fcaR, an R package to handle fuzzy implications: design of a recommendation system for medical diagnosis

D. López, A. Mora

Universidad de Málaga

Table of contents

UNIVERSIDAD
DE MÁLAGA

R Packages for Implications and Rules

- **arules**: Provides the infrastructure for representing, manipulating and analyzing transaction data and patterns (frequent itemsets and association rules). Also provides C implementations of the association mining algorithms Apriori and Eclat.
- **frbs**: An implementation of various learning algorithms based on fuzzy rule-based systems (FRBSs) for dealing with classification and regression tasks.
- **RKEEL**: `KEEL` is a popular Java software for a large number of different knowledge data discovery tasks. This package takes the advantages of `KEEL` and R, allowing to use `KEEL` algorithms in simple R code.

UNIVERSIDAD DE MÁLAGA

Objective

The main objective is to create an R package able to:

- Manage formal contexts and find concepts.
- Extract implications from a context.
- Provide tools to visualize the extracted knowledge.
- Compute closures and recommendations.
- Integrate with **arules**.

UNIVERSIDAD
DE MÁLAGA

Table of contents

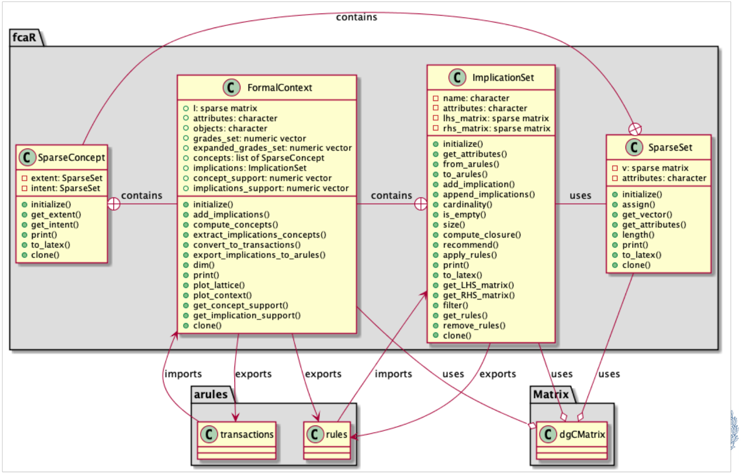
UNIVERSIDAD DE MÁLAGA

Class Structure

- `FormalContext` encapsulates the definition of a formal context (G, M, I), being G the set of objects, M the set of attributes and I the (fuzzy) relationship matrix, and provides methods to operate on the context using FCA tools.
- `ImplicationSet` represents a set of implications over a specific formal context.
- `SparseSet` is a class solely used for visualization purposes, since it encapsulates in sparse format a (fuzzy) set.
- `SparseConcept` encapsulates internally both *extent* and *intent* of a formal concept as `SparseSet`.

UNIVERSIDAD DE MÁLAGA

UML Diagram



UNIVERSIDAD DE MÁLAGA

Use of Formal Contexts

Table: A sample formal context. Attributes are P1 to P6 and objects are named O1 to O6.

	P1	P2	P3	P4	P5	P6
O1	0.0	1.0	0.5	0.5	1.0	0
O2	1.0	1.0	1.0	0.0	0.0	0
O3	0.5	0.5	0.0	0.0	0.0	1
O4	0.0	0.0	0.0	1.0	0.5	0
O5	0.0	0.0	1.0	0.5	0.0	0
O6	0.5	0.0	0.0	0.0	0.0	0

UNIVERSIDAD DE MÁLAGA

Use of Formal Contexts

```
R> fc <- formal_context$new(I)
R> print(fc)
FormalContext with 6 objects and 6 attributes.
Attributes' names are: P1, P2, P3, P4, P5, P6
Matrix:
    P1  P2  P3  P4  P5 P6
O1 0.0 1.0 0.5 0.5 1.0  0
O2 1.0 1.0 1.0 0.0 0.0  0
O3 0.5 0.5 0.0 0.0 0.0  1
O4 0.0 0.0 0.0 1.0 0.5  0
O5 0.0 0.0 1.0 0.5 0.0  0
O6 0.5 0.0 0.0 0.0 0.0  0
```

UNIVERSIDAD DE MÁLAGA

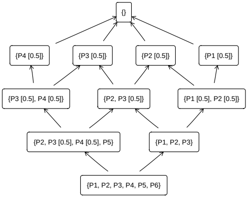
Importing from arules

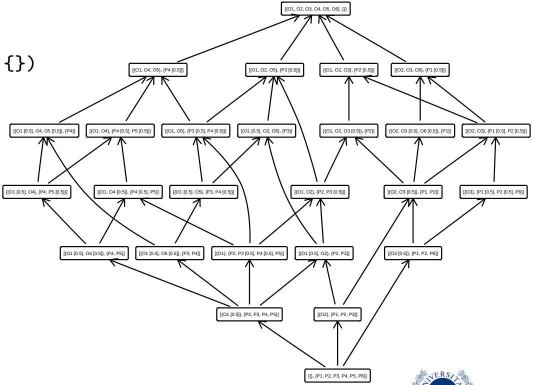
```
R> fc_mushroom <- formal_context$new(Mushroom)
R> fc_mushroom
Warning: Too many attributes, output will be truncated.
FormalContext with 8124 objects and 114 attributes.
Attributes' names are: Class=edible, Class=poisonous, CapShape=bell,
  CapShape=conical, CapShape=flat, CapShape=knobbed, ...
Matrix:
        Class=edible Class=poisonous CapShape=bell CapShape=conical
1                0               1             0                0
2                1               0             0                0
3                1               0             1                0
4                0               1             0                0
5                1               0             0                0
6                1               0             0                0
        CapShape=flat CapShape=knobbed CapShape=sunken
1                 0                0               0
2                 0                0               0
3                 0                0               0
4                 0                0               0
5                 0                0               0
6                 0                0               0
```

UNIVERSIDAD
DE MÁLAGA

Finding and Plotting Concepts

```
R> concepts <- fc$compute_concepts()
R> length(concepts)
[1] 26
R> concepts[[1]]
({O1, O2, O3, O4, O5, O6}, {})
```





UNIVERSIDAD DE MÁLAGA

Managing Implications

In order to build the Duquenne-Guigues basis of implications, by using NextClosure, the `extract_implications_concepts()` method is called from a `FormalContext` object, and the result is stored in `fc$implications`.

```
R> fc$extract_implications_concepts()
```

```
R> fc$implications
```

```
Implication set with 12 implications.
Rule 1: {P6 [0.5]} -> {P1 [0.5], P2 [0.5], P6}
Rule 2: {P5 [0.5]} -> {P4 [0.5]}
Rule 3: {P3 [0.5], P4 [0.5], P5 [0.5]} -> {P2, P5}
Rule 4: {P3 [0.5], P4} -> {P3}
Rule 5: {P2 [0.5], P4 [0.5]} -> {P2, P3 [0.5], P5}
Rule 6: {P2 [0.5], P3 [0.5]} -> {P2}
Rule 7: {P2, P3, P4 [0.5], P5} -> {P4}
Rule 8: {P1 [0.5], P4 [0.5]} -> {P1, P2, P3, P4, P5, P6}
Rule 9: {P1 [0.5], P3 [0.5]} -> {P1, P2, P3}
Rule 10: {P1 [0.5], P2} -> {P1}
Rule 11: {P1, P2 [0.5]} -> {P2}
Rule 12: {P1, P2, P3, P6} -> {P4, P5}
```

UNIVERSIDAD DE MÁLAGA

Importing/Exporting Binary Rules

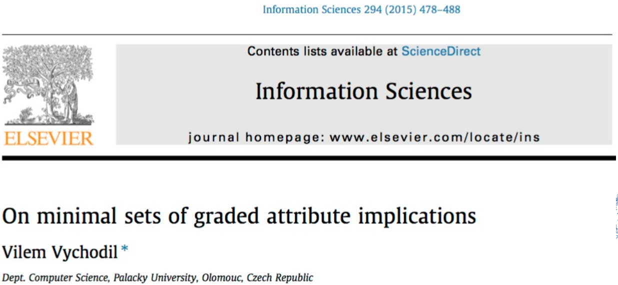
The `ImplicationSet` can interface with **arules**, importing and exporting sets of rules in **arules** format.

```
R> fc_mushroom$add_implications(mush_rules)
R> fc_mushroom$implications$cardinality()
[1] 2002
R> my_rules <- fc_mushroom$export_implications_to_arules(
                                             quality = TRUE)
R> class(my_rules)
[1] "rules"
attr(,"package")
[1] "arules"
```

UNIVERSIDAD DE MÁLAGA

Removing redundancy in graded attribute implications

- To obtain equivalent implicational sets with lower size (small number of implications) or with less attributes in the LHS or RHS.
- Widely studied in the classical setting.
- Approached by Vilem Vychodil in [IS2015].



Information Sciences 294 (2015) 478–488

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

ELSEVIER

On minimal sets of graded attribute implications

Vilem Vychodil *

Dept. Computer Science, Palacky University, Olomouc, Czech Republic

Removing redundancy in graded attribute implications

- We tackle this problem using the so-called *Fuzzy Attribute Simplification Logic*, **FASL**, which has been introduced in [Belohlavek 2016].
- This logic leads to the design of automatic reasoning methods for implications in data with grades.

International Journal of Approximate Reasoning 70 (2016) 51–67

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

www.elsevier.com/locate/ijar

ELSEVIER

Automated prover for attribute dependencies in data with grades

Radim Belohlavek [a], Pablo Cordero [b], Manuel Enciso [b], Ángel Mora [b], Vilem Vychodil [a]

[a] Dept. Computer Science, Palacky University, Olomouc, Czech Republic
[b] Universidad de Málaga, Spain

UNIVERSIDAD DE MÁLAGA

FASL

Axiomatic system

- Firstly, we proposed a new Simplification Rule adequate to remove redundancy in an automatic way.
- Simplification Rule turned into the *heart* of a novel logic: FASL - Fuzzy Attributes Simplification Logic.
- FASL becomes the *engine* of automated methods: redundancy removal, closure algorithm, etc.

UNIVERSIDAD DE MÁLAGA

FASL

Axiomatic system

The axiomatic system in FASL is defined as follows: for all $A, B, C, D \in L^\Omega$ and $c \in L$,

[Ax] infer $A \cup B \Rightarrow A$ (*Axiom*)

[Mul] from $A \Rightarrow B$ infer $c^* \otimes A \Rightarrow c^* \otimes B$ (*Multiplication*)

[Sim] from $A \Rightarrow B$ and $C \Rightarrow D$ infer $A \cup (C \smallsetminus B) \Rightarrow D$ (*Simplification*)

UNIVERSIDAD DE MÁLAGA

Removal of Redundancies with FASL

The axiomatic system is equivalent to the following, which are implemented in **fcaR**:

- **Reduction** rule: from $A \Rightarrow B$, deduce $A \Rightarrow B \smallsetminus A$, being $B \smallsetminus A$ the fuzzy set difference B minus A.
- **Generalization** rule: from $A \Rightarrow B$ and $C \Rightarrow D$, if $A \subseteq C$ and $D \subseteq B$, remove $C \Rightarrow D$.
- **Composition** rule: if $A \Rightarrow B$ and $A \Rightarrow C$, substitute both implications by this $A \Rightarrow B \cup C$.
- **Simplification** rule: if $A \Rightarrow B$ and $C \Rightarrow D$, with $A \subset C$ and $A \cap B = \emptyset$, substitute $C \Rightarrow D$ by $C \smallsetminus B \Rightarrow D \smallsetminus B$.

UNIVERSIDAD DE MÁLAGA

Removal of Redundancies

```
R> fc_mushroom$implications$apply_rules(c("reduction",
R+                                         "composition",
R+                                         "generalization",
R+                                         "simplification"))
Using parallel execution
Processing batch
--> reduction : from 2002 to 2002 in 0.199 secs.
--> composition : from 2002 to 961 in 2.719 secs.
--> generalization : from 961 to 961 in 0.07 secs.
--> simplification : from 961 to 961 in 18.788 secs.
Batch took 21.781 secs.
```

These set operations have been implemented in C.
If a parallel backend is available (via package **parallel**, for instance), the `ImplicationSet` is split into batches of a specific cardinality (defined by the `batch_size` parameter), and the `apply_rules` method distributes batches across workers, allowing to process all batches in parallel.

UNIVERSIDAD DE MÁLAGA

FASL for Computing Closures

```
R> # Generate a fuzzy set with attribute "CapColor=white"
R> A <- sparse_set$new(attributes = fc_mushroom$attributes)
R> A$assign(attributes = "CapColor=white", values = 1)
R>
R> closure <- fc_mushroom$implications$compute_closure(A,
R+                                                      reduce = TRUE)
R> closure$closure
```

```
{CapColor=white, GillAttached=free, ColorAboveRing=white,
        ColorBelowRing=white, VeilType=partial, VeilColor=white}
```

```
R> closure$implications$get_rules(1:4)
```

```
Implication set with 4 implications.
Rule 1: {Habitat=woods} -> {RingNumber=one}
Rule 2: {Habitat=woods} -> {RingNumber=one}
Rule 3: {StalkRoot=bulbous} -> {RingType=pendant}
Rule 4: {CapColor=yellow, StalkShape=enlarging} -> {GillSpace=close}
```

UNIVERSIDAD DE MÁLAGA

Table of contents

UNIVERSIDAD
DE MÁLAGA

Included Datasets

Two complete examples of the use of **fcaR** on real-world problems:

`cobre32`: Designing a diagnostic system from a formal context with (fuzzy) medical data.

`vegas`: Extracting knowledge about the features of tourist destinations given an user profile.

The datasets for this section are provided and documented in the package.

Medical Context

```
R> colnames(cobre32)
 [1] "COSAS_1"   "COSAS_2"   "COSAS_3"   "COSAS_4"
 [5] "COSAS_5"   "COSAS_6"   "COSAS_7"   "FICAL_1"
 [9] "FICAL_2"   "FICAL_3"   "FICAL_4"   "FICAL_5"
[13] "FICAL_6"   "FICAL_7"   "FICAL_8"   "FICAL_9"
[17] "SCIDII_10" "SCIDII_11" "SCIDII_12" "SCIDII_13"
[21] "SCIDII_14" "SCIDII_15" "SCIDII_16" "SCIDII_17"
[25] "SCIDII_18" "SCIDII_19" "SCIDII_20" "SCIDII_21"
[29] "SCIDII_22" "SCIDII_23" "dx_ss"     "dx_other"
```

- The *Simpson-Angus Scale*, 6 items to evaluate Parkinsonism-like alterations.
- *Calgary Depression Scale for Schizophrenia*, 9 items (attributes) assessing the level of depression in schizophrenia.
- The *Structured Clinical Interview for DSM-III-R Personality Disorders*, with 9 variables related to the presence of signs affecting personality.
- The diagnosis for each individual: it can be *schizophrenia strict* or *other diagnosis* (which includes schizoaffective and bipolar disorders).

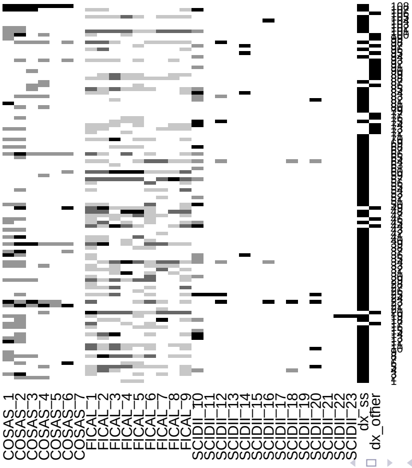
UNIVERSIDAD
DE MÁLAGA

Medical Context

In summary, the dataset consists in the previous 30 attributes related to signs or symptoms, and 2 attributes related to diagnosis. This makes a dataset with 105 objects (patients) and 32 attributes to explore. For a given attribute (symptom), the available grades range from *absent* to *extreme*, with *minimal*, *mild*, *moderate*, *moderate severe* and *severe* in between. These fuzzy attributes are mapped to values in the interval $[0, 1]$.

Medical Context

```
R> fc <- formal_context$new(cobre32)
```



UNIVERSIDAD DE MÁLAGA

Implications and Simplification

```
R> fc$extract_implications_concepts()
R> length(fc$concepts)

[1] 14686

R> fc$implications$cardinality()

[1] 985

R> fc$implications$apply_rules(rules = c("simplification"))

Using parallel execution
Processing batch
--> simplification : from 985 to 985 in 3.847 secs.
Batch took 3.849 secs.
```

UNIVERSIDAD DE MÁLAGA

Simple Diagnostic System

```
R> diagnose <- function(S) {
R+
R+   fc$implications$recommend(S = S,
R+                             attribute_filter = c("dx_ss",
R+                                                  "dx_other"))
R+
R+ }
```

UNIVERSIDAD DE MÁLAGA

Diagnosis Example

```
R> S1 <- sparse_set$new(attributes = fc$attributes)
R> S1$assign(attributes = c("COSAS_1", "COSAS_2", "COSAS_3", "COSAS_4",
R+                          "COSAS_5", "COSAS_6"),
R+           values = c(0.5, 1, 0.5, 0.166667, 0.5, 1))
R>
R> diagnose(S1) # Schizophrenia strict

dx_ss dx_other
    1        0

R> S2 <- sparse_set$new(attributes = fc$attributes)
R> S2$assign(attributes = c("FICAL_1", "FICAL_2", "COSAS_1"),
R+           values = c(0, 1, 0))
R>
R> diagnose(S2) # Not enough information

dx_ss dx_other
    0        0

R> S3 <- sparse_set$new(attributes = fc$attributes)
R> S3$assign(attributes = c("COSAS_4", "FICAL_3", "FICAL_5", "FICAL_8"),
R+           values = c(0.6666667, 0.5, 0.5, 0.5))
R>
R> diagnose(S3) # Other, not schizophrenia strict

dx_ss dx_other
    0        1
```

UNIVERSIDAD DE MÁLAGA

Data and Problem

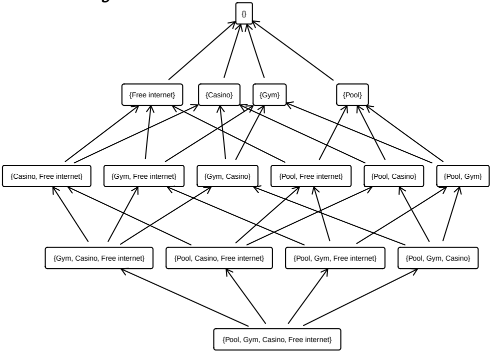
500 TripAdvisor reviews of hotels in Las Vegas Strip:

- `Period of Stay`: 4 categories are present in the original data, which produces as many binary variables: `Period of stay=Dec-Feb`, `Period of stay=Mar-May`, `Period of stay=Jun-Aug` and `Period of stay=Sep-Nov`.
- `Traveler type`: five binary categories are created from the original data: `Traveler type=Business`, `Traveler type=Couples`, `Traveler type=Families`, `Traveler type=Friends` and `Traveler type=Solo`.
- `Pool`, `Gym`, `Tennis court`, `Spa`, `Casino`, `Free internet`: binary variables for the services offered by each destination hotel.
- `Stars`: five binary variables are created, according to the number of stars of the hotel, `Stars=3`, `Stars=3.5`, `Stars=4`, `Stars=4.5` and `Stars=5`.
- `Score`, the score assigned in the review, from `Score=1` to `Score=5`, five variables are created.

UNIVERSIDAD DE MÁLAGA

Managing Knowledge

```
R> fc <- formal_context$new(vegas)
R> fc$extract_implications_concepts()
R> fc$plot_lattice(minsupp = 0.9,
R+                 object_names = FALSE)
```



Managing Knowledge

```
R> # Remove redundancies
R> fc$implications$apply_rules(c("simplification",
R+                                "composition",
R+                                "generalization"))

Using parallel execution
Processing batch
--> simplification : from 382 to 382 in 1.173 secs.
--> composition : from 382 to 382 in 0.004 secs.
--> generalization : from 382 to 382 in 0.012 secs.
Batch took 1.191 secs.

R> # Remove implications with no support
R> supp <- fc$get_implication_support()
R> idx_zero_supp <- which(supp == 0)
R> fc$implications$remove_rules(idx_zero_supp)
```

UNIVERSIDAD
DE MÁLAGA

Desirable Services in Destination

Question:
For a given couple, searching for a hotel in Las Vegas with Spa, which are the additional services that a destination should offer to get the highest score of 5?

Applying FASL

First, filter those rules related to couples:

```
R> base_implications <- fc$implications$filter(
R+                              "Traveler type=Couples")
```

Then, specify the minimum services (Spa)

```
R> S <- sparse_set$new(fc$attributes)
R> S$assign(attributes = c("Traveler type=Couples", "Spa"),
R+          values = c(1, 1))
```

And compute the closure by using the simplification logic, since we are interested in the knowledge that can be inferred from the condition given by the set :

```
R> cl <- base_implications$compute_closure(S, reduce = TRUE)
R> specific_implications <- cl$implications
```

UNIVERSIDAD
DE MÁLAGA

Inspecting the Rules

```
R> # Filter interesting rules
R> specific_implications$filter(rhs = c("Score=5"))
Implication set with 5 implications.
Rule 1: {Period of stay=Mar-May, Stars=4.5} -> {Score=5}
Rule 2: {Period of stay=Jun-Aug, Stars=4.5} -> {Score=5}
Rule 3: {Period of stay=Jun-Aug, Tennis court, Stars=3.5} -> {Score=5}
Rule 4: {Period of stay=Dec-Feb, Tennis court, Stars=3.5} -> {Score=5}
Rule 5: {Period of stay=Dec-Feb, Tennis court, Stars=3} -> {Score=5}
```

UNIVERSIDAD DE MÁLAGA

Table of contents

UNIVERSIDAD
DE MÁLAGA

GitHub

fcaR

lifecycle maturing | CRAN not published | build passing | codecov 92%

The goal of fcaR is to provide FCA tools inside the R environment.

Installation

The development version of this package can be installed with

```
remotes::install_github("neuroimaginador/fcaR", build_vignettes = TRUE)
```

- Installation instructions.
- Unit tests.
- Vignettes with demos.

UNIVERSIDAD DE MÁLAGA

Table of contents

UNIVERSIDAD DE MÁLAGA

Conclusions

Conclusions

- A package for fuzzy FCA in the R programming language has been presented.
- It is integrable with **arules**.
- Its main methods are related to the computation and plotting of the concept lattice and the calculation of implications and their management.
- The package is applicable to the creation of recommendation systems and to explore conceptual knowledge in a formal context.

Future Work

Future Work

- Include other theoretical aspects regarding computation of direct bases or positive and negative attributes.
- Implement more efficient algorithms for extracting concepts and implications.
- Make it available in CRAN.

UNIVERSIDAD DE MÁLAGA

fcaR, an R package to handle fuzzy implications: design of a recommendation system for medical diagnosis

D. López, A. Mora

Universidad de Málaga

UNIVERSIDAD DE MÁLAGA