

Augmented Visualization of Association Rules for Data Mining

Wilson Castillo Rojas

wilson.castillo@unap.cl

Alexis Peralta

alexisperalta@unap.cl

Claudio Meneses Villegas

cmeneses@ucn.cl



Outline

- Overview
- Visualization in Data Mining
- The VAM-MD Scheme
- Case study: Association Rules with SOM
- Experimental Prototype
- Evaluation and Analysis of Results
- Conclusions & Future Work

Overview

- The **Data Mining (DM) Process is complex:**
 - obstacles on data setting
 - several entities involved
 - research is needed
- The **outputs of the KDD process are focused on data:**
 - robust algorithms for DM models.
 - reliable data results.
 - massive data support.

Overview...

But...

- on a massive amount of data the pattern recognition task is hard.
 - is not easy to reach the entity data.
 - lack of understanding about the model structure.
- As a result; Output goals (knowledge) from the KDD process must be improved:
 - interaction with KDD entities.
 - how the entities are constructed.
 - better understanding of the underlying data mining model...

How? an option, **use visualization.**

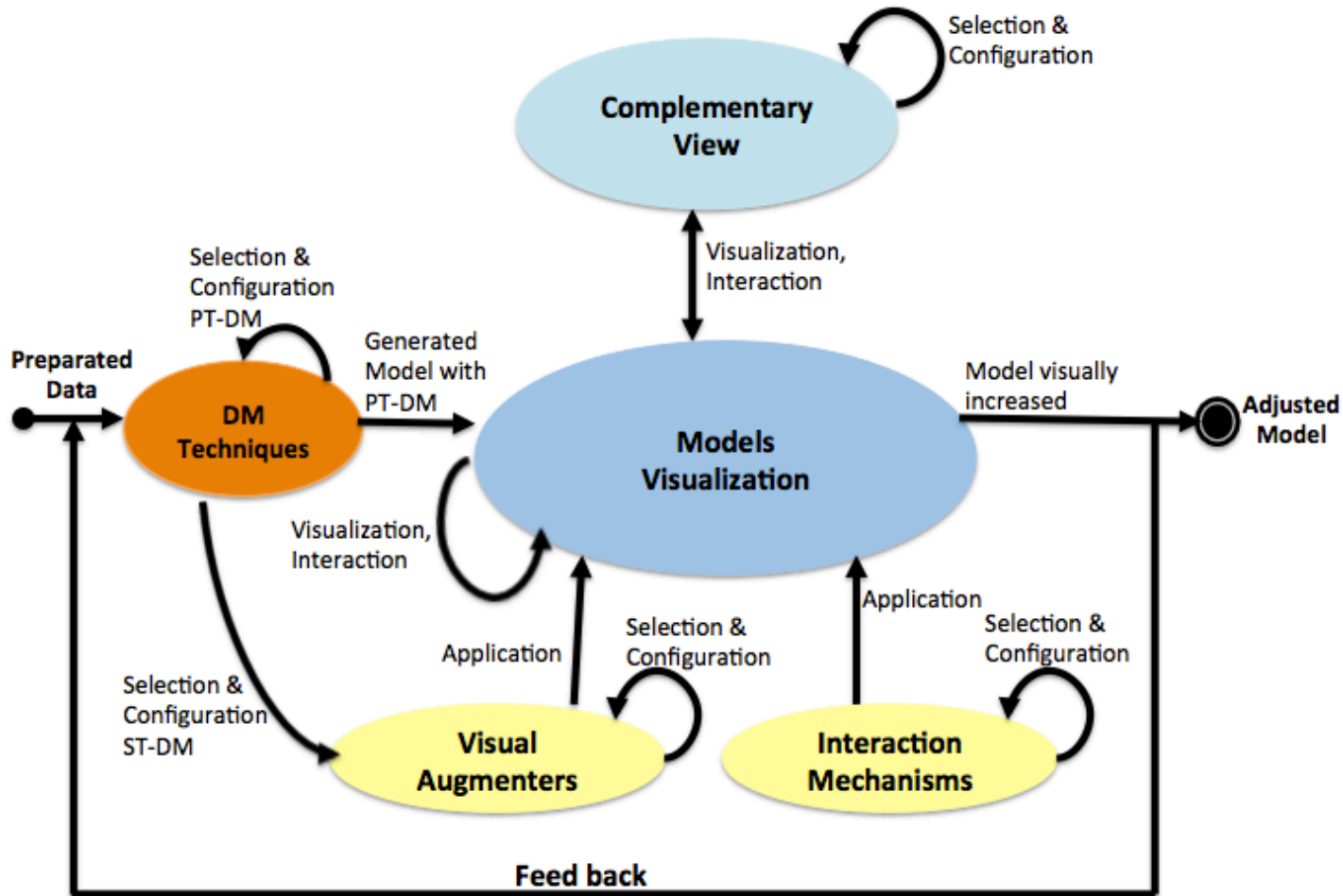
Visualization in Data Mining

- Visualization is slowly being considered on the KDD process.
- The role of visualization is not defined as a part of the KDD process.
- Visualization key factors:
 - preliminary data understanding.
 - DM tasks could be improved at the problem formulation phase.
 - issues with High-Dimensional datasets.
 - comparison and validation between similar models from the human visual perspective.
- Visualization in: Data, Models and Patterns.

The VAM-MD Scheme

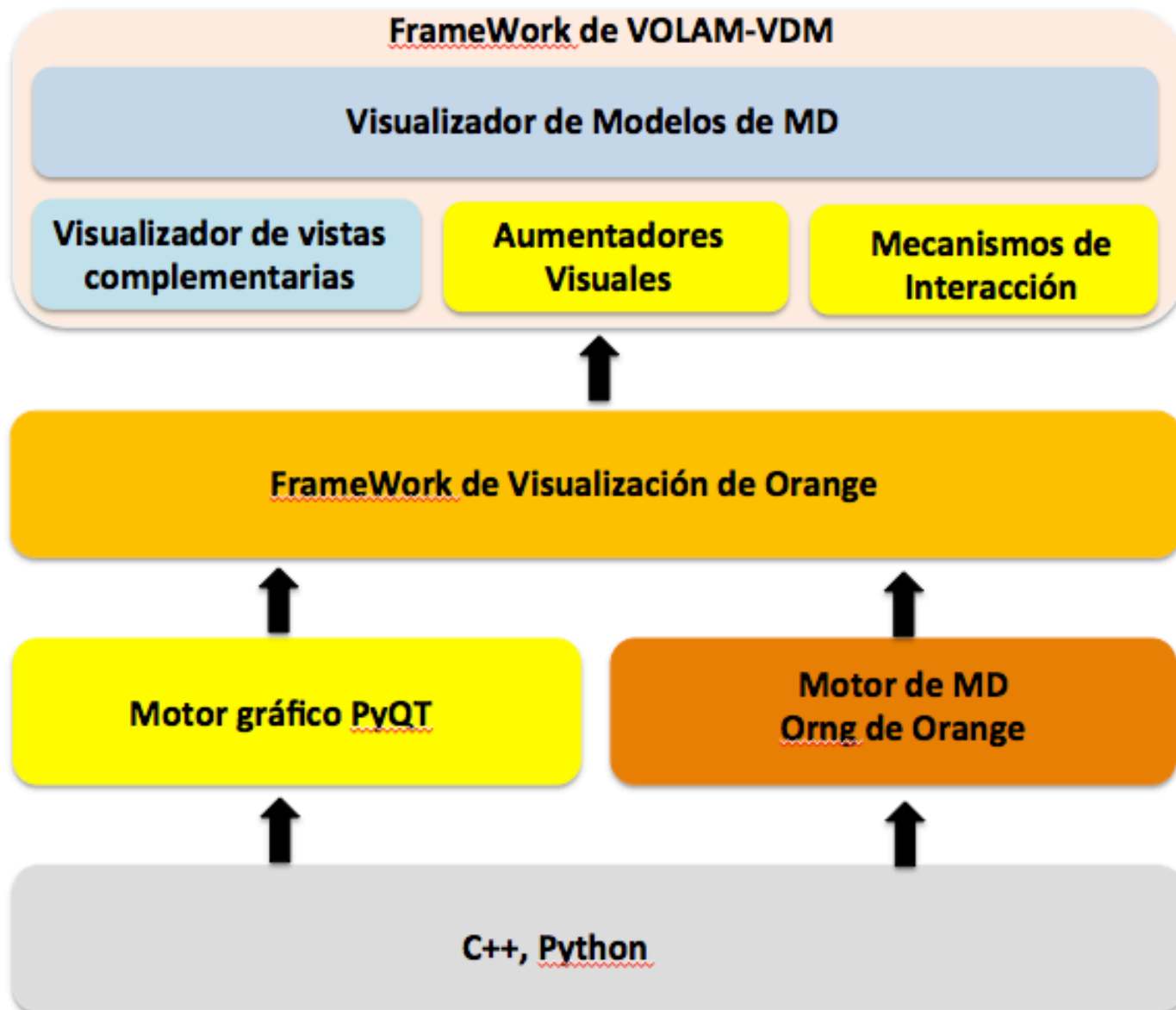
- VAM-MD: Augmented Visualization for Data Mining Models (*Spanish's acronym*)
- Research Goal:
 - **Improve the understandability of data mining models through the use of visualizations.**
- Features:
 - **visual perception model with user interaction.**
 - **focused on the step of adjusting** or tuning of the model DM.
 - **it brings the concept of '*Augmented Visualization*'** for DM models.
 - **use of traditional graphical artifacts** (applied to the data of the model components).
 - **combination of DM techniques** (multi-scheme approach on two layers for techniques):
 - PT-DM: Primary DM technique (model to adjust)
 - ST-DM: Secondary DM technique (exploratory model)

The VAM-MD scheme...



- Characteristics or conditions of visual augmenters:
 - provide a complementary viewing.
 - ad-hoc data domain.
 - the ST-DM should be descriptive.

The VAM-MD scheme...



Case Study: Association Rules with SOM

- Main problem:
 - **Lack of user understanding about inner-working DM model structure and mechanical.**
- In the case of Association Rules (AR) models is more complex, because of:
 - many rules.
 - massive amount of data.
 - high dimensionality in the data.
 - structure.

In General, the AR models analysis is:

Generated sets of large itemsets:

Size of set of large itemsets L(1): 8

Size of set of large itemsets L(2): 13

Size of set of large itemsets L(3): 2

Best rules found:

1. venomous=false tail=true 71 ==> backbone=true 71 conf:(1)
2. tail=true 75 ==> backbone=true 74 conf:(0.99)
3. backbone=true tail=true 74 ==> venomous=false 71 conf:(0.96)
4. backbone=true 83 ==> venomous=false 79 conf:(0.95)
5. breathes=true 80 ==> fins=false 76 conf:(0.95)
6. airborne=false 77 ==> feathers=false 73 conf:(0.95)
7. tail=true 75 ==> venomous=false 71 conf:(0.95)
8. tail=true 75 ==> backbone=true venomous=false 71 conf:(0.95)
9. breathes=true venomous=false 75 ==> fins=false 71 conf:(0.95)
10. breathes=true 80 ==> venomous=false 75 conf:(0.94)



Static rule; poor or no interaction
with components

Statistics & general information:
instances, rules, confidence,
support, etc.

Statistics & general information: + instances, rules, confidence, support, etc.

Data mining SOM technique, provides:

- Data dispersion in each rule,
- Spatial distribution &

Generated sets of large itemsets:

Size of set of large itemsets L(1): 8

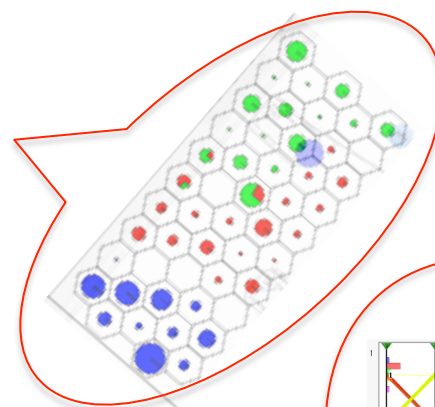
Size of set of large itemsets L(2): 13

Size of set of large itemsets L(3): 2

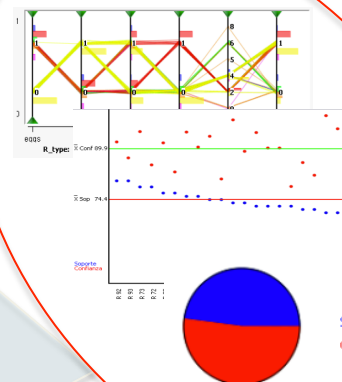
Best rules found:

```
1. venomous=false tail=true 71 ==> backbone=true 71    conf:(1)
2. tail=true 75 ==> backbone=true 74    conf:(0.99)
3. backbone=true tail=true 74 ==> venomous=false 71    conf:(0.96)
4. backbone=true 83 ==> venomous=false 79    conf:(0.95)
5. breathes=true 80 ==> fins=false 76    conf:(0.95)
6. airborne=false 77 ==> feathers=false 73    conf:(0.95)
7. tail=true 75 ==> venomous=false 71    conf:(0.95)
8. tail=true 75 ==> backbone=true venomous=false 71    conf:(0.95)
9. breathes=true venomous=false 71    conf:(0.95)
10. breathes=true 80 ==> venomous=false 75    conf:(0.94)
```

- Compares rules (quantitatively).
- Dynamic AR model (navigable)



Visual
augmenters



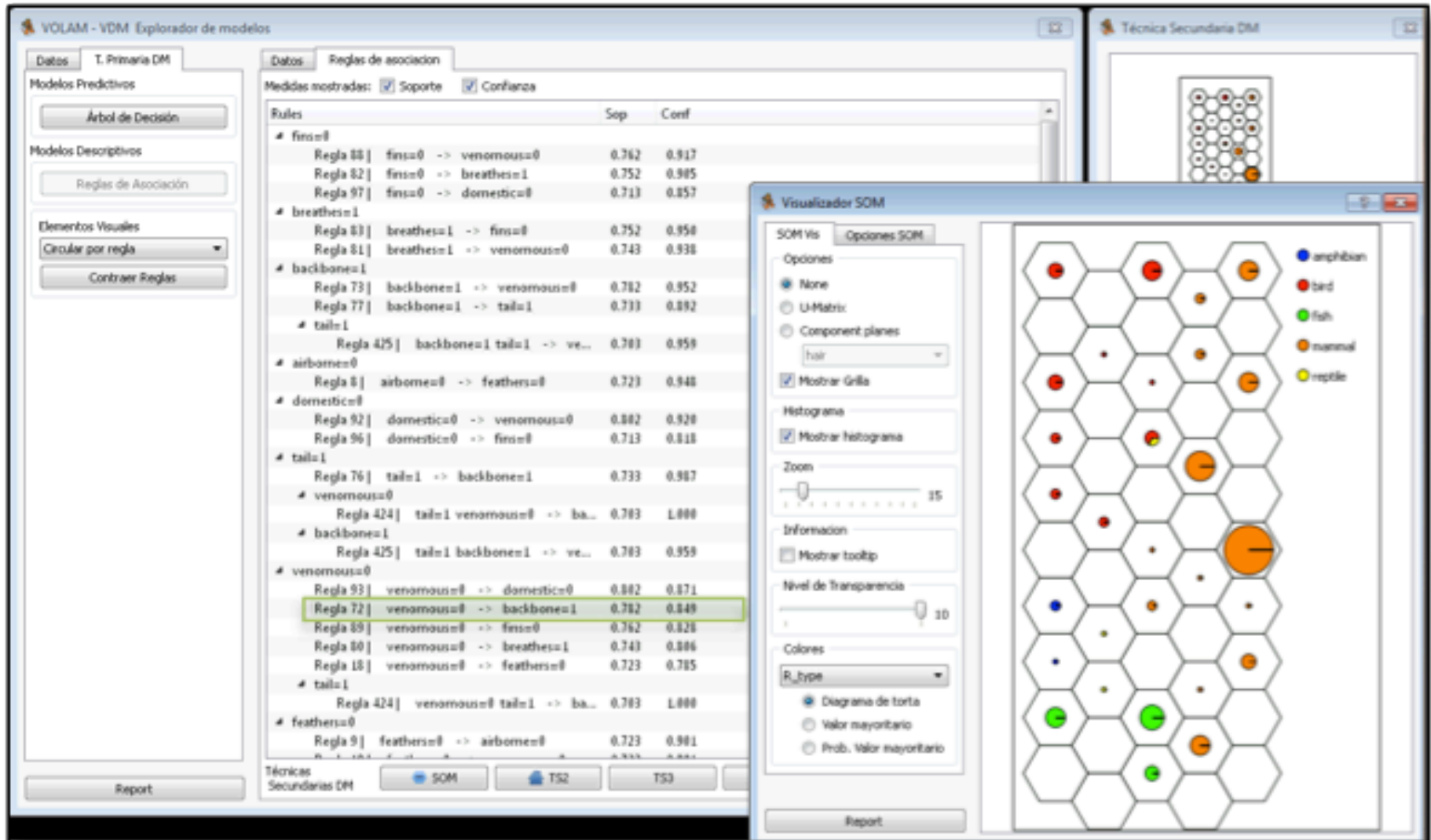
Visual elements or graphical artifacts (ad-hoc) applied to rule data.

Visualization & interaction
(with model and its components)

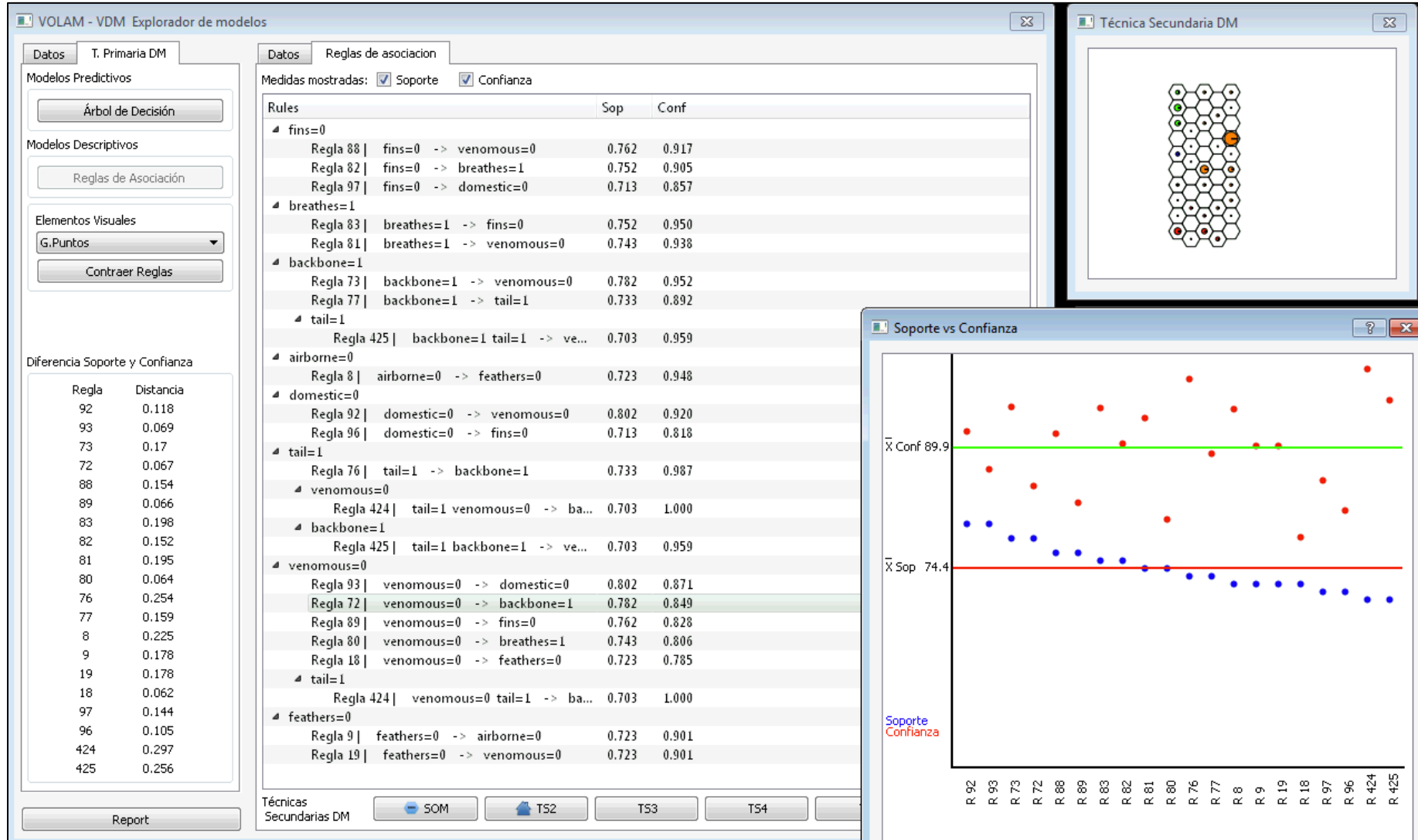
Experimental Prototype

- The prototype implements the VAM-MD scheme.
- Hierarchical structure is covered by the scheme with:
 - PT-DM: AR (Association Rules) , DT (Decision Trees), ...
 - ST-DM: SOM, clustering, ...
- Visual elements:
 - data table, pie chart, scatter plot, box-plot, etc.
- User actions:
 - Zoom (in/out), select, drag & drop, tuning, transparency for context, etc.

Experimental Prototype...



Experimental Prototype...



Evaluation and Analysis of Results

- The proposed approach **was tested on a controlled experiment.**
- A benchmark test **was developed to assess the usefulness of the proposed approach** to perform some data mining tasks.
- The experiment:
 - The test was performed on 17 users, each one with different levels of DM process.

Evaluation and Analysis of Results...

- Objectives of the controlled experiment:
 - **To assess usefulness provided by the combination and application of the SOM technique on the components of a AR.**
 - **To get a description of the level of effectiveness of the prototype software.**
 - **To check empirically if users are able to get a better understanding of the AR model, applying the SOM technique.**

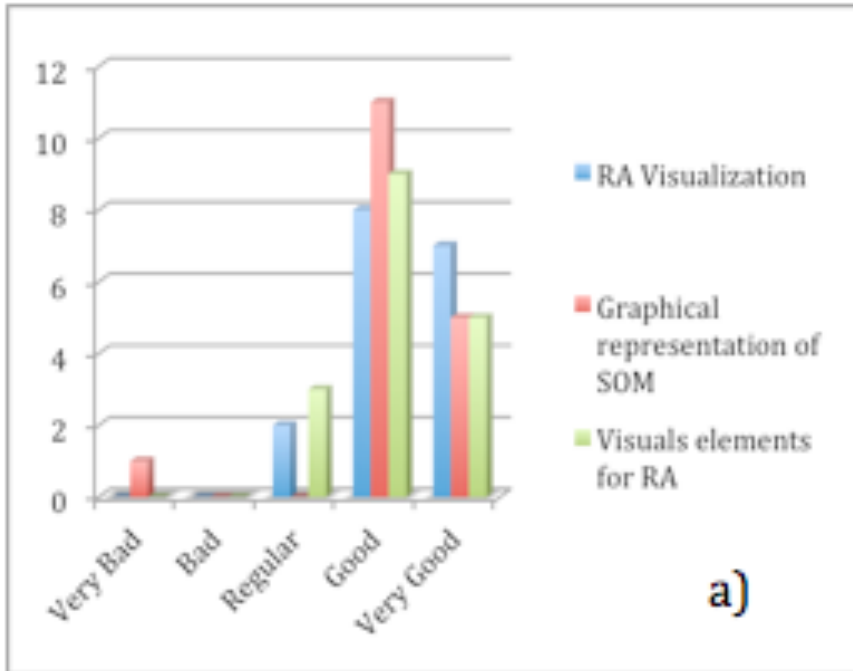
Evaluation and Results Analysis...

- The controlled experiment considers:
 - With both tools (Prototype & Weka), users should:
 - generate a AR model,
 - visualize and interact with the model, and
 - interpret patterns or rules obtained using the options available on the two software tools.
 - Focus on interaction and usefulness.

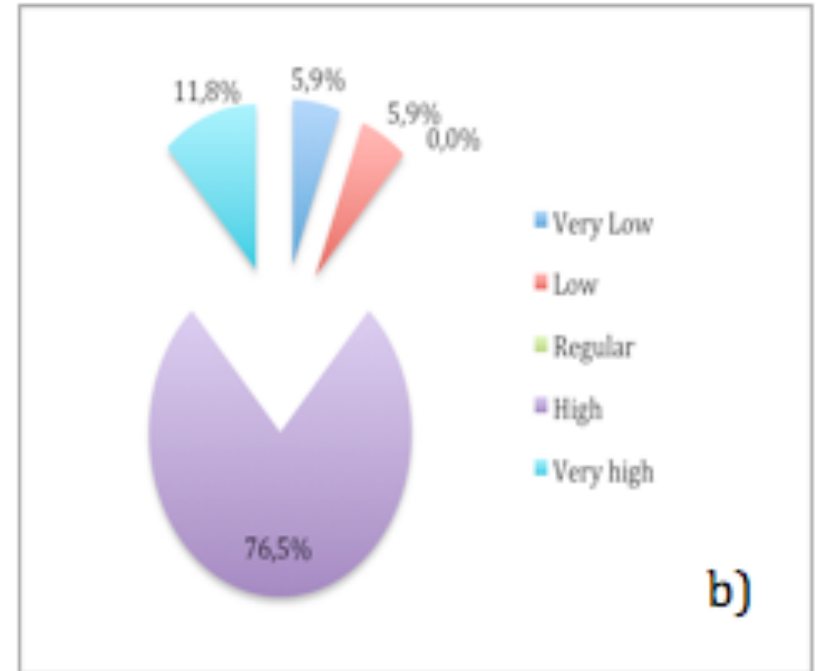
Evaluation and Analysis of Results

- Results:
 - The experimental prototype was **widely accepted** by the users.
 - The key factors were **usability and performance**.
 - The users consider the prototype successful on **showing relationships in data**.
 - Wide assessment of prototype software for the actual utility of the options and parameters offered, in order to get a better understanding of the AR model.
 - It was successful on **augmenting the model understanding** from user point of view.

Evaluation and Analysis of Results



a) Level of acceptance of views available to describe the AR model.



b) Ability to describe the data on the AR model using the SOM technique

Conclusions

- **The VAM-MD scheme and its components serve as a guide** for building visualizations that support the task of analysis and exploration of AR models, in their stage of refinement or adjustment.
- The positive perception obtained in the controlled experiment, giving rise to hold **the convenience and utility of combining a complementary technique with a previously generated model.**
- The **combination of visualization techniques can improve the understanding of AR models**, compared to other tools.
- **Visual elements or graphical artifacts** provided in the software prototype, **applied to the data in each rule** of the AR model, **meet to support the analysis and exploration of the generated model.**

Future Work

- **To formalize the VAM-DM abstract model**, developing formal specification to approach all components.
- **To explore (add) other data mining techniques** for:
 - PT-DM & ST-DM.
- **To include other visual elements**, some of them should be ad-hoc to each kind of model.
- **The interaction mechanisms must be improved** to enhance the model perception.



Thank you...
questions?

Wilson Castillo Rojas

wilson.castillo@unap.cl

Alexis Peralta

alexisperalta@unap.cl



Claudio Meneses Villegas

cmeneses@ucn.cl

