"Validaty" shiny app

Matthias Gomolka

Mariko Kawano

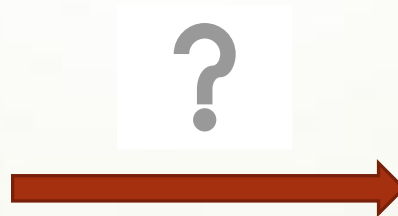Mark van der Loo

Marianna Mitratza

Kazumi Wada

14 September 2018

Problem

- quick test of the dataset quality : are specific rules followed or not ?

- domain experts set the rules + can communicate results of test

- no programming experience needed

R packages behind the idea

- "validate"

1) Basic data validation:

- Enter rules, and their metadata (descriptions)
- Apply rules to a dataset interactively. Update as rules are added
- Import and export rules from and to file
- Show an interactive dashboard with results

2) Analyze rule sets:

- Visualize the rules and the variables that they `touch'
- Interactively look for redundancies or inconsistencies and present it to the user

- "errorlocate"

Design - Features

CSV file

| Browse... | retailers.csv |
| Upload complete |

Select key variable

| id | ▲ |

- no key
- id
- size
- incl.prob
- staff
- turnover
- other.rev
- total.rev

Show [25 ▼] entries

Search: [ ]

id	size	incl.prob	staff	turnover	other.rev	total.rev	staff.costs	total.costs	profit
RET01	sc0	0.02	75			1130		18915	20045
RET02	sc3	0.14	9	1607		1607	131	1544	63
RET03	sc3	0.14		6886	-33	6919	324	6493	426
RET04	sc3	0.14		3861	13	3874	290	3600	274
RET05	sc3	0.14			37	5602	314	5530	72
RET06	sc0	0.02	1	25		25		22	3
RET07	sc3	0.14	5			1335	135	136	1
RET08	sc1	0.02	3	404	13	417		342	75
RET09	sc3	0.14	6	2596		2596	147	2486	110
RET10	sc2	0.05	5						
RET11	sc2	0.05	5	645		645	130	636	9
RET12	sc2	0.05	5	2872		2872	182	2652	220
RET13	sc3	0.14	13	5678	12	5690	326	5656	34
RET14	sc1	0.02		931397		931397	36872	841489	89908

Free text/YAML

Browse...	testrules.R

Upload complete

View Rules   Coverage   Imposed limits

Show 25 entries   Search:

name	label	rule
V1		$(staff - 0) >= -1e-08$
V2		$abs(turnover + other.rev - total.rev) < 1e-08$
name	label	rule

Showing 1 to 2 of 2 entries   Previous 1 Next

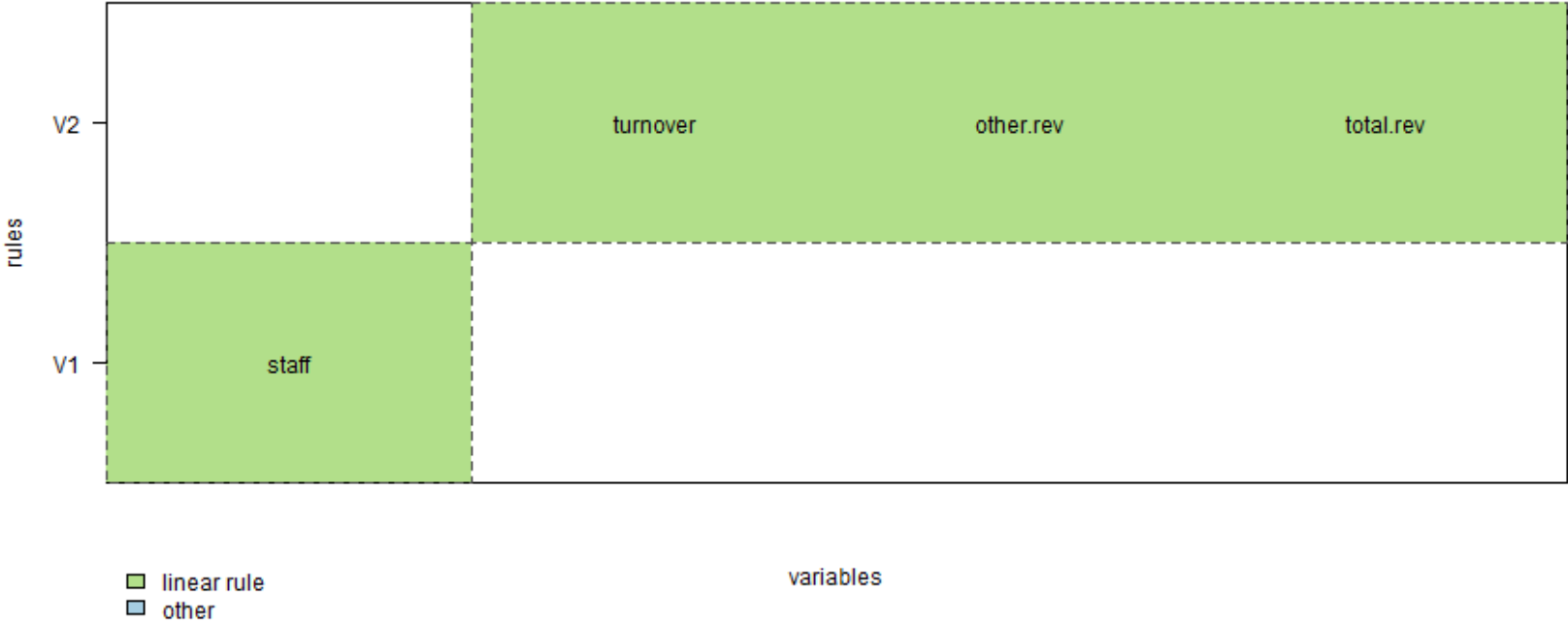⬇ Download Rules

Free text/YAML

Browse...   testrules.R

Upload complete

View Rules   Coverage   Imposed limits

Variables covered: staff, turnover, other.rev, total.rev
Variables not covered: id, size, incl.prob, staff.costs, total.costs, profit, vat

View Rules   Coverage   **Imposed limits**

Boundaries

Limits imposed by the rule set

Show [25 ▾] entries   Search: [   ]

variable	lowerbound	upperbound
staff	0	Inf
other.rev	-Inf	Inf
total.rev	-Inf	Inf
turnover	-Inf	Inf
[variable]	[lowerbound]	[upperbound]

Showing 1 to 4 of 4 entries   Previous **1** Next

Fixed values

Variables which can only take one value under the current rule set

Show [25 ▾] entries   Search: [   ]

variable	value
No data available in table	
[variable]	[value]

Validaty
Data Input   Rule Management   Confrontation   Error localization

Tolerance for equality

0.00000001

Go!

summary   plot   details

Show 25 entries   Search:

name	items	passes	fails	nNA	error	warning	expression
V1	60	54	0	6	FALSE	FALSE	(staff - 0) >= 0L
V2	60	19	4	37	FALSE	FALSE	turnover + other.rev == total.rev
name	items	passes	fails	nNA	error	warning	expression

Showing 1 to 2 of 2 entries

Previous   1   Next

Tolerance for equality

0.00000001

Go!

summary   plot   details

Results by rule

V2   turnover + other.rev == total.rev

V1   (staff - 0) >= 0L

0   10   20   30   40   50   60

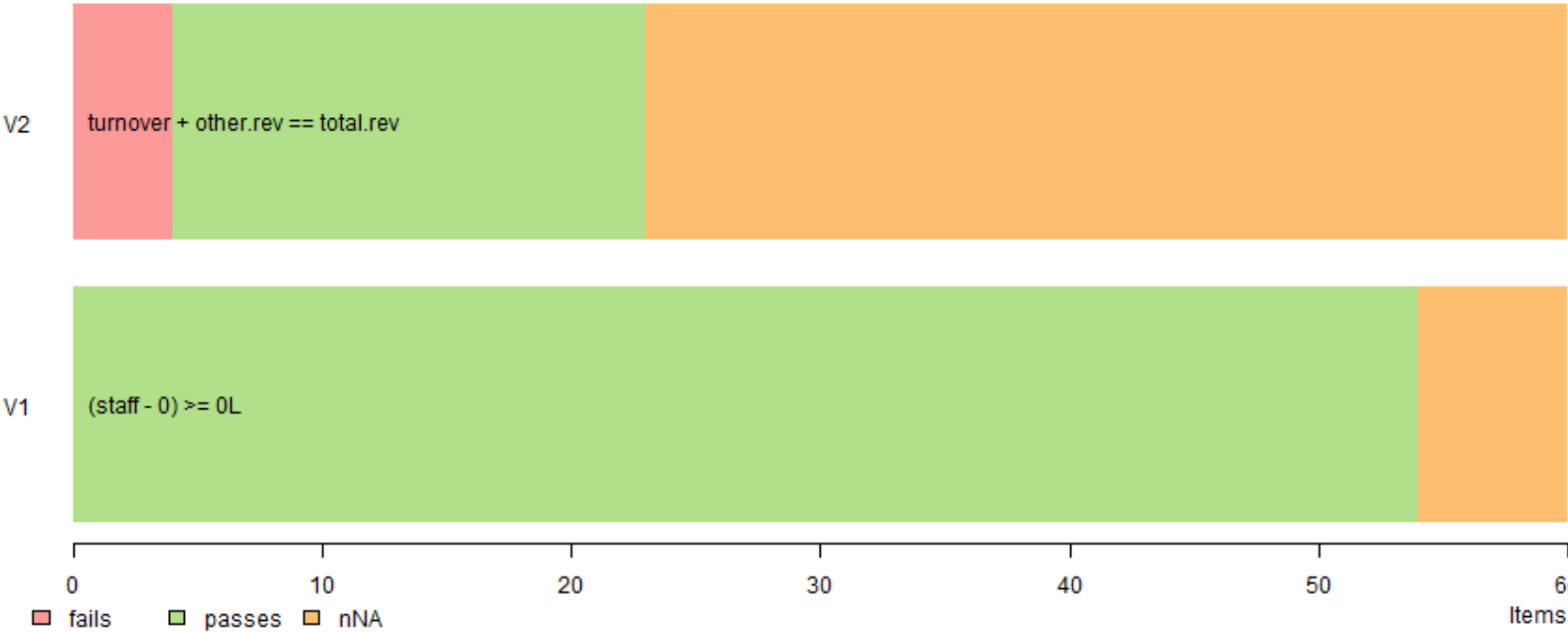Items

☐ fails   ☐ passes   ☐ nNA

⬇ Download Report

Tolerance for equality

0.00000001

Go!

summary   plot   details

Show 10 entries   Search:

name	value	expression
V1	TRUE	(staff - 0) >= 0L
V1	TRUE	(staff - 0) >= 0L
V1		(staff - 0) >= 0L
V1		(staff - 0) >= 0L
V1		(staff - 0) >= 0L
V1	TRUE	(staff - 0) >= 0L
V1	TRUE	(staff - 0) >= 0L
V1	TRUE	(staff - 0) >= 0L
V1	TRUE	(staff - 0) >= 0L
V1	TRUE	(staff - 0) >= 0L
name	value	expression

Showing 1 to 10 of 120 entries

Previous 1 2 3 4 5 ... 12 Next

⬇ Download Detailed Results

Summary   Detailed

Summary by variable

Show [10 ▼] entries   Search: [   ]

name	errors	missing
total.rev	2	2
turnover	1	4
other.rev	1	36
id	0	0
size	0	0
incl.prob	0	0
staff	0	6
staff.costs	0	10
total.costs	0	5
profit	0	5
name	errors	missing

Summary by record

Show [25 ▼] entries   Search: [   ]

errors	N
0	56
1	4
errors	N

Showing 1 to 2 of 2 entries

Previous   1   Next

Validaty   Data Input   Rule Management   Confrontation   Error localization

Summary   Detailed

Show [25 ▾] entries   Search: [   ]

id	size	incl.prob	staff	turnover	other.rev	total.rev	staff.costs	total.costs	profit	vat
false	false	false		true	false	false	false	false	false	
false	false	false	false	false	false	true	false	false	false	false
false	false	false	false	false	true	false	false	false	false	false
false	false	false	false	false	false	true	false	false	false	false
id	size	incl.prob	staff	turnover	other.rev	total.rev	staff.costs	total.costs	profit	vat

Showing 1 to 4 of 4 entries   Previous   1   Next