Tato prezentace je součástí wiki-prezentace Metoda GUHA, LISp-Miner a typové úlohy
Je dostupná z této adresy
Verse 20. 8. 2019

Typ úlohy: Stejná úloha v 4ft-Miner a arules v R

Data: Adult

Problém: *Nalézt všechna asociační pravidla $X \rightarrow Y$ tak, že $conf(X \rightarrow Y) \geq 0.95$ a $sup(X \rightarrow Y) \geq 0.45$.*

Jan Rauch
Katedra informačního a znalostního inženýrství
Vysoká škola ekonomická v Praze

© Jan Rauch

Stejná úloha ve 4ft-Miner a v arules v R

- Data Adult a vzorový příklad
- Vzorový příklad a arules v R-package
- Vzorový příklad a GUHA procedura 4ft-Miner

Data Adult

http://archive.ics.uci.edu/ml/datasets/Adult

UCI Machine Learning Repository

Center for Machine Learning and Intelligent Systems

About Citation Policy Donate a Data Set Contact

Search

Repository Web

View ALL Data Sets

Adult Data Set

Download: Data Folder, Data Set Description

Abstract: Predict whether income exceeds $50K/yr based on census data. Also known as "Census Income" dataset.

Data Set Characteristics:	Multivariate	**Number of Instances:**	48842	**Area:**	Social
Attribute Characteristics:	Categorical, Integer	**Number of Attributes:**	14	**Date Donated**	1996-05-01
Associated Tasks:	Classification	**Missing Values?**	Yes	**Number of Web Hits:**	647040

Source:

Donor:

Ronny Kohavi and Barry Becker
Data Mining and Visualization
Silicon Graphics.
e-mail: ronnyk '@' live.com for questions.

Data Set Information:

Extraction was done by Barry Becker from the 1994 Census database. A set of reasonably clean records was extracted using the following conditions: ((AAGE>16) && (AGI>100) && (AFNLWGT>1)&& (HRSWK>0))

Prediction task is to determine whether a person makes over 50K a year.

Data Adult – staženo

AdultUCI – Poznámkový blok

Soubor Úpravy Formát Zobrazení Nápověda

```
"33";45;"Private";386940;"Bachelors";13;"Divorced";"Exec-managerial";"Own-child";"White";"Male";0;1
"34";30;"Federal-gov";59951;"Some-college";10;"Married-civ-spouse";"Adm-clerical";"Own-child";"Whit
"35";22;"State-gov";311512;"Some-college";10;"Married-civ-spouse";"Other-service";"Husband";"Black"
"36";48;"Private";242406;"11th";7;"Never-married";"Machine-op-inspct";"Unmarried";"White";"Male";0;
"37";21;"Private";197200;"Some-college";10;"Never-married";"Machine-op-inspct";"Own-child";"White";
"38";19;"Private";544091;"HS-grad";9;"Married-AF-spouse";"Adm-clerical";"Wife";"White";"Female";0;0
"39";31;"Private";84154;"Some-college";10;"Married-civ-spouse";"Sales";"Husband";"White";"Male";0;0
"40";48;"Self-emp-not-inc";265477;"Assoc-acdm";12;"Married-civ-spouse";"Prof-specialty";"Husband";"
"41";31;"Private";507875;"9th";5;"Married-civ-spouse";"Machine-op-inspct";"Husband";"White";"Male";
"42";53;"Self-emp-not-inc";88506;"Bachelors";13;"Married-civ-spouse";"Prof-specialty";"Husband";"Wh
"43";24;"Private";172987;"Bachelors";13;"Married-civ-spouse";"Tech-support";"Husband";"White";"Male
"44";49;"Private";94638;"HS-grad";9;"Separated";"Adm-clerical";"Unmarried";"White";"Female";0;0;40;
"45";25;"Private";289980;"HS-grad";9;"Never-married";"Handlers-cleaners";"Not-in-family";"White";"M
"46";57;"Federal-gov";337895;"Bachelors";13;"Married-civ-spouse";"Prof-specialty";"Husband";"Black"
"47";53;"Private";144361;"HS-g                                                                  t";"Husband";"White";"Ma
"48";44;"Private";128354;"Mast                                                                  ed";"White";"Female";0;0
"49";41;"State-gov";101603;"As                                                                  ";"Husband";"White";"Ma
"50";29;"Private";271466;"Assoc-voc";11;"Never-married";"Prof-specialty";"Not-in-family";"White";"M
"51";25;"Private";32275;"Some-college";10;"Married-civ-spouse";"Exec-managerial";"Wife";"Other";"Fe
"52";18;"Private";226956;"HS-grad";9;"Never-married";"Other-service";"Own-child";"White";"Female";0
"53";47;"Private";51835;"Prof-school";15;"Married-civ-spouse";"Prof-specialty";"Wife";"White";"Fema
"54";50;"Federal-gov";251585;"Bachelors";13;"Divorced";"Exec-managerial";"Not-in-family";"White";"M
"55";47;"Self-emp-inc";109832;"HS-grad";9;"Divorced";"Exec-managerial";"Not-in-family";"White";"Mal
"56";43;"Private";237993;"Some-college";10;"Married-civ-spouse";"Tech-support";"Husband";"White";"M
"57";46;"Private";216666;"5th-6th";3;"Married-civ-spouse";"Machine-op-inspct";"Husband";"White";"Ma
"58";35;"Private";56352;"Assoc-voc";11;"Married-civ-spouse";"Other-service";"Husband";"White";"Male
"59";41;"Private";147372;"HS-grad";9;"Married-civ-spouse";"Adm-clerical";"Husband";"White";"Male";0
"60";30;"Private";188146;"HS-grad";9;"Married-civ-spouse";"Machine-op-inspct";"Husband";"White";"Ma
"61";30;"Private";59496;"Bachelors";13;"Married-civ-spouse";"Sales";"Husband";"White";"Male";2407;0
"62";32;NA;293936;"7th-8th";4;"Married-spouse-absent";NA;"Not-in-family";"White";"Male";0;0;40;NA;"
```

NA = missing information

Data Adult – atributy po transformaci

	Attribute	Categories / frequencies
1	*Age*	Young ⟨15; 25⟩/9627, Middle-aged ⟨26; 45⟩/24671, Senior ⟨46; 65⟩/12741, Old ⟨66; 99⟩/1803
2	*Hours_per_week*	Part-time ⟨1; 25⟩/5913, Full-time ⟨26; 40⟩/28577 Over-time ⟨41; 60⟩/12676, Workaholic ⟨61; 168⟩/1676
3	*Capital-gain*	None/44807, Low⟨1; 7268⟩/2345, High⟨7269; 99999⟩/1690
4	*Capital-loss*	None/46560, Low⟨1; 1887⟩/1166, High⟨1888; 4500⟩/1116

	Attribute	Categories / frequencies
1	*Education* 16 categories	Preschool/83, 1st-4th/247, 5th-6th/509, 7th-8th/955 9th/756, 10th/1389, 11th/1812, 12th/657, HS-grad/15784, Some-college/10878, Assoc-voc/2061, Assoc-acdm/1601, Bachelors/8025, Masters/2657, Prof-school/834, Doctorate/594
2	*Income* missings: 16281	Small/24720, Large/7841

1	*Sex*	Female/16192, Male/32650
2	*Marital_status* 16 categories	Divorced/6633, Married-AF-spouse/37, Married-civ-spouse/22379, Married-spouse-absent/628, Never-married/16117, Separated/1530, Widowed 1518
3	*Relationship* 7 categories	Husband/19716, Not-in-family/12583, Other-relative/1506, Own-child/7581, Unmarried/5125, Wife/2331
4	*Workclass* 8 categories missings: 2799	Federal-gov/1432, Local-gov/3136, Never-worked/10, Private/33906, Self-emp-inc/1695, Self-emp-not-inc/3862, State-gov/1981, Without-pay/21
5	*Occupation* 16 categories missings: 2809	Adm-clerical/5611, Armed-Forces/15, Craft-repair/6112, Exec-managerial/6086, Farming-fishing/1490, Handlers-cleaners/2072, Machine-op-inspct/3022, Other-service/4923, Priv-house-serv/242, Prof-specialty/6172, Protective-serv/983, Sales/5504, Tech-support/1446, Transport-moving/2355
6	*Race* 6 categories	Amer-Indian-Eskimo/470, Asian-Pac-Islander/1519, Black/4685, Other/406, White 41762
7	*Native_country* 6 categories missings: 857	United-States/43832, Mexico/951, Philippines/295, Germany/206, Puerto-Rico/184, Canada/182, El-Salvador/155, India/151, *additional 33 countries with frequency* < 150, *their total frequency = 1172*

ATR_Adult = množina třinácti atributů {Age, Capital_gain, … , Sex, Workclass}

Data Adult – vzorový příklad

AdultUCI – Poznámkový blok

Soubor Úpravy Formát Zobrazení Nápověda

```
"33";45;"Private";386940;"Bachelors";13;"Divorced";"Exec-managerial";"Own-child";"White";"Male";0;1
```

Úloha: Nalézt všechna asociační pravidla $X \rightarrow Y$ tak, že:

- Y je tvaru $A(a)$ kde A je atribut z ATR_Adult a a je jeho hodnota (kategorie)
- X je konjunkce 0 až 12-ti zbývajících booleovských atributů $A(a)$ kde A je atribut z ATR_Adult a a je jeho hodnota (kategorie),
- $conf\,(X \rightarrow Y) \geq 0.95$ a $sup(X \rightarrow Y) \geq 0.45$.

```
"51";25;"Private";32275;"Some-college";10;"Married-civ-spouse";"Exec-managerial";"Wife";"Other";"Fe
"52";18;"Private";226956;"HS-grad";9;"Never-married";"Other-service";"Own-child";"White";"Female";0
```

ATR_Adult = množina třinácti atributů {Age, Capital_gain, … , Sex, Workclass}

```
"56";43;"Private";237993;"Some-college";10;"Married-civ-spouse";"Tech-support";"Husband";"White";"M
"57";46;"Private";216666;"5th-6th";3;"Married-civ-spouse";"Machine-op-inspct";"Husband";"White";"Ma
"58";35;"Private";56352;"Assoc-voc";11;"Married-civ-spouse";"Other-service";"Husband";"White";"Male
"59";41;"Private";147372;"HS-grad";9;"Married-civ-spouse";"Adm-clerical";"Husband";"White";"Male";0
"60";30;"Private";188146;"HS-grad";9;"Married-civ-spouse";"Machine-op-inspct";"Husband";"White";"Ma
"61";30;"Private";59496;"Bachelors";13;"Married-civ-spouse";"Sales";"Husband";"White";"Male";2407;0
"62";32;NA;293936;"7th-8th";4;"Married-spouse-absent";NA;"Not-in-family";"White";"Male";0;0;40;NA;"
```

Systém R

https://www.r-project.org/

[Home]

Download

CRAN

R Project

About R
Logo
Contributors
What's New?
Mailing Lists
Bug Tracking
Development Site
Conferences
Search

R Foundation

Foundation
Board
Members
Donors
Donate

Documentation

Manuals
FAQs
The R Journal
Books

The R Project for Statistical Computing

Getting Started

R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS. To **download R**, please choose your preferred CRAN mirror.

If you have questions about R like how to download and install the software, or what the license terms are, please read our answers to frequently asked questions before you send an email.

News

- **Notice XQuartz users (Mac OS X)** A security issue has been detected with the Sparkle update mechanism used by XQuartz. Avoid updating over insecure channels.
- **R version 3.2.4 (Very Secure Dishes)** has been released on Thursday 2016-03-10.
- **R version 3.3.0 (Supposedly Educational) prerelease versions** will appear starting Monday 2016-03-14. Final release is scheduled for Thursday 2016-04-14.
- The **R Logo** is available for download in high-resolution PNG or SVG formats.
- **useR! 2016**, will take place at Stanford University, CA, USA, June 27 - June 30, 2016.
- **The R Journal Volume 7/2** is available.
- **R version 3.2.3 (Wooden Christmas-Tree)** has been released on 2015-12-10.
- **R version 3.1.3 (Smooth Sidewalk)** has been released on 2015-03-09.

Package arules

https://cran.r-project.org/web/packages/arules/

CRAN
Mirrors
What's new?
Task Views
Search

About R
R Homepage
The R Journal

Software
R Sources
R Binaries
Packages
Other

Documentation
Manuals
FAQs
Contributed

arules: Mining Association Rules and Frequent Itemsets

Provides the infrastructure for representing, manipulating and analyzing transaction data and patterns (frequent itemsets and association rules). Also provides interfaces to C implementations of the association mining algorithms Apriori and Eclat by C. Borgelt.

Version:	1.3-1
Depends:	R (≥ 3.2.0), Matrix (≥ 1.2-0)
Imports:	stats, methods, graphics, utils
Suggests:	pmml, XML, arulesViz, testthat
Published:	2015-12-14
Author:	Michael Hahsler [aut, cre, cph], Christian Buchta [aut, cph], Bettina Gruen [aut, cph], Kurt Hornik [aut, cph], Christian Borgelt [ctb, cph]
Maintainer:	Michael Hahsler <mhahsler at lyle.smu.edu>
BugReports:	https://github.com/mhahsler/arules/issues
License:	GPL-3
Copyright:	The code for apriori and eclat in src/rapriori.c was obtained from http://www.borgelt.net/ and is Copyright (C) 1996-2003 Christian Borgelt. All other code is Copyright (C) Michael Hahsler, Christian Buchta, Bettina Gruen and Kurt Hornik.
URL:	http://lyle.smu.edu/IDA/arules
NeedsCompilation:	yes
Classification/ACM:	G.4, H.2.8, I.5.1
Citation:	arules citation info
Materials:	README NEWS
In views:	MachineLearning
CRAN checks:	arules results

Downloads:

Reference manual:	arules.pdf
Vignettes:	Introduction to arules
Package source:	arules_1.3-1.tar.gz
Windows binaries:	r-devel: arules_1.3-1.zip, r-release: arules_1.3-1.zip, r-oldrel: arules_1.1-9.zip
OS X Snow Leopard binaries:	r-release: arules_1.3-1.tgz, r-oldrel: arules_1.1-6.tgz
OS X Mavericks binaries:	r-release: arules_1.3-1.tgz

Introduction to arules (1)

Introduction to arules – A computational environment for mining association rules and frequent item sets

Michael Hahsler
Southern Methodist University

Bettina Grün
Johannes Kepler Universität Linz

Kurt Hornik
Wirtschaftsuniversität Wien

Christian Buchta
Wirtschaftsuniversität Wien

Abstract

Mining frequent itemsets and association rules is a popular and well researched approach for discovering interesting relationships between variables in large databases. The R package **arules** presented in this paper provides a basic infrastructure for creating and manipulating input data sets and for analyzing the resulting itemsets and rules. The package also includes interfaces to two fast mining algorithms, the popular C implementations of Apriori and Eclat by Christian Borgelt. These algorithms can be used to mine frequent itemsets, maximal frequent itemsets, closed frequent itemsets and association rules.

Keywords: data mining, association rules, apriori, eclat.

R Console – summary(AdultUCI)

```
> dim(AdultUCI)
[1] 48842    13
> summary(AdultUCI)
          age                    workclass            education
 Young       : 9627   Private         :33906   HS-grad     :15784
 Middle-aged:24671   Self-emp-not-inc: 3862   Some-college:10878
 Senior      :12741   Local-gov       : 3136   Bachelors   : 8025
 Old         : 1803   State-gov       : 1981   Masters     : 2657
                      Self-emp-inc    : 1695   Assoc-voc   : 2061
                      (Other)         : 1463   11th        : 1812
                      NA's            : 2799   (Other)     : 7625
               marital-status            occupation            relationship
 Divorced              : 6633   Prof-specialty : 6172   Husband       :19716
 Married-AF-spouse     :   37   Craft-repair   : 6112   Not-in-family :12583
 Married-civ-spouse    :22379   Exec-managerial: 6086   Other-relative: 1506
 Married-spouse-absent:  628   Adm-clerical   : 5611   Own-child     : 7581
 Never-married         :16117   Sales          : 5504   Unmarried     : 5125
 Separated             : 1530   (Other)        :16548   Wife          : 2331
 Widowed               : 1518   NA's           : 2809
                race            sex        capital-gain capital-loss    hours-per-week
 Amer-Indian-Eskimo:  470   Female:16192   None:44807   none:46560   Part-time : 5913
 Asian-Pac-Islander: 1519   Male  :32650   Low : 2345   low : 1166   Full-time :28577
 Black             : 4685                  High: 1690   high: 1116   Over-time :12676
 Other             :  406                                            Workaholic: 1676
 White             :41762


      native-country    income
 United-States:43832   small:24720
 Mexico       :  951   large: 7841
 Philippines  :  295   NA's :16281
 Germany      :  206
 Puerto-Rico  :  184
 (Other)      : 2517
 NA's         :  857
```

arules – zadání parametrů a protokol

```
> rules <- apriori(Adult,parameter = list(support = 0.45, confidence = 0.95))
Apriori

Parameter specification:
 confidence minval smax arem  aval originalSupport support minlen maxlen target   ext
       0.95    0.1    1 none FALSE            TRUE    0.45      1     10  rules FALSE

Algorithmic control:
 filter tree heap memopt load sort verbose
    0.1 TRUE TRUE  FALSE TRUE    2    TRUE

Absolute minimum support count: 21978

set item appearances ...[0 item(s)] done [0.00s].
set transactions ...[115 item(s), 48842 transaction(s)] done [0.09s].
sorting and recoding items ... [10 item(s)] done [0.02s].
creating transaction tree ... done [0.09s].
checking subsets of size 1 2 3 4 5 done [0.00s].
writing ... [21 rule(s)] done [0.00s].
creating S4 object  ... done [0.02s].
```

arules – zadání parametrů a protokol, komentář

```
> rules <- apriori(Adult,parameter = list(support = 0.45, confidence = 0.95))
Apriori
```

Úloha: Nalézt všechna asociační pravidla $X \rightarrow Y$ tak, že:

- Y je tvaru $A(a)$ kde A je atribut z ATR_Adult a a je jeho hodnota (kategorie)
- X je konjunkce 0 až 12-ti zbývajících booleovských atributů $A(a)$ kde A je atribut z ATR_Adult a a je jeho hodnota (kategorie),
- $conf\,(X \rightarrow Y) \geq 0.95$ a $sup(X \rightarrow Y) \geq 0.45$.

```
set transactions ...[115 item(s), 48842 transaction(s)] done [0.09s].
```

ATR_Adult = množina třinácti atributů {Age, Capital_gain, … , Sex, Workclass}

```
checking subsets of size 1 2 3 4 5 done [0.00s].
writing ... [21 rule(s)] done [0.00s].
creating S4 object  ... done [0.02s].
```

Nalezeno 21 pravidel

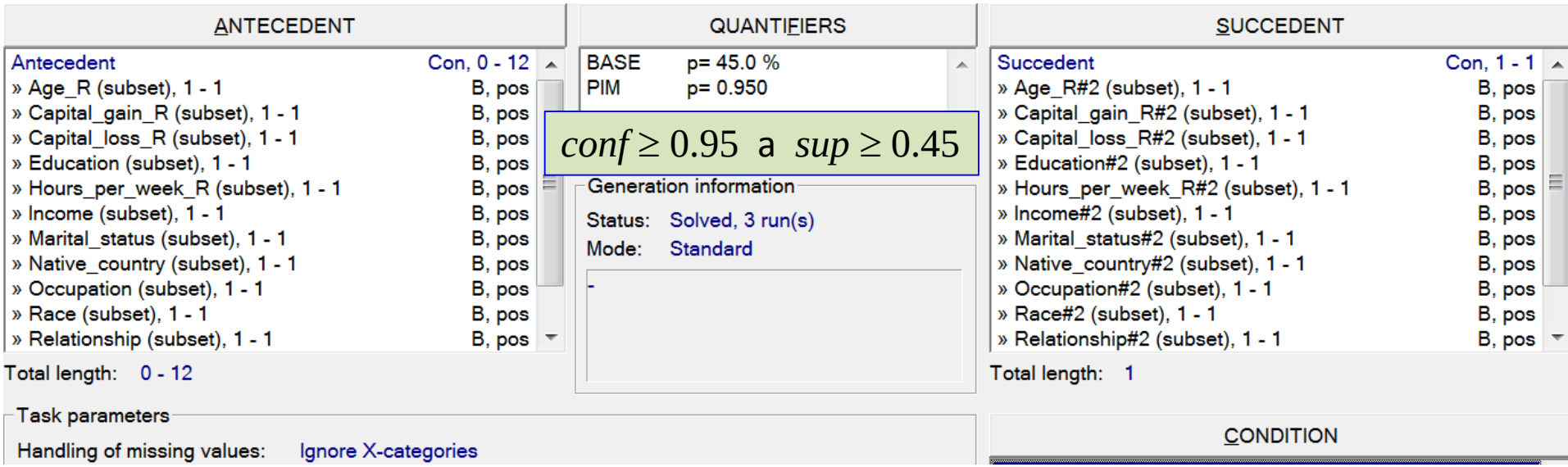
arules – ukázka výstupu, 20 z 21 nalezených pravidel

```
> inspect(rules)
   lhs                                  rhs                     support confidence      lift
1  {}                                => {capital-loss=none} 0.9532779  0.9532779 1.0000000
2  {age=Middle-aged}                 => {capital-loss=none} 0.4800786  0.9504276 0.9970100
3  {income=small}                    => {capital-gain=None} 0.4849310  0.9581311 1.0444135
4  {income=small}                    => {capital-loss=none} 0.4908480  0.9698220 1.0173549
5  {hours-per-week=Full-time}        => {capital-loss=none} 0.5606650  0.9582531 1.0052191
6  {workclass=Private}               => {capital-loss=none} 0.6639982  0.9564974 1.0033773
7  {race=White}                      => {capital-loss=none} 0.8136849  0.9516307 0.9982720
8  {native-country=United-States}    => {capital-loss=none} 0.8548380  0.9525461 0.9992323
9  {capital-gain=None,
    income=small}                    => {capital-loss=none} 0.4696573  0.9685033 1.0159716
10 {capital-loss=none,
    income=small}                    => {capital-gain=None} 0.4696573  0.9568282 1.0429934
11 {race=White,
    hours-per-week=Full-time}        => {capital-loss=none} 0.4618566  0.9572671 1.0041847
12 {hours-per-week=Full-time,
    native-country=United-States}    => {capital-loss=none} 0.4959052  0.9574275 1.0043529
13 {capital-gain=None,
    hours-per-week=Full-time}        => {capital-loss=none} 0.5191638  0.9550659 1.0018756
14 {workclass=Private,
    race=White}                      => {capital-loss=none} 0.5674829  0.9549683 1.0017732
15 {workclass=Private,
    native-country=United-States}    => {capital-loss=none} 0.5897179  0.9554818 1.0023119
16 {workclass=Private,
    capital-gain=None}               => {capital-loss=none} 0.6111748  0.9529145 0.9996188
17 {race=White,
    native-country=United-States}    => {capital-loss=none} 0.7490480  0.9504325 0.9970152
18 {capital-gain=None,
    hours-per-week=Full-time,
    native-country=United-States}    => {capital-loss=none} 0.4580689  0.9540725 1.0008335
19 {workclass=Private,
    race=White,
    native-country=United-States}    => {capital-loss=none} 0.5181401  0.9535418 1.0002768
20 {workclass=Private,
    race=White,
    capital-gain=None}               => {capital-loss=none} 0.5204742  0.9511000 0.9977153
```

Stejná úloha ve 4ft-Miner a v arules v R

- Data Adult a vzorový příklad
- Vzorový příklad a arules v R-package
- Vzorový příklad a GUHA procedura 4ft-Miner

4ft-Miner – zadání úlohy



$conf \geq 0.95$ a $sup \geq 0.45$

Úloha: Nalézt všechna asociační pravidla $X \rightarrow Y$ tak, že:

- Y je tvaru $A(a)$ kde A je atribut z ATR_Adult a a je jeho hodnota (kategorie)
- X je konjunkce 0 až 12-ti zbývajících booleovských atributů $A(a)$ kde A je atribut z ATR_Adult a a je jeho hodnota (kategorie),
- $conf\,(X \rightarrow Y) \geq 0.95$ a $sup(X \rightarrow Y) \geq 0.45$.

ATR_Adult = množina třinácti atributů {Age, Capital_gain, … , Sex, Workclass}

4ft-Miner – ukázka výstupu

Task run

Start: 28.1.2019 10:06:14 Total time: 0h 0m 0s

Number of verifications: 509

Number of hypotheses: 21 Mode: Standard

Nalezeno 21 pravidel

Konfidence pravidla, zde zobrazeno 15 pravidel s nejvyšší konfidencí

Actual group of hypotheses: All hypotheses

Hypotheses in group: 21 Shown hypotheses: 21 Highlighted: 0

Nr.	Id	PIM	Hypothesis
1	16	0.970	**Income**(*small*) >÷< **Capital_loss**(*None*)
2	19	0.969	**Capital_gain**(*None*) & **Income**(*small*) >÷< **Capital_loss**(*None*)
3	9	0.958	**Hours_per_week**(*Full-time*) >÷< **Capital_loss**(*None*)
4	15	0.958	**Income**(*small*) >÷< **Capital_gain**(*None*)
5	14	0.957	**Hours_per_week**(*Full-time*) & **Native_country**(*United-States*) >÷< **Capital_loss**(*None*)
6	20	0.957	**Hours_per_week**(*Full-time*) & **Race**(*White*) >÷< **Capital_loss**(*None*)
7	18	0.957	**Capital_loss**(*None*) & **Income**(*small*) >÷< **Capital_gain**(*None*)
8	5	0.956	**Workclass**(*Private*) >÷< **Capital_loss**(*None*)
9	7	0.955	**Native_country**(*United-States*) & **Workclass**(*Private*) >÷< **Capital_loss**(*None*)
10	12	0.955	**Capital_gain**(*None*) & **Hours_per_week**(*Full-time*) >÷< **Capital_loss**(*None*)
11	8	0.955	**Race**(*White*) & **Workclass**(*Private*) >÷< **Capital_loss**(*None*)
12	21	0.954	**Capital_gain**(*None*) & **Hours_per_week**(*Full-time*) & **Native_country**(*United-States*) >÷< **Capital_loss**(*None*)
13	13	0.954	**Native_country**(*United-States*) & **Race**(*White*) & **Workclass**(*Private*) >÷< **Capital_loss**(*None*)
14	1	0.953	(empty) >÷< **Capital_loss**(*None*)
15	6	0.953	**Capital_gain**(*None*) & **Workclass**(*Private*) >÷< **Capital_loss**(*None*)

4ft-Miner – detailní výstupu prvního pravidla

Antecedent: Income(small)
Succedent: Capital_loss(None)
Condition: (empty)

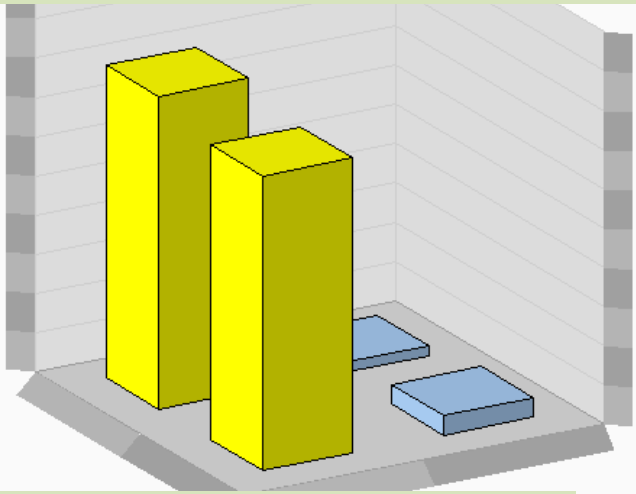
Income(small) $\rightarrow_{0.97,0.49}$ Capital_loss(None)

TEXT | DATA | FREQUENCIES | 9-FOLD TABLE | FREQDIFF | F

	Succedent	¬Succedent
Ant.	23974	746
¬Ant.	22586	1536

$$\text{konfidence} = \frac{23\,974}{23\,974 + 746} = 0.97$$

$$\text{podpora} = \frac{23\,974}{23\,974 + 746 + 22\,586 + 1536} = 0.49$$



Platí zároveň:

- 97 % osob splňujících Income(small) splňuje i Capital_loss(None)
- 23 974 osob splňuje zároveň Income(small) i Capital_loss(None)

Částečné porovnání výstupu arules a 4ft-Miner

1	16	0.970	**Income**(*small*) >÷< **Capital_loss**(*None*)	4
2	19	0.969	**Capital_gain**(*None*) & **Income**(*small*) >÷< **Capital_loss**(*None*)	9
3	9	0.958	**Hours_per_week**(*Full-time*) >÷< **Capital_loss**(*None*)	5
4	15	0.958	**Income**(*small*) >÷< **Capital_gain**(*None*)	3
5	14	0.957	**Hours_per_week**(*Full-time*) & **Native_country**(*United-States*) >÷< **Capital_loss**(*None*)	12
6	20	0.957	**Hours_per_week**(*Full-time*) & **Race**(*White*) >÷< **Capital_loss**(*None*)	11
7	18	0.957	**Capital_loss**(*None*) & **Income**(*small*) >÷< **Capital_gain**(*None*)	10

```
> inspect(rules)
   lhs                                  rhs                     support confidence      lift
1  {}                                => {capital-loss=none} 0.9532779  0.9532779 1.0000000
2  {age=Middle-aged}                 => {capital-loss=none} 0.4800786  0.9504276 0.9970100
3  {income=small}                    => {capital-gain=None} 0.4849310  0.9581311 1.0444135
4  {income=small}                    => {capital-loss=none} 0.4908480  0.9698220 1.0173549
5  {hours-per-week=Full-time}        => {capital-loss=none} 0.5606650  0.9582531 1.0052191
6  {workclass=Private}               => {capital-loss=none} 0.6639982  0.9564974 1.0033773
7  {race=White}                      => {capital-loss=none} 0.8136849  0.9516307 0.9982720
8  {native-country=United-States}    => {capital-loss=none} 0.8548380  0.9525461 0.9992323
9  {capital-gain=None,
    income=small}                    => {capital-loss=none} 0.4696573  0.9685033 1.0159716
10 {capital-loss=none,
    income=small}                    => {capital-gain=None} 0.4696573  0.9568282 1.0429934
11 {race=White,
    hours-per-week=Full-time}        => {capital-loss=none} 0.4618566  0.9572671 1.0041847
12 {hours-per-week=Full-time,
    native-country=United-States}    => {capital-loss=none} 0.4959052  0.9574275 1.0043529
```