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
Tomas Kliegr
Associate professor, Dep. of Information and Knowledge Engineering at VSE Praha

Software projects
• EasyMiner.eu – web-based rule learning system
• Inbeat.eu – rule-based recommender system
• R packages for building and explaining association rule classifiers (arc, qCBA, contributor: rCBA, arulesExplain – coming)

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

**IF milk and diapers THEN beer**

• 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)
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
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_
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 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.

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)

1. Discretization
2. Frequent item sets
3. Association rules
4. 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}
  \begin{align*}
  \text{Humidity} &= (40; 60] \\
  \text{Temperature} &= 30, 35, 40, 45
  \end{align*}
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)

Confidence of association rule

Discretization

Frequent item sets

Association rules

Classification rule lists

\[
\text{Temp}= (25;30] \text{ AND } \text{Hum} = (40;60] \implies \text{Comf}=4
\]

Support = 3; \textbf{Confidence} = 0.6 = 3/5

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

\[
\text{Conf}(X \rightarrow Y) = \frac{\text{Number of rows matching } X \text{ and } Y}{\text{Number of rows matching } X}
\]
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

\{\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\}
Classification based on Associations (CBA)

the core of CBA is effective choice of rules

Discretization

Frequent item sets

Association rules

Classification rule lists

Input: Discovered association rules
... recommended is 70,000 rules

Principle:
Rules selected using heuristics

Objectives
- Improve accuracy
- Reduce number of rules

Output: Classification rule list
... on average about 70 rules
Classification based on Associations (CBA)

rule list is used to create the classifier

Discretization

Frequent item sets

Association rules

Classification rule lists

The last rule in the classifier is called default rule (light green), it ensures that all conceivable instances are covered by the classifier.
The first rule in the order of confidence, support and length (more general rules are preferred)

## [1] {Humidity=(80;100]}                    => {Comfort=1} 0.11  0.80  1
## [2] {Temperature=(30;35]}                  => {Comfort=4} 0.14  0.64  1
## [3] {Temperature=(25;30],Humidity=(40;60]} => {Comfort=4} 0.08  0.60  2
## [4] {Temperature=(15;20]}                  => {Comfort=2} 0.11  0.57  1
## [5] {Temperature=(25;30]}                  => {Comfort=4} 0.14  0.50  1
## [6] {}                                     => {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>

```r
## [1] {Humidity=(80;100]} => {Comfort=1} 0.11 0.80 1
## [2] {Temperature=(30;35]} => {Comfort=4} 0.14 0.64 1
## [3] {Temperature=(25;30],Humidity=(40;60]} => {Comfort=4} 0.08 0.60 2
## [4] {Temperature=(15;20]} => {Comfort=2} 0.11 0.57 1
## [5] {Temperature=(25;30]} => {Comfort=4} 0.14 0.50 1
## [6] {} => {Comfort=2} 0.28 0.28 x
```
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 algorithm, generating always the same output with given inputs. Some algorithms use stochastic methods, such as genetic or evolitional 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>no</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</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 rule-based 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 (14)</td>
<td>0.94 (40)</td>
<td>0.95 (37)</td>
<td>0.99 (24)</td>
<td>0.96 (27)</td>
</tr>
<tr>
<td>australian</td>
<td>0.85 (4)</td>
<td>0.86 (9)</td>
<td>0.86 (6)</td>
<td>0.86 (9)</td>
<td>0.85 (109)</td>
</tr>
<tr>
<td>autos</td>
<td>0.79 (15)</td>
<td>0.79 (32)</td>
<td>0.78 (22)</td>
<td>0.78 (22)</td>
<td>0.79 (57)</td>
</tr>
<tr>
<td>breast-w</td>
<td>0.96 (6)</td>
<td>0.94 (10)</td>
<td>0.96 (10)</td>
<td>0.96 (16)</td>
<td>0.95 (51)</td>
</tr>
<tr>
<td>diabetes</td>
<td>0.75 (4)</td>
<td>0.74 (8)</td>
<td>0.74 (11)</td>
<td>0.75 (8)</td>
<td>0.76 (30)</td>
</tr>
<tr>
<td>glass</td>
<td>0.67 (8)</td>
<td>0.65 (15)</td>
<td>0.69 (16)</td>
<td>0.72 (15)</td>
<td>0.71 (28)</td>
</tr>
<tr>
<td>hepatitis</td>
<td>0.79 (4)</td>
<td>0.81 (4)</td>
<td>0.78 (6)</td>
<td>0.81 (8)</td>
<td>0.79 (32)</td>
</tr>
<tr>
<td>hypothyroid</td>
<td>0.99 (5)</td>
<td>1 (12)</td>
<td>0.99 (8)</td>
<td>1 (14)</td>
<td>0.98 (29)</td>
</tr>
<tr>
<td>ionosphere</td>
<td>0.91 (6)</td>
<td>0.87 (7)</td>
<td>0.88 (5)</td>
<td>0.89 (11)</td>
<td>0.92 (53)</td>
</tr>
<tr>
<td>iris</td>
<td>0.92 (4)</td>
<td>0.94 (4)</td>
<td>0.93 (5)</td>
<td>0.93 (5)</td>
<td>0.92 (6)</td>
</tr>
<tr>
<td>labor</td>
<td>0.88 (3)</td>
<td>0.71 (4)</td>
<td>0.84 (5)</td>
<td>0.74 (6)</td>
<td>0.84 (11)</td>
</tr>
<tr>
<td>lymph</td>
<td>0.77 (8)</td>
<td>0.74 (8)</td>
<td>0.78 (11)</td>
<td>0.87 (16)</td>
<td>0.81 (38)</td>
</tr>
<tr>
<td>sonar</td>
<td>0.74 (6)</td>
<td>0.68 (7)</td>
<td>0.73 (7)</td>
<td>0.79 (10)</td>
<td>0.74 (44)</td>
</tr>
<tr>
<td>vehicle</td>
<td>0.67 (21)</td>
<td>0.72 (44)</td>
<td>0.73 (35)</td>
<td>0.72 (24)</td>
<td>0.69 (147)</td>
</tr>
</tbody>
</table>

| average    | 0.83 (8)   | 0.81 (5)    | 0.83 (13)   | 0.84 (13)   | 0.84 (47)   |

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

- Neural networks, deep learning
- Random forest
- Support vector machines
- Decision trees

Interpretability (explainability, comprehensibility)

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
Combinatorial explosion

Recovering information lost during discretization

Quantitative CBA performs additional optimization of the list of rules generated by CBA. QCBA achieves consistent reduction of model size by 50% without reduction of accuracy.


Tutorial on QCBA: https://nb.vse.cz/~klit01/qcba/tutorial.html
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

<table>
<thead>
<tr>
<th>Assume</th>
</tr>
</thead>
<tbody>
<tr>
<td>m=70 binary attributes</td>
</tr>
</tbody>
</table>

<p>| |</p>
<table>
<thead>
<tr>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>140 rules of length 1</td>
</tr>
</tbody>
</table>

<p>| |</p>
<table>
<thead>
<tr>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>9660 rules of length 2</td>
</tr>
</tbody>
</table>

<p>| |</p>
<table>
<thead>
<tr>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>2.5 * 10^33 rules of length 70</td>
</tr>
</tbody>
</table>
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 arc</td>
<td>2015</td>
<td>Apache 2.0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2016</td>
<td>AGPL-3</td>
<td></td>
</tr>
<tr>
<td><strong>R implementations</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<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>
<tr>
<td>Other implementations</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Software from our group:
- arc (R Package with CBA implementation)
- pyARC (Python version of arc)
- 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 and challenge
  + 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

• Goal: achieving the right balance between speed, explainability and accuracy of models
Publications


Thanks for your attention