Financial Self-Organizing Maps

Marina Resta

DIEC, via Vivaldi 5, 16126
University of Genova, Italy
resta@economia.unige.it

Abstract. This paper introduces Financial Self–Organizing Maps (FinSOM) as a SOM sub–class where the mapping of inputs on the neural space takes place using functions with economic soundness, that makes them particularly well–suited to analyze financial data. The visualization capabilities as well as the explicative power of both the standard SOM and the FinSOM variants is tested on data from the German Stock Exchange. The results suggest that, dealing with financial data, the FinSOM seem to offer superior representation capabilities of the observed phenomena.

Keywords: Self–Organizing Maps, Value at Risk, Granger Causality.

1 Introduction

The 2008 great crisis dramatically highlighted the poor forecasting performance of existing Earling Warning Systems (EWS) i.e. those automatic systems that looking at proper combinations of macroeconomic variables would have alerted both policy makers and investors, hence stemming most dramatic aspects of the flooding financial wave. From the perspective of automatic systems design, this experience suggested the importance to develop new systems that are able to offer more readable and intuitive results, to facilitate the task of monitoring and regulating the overall level of risk. This rationale has recently inspired the development of EWS based on the paradigm of Kohonen maps [5]: an EWS based on Self–Organizing Maps (SOM) was suggested in [8] to measure the economic vulnerability of countries, and to estimate the probability of future crises; [13] and [14] discussed a fuzzified version of SOM, particularly well suited to apply on macroeconomic variables and to pickup alerting signal for upcoming financial shocks; [9] and later [10] analyzed a hybrid structure joining SOM to graphs in order to enhance clusters visualization capabilites of Kohonen maps, and used it to analyze the topological structure of various financial markets. A common aspect of the cited works is that in substance they leave unchanged the backbone of the original Kohonen's algorithm. However, a not negligible issue concerns the way SOM perform the mapping task. Obviously, depending on the metric in use, results can consistently vary: [6], for instance, discovered that hyperbolic space is ideally suited to embed large hierarchical structures, as later proved by the Hyperbolic Self-Organizing Map [12]. Moreover [1] show that in high dimensional

S. Wermter et al. (Eds.): ICANN 2014, LNCS 8681, pp. 781–788, 2014.

space, the concept of proximity, distance or nearest neighbor may not even be qualitatively meaningful.

The main point, however, is that despite of the importance of information retrieval from financial markets, there is a lack of metrics specifically thought to manage financial data.

Moving from this point, in this paper we discuss an alternative approach, replacing the similarity measure which represents the core of the SOM algorithm by way of alternative functions with greater financial soundness, thus originating Financial Self Organizing Maps (FinSOM). FinSOM would allow to make the SOM algorithm most suitable to analyze financial data and to capture relevant information. In our view rather than a single algorithmic procedure FinSOM must be intended as a family of different SOM, whose members are characterized depending on the function used to manage the similarity between inputs and neurons in the topological grid. In this respect, we hereafter discuss two SOM variants obtained incorporating into the learning procedure, respectively, Value at Risk (VaR–SOM), and Linear Granger Causality (LGC–SOM). The structure of the paper is therefore as follows. Section 2 is divided into two subsections, to provide the reader with basic understanding of the proposed algorithms. Section 3 discusses an application on financial data, while Section 4 concludes.

2 The FinSOM Framework

As widely known, Self Organizing Maps [5] (SOM) assume to order a set of neurons, often (but not exclusively) arranged either in a mono–dimensional or in a 2–D rectangular/hexagonal grid, to form a discrete topological mapping of an input space $X \subset \mathbb{R}^n$.

Assuming for sake of simplicity a map made by M nodes, if we denote by $\boldsymbol{w}_i \in \mathbb{R}^n$ $(i = 1, \ldots, M)$ the weight vector associated to neuron r_i, the algorithm works as shown in the Box 1.

Algorithm 1. The basic SOM algorithm explained.

Assume T as the size of input space X.
Set M as the map size.
At $t = 0$ initialize the weights at random.
for $1 \leq t \leq T$ **do**
  (i) Present an input $\boldsymbol{x}(t) \in X$
  (ii) Select the winner: $\nu(t) = \arg\min_i ||\boldsymbol{x}(t) - \boldsymbol{w}_i(t)||$
  (iii) Update the weights of the winner and its neighbors: $\Delta\boldsymbol{w}_i(t) = h(t)\eta(\nu, i, t)\left(\boldsymbol{x}(t) - \boldsymbol{w}_\nu\right)$.
end for

Here $||\cdot||$ denotes a distance (usually the Euclidean distance or, more generally, a function in the family of either Minkowsky or Riemann norms), while

$\eta(\nu, i, t) = \exp\left(-\frac{||r_\nu - r_i||^2}{2s^2}\right)$ is the neighborhood function among the leader node r_ν and all the grid neurons r_i $(i = 1, \ldots, M)$, s is the effective range of the neighborhood; finally $h(t)$ is the so called learning rate, that is a scalar–valued function, decreasing monotonically, and satisfying: $(i) 0 < h(t) < 1$; $(ii) \lim_{t \to 0} h(t) \to +\infty$;$(iii) \lim_{t \to \infty} h(t) \to 0$ [5,11].

Next paragraphs are devoted to provide insights on how to modify the standard SOM backbone, thus making it more suitable to deal with financial and economic data.

2.1 The Value at Risk SOM

The Value at Risk SOM (VaR–SOM) is a SOM based on the key concept of Value at Risk.

In a quite informal way, assuming the level of confidence α, $VaR_{1-\alpha}$ is a smallest value such that probability that loss exceeds or equals to this value is bigger than or equals to α:

$$VaR_{1-\alpha} = -x_\alpha \tag{1}$$

where x_α is the left–tail α percentile of a normal distribution: $N(\mu, \sigma^2)$ with mean μ and variance σ^2; x_α is described in the expression: $P[R < x_\alpha] = \alpha$, where R is the expected return. Using a standard normal distribution enables to replace x_α by z_α through the following permutation: $z_\alpha = (x_\alpha - \mu)/\sigma$, which yields: $x_\alpha = \mu + z_\alpha \cdot \sigma$, being z_α the left-tail α percentile of a standard normal distribution. Consequently, it is possible to re-write (1) as:

$$VaR_{1-\alpha} = -(\mu + z_\alpha \cdot \sigma) \tag{2}$$

In order for VaR to be meaningful, the confidence level is generally set equal to 95% or 99%: the higher the confidence level, the higher the VaR, as it travels downwards along the tail of the distribution (further left on the x-axis).

Incorporating VaR into the SOM algorithm means to pair any input to a node in the map having similar behavior in the left hand side of the sampled distribution; in the case of financial data this means to match patterns sharing similar losses profile.

In Box 2 some pseudo–code is provided, explaining how the Kohonen's algorithm is modified to take the VaR information into account. Note that $VaR_{1-\alpha}$ $(\boldsymbol{z}(t))$ and $Var_{1-\alpha}(\mathbf{w_i}(\mathbf{t}))$, $(i = 1, \ldots, M)$ indicate the Value at Risk associated to the normalized input $\boldsymbol{z}(t)$ and the normalized map nodes, respectively at the level $(1 - \alpha)$%. As VaR is generally negative (it represents a loss!), here we considered its absolute value.

2.2 The Linear Granger Causality SOM

The Linear Granger Causality [3] (LGC) is a statistical measure of causality based on forecast power. Given two stationary time–series A and B, (for sim-

Algorithm 2. The VaR–SOM algorithm.

Assume T as the size of input space X.
Set M as the map size.
Set the confidence level α.
At $t = 0$ initialize the weights at random.
for $1 \leq t \leq T$ **do**
 (i) Convert the input $\boldsymbol{x}(t) \in X$ into $\boldsymbol{z}(t)$
 (ii) Select the winner: $\nu(t) = \arg\min_i \left| |VaR_{1-\alpha}(\boldsymbol{z}(t))| - |VaR_{1-\alpha}(\boldsymbol{w}_i(t))| \right|$
 (iii) Update the weights of the winner and its neighbors: $\Delta\boldsymbol{w}_i(t) = h(t)\eta(\nu, i, t)\left(\boldsymbol{z}(t) - \boldsymbol{w}_\nu\right)$.
end for

plicity assume that they have zero mean), we can represent their linear interrelationships with the following model:

$$\begin{aligned} A_t &= \sum_{j=1}^{p} a_j A_{t-j} + \sum_{j=1}^{q} b_j B_{t-j} + \varepsilon_t \\ B_t &= \sum_{j=1}^{p} c_j B_{t-j} + \sum_{j=1}^{q} d_j A_{t-j} + \theta_t \end{aligned} \tag{3}$$

where ε_t and θ_t are two uncorrelated white noise processes; p, q are the maximum lags considered; a_j, b_j, c_j and d_j are the real valued model coefficients. The definition of causality implies that B causes A when b_j is different from zero; likewise, A causes B when d_j is different from zero. The causality is based on the F–test [7] of the null hypothesis that coefficients b_j or d_j are equal to zero according to the direction of the Granger causality.

Incorporating Granger causality into the SOM algorithm, means testing whether some causality is present or not among the input patterns and the map nodes. Under the F–test the following situations can therefore occur: (i) the causality is not significant; (ii) the causality is significant towards one direction (either from X to M or from M to X); (iii) the causality is significant towards both direction. Clearly, the most desiderable situations are either (ii), in the direction from M to X, or (iii). In both cases, in fact, the nodes behavior should increase the prediction (and hence the knowledge) of input patterns.

Box 3 shows how to build a SOM incorporating such information.

Note that selecting the winner implies now to choose either the node for which the causality is highest, if the F–score is significant, or the neuron whose non–causality is lowest, if the null hypothesis cannot be rejected. Clearly the results are conditioned by the choice of the lags amplitude p and q. A way to stem this issue is to run a bunch of LGC–SOM varying the couple $\{p, q\}$, and hence selecting the map that assures the best performance under the Akaike Information Criterion [2].

Algorithm 3. The LGC–SOM algorithm.

Assume T as the size of input space X.
Set M as the map size.
Set p, q as the lag amplitudes.
At $t = 0$ initialize the weights at random with zero mean.
for $1 \leq t \leq T$ **do**
  Present an input $\boldsymbol{x}(t) \in X$
  for $1 \leq i \leq M$ **do**
    (i) $\boldsymbol{x}(t) \rightarrow A_t$
    (ii) $\boldsymbol{w}_i(t) \rightarrow B_t$
    (iii) Apply (3)
    (iv) Run the F–test $\rightarrow F_i$
  end for
  Select the winner: $\nu(t)$ as the node having the best F test score F_ν.
  Update the weights of the winner and its neighbors: $\Delta \boldsymbol{w}_i(t) = h(t)\eta(\nu, i, t)\left(\boldsymbol{x}(t) - \boldsymbol{w}_\nu\right)$.
end for

3 Case study

The FinSOM class has been tested on a data sample made up by daily quotations of 207 German companies, in the period: October 2012-December 2013. for an overall number of 301 observations for each stock. The resulting 207×301 input matrix of price levels $\ell(t)$ at time t has been then turned in the correspondent 207×300 matrix of log-returns X, where the log-return at time t for the i–th stock is given by:

$$lr_i(t) = \log \ell(t) - \log \ell(t-1),\ t = 2, \ldots 301 \tag{4}$$

We then launched both SOM and FinSOM procedures: in order to choose optimal map dimensions we run extensive simulations, and motivated by the robustness of the results, we are now going to discuss the results obtained by way of equally sized maps composed by 96 neurons, arranged into a 8×12 grid. VaR computations have been made at both 95% and 99% levels of confidence, while LGC–SOM assumed: $p = 5 = q$, as to say, we assumed that causality can affect log–returns on a five days basis. This *magic* number corresponds to the value of the fractal dimension estimated on data by way of the False Nearest Neighbor (FNN) method [4]. The final maps are shown in Figure 1.

Clearly there is no room enough for a deeper investigation of the results, however, from a visual perspective, it is possible to observe that the number of maps clusters is quite different: six in the case of SOM, nine for VaR–SOM with $\alpha = 99\%$, and five for both VaR–SOM with $\alpha = 95\%$ and LGC–SOM.

Furthermore, it should be noted that in the case of standard SOM the nodes coloring has been made uniquely by referring to the unified distance matrix (UMatrix), while VaR–SOM nodes were colored by considering also the losses profile associated to each neuron. Finally, color shades in the LGC–SOM take also the causality significance of each node into account.

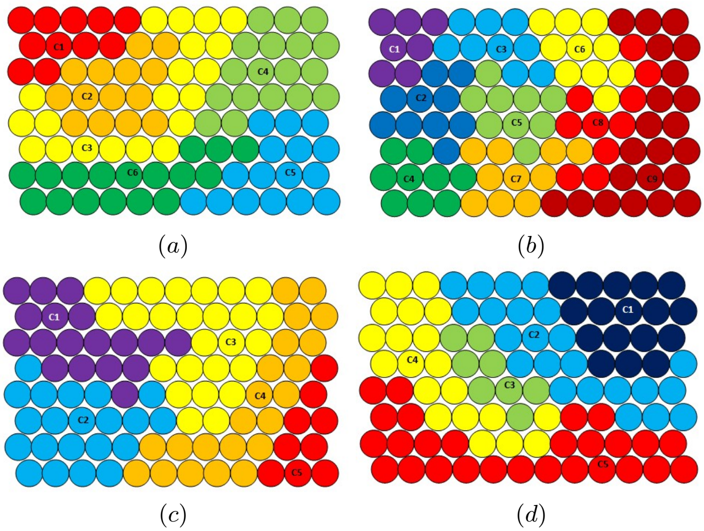

Fig. 1. From left to right and from top to bottom: SOM (a), VaR–SOM with $\alpha = 99\%$ (b), VaR–SOM with $\alpha = 95\%$ (c), and LGC–SOM (d), trained on data in the range October 2012–December 2013

The groups highlighted in the standard SOM (Fig.1a), are strongly connoted by sectors: C1 mainly embodies Banking and Finance stocks, C2 groups Heavy Industry companies, C3 contains High–Tech firms, while Health–Care and Energy Commodities are equally shared between C4 and C5.

In the case of VaR–SOM, nodes (and hence firms) with similar probability of losses exposition are highlighted. In particular, when $\alpha = 99\%$ (Fig.1b), lowest VaR is associated to clusters C1 to C4, where we can find stocks of companies that operating mainly at international rather than at national level have had greater opportunities to hedge from local crisis effect. On the other hand, the clusters C7 to C9 gather stocks/companies more sensitive to possible defaults and characterized by highest volatility levels, as well as by higher VaR values. Clusters C5 and C6, are of dubious interpretation, and refers to borderline situations with respect to those highlighted for both the groups C1 to C4, and C7 to C9. For what is concerning the VaR–SOM with $\alpha = 95\%$ (Fig.1c), the lower number of clusters probably depends on the variation in the confidence level. Now clusters C1 and C2 enclose German firms mainly projected at the international level and with lower VaR values, cluster C3 and C4 contain German stocks with middle–high levels of exposure, while C5 groups stocks with highest loss probability.

Moving to LGC–SOM, the visual inspection of the map needs to be coupled to the analysis of the regression coefficients that the procedure associates to every pair (node, stock), and *a fortiori* to the values of the related F–statistics. In this case, the analysis reveals that in three clusters of five (C1, C2 and C3) the nodes exhibit Linear Granger Causality (LGC) towards the input space;

in the remaining clusters the LGC assumption has been either weakly (C4) or hardly (C5) rejected. From an economical perspective, the former information is quite important, because it can be read in a forecasting key, that is the patterns associated to the map nodes can be helpful to forecast the behaviour of related stocks in the input space. However, it is difficult to get an interpretation key for the results corresponding to nodes in clusters C4 and C5.

4 Conclusion

In this work we discussed an enhancement of Self Organizing Maps (SOM), thought to improve the capability of the original algorithm to exploit meaningful patterns from financial data. The main issue, in fact, is that similarity measures do not take enough into account the intrinsic complexity of this kind of data.

According to this rationale, we introduced Financial SOM (FinSOM) as a family of SOM whose members modify the original SOM algorithm by evaluating the similarity (and hence the proximity) among inputs and neurons by way of functions with more economic soundness. In particular, we introduced a risk–oriented SOM based on Value at Risk (VaR–SOM), and a SOM based on Linear Granger Causality (LGC–SOM).

Financial SOM have been compared to standard SOM in an application on stocks data from the German market, observed in the period: October 2012-December 2013. By comparison to the classical SOM, FinSOM seem to provide more meaningful economic taxonomies. This seems particularly true in the case of VaR–SOM, that offer a quite intuitive intepretation of the generated clusters.

Clearly the results are sensitive to the parameters in use, and this can be a not negligible issue. Despite our general impression that FinSOM can be effective tools to inspect financial data, and to exploit significant patterns, future research must be oriented in search of further improvements of the technique. Towards this direction we think there is great room for improvements, due to the wide basin from which functions to replace standard metrics can be drawn.

References

1. Aggarwal, C.C., Hinneburg, A., Keim, D.A.: On the surprising behavior of distance metrics in high dimensional spaces. In: Van den Bussche, J., Vianu, V. (eds.) ICDT 2001. LNCS, vol. 1973, pp. 420–434. Springer, Heidelberg (2000)
2. Akaike, H.: A new look at the statistical model identification. IEEE Transactions on Automatic Control 19(6), 716–723 (1974)
3. Granger, C.W.J.: Investigating Causal Relations by Econometric Models and Cross-spectral Methods. Econometrica 37(3), 424–438 (1969)
4. Kennel, M.B., Brown, R., Abarbanel, H.D.I.: Determining embedding dimension for phase-space reconstruction using a geometrical construction. Phys. Rev. A 45, 3403 (1992)
5. Kohonen, T.: Self-Organizing Maps. Springer, Heidelberg (2002)
6. Lamping, J., Rao, R.: Laying out and visualizing large trees using a hyperbolic space. In: ACM Symposium on User Interface Software and Technology, pp. 13–14 (1994)

7. Lomax, R.G., Hahs-Vaughn, D.L.: Statistical Concepts: A Second Course. Routledge/Taylor & Francis (2007)
8. Resta, M.: Early Warning Systems: An Approach via Self Organizing Maps with Applications to Emergent Markets. In: Apolloni, B., Bassis, S., Marinaro, M. (eds.) Proceedings of the 2009 Conference on New Directions in Neural Networks: 18th Italian Workshop on Neural Networks, WIRN 2008, pp. 176–184. IOS Press, Amsterdam (2009)
9. Resta, M.: The Shape of Crisis. Lessons from Self Organizing Maps. In: Kahraman, C. (ed.) Computational Intelligence Systems in Industrial Engineering. Springer Atlantis series in Computational Intelligence Systems, vol. 6, pp. 535–555 (2012)
10. Resta, M.: On a Data Mining Framework for the Identification of Frequent Pattern Trends. In: Perna, C., Sibillo, M. (eds.) To appear in Mathematical and Statistical Methods for Actuarial Sciences and Finance (2014), doi:10.1007/978-3-319-05014-0_39
11. Ritter, H., Schulten, K.: Convergence properties of Kohonen's topology conserving maps: fluctuations, stability, and dimension selection. Biological Cybernetics 60, 59–71 (1988)
12. Ritter, H.: Self–organizing maps in non–euclidian spaces. In: Oja, E., Kaski, S. (eds.) Kohonen Maps, pp. 97–110. Elsevier (1999)
13. Sarlin, P., Eklund, T.: Fuzzy Clustering of the Self-Organizing Map: Some Applications on Financial Time Series. In: Laaksonen, J., Honkela, T. (eds.) WSOM 2011. LNCS, vol. 6731, pp. 40–50. Springer, Heidelberg (2011)
14. Sarlin, P.: Visual tracking of the millennium development goals with a fuzzifieed self–organizing neural network. International Journal of Machine Learning and Cybernetics 3, 233–245 (2012)