A Cluster Analysis of Stock Market Data Using Hierarchical SOMs

César A. Astudillo[1], Jorge Poblete[1], Marina Resta[2], and B. John Oommen[3](✉)

[1] Department of Computer Science, Universidad de Talca, Km 1. Camino a los Niches, Curicó, Chile
castudillo@utalca.cl

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

[3] School of Computer Science, Carleton University, Ottawa K1S 5B6, Canada
oommen@scs.carleton.ca

Abstract. The analysis of stock markets has become relevant mainly because of its financial implications. In this paper, we propose a novel methodology for performing a *structured* cluster analysis of stock market data. Our proposed method uses a tree-based neural network called the TTOSOM. The TTOSOM performs self-organization to construct tree-based clusters of vector data in the multi-dimensional space. The resultant tree possesses interesting mathematical properties such as a succinct representation of the original data distribution, and a preservation of the underlying *topology*. In order to demonstrate the capabilities of our method, we analyze 206 assets of the Italian stock market. We were able to establish topological relationships between various companies traded on the Italian stock market and visually inspect the resultant taxonomy. The results that we obtained, briefly reported here (but more elaborately in [10]), were amazingly accurate and reflected the real-life relationships between the stocks.

Keywords: TTOSOM · Stock market · Clustering · Hierarchical SOM · Tree-based SOM

1 Introduction

The focus of this paper is the cluster analysis of stock market data. There are several reasons motivating such an analysis. One of these reasons is that an *a priori* knowledge of the patterns that govern the movements of stocks provide a lucrative advantage to an intelligent investor. Also, the financial scenario

C.A. Astudillo—Assistant Professor. IEEE Member. The work of this author is partially supported by the FONDECYT grant 11121350, Chile.

B.J. Oommen—*Chancellor's Professor*; *Fellow: IEEE* and *Fellow: IAPR*. This author is also an *Adjunct Professor* with the University of Agder in Grimstad, Norway. The work of this author was partially supported by NSERC, the Natural Sciences and Engineering Research Council of Canada.

B.H. Kang and Q. Bai (Eds.): AI 2016, LNAI 9992, pp. 101–112, 2016.
DOI: 10.1007/978-3-319-50127-7_8

encountered in the stock market is not too different from those encountered in other application domains in which data-driven decisions complement the views of human experts. Indeed, more recently, data mining has demonstrated to be a powerful tool for the prediction and forecasting of financial patterns.

A particular problem that arises in finance is to find the relationships between different stocks. The literature records solutions that simplistically assume an independent relationship between them except that the prices are modeled using various time series. From an opposing perspective, one can assume that all the stocks are dependent on each other, implying the creation of a completely connected graph that, in most cases, can be very complex to analyze.

The problem can be solved by using an alternate paradigm. One may opt to consider a stock as being a collection of features, and thus attempt to discover common patterns. From a Bayesian perspective, one could interpret those stocks as being pattern whose features are independent of each other [6]. On the other hand, if the features are considered to be dependent, the *forms* of the corresponding covariance matrices can render the problem to be intractable.

Our proposal is to apply a Neural Network (NN) processing paradigm, and to thus create a tree-based model from the data. The goal is to capture the correlation between the different commodities (*and not the features*)in the stock market. Unlike previous methods reported in the literature, our plan is to take advantage of the properties of Self-Organizing Maps (SOMs) to capture the stochastic and topological structure of the data. However, we achieve this goal without the necessity of generating the traditional SOM-grid. Instead, the method directly learns a Tree-Based SOM. The algorithm that we use is called the Tree-based Topology-Oriented Self-Organizing Map (TTOSOM) [2], explained below.

The algorithms, literature survey and results presented here are, by necessity, brief. More details of each of these aspects are found in [10].

2 Literature Review

The SOM [8] is a neural network that is typically trained using (un)supervised learning, so as to produce a neural representation in a space whose dimension is usually smaller than that in which the training samples lie. Further, the neurons attempt to preserve the topological properties of the input space.

The SOM concentrates all the information contained in a set of n input samples belonging to the d-dimensional space, say $\mathcal{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}, \mathbf{x} \in \mathbb{R}^d$, utilizing a much smaller set of neurons, $\mathcal{C} = \{\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_m\}$, each of which is represented as a vector. Each of the m neurons contains a weight vector $\mathbf{w} = [w_1, w_2, \ldots, w_d]^t \in \mathbb{R}^d$ associated with it. These vectors are synonymously called "weights", "prototypes" or "codebook" vectors. The vector $\mathbf{w}_i$ may be perceived as the *position* of neuron $\mathbf{c}_i$ in the feature space. During the training phase, the values of these weights are adjusted simultaneously so as to represent the data's distribution and its structure. In each training step, an input $\mathbf{x}$ is presented to the network, and the neurons compete between themselves so as to identify which is the "winner", the Best Matching Unit (BMU), represented by:

$$s = s(\mathbf{x}) = \arg\min_{c \in \mathcal{C}} \parallel \mathbf{x} - \mathbf{w}_c \parallel, \tag{1}$$

where $\parallel . \parallel$ denotes the appropriate norm, which is here the Euclidean norm.

Once the BMU has been identified, it is migrated towards the input signal as per the so-called Update rule, given by the following equation:

$$\mathbf{w}(t+1) = \mathbf{w}_i(t) + h_{si}(t)[\mathbf{x}(t) - \mathbf{w}_i(t)], \tag{2}$$

where $h(t) = h(|c_i - c_j|; t)$ is a kernel function called the Bubble of Activity (BoA) and t is the time variable. $h(t)$ is a decreasing function that depends on the distance between two neurons.

In brief, the SOM algorithm can be summarized in the following steps:

1. Obtain a sample $\mathbf{x}$ from $\mathcal{X}$.
2. Find the Winner neuron, the BMU, i.e., the one which is most similar to $\mathbf{x}$.
3. Determine a subset of neurons close to the winner, the BoA.
4. Migrate the closest neuron and its neighbors in the BoA towards $\mathbf{x}$.
5. Modify the learning factor and radius as per the pre-defined schedule.

Although the SOM has demonstrated an ability to solve problems over a wide spectrum, it possesses some fundamental drawbacks. One of these drawbacks is that the user must specify the lattice *a priori*, which has the effect that he must run the NN a number of times to obtain a suitable configuration. Other handicaps involve the size of the maps, where a lesser number of neurons often represent the data inaccurately. The state-of-the-art approaches attempt to render the topology more flexible, so as to represent complicated data distributions in a better way and/or to make the process faster by, for instance, speeding up the task of determining the BMU. Other drawbacks concentrate on the quality of the resultant map [2] and the time necessary to achieve convergence [5]. Researchers have also focused on how to tackle the above mentioned problems, and as a result, different structured variations of the original SOM have been proposed[1]. A comprehensive comparison of selected variants can be found in [4].

The present paper applies the Tree-based Topology Oriented SOM (TTOSOM) [2] to analyze the stock market. The TTOSOM is a tree-structured SOM which aims to discover the underlying distribution of the input data set $\mathcal{X}$, while also attempting to perceive the topology of $\mathcal{X}$ as viewed through the user's desired perspective. It works with an imposed tree-structured topology, where the codebook vectors are adjusted using a VQ-like strategy. Besides, by defining a user-preferred neighborhood concept, as a result of the learning process, it also learns the topology and preserves the prescribed relationships between the

[1] A paper, written by two of these present authors, which reported the preliminary results of a *dynamic* Tree SOM, won the Best Paper Award in a well-known international AI conference [1].

neurons as per this neighborhood. Thus, the primary consideration is that the concept of neurons being "near each other" is not prescribed by the metric in the space, but rather by the structure of the imposed tree. The BoA of the TTO-SOM defines a distinct scheme, differing from previous methods by two main aspects: The use of a tree structure and a semi-supervised learning paradigm.

The first task is to declare the user-defined tree, which, as explained in [2], is done in a recursive manner. The TTOSOM considers the concept of neural distance which is given by the minimum *number* of unweighted connections that separate them in the user-defined tree. Unlike other methods (c.f., [9]), this notion of distance is not dependent on whether the nodes are leaves or not. As in the case of the traditional SOM, the TTOSOM requires the identification of the BMU, i.e. the closest neuron to a given input signal, and involves a distance in the feature space (and not in the tree space) as per Eq. (1).

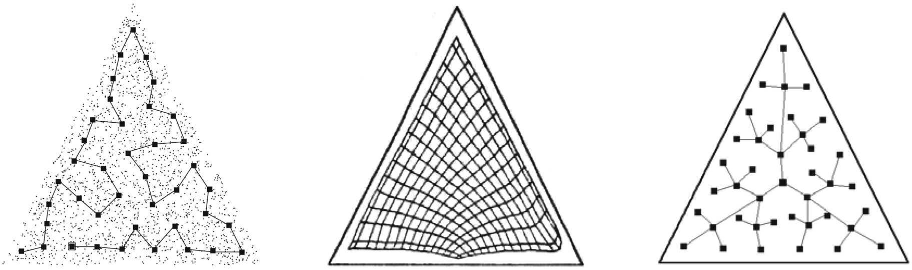
Intricately related to the notion of inter-node distance, is the concept of the BoA which is the subset of nodes "close" to the unit being currently examined. These nodes are essentially those which are to be moved toward the input signal presented to the network. This concept involves the consideration of a quantity, the so-called *radius*, which determines how big the BoA is. In particular non-leaf nodes can be part of the BoA. The question of whether or not a neuron should be part of the current bubble, depends on the number of connections that separate the nodes rather than the distance that separate the networks in the solution space (for instance, the Euclidean distance). This concept of neighborhood is distinct and different from the ones used in other approaches such as the ET [9] or the SOTM [7].

The training process of the TTOSOM involves positioning the neurons which describe the user-defined tree in the feature space so as to capture the distribution and topology of the data points. This process involves a loop of training steps which terminates when the convergence is acceptable to the user. The *Training step* involves requesting an input sample from the dataset, locating the BMU, computing the nodes within the *current* BoA, and migrating those neurons toward the input signal using a SOM-like philosophy.

3 Proposed Methodology

A fundamental question in our study is whether we should model the Stock Market's (SM's) topology using the SOM and then invoke a Maximum Spanning Tree (MST)-based post-processing phase, or whether we should rather utilize a neural strategy that inherently captures the structure of the underlying system. We propose that the neural scheme should be the TTOSOM over the SOM.

The first argument to support our assumption is the fact that unsupervised learning usually demands the deduction of the structure of the data. The tree topology used by the TTOSOM will "absorb" and display the properties of the input set. We shall now show how the representation obtained by the TTO-SOM is superior to both the ones obtained by using the SOM with a linear ordering of the neurons, and the one obtained by invoking a grid representation.



(a) (b) (c)

Fig. 1. How a triangle-shaped distribution is learned through unsupervised learning. (a) The positions of the SOM's neurons learned by a linear ordering, (b) The positions of the SOM's neurons learned by a grid-based ordering, and (c) The tree learned by the TTOSOM.

To demonstrate this, consider, for example, the following: A user may want to devise an algorithm that is capable of learning a triangle-shaped distribution as the one depicted in Fig. 1. In this case, we have displayed three different images of the same, explained below.

In Fig. 1a, we have not shown the boundary of the triangle, but rather the actual set of points from which the data points are sampled. The neurons in this case are arranged *linearly*, but not superimposed on any grid-like structure. Initially, the codebook vectors were randomly placed within the triangle and the so linear topology was completely lost due to the randomness of the data points. At the end of the training process, the list represents the triangle very effectively, as also reported in [8]. The SOM is capable of not only representing the whole distribution of the input samples, but also of preserving *this* linear topology. Indeed, on termination, the indices of the codebook vectors are arranged in an increasing order as seen in Fig. 1a. In this case, the one-dimensional list of neurons is evenly distributed over the triangle preserving the original properties of the 2-dimensional object, yielding a shape similar to the so-called Peano curve.

In the second case, in Fig. 1b, the SOM operates on an grid, and attempts to fit the grid within the overall shape (duplicated from [8]). In this case, we have shown only the boundary of the triangle. From our perspective, a grid does not naturally fit the triangular-shaped distribution, experiencing a deformation of the original lattice during training.

As opposed to the above two SOM solutions, Fig. 1c, shows the result of applying the TTOSOM [2] to the same dataset[2]. As one can observe from Fig. 1c, a 3-ary tree seems to be a far more superior choice for representing the particular shape in question. On closer inspection, Fig. 1c depicts how the complete tree fills in the triangle formed by the set of stimuli, and further, seems to do it *uniformly*. The final position of the nodes of the tree suggests that the underlying structure of the data distribution corresponds to the triangle. Additionally, the root of the

[2] Other examples of applying the TTOSOM are found in [2].

tree is placed roughly in the center of mass of the triangle. Of course, the triangle of Fig. 1c serves only as a very simple *prima facie* example to demonstrate to the reader, in an informal manner, how both techniques will try to learn the set of stimuli. We believe that the TTOSOM's topology-learning phenomenon (over the SOM's) makes it a superior choice in modeling the SM.

Recent studies have also shown other desirable properties of the TTOSOM. First of all, it possesses a holograph-like property. Figure 1c shows that each of the three main branches of the tree, cover the areas directed towards a vertex of the triangle respectively, and their sub-branches fill in the surrounding space around them in a recursive manner, which we identify as being a holograph-like behavior. This phenomenon is further investigated in [2], where more examples are available. The representation with a fewer number of neurons mimics the representation less accurately (with lesser resolution) than with a larger number.

The consequence of this to modeling the SM is immediate. It is, of course, obvious that we cannot model the SM using all the commodities and traded stocks. However, if we model it using a properly-sampled subset of the stocks and with fewer neurons, the model will be a lower-resolution model of the true model that would have been obtained by including all the stocks and if a larger NN was invoked. Such a conclusion cannot be drawn from a SOM-related strategy.

The TTOSOM has been shown to be useful for performing classification even in environments where the data contains missing values [3]. This, again, is crucial in modeling the SM because of the randomness of the task and the fact that all the transactions involving every single stock are not readily available.

Finally, the literature accepts the fact that the most time consuming task inherited from the SOM is the process of searching for the BMU. However, in the case of the TTOSOM, this task can be alleviated by incorporating tree-based data structures [5]. Such a scheme could be logarithmic, as opposed to linear, in the number of data points. This aspect could be crucial if the aim is to process even more complex economic data, such as intra-day time series involving numerous stocks, that can be in the order of Terabytes.

By virtue of all these arguments, we believe that the ideal neural strategy to model the SM is the TTOSOM.

4 Validation of the Model

To confirm the hypothesis, i.e., that the model obtained by the TTOSOM is truly a viable model, we performed experiments by examining data from the Italian market (the details of the companies, their prices etc. are found in [10]). This involved companies incorporated into the Mibtel index for the period between February 2013 and March 2014. This dataset included a total of 287 days and 206 stocks, which were classified into 37 productive sectors.

For the implementation of the proposed analysis, we selected all the available price levels available for the aforementioned period. For the generic asset i ($i = 1, 2, \ldots, N$, where $N = 206$ is the total number of companies available for the Mibtel index), we built the corresponding time-series of log-returns $S^{(i)} = \left\{pv_k^{(i)}\right\}$ with length $T-1$ (where $T = 287$ is the total number of days), being:

$$pv_k^{(i)} = \log \frac{pl_k^{(i)}}{pl_{k-1}^{(i)}},\ k = 2, ..., T, \tag{3}$$

where $pl_k^{(i)}$ is the price level at time k for the stock i.

By invoking such a transformation, we observed that it was possible to avoid trend effects on the data. The transformed data was then processed using the TTOSOM. In particular, our procedure consisted of three steps:

1. We defined a set of topologies for the TTOSOM tree. In this way it is possible to consider tree structures possessing different features.
2. We established a mechanism for optimizing the parameters, including the total number of iterations, the learning rate, the tree topology and the number of neurons. This stage allowed us to select the set of values for the parameters that produced the best results in the search space.
3. We utilized the best set of parameters to obtain a set of clusters for companies in the Italian SM.

The experiments considered all the combinations among the following values for the parameters: The total number of iterations which was either $50,000$ and $100,000$; The radius for the BoA equal to the maximum level of the depth in the tree (h) and also twice this level $(2h)$; The learning rate of 0.5 or 0.9; Five different tree configurations that differed in the number of nodes, depth and branching factor, were also considered.

The validation of the experiments on our dataset was achieved using the TTOSOM and by measuring a modified version of the quantization error that considered the tree structure. Additionally, our method provided ways by which we could control the total number of clusters to be considered by the prior establishment of tree topologies.

For each configuration, we averaged 10 replications for each combination. Subsequently, we performed a comparison of the means, the aim of which was to detect significant differences between the results obtained. As a consequence of these tasks, we were able to obtain a hierarchical structure that relates the market sectors, which is equivalent to the skeleton framework presented in [11].

5 Experimental Results

In our experiments, each data point corresponded to the stocks of a company quoted in the Italian market. Also, in our model, each neuron of the tree corresponded to a single cluster. The best results were obtained by using a complete tree with 3 children per node and 3 levels of depth (resulting in a total of 13 nodes). This implied that we considered partitioning the stocks into 13 groups. Each of the resulting clusters described a representative time series which is

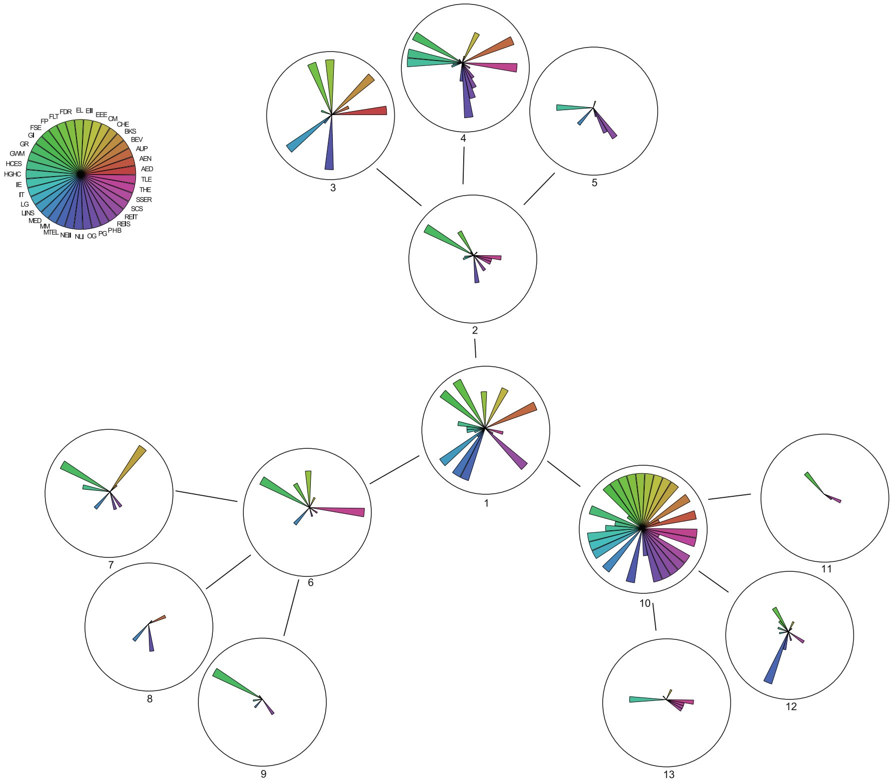


Fig. 2. The organization of the Italian market, as a result of using a TTOSOM trained with 50,000 iterations, a learning factor of 0.5 and an initial radius that equals 6.

equivalent to an average trend. Thus, essentially, we associated each stock to the BMU whose average trend was the closest. According to our observations, the best results were obtained when the TTOSOM was trained using an initial radius equal to twice the depth of the tree. We performed a total number of $50,000$ iterations with an initial learning rate of 0.5. In a subsequent phase, we assigned each data point to its closest BMU.

Based on our algorithm[3], the 206 stocks were distributed in 13 clusters. Figure 2 shows a pictorial representation of the clusters. The pie chart within each cluster represents the component sectors by different coloured bars, and the cardinality of the sector by means of the length of the bars themselves. The number below each of the clusters, labels them in the correct order: The root of the tree corresponds to the Cluster identified by the index 1 and lies in the center of the figure; three children originated from it (namely: Clusters labelled

[3] The results presented here are brief in the interest of space. Additional results can be found in [10], and this paper can be sent to the Referees is required.

by indices 2, 6 and 10), and are symmetrically distributed in the figure. One should observe the holographic properties of the TTOSOM that emerged.

5.1 Discussion

Consider Fig. 2 which details the composition of the clusters. To demonstrate the salient aspects of our scheme, it would be helpful to record the following:

1. At a first glance, we observe that both the root of the tree and Cluster 10 (located at the second hierarchical level) are those with the largest numbers of companies and diversity of sectors.
2. Cluster 3, abbreviated as CL03, contains 12 assets belonging to the Banking Sector (BKS).
3. Cluster 1 (CL01) is strongly dominated by the Construction and Materials (CM), Automobile and Parts (AUP), and General Industries (GI).
4. CL02 has a homogeneous distribution in the Industrial Areas (IE and IT), the area of Technological Services (SCS and SSER), as well as in the Retail (GR) area.
5. CL03 and CL07 are composed mainly by the Bank Sector (BKS).
6. In CL05 and CL09, the prominent areas are Industrial Engineering (IE) and Household Goods and Home Construction (HGHC).
7. In CL04, we identify companies and services associated to the home, such as Personal Goods (PG), Household Goods and Home Construction (HGHC), Automobile and Parts (AUP) and health (HCES).
8. The clusters CL06 and CL09 constitute the more heterogeneous groups, including companies from the Electricity sector (EL), Media (MED), Financial Services (FSE) and Retail (GR).
9. CL08 relates to Media (MED), Oil (OG), Vehicles (AUP) and Banking (BKS).
10. We also record that CL10, CL11 and CL12 have an important percentage of companies related to Financial Services (FSE) and Technological Support (SSER).
11. Finally cluster CL13 groups a variety of companies, including Software (SCS), Hardware (THE) and Construction (CM and HGHC).
12. The analysis of the composition of the clusters reveals a clear dominance of the banking sector in CL03. We also identify a strong presence of the financial services in CL10, CL11 and CL12, while personal products are mostly grouped in CL04. Additionally, the construction area was present in CL01, the electricity and media in CL06 and media in CL08 and CL09.

The fact that the respective clusters collect stocks of similar types – without any prior knowledge of its constituent commodity type – is, in our opinion, truly amazing.

Additional insight can be gleaned by combining the analysis of the composition of the clusters to main statistics. The statistical analysis is detailed in Table 1, where we provided, for each cluster, the corresponding mean (mu),

Table 1. Statistics for the clusters of the Italian Stock Market between February 2013 and March 2014.

Cluster	mu	sd	sk	SR
1	−0,0007938	0,0039923	0,3311401	−19,883
2	−0,0006634	0,0050274	−0,4954843	−13,195
3	−0,0004778	0,0070769	0,4920971	−6,752
4	−0,0004604	0,0040726	0,5135600	−11,305
5	−0,0002265	0,0083026	−1,2349395	−2,728
6	−0,0007216	0,0052765	−0,1541048	−13,675
7	−0,0007557	0,0063180	−2,3810895	−11,961
8	−0,0001645	0,0083680	−0,2360082	−1,965
9	−0,0004073	0,0058753	−0,7328621	−6,932
10	−0,0004129	0,0019768	0,1761923	−20,885
11	−0,0008235	0,0045209	0,0844924	−18,216
12	−0,0002758	0,0033300	−0,3934497	−8,282
13	−0,0004909	0,0062031	−1,2187910	−7,914

standard deviation (sd), and skewness (sk). In addition, we included the Sharpe Ratio, SR, [12], which is a general indicator of the profitability for the group of companies included in each cluster: the higher the value of SR[4] the more profitable is the cluster.

As a results of such an analysis, we can report the following: All the groups exhibited negative average returns, and the variability was very low. *By combining these indications, we can obtain a simple snapshot of the Italian market in 2013, namely that it represented an economy that was deep into a stagnation phase.* The most alerting signals are those concerning the cluster with worst returns, i.e. mainly CL10 and CL11, predominantly containing companies dealing with financial services. Similar remarks can be made about companies in the construction area too, primarily grouped into CL01. It is pertinent to mention that these observations fit well with the actual situation that the financial services, the buildings and the industrial sectors in Italy were encountering – *all were going through very hard times.* The fact that we can accurately infer such economical facts from the TTOSOM is quite astonishing.

Finally, despite being negative, CL08 and CL09 characterized by the presence of companies in the media sector, were among the most attractive groups in the Italian market. It is worth mentioning that the banks (with their high presence in CL03) still maintained their (relative) attractiveness.

We hope that the reader appreciates the power of utilizing the TTOSOM in the clustering and analysis of such complex SM-based data.

[4] In this table, the SR values have been multiplied by 100 to make the results more readable.

6 Conclusions

In this paper, we have proposed the use of a self-organizing Neural Network (NN), based on a hierarchical structure, for analyzing the behavior of the Stock Market (SM). We successfully applied this NN for performing a static analysis of the Italian market for the period between February 2013 and March 2014, which was a critical economic period.

The main novelty of the paper was in suggesting the TTOSOM, a variant of the SOM that uses a tree-based topology, for the neurons. This NN was capable of defining clusters that are hierarchically connected. The main advantage of this tree-based SOM is that it produces a tree-based structure for the market topology. In doing this, it incorporates the tree structure into the self-organizing process, without any further processing, as one would have done if we would have worked with a standard SOM network.

With the proposed methodology we were able to visualize the relationships between the various companies traded on the Italian Stock Market. The automatic clustering of the companies into their relevant groups and sub-groups, achieved by identifying common winning neurons, was truly amazing. More specifically, the features of the clusters can lead the analyst to conclude that during the period under examination, the Italian market was moving towards a stagnation phase, with negative average returns and low dispersion. Without any external economic feedback, the method also highlighted that the clusters with higher concentration of companies from the sectors of both construction and financial services were the least attractive, while the media sector happened to be less negative.

To conclude, the technique that we have discussed is extremely promising, because it allows the analyst to infer the topology of the SM as the output of the learning procedure. The avenues for future research is vast. We recommend that future efforts be devoted to deepen the potential of tree-based SOMs to analyze markets topology, and those devoted to more risk-related stocks and derivatives. One could also apply our techniques to markets of different countries, and to the quotations of stocks sampled at higher time frequencies, i.e., moving from daily data to intra-day data.

References

1. Astudillo, C.A., Oommen, B.J. On using adaptive binary search trees to enhance self organizing maps. In: Nicholson, A., Li, X. (eds.) 22nd Australasian Joint Conference on Artificial Intelligence (AI 2009), pp. 199–209 (2009)
2. Astudillo, C.A., Oommen, B.J.: Imposing tree-based topologies onto self organizing maps. Inform. Sci. **181**(18), 3798–3815 (2011)
3. Astudillo, C.A., Oommen, B.J.: On achieving semi-supervised pattern recognition by utilizing tree-based SOMs. Pattern Recogn. **46**(1), 293–304 (2013)
4. Astudillo, C.A., Oommen, B.J.: Topology-oriented self-organizing maps: a survey. Pattern Anal. Appl. (2014). http://dx.doi.org/10.1007/s10044-014-0367-9

5. Astudillo, C.A., Oommen, B.J.: Fast BMU search in SOMs using random hyperplane trees. In: Pham, D.-N., Park, S.-B. (eds.) PRICAI 2014. LNCS (LNAI), vol. 8862, pp. 39–51. Springer, Heidelberg (2014). doi:10.1007/978-3-319-13560-1_4
6. Gámez, J.A., Rumí, R., Salmerón, A.: Unsupervised Naive Bayes for data clustering with mixtures of truncated exponentials. In: Studený, M., Vomlel, J. (eds.) Third European Workshop on Probabilistic Graphical Models, 12-15 September 2006, Prague, Czech Republic, Electronic Proceedings, pp. 123–130 (2006)
7. Guan, L.: Self-organizing trees and forests: a powerful tool in pattern clustering and recognition. In: Campilho, A., Kamel, M.S. (eds.) ICIAR 2006. LNCS, vol. 4141, pp. 1–14. Springer, Heidelberg (2006). doi:10.1007/11867586_1
8. Kohonen, T.: Self-Organizing Maps. Springer-Verlag New York Inc., Secaucus (1995)
9. Pakkanen, J., Iivarinen, J., Oja, E.: The evolving tree – a novel self-organizing network for data analysis. Neural Process. Lett. **20**(3), 199–211 (2004)
10. Poblete, J.A., Resta, M., Oommen, B.J., Astudillo, C.A.: On representing the stock market topologically and inferring its economic indicators using tree-based SOMs (2016, to be submitted)
11. Resta, M.: The shape of crisis lessons from self organizing maps. In: Kahraman, C. (ed.) Computational Intelligence Systems in Industrial Engineering. Atlantis Computational Intelligence Systems, vol. 6, pp. 535–555. Atlantis Press, Amsterdam (2012)
12. Sharpe, W.F.: The sharpe ratio. J. Portfolio Manage. **21**(1), 4958 (1994)