IJCSS
International Journal of Complex Systems in Science

Int. J. Complex Systems in Science
vol. **4(1)** (2014), pp. 21–26

Extracting comprehensible patterns from Venezuelan Assets by means of annotated Traffic Light Pannel

Karina Gibert[1,2,†] and Dante Conti[3]

[1] *Dep. of Statistics and Operation Research, Universitat Politècnica de Catalunya-BarcelonaTech, Spain*

[2] *Knowledge Engineering and Machine Learning Group, UPC-Barcelona-Tech,Spain*

[3] *Dep. of Operations Research, Universidad de Los Andes, Mérida, Venezuela*

Abstract. In this work the relationship between Venezuelan assets and the general Index of the Venezuelan Stock Exchange is analyzed by means of data mining methods. In particular, clustering followed by the Traffic Lights Panel visualization of cluster prototypes is used to understand the meaning of the discovered patterns. Also, associations between Venezuelan index and Dow Johnes (NY) or BOVESPA (Brazil) show the association with international context. The work confirms a well-known fact, that is the self-behaviour or Venezuelan stock market, often disconnected from international behaviour. Also, intrinsic uncertainty associated with the class prototypes is introduced into the visualization through annotated-TLP and more reliable or stable patterns can be distinguished from more variable, or volatile.

Keywords: Clustering, Knowledge Discovery in Databases (KDD), Data Mining, Patterns Interpretation, Post-processing, Traffic Lights Panel (TLP), annotated-Traffic Lights Panel, Financial assets, Venezuela Stock Exchange.
MSC 2000: 91C20, 62H30, 68T10

† **Corresponding author:** Karina Gibert karina.gibert@upc.edu
Received: December 29th, 2013
Published: March 1st, 2014

1. Introduction

It is well known that Knowledge Discovery in Databases (KDD) provides a framework to support analysis and decision-making regarding complex phenomena [4]. Mining of financial data is currently being used to complement classical financial modelling [14], [13] which has shown poor performance in some complex phenomena, like Venezuelan Stock Exchange (Bolsa de Valores de Caracas) [2] [5] [11]. One of the most popular KDD methods is Clustering [12]. Thus, in this paper, clustering techniques are used to find patterns on financial data related to assets of Venezuelan Stock Exchange, which is a particularly complex market due to the governmental interventions, among others.

However, often there is a gap between the raw data mining results and a real comprehension from the end-user point of view that might decrease the power of data mining to support complex decisions [9]. In the context of clustering, postprocessing techniques might contribute to bridge this gap. In this work, the use of Traffic Light Pannel [6] and some extensions is proposed to provide understandability of the financial discovered profiles.

2. Case Study and previous work

This work is a collaboration between Bolsa de Valores de Caracas (Venezuela), Universidad de los Andes (Venezuela) and Universitat Politècnica de Catalunya-BarcelonaTech (Spain).

As said before, hierarchical clustering is applied to find patterns on data provided by Bolsa de Valores de Caracas. The data refers the weekly variations in the price of 4 financial assets from Venezuela Stock Exchange (Bolsa de Valores de Caracas) considered relevant for the Venezuelan index (IGBC) and the two major indexes related to this market (Dow Jones-USA and BOVESPA-Brazil). Collected data corresponds to the period from January 1998 to April 2008. Multivariate patterns describing the relationships among the Venezuelan assets and both the internal and international indexes have been found, and, as usual in hierarchical clustering, the number of classes can be determined a posteriori, by optimizing the Calinski-Harabaz index [1]. A clustering in 5 classes was found.

3. Understanding patterns by means of post-processing

Performing a clustering over a set of data requires an important process of understanding the underlying genesis of the clusters to be able to find the

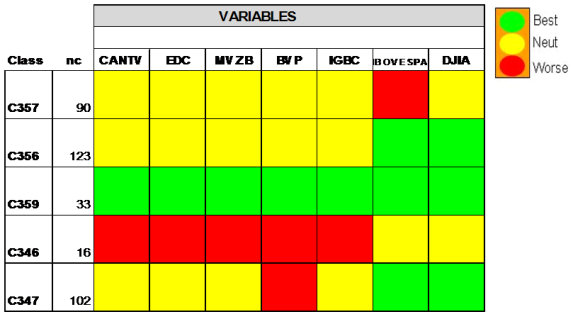

Figure 1: TLP describing the 5 clusters found.

right decision/action to be associated to every cluster. Till now, only few works can be found addressing these issues.

The Traffic Light Panel (TLP), introduced by Gibert [8], is a symbolic postprocessing of clustering oriented to help the expert to better understand the clusters and to identify domain concepts referred to the discovered classes. TLP was conceived to support the cluster's conceptualization, as a first bridge between clustering and effective decision-making. TLP proved to be extremely useful and well-accepted by domain experts in real applications from several domains, like mental health [8], health policy [10], water management [6], tourism or financial assets. In [3], TLP proved as a reliable goodness-of-clustering indicator.

The resulting TLP shows 5 clear patterns giving a realistic picture of the behaviour of the Venezuelan stock market (Fig 1). The patterns are directly understood by the experts, just looking at the TLP, as the color-coding is highly symbolic and do not requires extra explanation for non-technical users. A set of independent experts validated the results.

However, the TLP is giving a prototypical description of the patterns, which might be perturbed by some deviation of individual objects around those patterns (in this case, this means that the uptrend of BOVESPA index for some particular day of pattern c357 might be higher or slightly lower and it is not always exactly the same). The homogeneity of the pattern is not represented in the TLP as it was conceived in [8]. Thus, the annotated TLP (a-TLP) was introduced in [7] moving to a 9 colours scale, where desaturation of the three basic colors was associated to the impurity of the pattern itself, in that case, measured as the variation coeficient (VC) of each variable inside the classes. In this work impurity is measured as the quotient between interquartilic range and the median (named P), as an equivalent robust version of VC. Thus, figure 2 shows the enrichment of the original TLP with the information about which cells are more reliable in the patterns, according to their purity. Upon

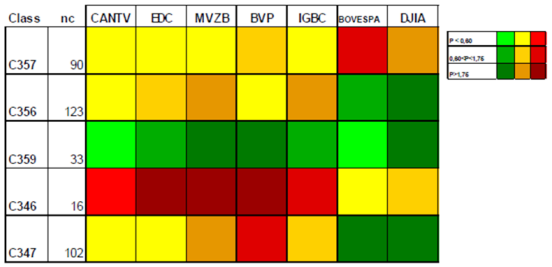

Figure 2: a-TLP introducing information about the purity of the patterns.

this new visualization one can learn what is really realiable in pattern c359 is that CANTV and BOVESPA keep in down trend, and in general, most of the assets and indexes behave around downtrend, whereas for some days, DJIA, MVZB and BVP might follow quite different situations. Similarly, pattern c357 shows as most reliable the stability of CANTV, EDC, MVZB and IGBC, whereas BVP might show more variability, DJIA might be even unstable and BOVESPA moves around uptrend but with some heterogeneity. This kind of information is refining the one provided by the original TLP by informing the user how homogeneous is the pattern itself and giving the user a more complete picture about how risky might be the decisions made upon a certain pattern.

4. Discussion and Conclusions

In this work Clustering is used to find patterns in the Venezuelan Stock market describing the relashionships between some assets and the general Venezuelan index, as well as some international indexes supposed to be relevant for the Venezuelan market. Post-processing techniques are used to understand the patterns proposed in the clustering process. The TLP is providing a symbolic view of the general trend of every asset and index for every pattern. The TLP is visualizing the prototypical behaviour of each class. In this work an enrichment of the original TLP, named annotated-TLP is introducing into the picture information about the homogeneity of the pattern by dessaturating the basic color of each cell according to a measure of variability. In this case, a ratio between interquartilic range and the median is used as a robust measure of purity. The patterns identify both most frequent and most particular situations and provides a good insight on the behavior of the Venezuelan stock market where, the national index IGBC frequently evolves independently from Brazil and NY (all clusters except c359, the smallest one), and some internal assets can have less influence than expected to the national index (BPV). This

is part of the know-how of specialized financial experts in Venezuela, and presumably, one of the reasons why the classical financial models perform poorly for this particular market.

Currently, temporal relationships among the discovered patterns are modelled to provide a qualitative framework to describe the dynamics of the process.

Acknowledgements Thanks to the Gerencia de Comunicación de la Bolsa de Balores de Caracas, Venezuela for its collaboration. To ME Naranjo and A Sánchez for collaboration in preprocessing and association rules mining. This research has been partially supported by project AGAUR-2009-SGR 1365

References

[1] T. Calinski and J. Harabasz, A dendrite method for cluster analysis, *Communications in Statistics - Simulation and Computation* **3(1)**, 1-27, Taylor and Francis(1974).

[2] D. Conti, M. Bencomo and A. Rodriguez, Optimal investment portfolio determination by using non linear programming, *Ciencia & Ingenieria* **26(2)**, 43-50 (2005).

[3] D. Conti and K. Gibert, The use of Traffic Lights Panel as a Goodness-of-Clustering Indicator: An Application to Financial Assets, *In Artificial Intelligence Research and Development. Frontiers in Artificial Intelligence and Applications* **248**, 19-28, IOSPress(2012).

[4] U. Fayyad, G. Piatetsky-Shapiro and P. Smyth, *From Data Mining to Knowledge Discovery: An overview. In: Advances in Knowledge Discovery and Data Mining*, AAAI/MIT Press.(1996).

[5] T. Fu, F. Chung, R. Luk and C. Ng, Stock time series pattern matching: Template-based vs. rule-based approaches, *Engineering Applications of Artificial Intelligence* **20(3)**, 347-364 (2007).

[6] K. Gibert, D. Conti and D. Vrecko, Assisting the end-user in the interpretation of profiles for decision support. An application to wastewater treatment plants, *Environmental Engineering and Management Journal* **11(5)**, 931-944 (2012).

[7] K. Gibert, D. Conti and M. Sanchez-Marre, Decreasing Uncertainty when Interpreting Profiles through the Traffic Lights Panel, *Communications in Computer and Information Science, CCIS* **298**, 137-148(2012).

[8] K. Gibert, A. García-Rudolph and G. Rodríguez-Silva, The role of KDD Support-Interpretation tools in the conceptualization of medical profiles: An application to neurorehabilitation, *Acta Informatica Medica* **16(4)**, 178-182 (2008).

[9] K. Gibert, G. Rodríguez-Silva and R. Annicchiarico, Post-processing: bridging the gap between modelling and effective decision-support. The Profile Assessment Grid in Human Behaviour, *Mathematical and Computer Modelling* **57(7-8)**, 1633-1639, Elsevier (2013).

[10] K. Gibert and L. Salvador-Carulla, Applying Clustering based on rules for finding patterns of functional dependency in schizophrenia. In JC Cortés, L Jódar et al Eds Mathematical Modelling in Social Sciences and Engineering, 2014, Nova (USA). ISBN:978-1-63117-335-6

[11] E. Hajizadeh, H. Ardakani and S. Jamal, Application of data mining techniques in stock markets:A survey, *Journal of Economics and International Finance* **2(7)**, 109-118,(2010).

[12] www.kdnuggets.com

[13] B. Kovalerchuk and E. Vityaev, Data Mining for Financial Applications, *In The Data Mining and Knowledge Discovery Handbook 2nd ed*, 1203-1224, Springer(2010).

[14] A. Wiegend, Data Mining in Finance, *Report from the Post-NNCM-96 Wksh. on Teaching Computer Intensive Methods for Financial Modelling and Data Analysis, Proc. 4th NNCM-96*, 399-411 (1997).