification based on Association rules
• Algoritmus Classification based on Associations (CBA)
  • Data preparation
  • Training phases
  • Prediction
• Evaluation and comparison with other algorithms
• Extensions and implementations
• Summary
Illustration problem

Dataset contains historical data on worker’s comfort
- Two predictors: temperature (Y axis) and room humidity (X axis)
- One target attribute: worker’s comfort (1 = worst, 4 = best)

The dataset was designed to allow visualization in 2D
Classification based on Associations
principle of the CBA algorithm (Liu, 1998)

- Discretization
- Frequent item sets
- Association rules
- Classification rule lists
Classification based on Associations (CBA) only nominal attributes are on the input

- Algorithms for association rule mining accept only nominal attributes on the input.
- For discretization – conversion of numerical attributes to intervals – one typically uses equidistant method or the entropy-based MDLP algorithm (Fayyad, 93)
- Item is a tuple: \texttt{attribute=value} 
  \texttt{Humidity=(40;60]}
Classification based on Associations (CBA) support of item set

Discretization

Frequent item sets

Association rules

Classification rule lists

Item set = conjunction of conditions

Temp=(25;30] AND Hum=(40;60] AND Comf=4;

**support** = 3

**Minimum support:** algorithm finds all combinations of items, which are *frequent* - they appear in at least user-set minimum number of input rows.
Classification based on Associations (CBA)

classification of association rule

Discretization

Frequent item sets

Association rules

Classification rule lists

Temp=(25;30] AND Hum=(40;60] => Comf=4
Support = 3; Confidence = 0.6 = 3/5

Discovered rules must comply to user-set threshold for minimum confidence:

\[
\text{Conf}(X \rightarrow Y) = \frac{\text{Number of rows matching } X \text{ i } Y}{\text{Number of rows matching } X}
\]
Classification based on Associations (CBA) rules are created from frequent item sets

Discretization

Frequent item sets

Association rules

Classification rule lists

Discovered rules, colours – predicted comfort minimum confidence = 0.5

1 = red,  2 = green,  3 = unassigned,  4 = blue

\[
\begin{align*}
\{\text{Humidity}=(80;100]\} &\Rightarrow \{\text{Comfort}=1\} \\
\{\text{Temperature}=(30;35]\} &\Rightarrow \{\text{Comfort}=4\} \\
\{\text{Temperature}=(25;30],\text{Humidity}=(40;60]\} &\Rightarrow \{\text{Comfort}=4\} \\
\{\text{Temperature}=(15;20]\} &\Rightarrow \{\text{Comfort}=2\} \\
\{\text{Temperature}=(25;30]\} &\Rightarrow \{\text{Comfort}=4\}
\end{align*}
\]
Classification based on Associations (CBA)
the core of CBA is effective choice of rules

Discretization

Frequent item sets

Association rules

Classification rule lists

Part of the algorithm called Classifier Builder (CBA-CB) selects subset from input rules to create the output classifier.

```
R = sort(R);
foreach pravidlo r ∈ R do
  temp = ∅;
  foreach instance d ∈ D do
    if d splňuje podmínky r then
      ulož d.id v temp a označ r pokud správně klasifikuje d
    end
  end
  if r je označeno then
    vlož r na konec C;
    z D ostraň všechny instance jejichž id je v temp ;
    vyber výchozí třídu pro aktuální C;
    vypočíjte celkový počet chyb C;
  end
end
nalozí první pravidlo p v C které má nejnižší celkový počet chyb a z C vymaž všechna pravidla, která jsou pod p ;
přidej výchozí třídu asociovanou s p na konec C a vrať C;
```

Algorithm CBA-CB in version M1
Classification based on Associations (CBA) rule list is used to create the classifier.

- CBA achieves best result when rules are selected from at least 60,000 input rules.
- This number can be generated even on small dataset.
- The last rule in the classifier is called default rule (light green), it ensures that all conceivable instances are covered by the classifier.
Classification based on Associations (CBA)
use for prediction

- The first rule in the order of confidence, support and length (more general rules are preferred)

<table>
<thead>
<tr>
<th>Temperature</th>
<th>Humidity</th>
<th>Comfort</th>
</tr>
</thead>
<tbody>
<tr>
<td>27</td>
<td>48</td>
<td>?</td>
</tr>
</tbody>
</table>

### 1
lhs

{Humidity=(80;100] => {Comfort=1} 0.11 0.80 1

### 2
lhs

{Temperature=(30;35]} => {Comfort=4} 0.14 0.64 1

### 3
lhs

{Temperature=(25;30],Humidity=(40;60]} => {Comfort=4} 0.08 0.60 2

### 4
lhs

{Temperature=(15;20]} => {Comfort=2} 0.11 0.57 1

### 5
lhs

{Temperature=(25;30]} => {Comfort=4} 0.14 0.50 1

### 6
lhs

{} => {Comfort=2} 0.28 0.28 x
Classification based on Associations (CBA) use for prediction

- The first rule in the order of confidence, support and length (more general rules are preferred)

<table>
<thead>
<tr>
<th>Temperature</th>
<th>Humidity</th>
<th>Comfort</th>
</tr>
</thead>
<tbody>
<tr>
<td>27</td>
<td>48</td>
<td>4</td>
</tr>
</tbody>
</table>

## 

<table>
<thead>
<tr>
<th>rule</th>
<th>lhs</th>
<th>rhs</th>
</tr>
</thead>
<tbody>
<tr>
<td>[1]</td>
<td>{Humidity=(80;100]}</td>
<td>{Comfort=1}</td>
</tr>
<tr>
<td>[2]</td>
<td>{Temperature=(30;35]}</td>
<td>{Comfort=4}</td>
</tr>
<tr>
<td>[3]</td>
<td>{Temperature=(25;30],Humidity=(40;60]}</td>
<td>{Comfort=4}</td>
</tr>
<tr>
<td>[4]</td>
<td>{Temperature=(15;20]}</td>
<td>{Comfort=2}</td>
</tr>
<tr>
<td>[5]</td>
<td>{Temperature=(25;30]}</td>
<td>{Comfort=4}</td>
</tr>
<tr>
<td>[6]</td>
<td>{}</td>
<td>{Comfort=2}</td>
</tr>
</tbody>
</table>
Outline

• Association rules
• Classification based on Association rules
• CBA algorithm
• Evaluation and comparison with other algorithms
  • Association rule classification
  • Other rule-based classifiers and decision trees
  • Other frequently used classifiers
• Extensions and implementations
• Summary
Evaluation - other association classifiers

• In last 20 years multiple algorithms derived from CBA were proposed
• The design goal was typically achieving higher model accuracy, using one of the following methods:
  • Instead of classification with one strongest rule in CBA (single), some methods combine multiple rules to classify each instance
  • Instead of crisp rules in CBA, use probabilistic approach with fuzzy rules
  • CBA is a deterministic (det) algorithm, generating always the same output with given inputs. Some algorithms use stochastic methods, such as genetic or evolutorial algorithms.

Categories single, crisp and det are used to compare interpretability of algorithms on the next slide.
Evaluation - other association classifiers

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Year</th>
<th>Single</th>
<th>Crisp</th>
<th>Det</th>
<th>Assoc</th>
<th>Acc</th>
<th>Rules</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>CBA</td>
<td>1998</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>.80</td>
<td>185</td>
<td>35s</td>
</tr>
<tr>
<td>CBA 2</td>
<td>2001</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>.79</td>
<td>184</td>
<td>2 m</td>
</tr>
<tr>
<td>2SLAVE</td>
<td>2001</td>
<td>no?</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>.77</td>
<td>16</td>
<td>22m</td>
</tr>
<tr>
<td>CMAR</td>
<td>no</td>
<td>2001</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>.79</td>
<td>1419</td>
<td>6m</td>
</tr>
<tr>
<td>CPAR</td>
<td>no</td>
<td>2003</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>.82</td>
<td>788</td>
<td>11s</td>
</tr>
<tr>
<td>LAFAR</td>
<td>2003</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>.75*</td>
<td>47*</td>
<td>5h*</td>
</tr>
<tr>
<td>FH-GBML</td>
<td>2005</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>.77</td>
<td>11</td>
<td>3h</td>
</tr>
<tr>
<td>CFAR</td>
<td>2008</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>.71*</td>
<td>47*</td>
<td>17m*</td>
</tr>
<tr>
<td>SGERD</td>
<td>2008</td>
<td>no?</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>.74</td>
<td>7</td>
<td>3s</td>
</tr>
<tr>
<td>FARC-HD</td>
<td>2011</td>
<td>no?</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>.84</td>
<td>39</td>
<td>1h 20m</td>
</tr>
</tbody>
</table>

**single** denotes one rule classification

**crisp** do conditions in the rules comprising the classifier have crisp boundaries (as opposed to fuzzy)

**det.** Is algorithm deterministic without any random element, such as genetic algorithm

**assoc** is the algorithm based on association rules

**acc, rules, time** average accuracy, number of rules and train time on across 26 datasets in Alcala, 2011.
Evaluation - other association classifiers

<table>
<thead>
<tr>
<th>algorithm</th>
<th>year</th>
<th>single</th>
<th>crisp</th>
<th>det</th>
<th>assoc</th>
<th>acc</th>
<th>rules</th>
<th>time</th>
</tr>
</thead>
<tbody>
<tr>
<td>CBA</td>
<td>1998</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>.80</td>
<td>185</td>
<td>35s</td>
</tr>
<tr>
<td>CBA 2</td>
<td>2001</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>.79</td>
<td>184</td>
<td>2m</td>
</tr>
<tr>
<td>2SLAVE</td>
<td>2001</td>
<td>no?</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>.77</td>
<td>16</td>
<td>22m</td>
</tr>
<tr>
<td>CMAR</td>
<td>no</td>
<td>2001</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>.79</td>
<td>1419</td>
<td>6m</td>
</tr>
<tr>
<td>CPAR</td>
<td>no</td>
<td>2003</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>.82</td>
<td>788</td>
<td>11s</td>
</tr>
<tr>
<td>LAFAR</td>
<td>2003</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>.75*</td>
<td>47*</td>
<td>5h*</td>
</tr>
<tr>
<td>FH-GBML</td>
<td>2005</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>.77</td>
<td>11</td>
<td>3h</td>
</tr>
<tr>
<td>CFAR</td>
<td>2008</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>.71*</td>
<td>47*</td>
<td>17m*</td>
</tr>
<tr>
<td>SGERD</td>
<td>2008</td>
<td>no?</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>.74</td>
<td>7</td>
<td>3s</td>
</tr>
<tr>
<td>FARC-HD</td>
<td>2011</td>
<td>no?</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>.84</td>
<td>39</td>
<td>1h 20m</td>
</tr>
</tbody>
</table>

- Best algorithm FARC—HD, has on average 4% higher accuracy, but generates less understandable fuzzy rules
- CBA creates more understandable models than other algorithms for classification on the basis of association rules.
## Evaluation - other association classifiers

<table>
<thead>
<tr>
<th>dataset</th>
<th>RIP</th>
<th>J48</th>
<th>PART</th>
<th>FURIA</th>
<th>CBA</th>
</tr>
</thead>
<tbody>
<tr>
<td>anneal</td>
<td>0.94</td>
<td>0.94</td>
<td>0.95</td>
<td>0.99</td>
<td>0.96</td>
</tr>
<tr>
<td>australian</td>
<td>0.85</td>
<td>0.86</td>
<td>0.86</td>
<td>0.86</td>
<td>0.85</td>
</tr>
<tr>
<td>autos</td>
<td>0.79</td>
<td>0.79</td>
<td>0.78</td>
<td>0.78</td>
<td>0.79</td>
</tr>
<tr>
<td>breast-w</td>
<td>0.96</td>
<td>0.94</td>
<td>0.96</td>
<td>0.96</td>
<td>0.95</td>
</tr>
<tr>
<td>diabetes</td>
<td>0.75</td>
<td>0.74</td>
<td>0.74</td>
<td>0.75</td>
<td>0.76</td>
</tr>
<tr>
<td>glass</td>
<td>0.67</td>
<td>0.65</td>
<td>0.69</td>
<td>0.72</td>
<td>0.71</td>
</tr>
<tr>
<td>hepatitis</td>
<td>0.79</td>
<td>0.81</td>
<td>0.78</td>
<td>0.81</td>
<td>0.79</td>
</tr>
<tr>
<td>hypothyroid</td>
<td>0.99</td>
<td>1</td>
<td>0.99</td>
<td>1</td>
<td>0.98</td>
</tr>
<tr>
<td>ionosphere</td>
<td>0.91</td>
<td>0.87</td>
<td>0.88</td>
<td>0.89</td>
<td>0.92</td>
</tr>
<tr>
<td>iris</td>
<td>0.92</td>
<td>0.94</td>
<td>0.93</td>
<td>0.93</td>
<td>0.92</td>
</tr>
<tr>
<td>labor</td>
<td>0.88</td>
<td>0.71</td>
<td>0.84</td>
<td>0.74</td>
<td>0.84</td>
</tr>
<tr>
<td>lymph</td>
<td>0.77</td>
<td>0.74</td>
<td>0.78</td>
<td>0.87</td>
<td>0.81</td>
</tr>
<tr>
<td>sonar</td>
<td>0.74</td>
<td>0.68</td>
<td>0.73</td>
<td>0.79</td>
<td>0.74</td>
</tr>
<tr>
<td>vehicle</td>
<td>0.67</td>
<td>0.72</td>
<td>0.73</td>
<td>0.72</td>
<td>0.69</td>
</tr>
<tr>
<td>average</td>
<td>0.83</td>
<td>0.81</td>
<td>0.83</td>
<td>0.84</td>
<td>0.84</td>
</tr>
</tbody>
</table>

- CBA is fast and gives equally good result as other rule based classifiers, but it is often faster
- CBA generates more rules
Comparison with other classifiers

Based on:

Comparison with other classifiers

Based on:

Outline

• Association rules
• Classification based on Association rules
• CBA algorithm
• Evaluation and comparison with other algorithms
• Extensions and implementations
  • Reducing the size of the model
  • Combinatorial explosion and its solution
• Software
• Summary
Reducing number of rules on the output of CBA

- CBA generates more rules than other rule learning algorithms based on "separate and conquer"
- *Quantitative CBA* performs additional optimization of the list of rules generated by CBA
- It is based on recovering information lost during discretization
- QCBA achieves consistent reduction of model size by 50% without reduction of accuracy

CBA Drawbacks – Combinatorial explosion

Sensitivity to thresholds of minimum support and confidence

Let’s assume that input dataset contains $m$ attributes $A_1 \ldots A_m$

Let $K_{A_1}, \ldots, K_{A_m}$ denote number of unique values of each of $m$ attributes

- Number of combinations of length 1:
  \[
  \sum_{i=1}^{m} K_{A_i}
  \]

- Number of combinations of length 2:
  \[
  \sum_{i,j=1, i \neq j}^{m} (K_{A_i} \times K_{A_j})
  \]

- Total number of combinations:
  \[
  \prod_{j=1}^{m} (1 + K_{A_j}) - 1
  \]

Assume $m=70$ binary attributes

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>140</td>
</tr>
<tr>
<td></td>
<td>9660</td>
</tr>
<tr>
<td></td>
<td>$2.5 \times 10^{33}$</td>
</tr>
</tbody>
</table>

(Berka, 2003)
Solution to combinatorial explosion

Automatic tuning of metaparameters

- Incorrect setting of minimum confidence and support thresholds affects quality of classifier
- We can’t use grid search, because of the risk of combinatorial explosion

Solution 1: Generic algorithm

Implemented in R Package `rCBA`

Solution 2: Set of heuristics combined with „time outs“

Implemented in R Package `arc`

```plaintext
Algorithm 1: Mine predefined number of rules \texttt{topRules()}

Require: train training data, \texttt{targetRuleCount = 1000}, \texttt{initSupport = 0.00, initConf = 0.5, confStep = 0.05, suppStep = 0.05, minLen = 2, initMaxLen = 3, iterationTimeout = 2, totalTimeout = 100, maxIterations = 40}

Ensure: \texttt{rules : rule list}

1: \texttt{startTime$\leftarrow currentTime(), support$\leftarrow initSupport, conf$\leftarrow initConf, maxLen$\leftarrow initMaxLen, iterations$\leftarrow 0, maxLenDecreasedDueToTIMEOUT$\leftarrow false, lastRuleCount$\leftarrow -1}
2: \texttt{MAXRULELEN$\leftarrow$ number of explanatory attributes
3: \texttt{while do}
4: \texttt{iterations$\leftarrow iterations + 1
5: \texttt{if iterations$\leftarrow maxIterations then
6: \texttt{break}
7: \texttt{end if}
8: \texttt{rulesCurrent$\leftarrow apriiors(minLen,maxLen,support,conf,iterationTimeout)
9: \texttt{if rulesCurrent not finished within iterationTimeout then
10: \texttt{if currentTime(-startTime > totalTimeout then
11: \texttt{break}
12: \texttt{else if maxLen > minLen then
13: \texttt{maxLen$\leftarrow maxLen - 1
14: \texttt{maxLenDecreasedDueToTIMEOUT$\leftarrow true
15: \texttt{else
16: \texttt{break [All options exhausted]
17: \texttt{end if}
18: \texttt{else
19: \texttt{rules$\leftarrow rulesCurrent
20: \texttt{if rulesCount \geq targetRuleCount then
21: \texttt{break [Target rule count satisfied]
22: \texttt{else if currentTime() - startTime > totalTimeout then
23: \texttt{break [Max execution time exceeded]
24: \texttt{else if maxLen < MAXRULELEN and lastRuleCount \neq count(rules) and (maxLenDecreasedDueToTIMEOUT = false then
25: \texttt{maxLen$\leftarrow maxLen + 1
26: \texttt{lastRuleCount$\leftarrow count(rules)
27: \texttt{else if maxLen < MAXRULELEN and maxLenDecreasedDueToTIMEOUT = true and support \leq (1-suppStep) then
28: \texttt{support$\leftarrow support + suppStep
29: \texttt{maxLen$\leftarrow maxLen + 1
30: \texttt{lastRuleCount$\leftarrow rulesCount
31: \texttt{maxLenDecreasedDueToTIMEOUT$\leftarrow false
32: \texttt{else if conf > confStep then
33: \texttt{conf$\leftarrow conf - confStep
34: \texttt{else
35: \texttt{break [All options exhausted]
36: \texttt{end if}
37: \texttt{end if}
38: \texttt{end while}
39: \texttt{return first targetRuleCount rules from \texttt{rules}}
```
Availability of implementations

<table>
<thead>
<tr>
<th>software name</th>
<th>1st release</th>
<th>license</th>
<th>note</th>
</tr>
</thead>
<tbody>
<tr>
<td>arulesCBA</td>
<td>2016</td>
<td>GPL-3</td>
<td>from author of popular arules R package</td>
</tr>
<tr>
<td>rCBA</td>
<td>2015</td>
<td>Apache 2.0</td>
<td></td>
</tr>
<tr>
<td>arc</td>
<td>2016</td>
<td>AGPL-3</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>software name</th>
<th>1st release</th>
<th>license</th>
<th>note</th>
</tr>
</thead>
<tbody>
<tr>
<td>DM-II</td>
<td>2001?</td>
<td>commercial</td>
<td>original implementation Liu et al. (1998)</td>
</tr>
<tr>
<td>LUCS-KDD</td>
<td>2004</td>
<td>not stated</td>
<td>endorsed by author of original impl.</td>
</tr>
<tr>
<td>KEEL</td>
<td>2010?</td>
<td>GPLv3</td>
<td>not available in RKEEL</td>
</tr>
</tbody>
</table>

Software from our group:
- arc (R Package with CBA implementation)
- qCBA (postprocess CBA models with Quantitative CBA)
- EasyMiner (Web framework with user interface, with CBA backend)
Outline

• Association rules
• Classification based on Association rules
• CBA algorithm
• Evaluation and comparison with other algorithms
• Extensions and implementations
• Summary
Summary

• We introduced principles of association rule classification algorithms composed of association rules

• High number of input rules is a strength, but also a problem when not addressed
  + Candidate rules are fast to generate
  + High number of candidates to select from
  - Sensitivity to minimum support
  - More rules on the output than for other rule models

• There are multiple algorithms and implementations that reduce or remove these limitations

• Challenge is achieving the right balance between speed, explainability and accuracy of models
Publications


Thanks for your attention