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

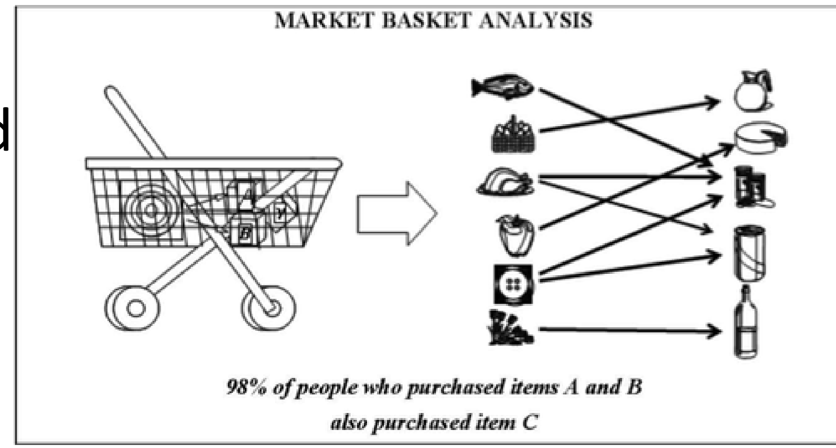
- Association rules
- Classification based on Association rules
- CBA algorithm
- Evaluation and comparison with other algorithms
- Extensions and implementations
- Summary

Outline

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

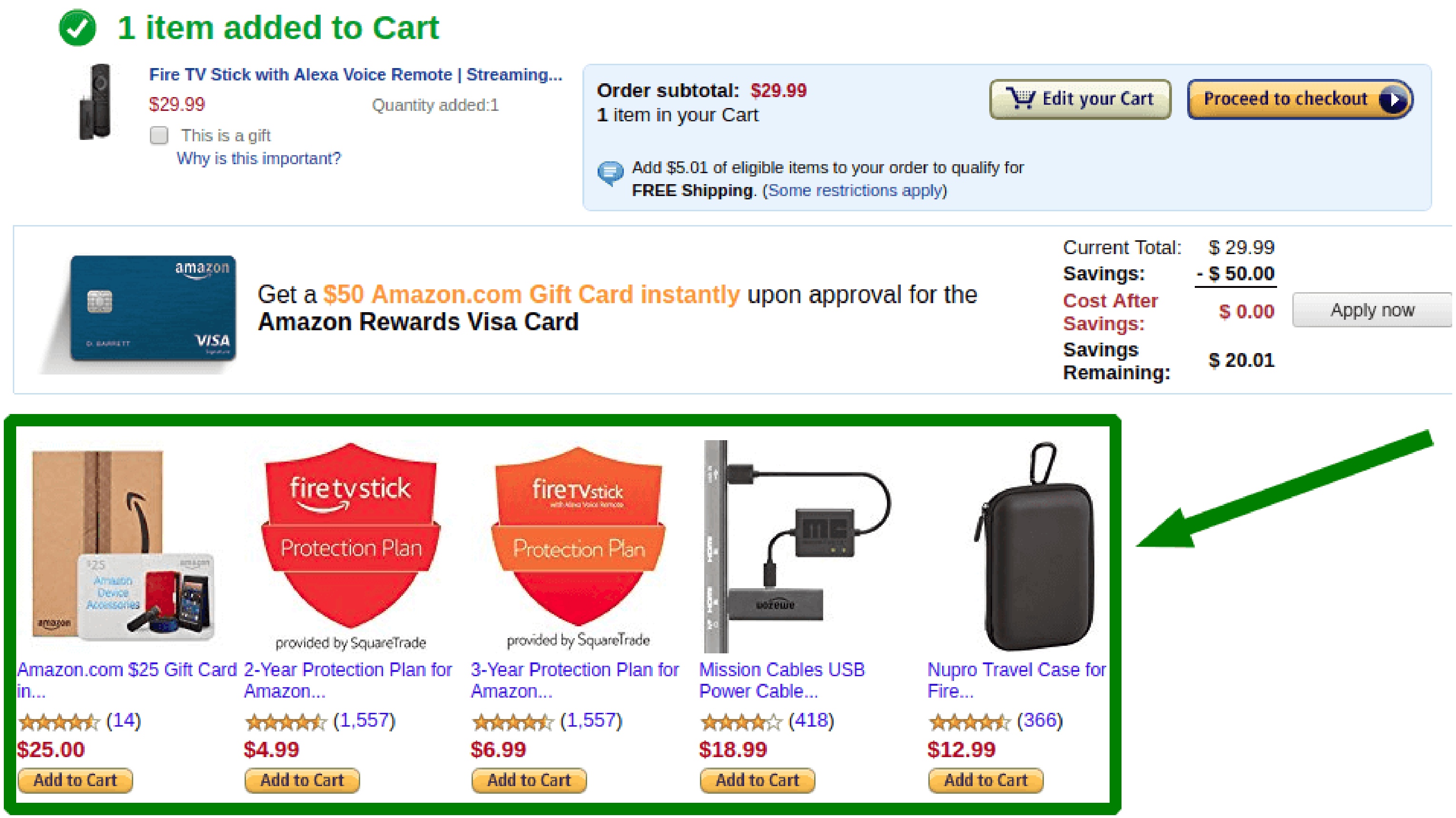
Association rules - introduction

- 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 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*:*

Bing Liu, Wynne Hsu, and Yiming Ma. 1998. Integrating classification and association rule mining. In Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining (KDD'98), Rakesh Agrawal and Paul Stolorz (Eds.). AAAI Press 80-86.

Outline

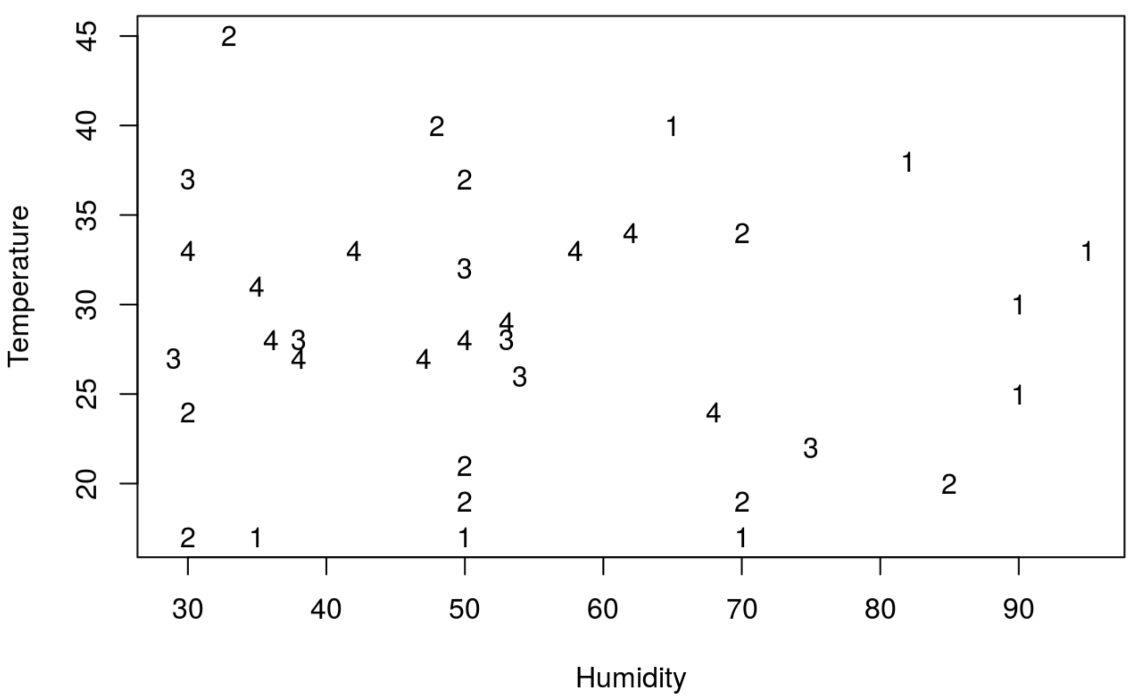
- Association rules
- Classification based on Association rules
- Algoritmus Classification based on Associations (CBA)
  - Data preparation
  - Training phases
  - Prediction
- Evaluation and comparison with other algorithms
- Extensions and implementations
- Summary

Illustration problem

Dataset contains historical data on worker's comfort

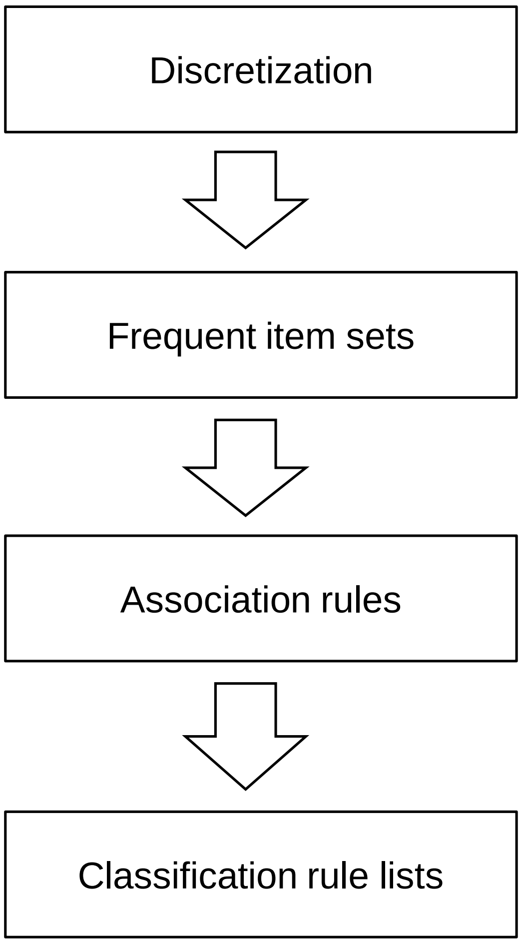
- Two predictors: temperature (Y axis) and room humidity (X axis)
- One target attribute: worker's comfort (1 = worst, 4 = best)

The dataset was designed to allow visualization in 2D



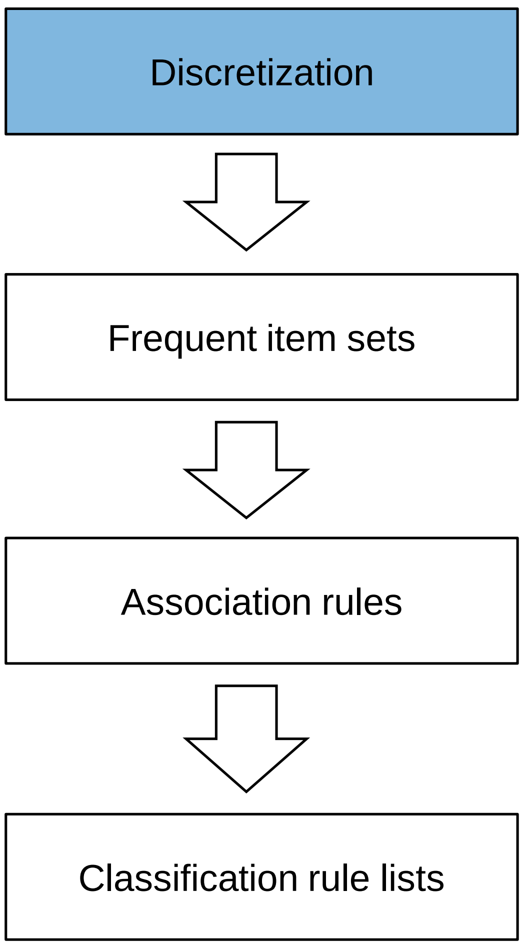
Classification based on Associations

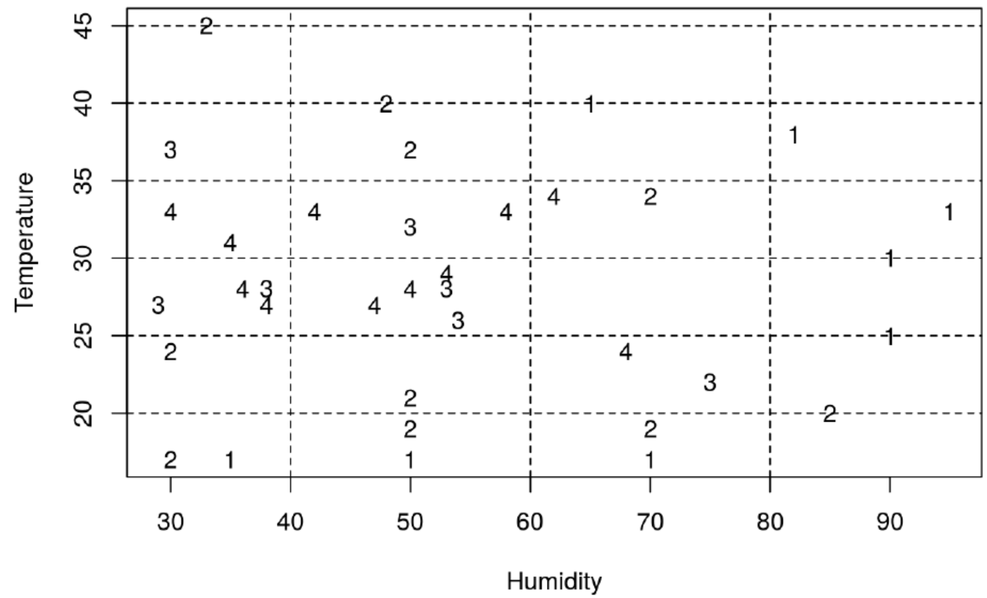
principle of the CBA algorithm (Liu, 1998)



Classification based on Associations (CBA)

only nominal attributes are on the input



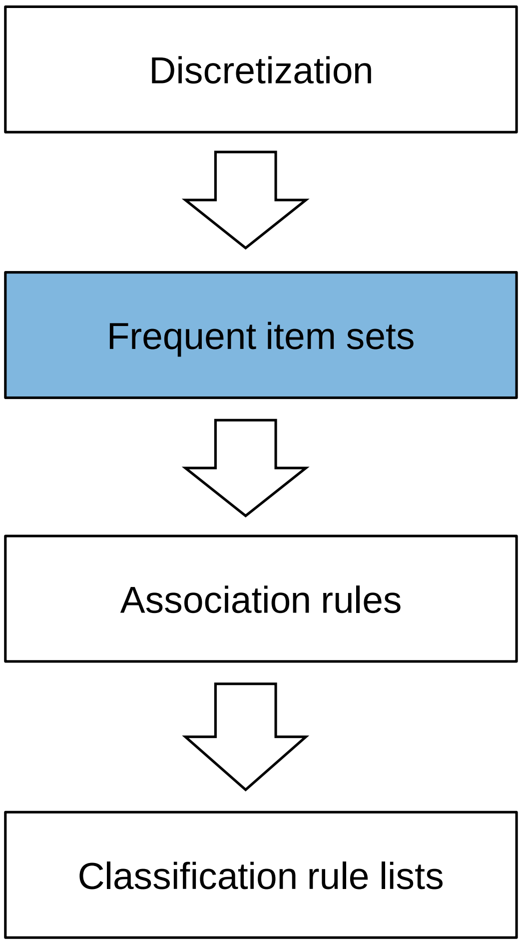


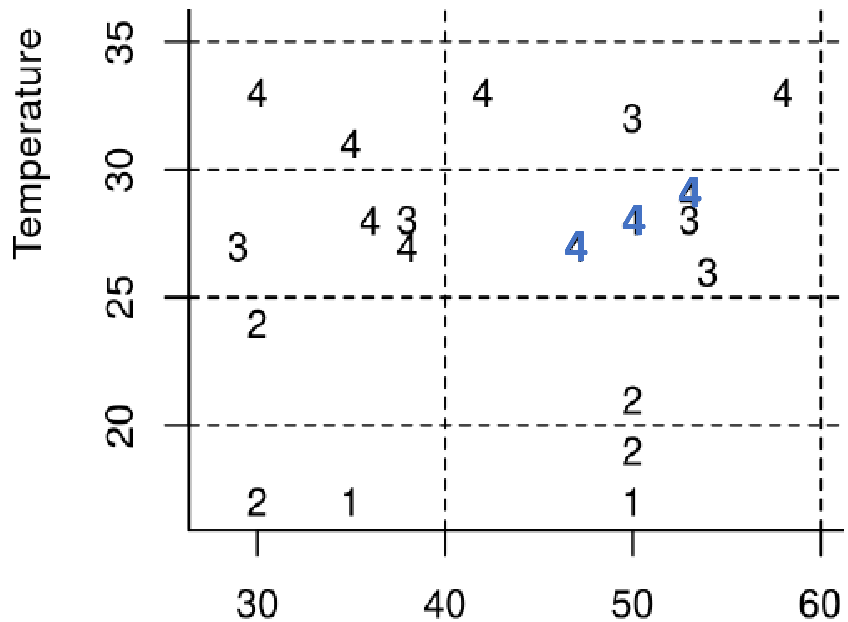
- Algorithms for association rule mining accept only nominal attributes on the input.
- For discretization – conversion of numerical attributes to intervals – one typically uses equidistant method or the entropy-based MDLP algorithm (Fayyad, 93)
- Item is a tuple: `attribute=value`

  `Humidity=(40;60]`

Classification based on Associations (CBA)

support of item set





Item set = conjunction of conditions

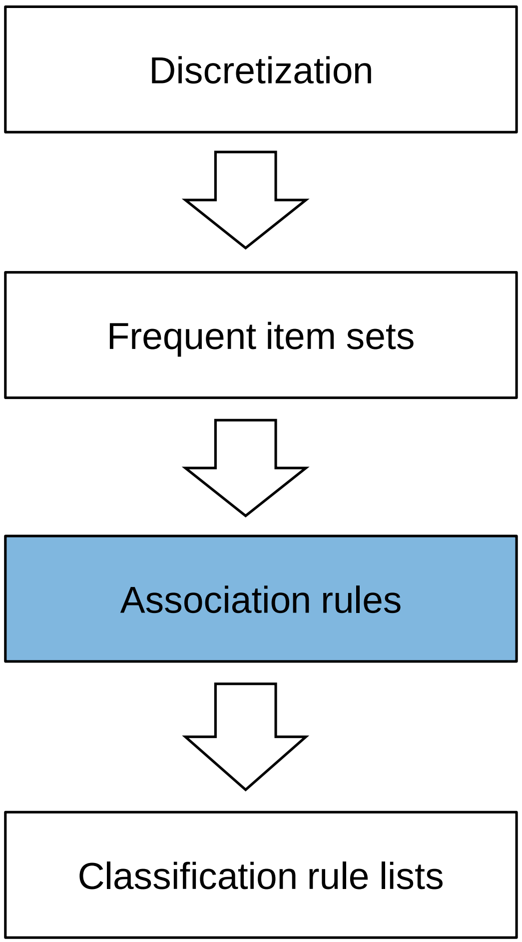
`Temp=(25;30] AND Hum=(40;60] AND Comf=4;`

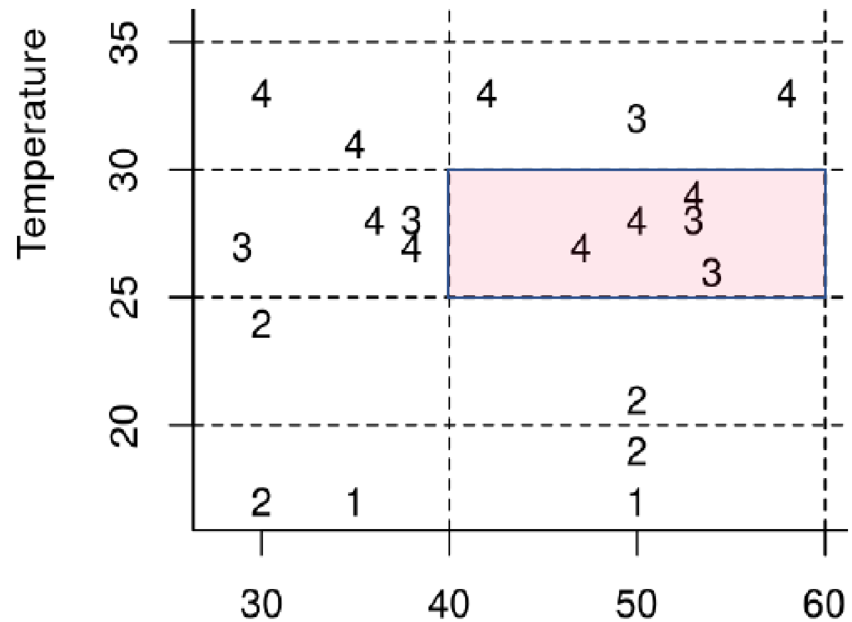
`support = 3`

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





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

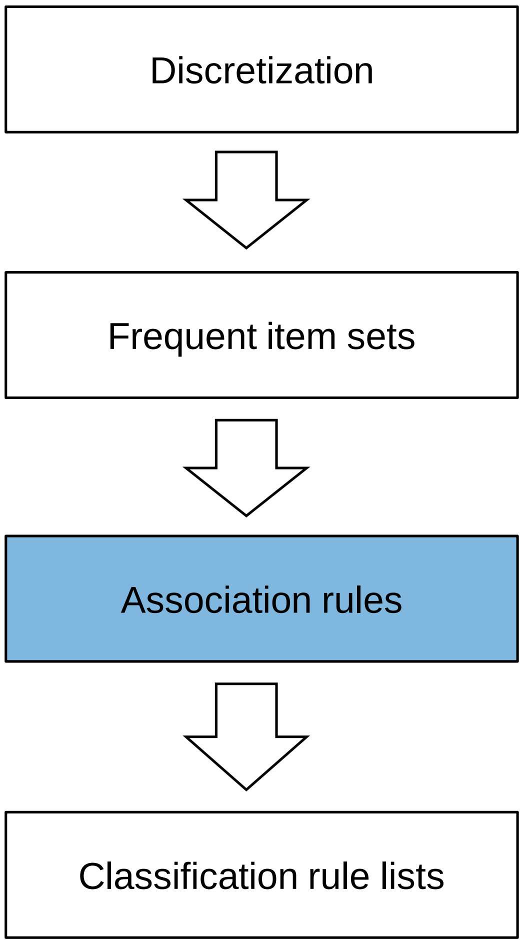
`Support = 3;` **`Confidence = 0.6 = 3/5`**

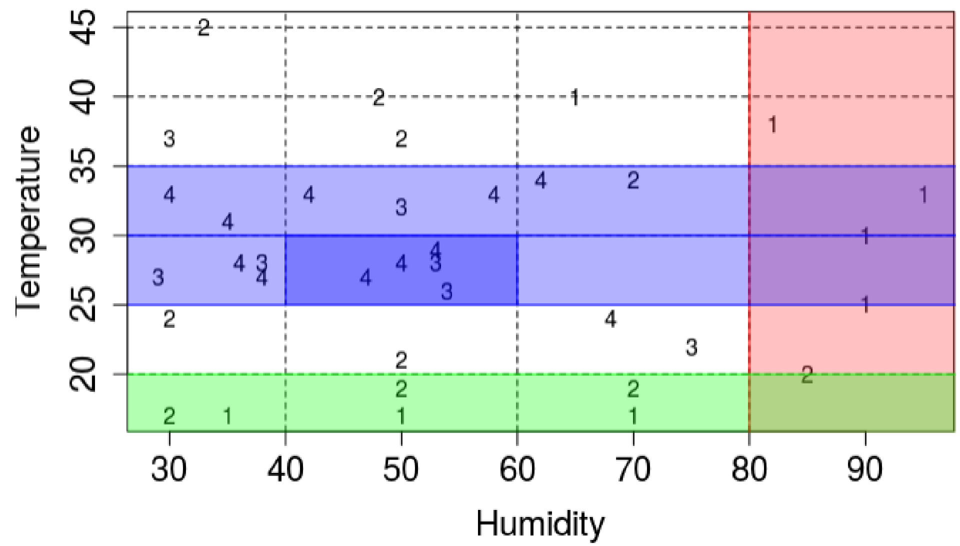
Discovered rules must comply to user-set threshold for **minimum confidence:**

$$\mathrm{Conf}(X \rightarrow Y) = \frac{\text{Number of rows matching } X \text{ i } Y}{\text{Number of rows matching } X}$$

Classification based on Associations (CBA)

rules are created from frequent item sets





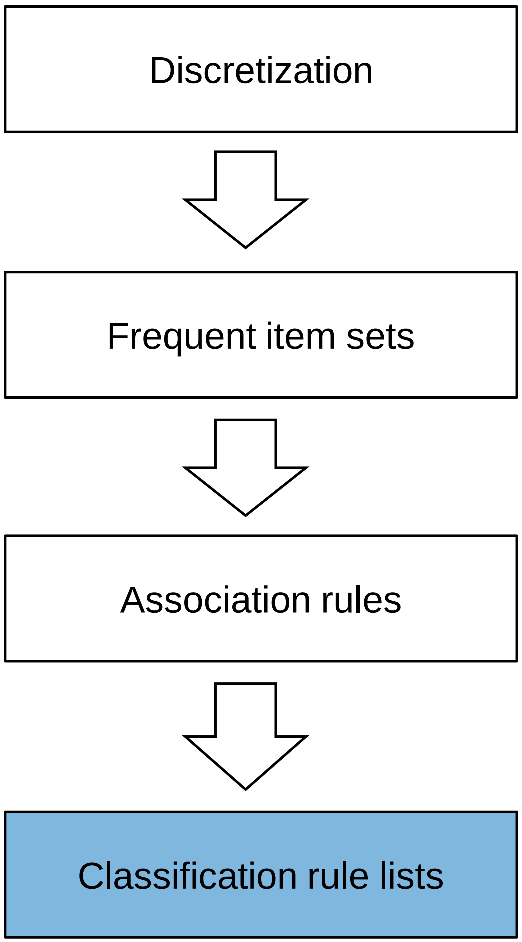
Discovered rules, colours – predicted comfort
minimum confidence = 0.5

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

```
{Humidity=(80;100]}                          => {Comfort=1}
{Temperature=(30;35]}                        => {Comfort=4}
{Temperature=(25;30],Humidity=(40;60]}=> {Comfort=4}
{Temperature=(15;20]}                        => {Comfort=2}
{Temperature=(25;30]}                        => {Comfort=4}
```

Classification based on Associations (CBA)

the core of CBA is effective choice of rules



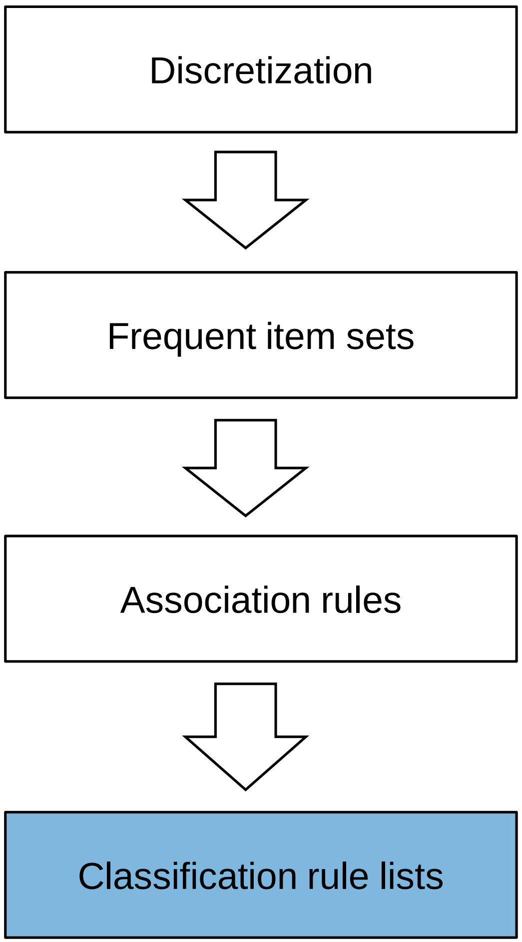
Part of the algorithm called Classifier Builder (CBA-CB) selects subset from input rules to create the output classifier.

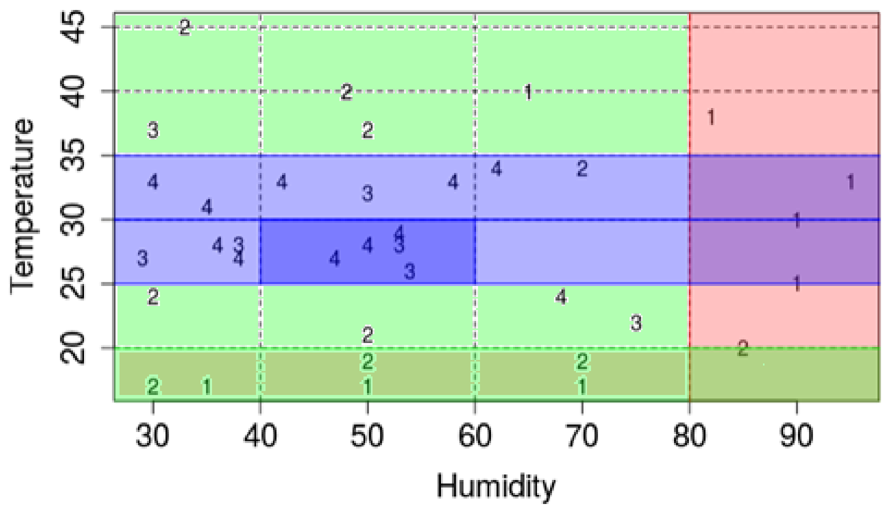
```
R = sort(R);
foreach pravidlo r ∈ R do
    temp = ∅;
    foreach instance d ∈ D do
        if d splňuje podmínky r then
            ulož d.id v temp a označ r pokud správně klasifikuje d
        end
    end
    if r je označeno then
        vlož r na konec C;
        z D ostraň všechny instance jejichž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

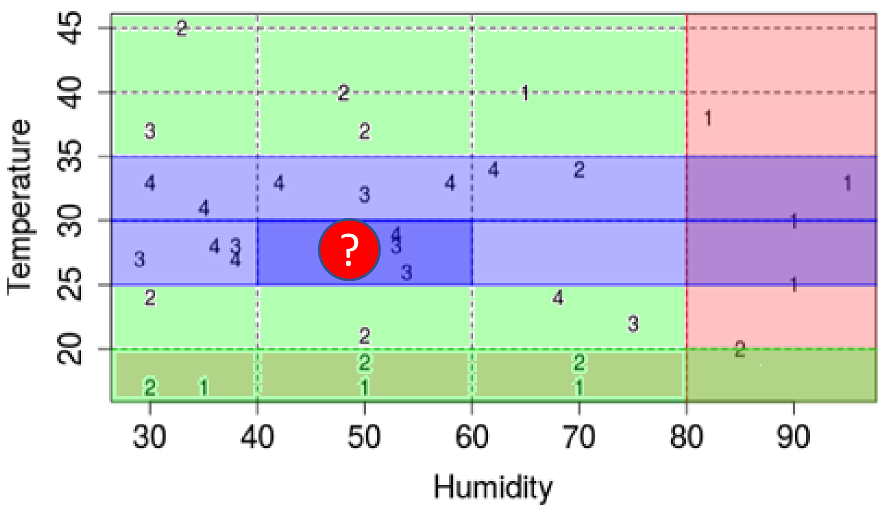




- CBA achieves best result when rules are selected from at least 60.000 input rules.
- This number can be generated even on small dataset.
- The last rule in the classifier is called default rule (light green), it ensures that all conceivable instances are covered by the classifier.

Classification based on Associations (CBA)

use for prediction



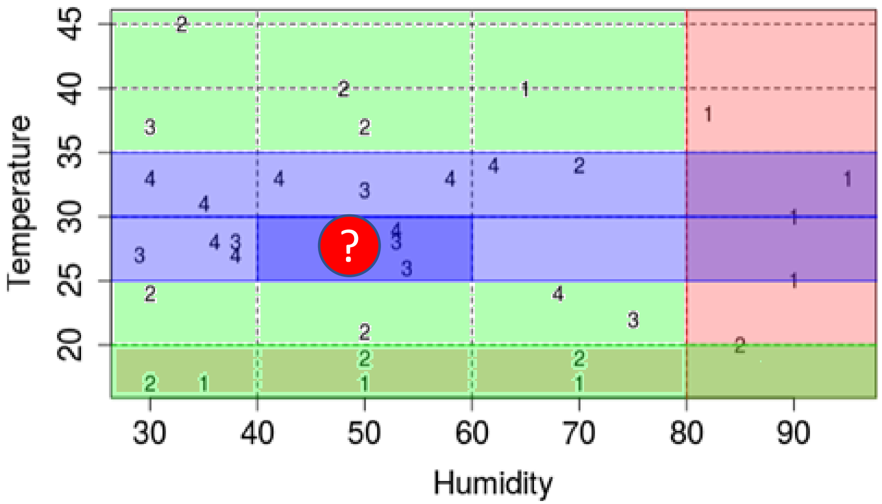
- The first rule in the order of confidence, support and length (more general rules are preferred)

Temperature	Humidity	Comfort
27	48	?

```
##     lhs                                           rhs           sup  conf  len
## [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)

Temperature	Humidity	Comfort
27	48	4

```
##     lhs                                   rhs            sup  conf  len
## [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 (**det**) algorithm, generating always the same output with given inputs. Some algorithms use stochastic methods, such as genetic or evolutional algorithms.

Categories **single**, **crisp** and **det** are used to compare interpretability of algorithms on the next slide.

Evaluation - other association classifiers

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

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

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

Zdroj: autor

- 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

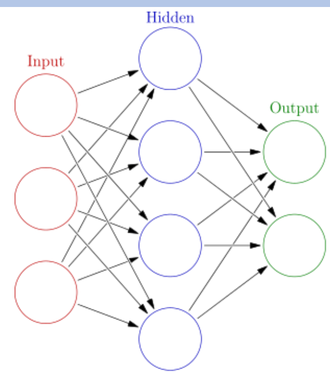
dataset	RIP	J48	PART	FURIA	CBA
anneal	0.94 (14)	0.94 (40)	0.95 (37)	0.99 (24)	0.96 (27)
australian	0.85 (4)	0.86 (9)	0.86 (6)	0.86 (9)	0.85 (109)
autos	0.79 (15)	0.79 (32)	0.78 (22)	0.78 (22)	0.79 (57)
breast-w	0.96 (6)	0.94 (10)	0.96 (10)	0.96 (16)	0.95 (51)
diabetes	0.75 (4)	0.74 (8)	0.74 (11)	0.75 (8)	0.76 (30)
glass	0.67 (8)	0.65 (15)	0.69 (16)	0.72 (15)	0.71 (28)
hepatitis	0.79 (4)	0.81 (4)	0.78 (6)	0.81 (8)	0.79 (32)
hypothyroid	0.99 (5)	1 (12)	0.99 (8)	1 (14)	0.98 (29)
ionosphere	0.91 (6)	0.87 (7)	0.88 (5)	0.89 (11)	0.92 (53)
iris	0.92 (4)	0.94 (4)	0.93 (5)	0.93 (5)	0.92 (6)
labor	0.88 (3)	0.71 (4)	0.84 (5)	0.74 (6)	0.84 (11)
lymph	0.77 (8)	0.74 (8)	0.78 (11)	0.87 (16)	0.81 (38)
sonar	0.74 (6)	0.68 (7)	0.73 (7)	0.79 (10)	0.74 (44)
vehicle	0.67 (21)	0.72 (44)	0.73 (35)	0.72 (24)	0.69 (147)
average	0.83 (8)	0.81 (5)	0.83 (13)	0.84 (13)	0.84 (47)

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

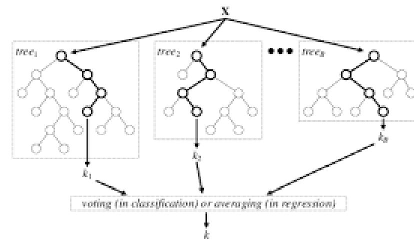
Comparison with other classifiers

Accuracy

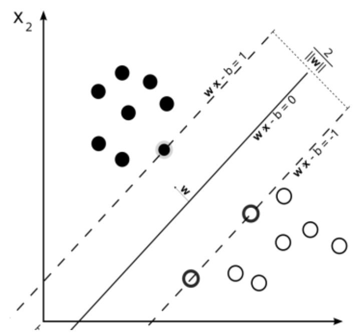
Neural networks, deep learning



Random forest



Support vector machines



Decision trees and rules

job(General Labourer) → survived(FALSE)
Confidence: 0.969 | Support: 0.071

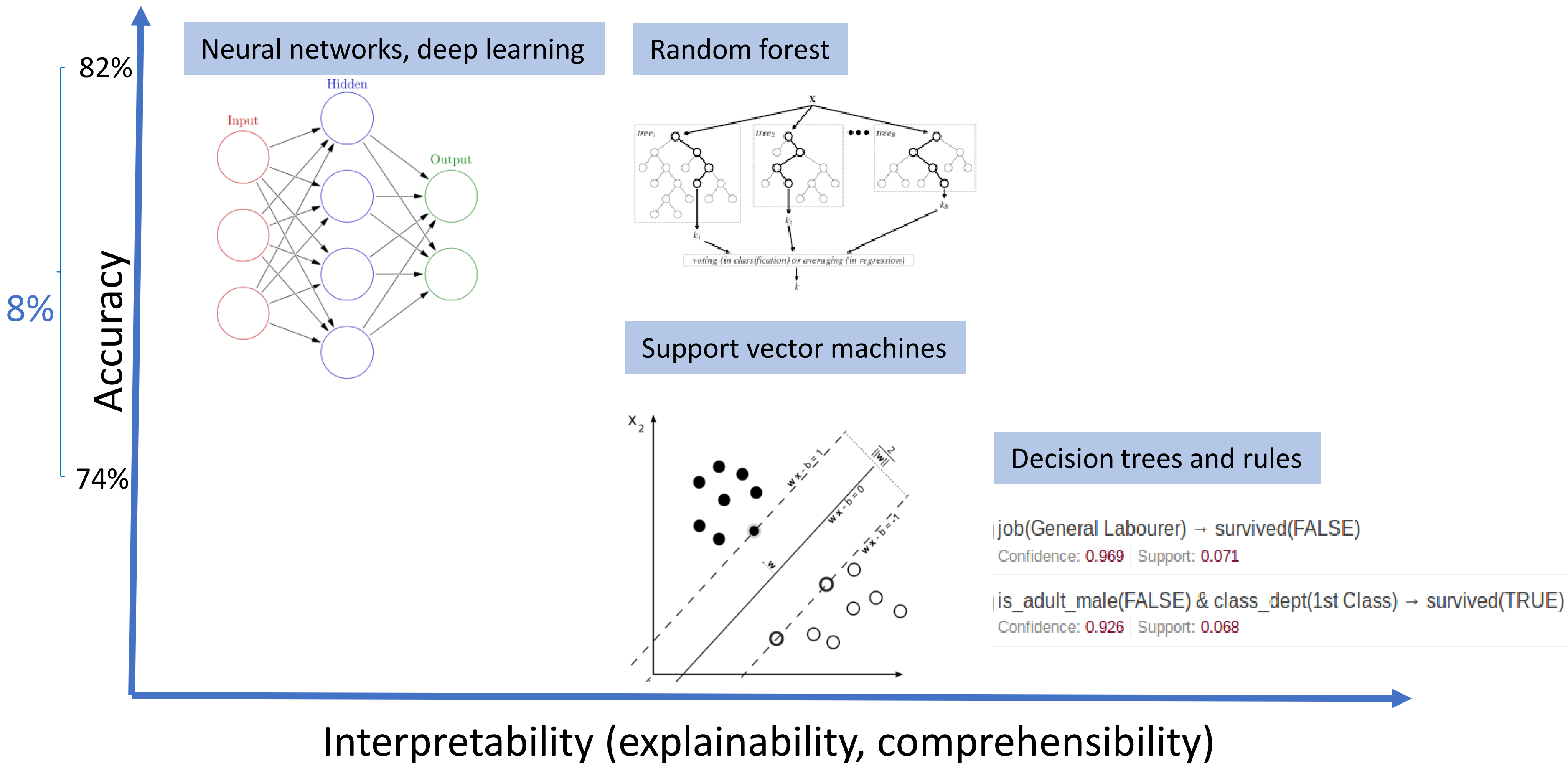
is_adult_male(FALSE) & class_dept(1st Class) → survived(TRUE)
Confidence: 0.926 | Support: 0.068

Interpretability (explainability, comprehensibility)

Based on:

Explainable Artificial Intelligence – Program Update, DARPA, US, 2017.

Comparison with other classifiers



Based on:

Explainable Artificial Intelligence – Program Update, DARPA, US, 2017.

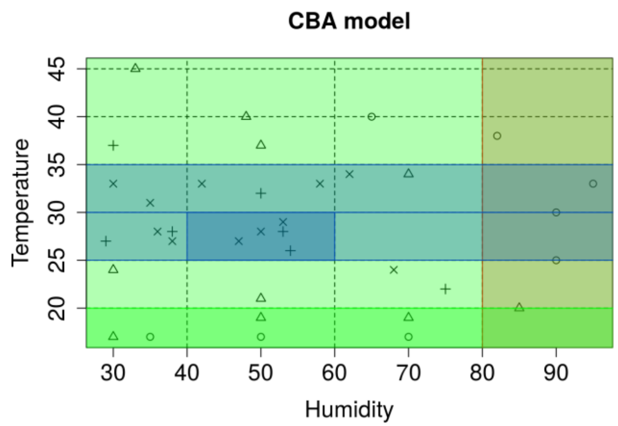
Fernández-Delgado, Manuel, et al. "Do we need hundreds of classifiers to solve real world classification problems?." The Journal of Machine Learning Research 15.1 (2014): 3133-3181.

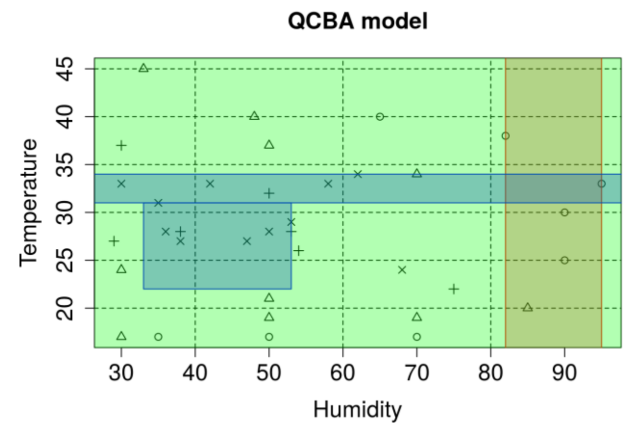
Outline

- 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





Kliegr, Tomas. "Quantitative CBA: Small and Comprehensible Association Rule Classification Models." arXiv preprint arXiv:1711.10166 (2017).

CBA Drawbacks – Combinatorial explosion

Sensitivity to thresholds of minimum support and confidence

Let's assume that input dataset contains *m* attributes $A_1 \dots A_m$

Let $K_{A1},\dots K_{Am}$ denote number of unique values of each of m attributes

- Number of combinations of length 1: $\sum_{i=1}^{m} K_{Ai}$
- Number of combinations of length 2: $\sum_{i,j=1,\, i \neq j}^{m} (K_{Ai} \times K_{Aj})$
- Total number of combinations: $\prod_{j=1}^{m} (1 + K_{Aj}) - 1$

Assume m=70 binary attributes
140
9660
2.5 * 10^33

(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

<u>Solution 1: Generic algorithm</u>

Implemented in R Package rCBA

<u>Solution 2: Set of heuristics</u> combined with „time outs"

Implemented in R Package arc

Algorithm 1 Mine predefined number of rules *topRules()*

Require: *train* training data, $targetRuleCount$ = 1000, $initSupport$ = 0.00, $initConf$ = 0.5, $confStep$ = 0.05, $suppStep$ = 0.05, $minLen$ = 2, $initMaxlen$ = 3, $iterationTimeout$ = 2, $totalTimeout$ = 100.0, $maxIterations$ = 40

Ensure: *rules* - rule list

```
startTime ← currentTime(), support ← initSupport, conf ← initConf, maxLen ← initMaxlen, iterations ← 0, maxLenDecreasedDueToTIMEOUT ← false, lastRuleCount ← -1
MAXRULELEN ← number of explanatory attributes
while
 do
  iterations ← iterations + 1
  if iterations = maxIterations then
    break
  end if
  rulesCurrent ← apriori(minLen,maxLen,support,conf,iterationTimeout)
  if apriori not finished within iterationTimeout then
    if currentTime()-startTime > totalTimeout then
      break
    else if maxLen > minLen then
      maxLen ← maxLen - 1
      maxLenDecreasedDueToTIMEOUT ← true
    else
      break {All options exhausted}
    end if
  else
    rules ← rulesCurrent
    if rulecount ≥ targetRuleCount then
      break {Target rule count satisfied}
    else if currentTime() - startTime > totalTimeout then
      break {Max execution time exceeded}
    else if maxLen < MAXRULELEN and lastRuleCount != count(rules) and (maxLenDecreasedDueToTIMEOUT = false) then
      maxLen ← maxLen + 1
      lastRuleCount ← count(rules)
    else if maxLen < MAXRULELEN and maxLenDecreasedDueToTIMEOUT = true and support ≤ (1-suppStep) then
      support ← support + suppStep
      maxLen ← maxLen + 1
      lastRuleCount ← rulecount
      maxLenDecreasedDueToTIMEOUT ← false
    else if conf > confStep then
      conf ← conf - confStep
    else
      break {All options exhausted}
    end if
  end if
end while
return first targetRuleCount rules from rules
```

Availability of implementations

software name	1st release	license	note
		R implementations	
arulesCBA	2016	GPL-3	from author of popular arules R package
rCBA	2015	Apache 2.0	
arc	2016	AGPL-3	
		Other implementations	
DM-II	2001?	commercial	original implementation Liu et al. (1998)
LUCS-KDD	2004	not stated	endorsed by author of original impl.
KEEL	2010?[7]	GPLv3	not available in RKEEL

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
  - - Sensititivity 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

- Fürnkranz, Johannes, and Tomáš Kliegr. "The Need for Interpretability Biases." International Symposium on Intelligent Data Analysis. Springer, Cham, 2018.
- Vojíř, S., Zeman, V., Kuchař, J., & Kliegr, T. (2018). EasyMiner. eu: Web framework for interpretable machine learning based on rules and frequent itemsets. Knowledge-Based Systems, 150, 111-115.
- Fürnkranz, Johannes, Tomáš Kliegr, and Heiko Paulheim. "On Cognitive Preferences and the Plausibility of Rule-based Models." arXiv preprint arXiv:1803.01316 (2018).
- Kliegr, Tomáš, Štěpán Bahník, and Johannes Fürnkranz. "A review of possible effects of cognitive biases on interpretation of rule-based machine learning models." arXiv preprint arXiv:1804.02969 (2018).
- Kliegr, Tomas. "Quantitative CBA: Small and Comprehensible Association Rule Classification Models." arXiv preprint arXiv:1711.10166 (2017).

Thanks for your attention