Augmented Visualization of Association Rules for Data Mining

Wilson Castillo-Rojas[1], Alexis Peralta[1], and Claudio Meneses[2]

[1]Faculty of Engineering and Architecture, Arturo Prat University, Iquique - Chile

wilson.castillo@unap.cl,alexisperalta@unap.cl

[2]Faculty of Engineering and Geological Sciences, North Catholic University, Antofagasta - Chile

cmeneses@ucn.cl

Abstract. This paper describes a proposal for enhanced visualization of a data-mining model generated with Association Rule (AR) techniques by applying Self-Organizing Maps (SOM). A representation of visual perception model of AR based on a method called AVM-DM (Augmented Visualization Models for Data Mining) is established, together with data and patterns, which support the visual exploration stage, thus fitting in the context of the KDD (Knowledge Discovery in Databases) process. This methodology seeks to answer generic user questions regarding the inner workings of the model, and to support understanding the generated model. The use of the SOM technique as a visual enhancer applied to an AR model, serves a dual purpose: to obtain the spatial distribution of the subset of data associated with a rule, and to display this subset using a map. The visualization of the RA model, proposed in this work, is implemented through a software tool giving users different interaction mechanisms. Results of user experiments demonstrate the usefulness of the proposed SOM technique in visually enhancing and helping to understand the AR model.

Keywords: Data mining, visual data mining, visualization of data mining models, visualization of association rules.

1 Introduction

In a KDD process, the utility of a Data Mining (DM) model depends mainly on two factors: the ability of the model to discover interesting patterns and the ease with which the model structure can be understood and adjusted by users. Thus, along with the predictive and descriptive power of a DM model, its structure should be well understood and interpreted by the users, because the classification or description of the data without an explanation model induced from data, can reduce the credibility of the

results of the KDD process [1]. In this regard, appropriate visualizations of DM models can transform them into understandable tools that convert data into knowledge. In addition, appropriate visualizations of patterns can facilitate the task of discovering knowledge to interpret and evaluate these patterns visually [2, 3].
This work proposes to visually enhance the DM model generated by an Association Rule mining (AR) technique, by combining the SOM technique and creating complementary views of the different rules or model components. This method seeks to answer generic user questions regarding the inner workings of the model. This approach is based on the Augmented Visualization Models for Data Mining (AVM-DM) [4] method that proposes a model of visual perception and user interaction, focusing on the stage of adjustment or refinement of the DM model generated within the wider context of the entire KDD process.
The proposed work includes the implementation of part of the AVM-DM method in a prototype tool that accepts a set of appropriate data and an AR model. Finally, a subjective evaluation of the prototype is presented through the user evaluation experiment, consisting of a survey . Participants provided information about the performance, usability, management views and support provided by the developed tool in understanding a previously generated DM model.

2 Visualization of Association Rules

AR represent the relationships between several variables, i.e., consider that AR is an implication of the form $X \rightarrow Y$, where X is a set of items called antecedents, and Y is the set of consequent items. At least five parameters should be considered in the visualization of an AR: the set of antecedent items, consequent items, associations between antecedents and consequents, the rule's support, and its confidence [5].
Research on visualization of AR can be categorized into three main groups, depending on whether they are based on tables, matrices, or graphs. Tabl-based techniques are the most common and traditional approach to represent AR. The columns of a table generally represent the items of the AR model while each row represents a rule. Examples of techniques based on tables can be found in several commercial systems, including SAS Enterprise Miner and DB Miner [6]. Matrix-based techniques such as those implemented in MineSet [7] and InfoVis [8] use a coordinate axes grid that represents the antecedents and consequents. The last group consists of the techniques that are based on graphs using nodes to represent the items and edges to represent the associations of

items in the rules. Some of these techniques have proposed several types of representations known to study a large set of data, such as hyperbolic trees [9].
In summary, although these efforts to improve the visualization of ARs were able to supplement the rule mining with graphics that allow us to observe each rule in detail, we failed to find visualization tools that allow an interaction with each rule, while also visualizing how the data in each rule are spatially distributed. A comparative review of visualization tools for DM (including AR) techniques, by Castillo [5], concluded that: a) most research recommends using a combination of DM techniques with appropriate views, b) it is essential to consider in the design of views, the mechanisms for user interaction, and c) the role of visualization in the KDD process must be extended in all its stages.

3 The AVM-DM Scheme

The AVM-DM scheme proposed by Castillo in [4] considers the characteristics of the analyzed models of perception, and includes the most relevant aspects of each, particularly with regard to the integration of the display in step adjustment or refinement and evaluation of DM models. AVM-DM brings the concept of "Augmented Visualization" for DM models, and suggests that, given a DM technique to be visualized, called Primary DM Technique (PT-DM), should allow the user to incorporate in this display, different visuals regarding the type of model and data domain, and in turn need to apply another DM technique, called Secondary DM Technical (ST-DM), as a visual enhancer that allows exploring the PT-DM. The selected ST-DM technique must meet the requirements of being a descriptive DM technique that is appropriate to the domain data being analyzed within the PT-DM.

4 Augmented Visualization of AR Model using SOM

In the case of the AR techniques, several visualization methods analyzed in [5] propose a static display, without any possibility for the user to interact with each rule. Most DM visualization tools delivered an overview of the ARs but cannot combine DM techniques to provide information on model rules and instances supporting each rule, and only a few tools provide interaction mechanisms for the user. The proposed use of the SOM technique as an AR visual augmenter, serves a dual purpose: to obtain the spatial distribution of the subset of data associated with each rule, and to display this partition using a map.

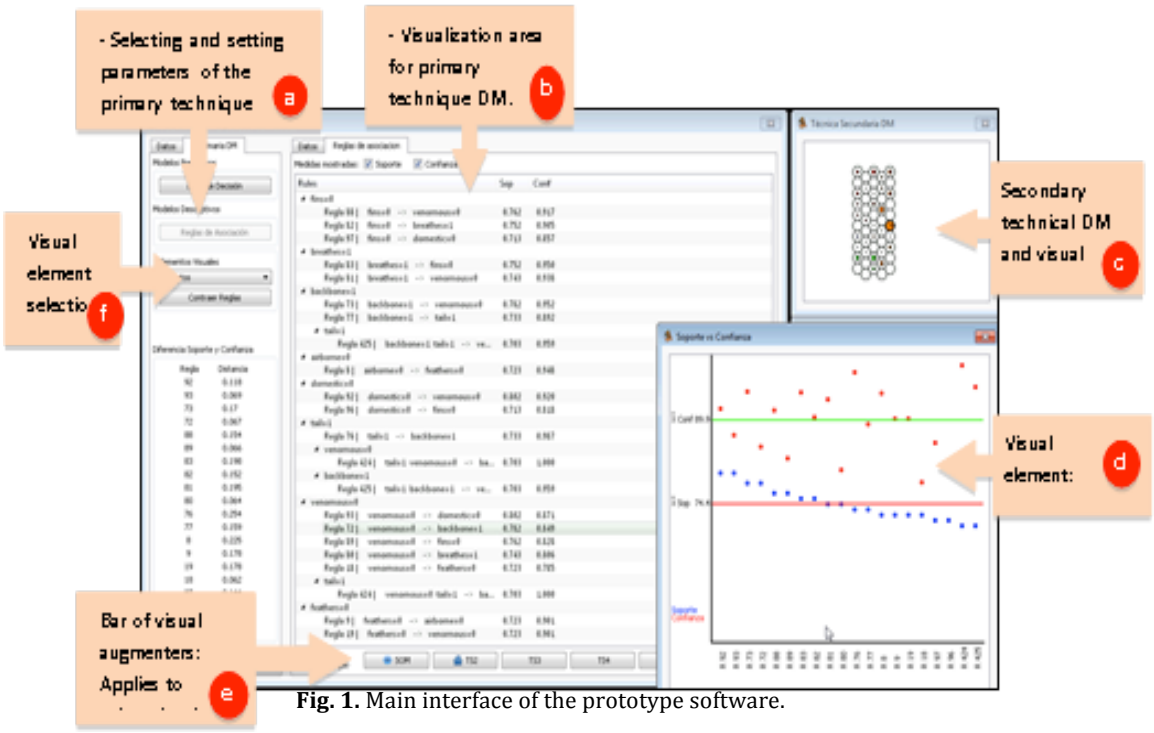

Fig. 1. Main interface of the prototype software.

The prototype implements the AVM-MD scheme for hierarchical structure techniques in DM (decision tree & AR), and in this paper, we concentrate on the AR mining technique. It incorporates a set of visual elements; data table, pie chart (by rule and general), dot plot, and parallel coordinates plot. Also, available interaction mechanisms include zoom, selection rules, and setting of the parameters. Figure 1 shows the main interface of the experimental prototype, where ARs are displayed in the central part, together with complementary views and visual elements on the right side. In this tool all architecture components of the proposed AVM-DM scheme are implemented. The user can maximize the image located in the c) section of the interface by clicking with the mouse, opening a window that presents a detailed view of this technique. They can re-configure their initial parameters on a selected rule and apply the SOM technique. Also, the user can see the shape of the distribution of the instances covered by this rule.

5 Controlled Experiment: Evaluations & Analysis

The following controlled experiment provides a comparison and subjective evaluation of the visualization of ARs obtained through a DM task to be performed by a set of users, whose aim is to check if the SOM-based visualization enhanced AR mining along with the set of visual elements provided by the prototype software, can improve the understanding of the model, such as looking at the distribution of data in each rule, compared with the visualization provided by another DM tool, that does not have this focus or visualization scheme. This experiment was conducted with 17 users of varying levels of expertise in DM processes, and the use of DM tools. We asked participants to perform a generic task description and could answer questions about the model and its components, and to relate the model to the characteristics of the data from which the model was generated.

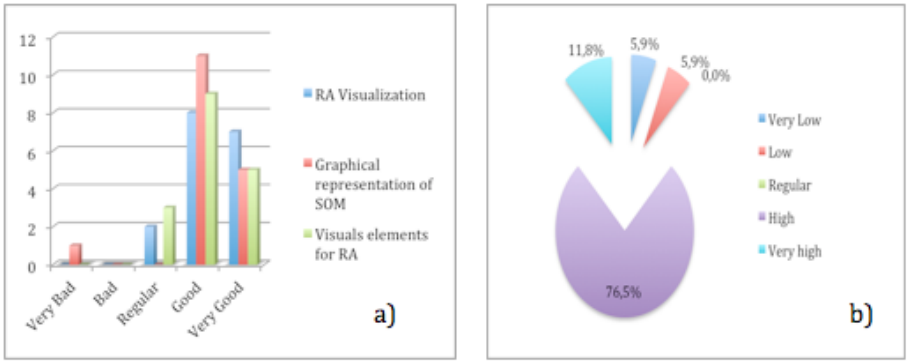

Fig. 2. a) Level of acceptance of views available to describe the AR model. b) Ability to describe the data on the AR model using the SOM technique.

Subsequently, once the DM task was prepared for this experiment, the users had to answer a survey designed to gather the subjective opinion of the group, regarding the performance of both tools, the visualization of the generated AR model, usability, utility of visual elements, the desirability of combining the SOM technique to achieve a visually augmented model, and the efficiency in understanding of the model. Users mostly stated that both the combination of the SOM technique applied to the AR model, and the use of graphic elements on the data rules, allowed them to improve their understanding of the generated AR model, achieving a score distribution of 54, 9% good and 33.33% very good, which can be seen from the graph in Figure 2 a). Also, as shown in Figure 2 b), users expressed mostly positive ability (76.5% high and 11.8% very high) to

obtain an augmented visualization of the AR model.

6 Conclusions and Future Work

The preliminary results of the presented study allow us to confirm the suitability and utility of combining the AR mining technique with the SOM technique for achieving augmented visualization for the AR model, and for visualizing the spatial distribution of the data covered by each rule, thus helping improve the understanding of their inner workings. Also the visual tools provided in the prototype software support the analysis and examination of the AR model. As future work, we are evaluating other descriptive DM techniques that can provide alternative views for visually enhanced AR models.

7 Acknowledgements

We thank the anonymous reviewers for their helpful suggestions. In particular, we thank all the effort in the editing phase which substantially improved the readability of the paper.

8 References

1. Keim, D.A., (1997). Visual Techniques for Exploring Databases. Third International Conference on KDD & Data Mining. Newport Beach, CA, August.
2. Meneses, C. J. & Grinstein, G. G., (2001). Visualization for Enhancing the Data Mining Process. In Proceedings of the Data Mining & KDD: Theory, Tools, and Technology. III Conference. Orlando-FL, April.
3. Thearling, K., Becker, B., Mawby, B., Pilote, M., Sommerfield, D. (1998). Visualizing Data Mining Models. In Proceedings of the Integration of Data Mining and Data Visualization Workshop, Springer-Verlag.
4. Castillo-Rojas, W., Meneses, C., & Medina, F. (2013). Augmented Decision Tree Models Using SOM. 6th Latin American Conference on Human Computer Interaction, Costa Rica. Proceedings pp. 148-155. Springer LNCS 8278, ISBN 978-3-319-03067-8.
5. Castillo-Rojas, W., & Meneses, C. (2012) Comparative Review of Schemes of Multidimensional Visualization for Data Mining Techniques. III International Congress of Informatics. August, Arica–Chile.
6. Han, J., Kamber, M. (2001). Data Mining Concepts and Techniques. Morgan Kaufmann.
7. Brunk, C., Kelly, J. and Kohavi, R. (1997). MineSet: An Integrated System for Data Mining. Proc. of Third Intel: Knowledge Discovery and Data Mining, pages 135-138.
8. Wong, P.C., Whitney, P., (1999). Visualizing association rules for text mining. INFOVIS. Pages 120–123.
9. Lamping, J., Rao, R., and P. Pirolli. (1995). A focus+context technique based on hyperbolic geometry for visualizing large hierarchies. In Proceedings of the ACM conference on Human Factors in Computing Systems, ACM Press. USA, pages 401–408.