Multi-Dimensional Pattern Discovery in Financial Time Series using SAX-GA with Extended Robustness

António Canelas Rui Neves Nuno Horta

Instituto de Telecomunicações
Instituto Supertior Técnico – Torre Norte – Piso 10
Av. Rovisco Pais, 1
1049-001 Lisboa – Porugal
Phone : +351 21 841 84 54

antonio.canelas@lx.it.pt rui.neves@ist.utl.pt nuno.horta@lx.it.pt

ABSTRACT

This paper proposes a new Multi-Dimensional SAX-GA approach to pattern discovery using genetic algorithms (GA). The approach is capable of discovering patterns in multi-dimensional financial time series. First, the several dimensions of data are converted to a Symbolic Aggregate approXimation (SAX) representation, which is, then, feed to a GA optimization kernel. The GA searches for profitable patterns occurring simultaneously in the multi-dimensional time series. Based on the patterns found, the GA produces more robust investment strategies, since the simultaneity of patterns on different dimensions of the data, reinforces the strength of the trading decisions implemented. The proposed approach was tested using stocks from S&P500 index, and is compared to previous reference works of SAX-GA and to the Buy & Hold (B&H) classic investment strategy.

Categories and Subject Descriptors

I.5.M [Pattern Recognition]: Miscellaneous

General Terms

Algorithms, Performance, Economics, Experimentation.

Keywords

Pattern discovery, financial market, time series, genetic algorithm, SAX representation, multi-dimensional time series analysis.

1. INTRODUCTION

Unsupervised pattern discovery has a major role in forecasting the evolution of several events that develop over time. The basis of these patterns lies in the data, usually long time series, which need to be collected and analyzed. In the financial world one of the first attempts used to forecast the market behavior was watching the financial graphical chart, and trying to identify some known graphic formations that, for most of the time, had the same outcome. Analysts employing this approach are referred to as *Chartists*. Unfortunately, this method only worked with trivial chart patterns formations. In cases where the data was extended over several dimensions, like in [4] where a 100-dimensional set of weather data was analyzed, this becomes a hard and very complex task. This huge and complex task, makes the soft computing and in general the computational intelligence domains especially appropriate for addressing the problem.

GECCO'13 Companion, July 6–10, 2013, Amsterdam, The Netherlands.
ACM 978-1-4503-1964-5/13/07.

The pattern discovery method presented in this paper, will be applied and tested with financial data. The Multi-Dimensional SAX-GA combines the Symbolic Aggregate approXimation (SAX) technique, which will be used to represent several data dimensions, together with a Genetic Algorithm (GA) optimization kernel, that is responsible for the discovery of meaningful patterns across several dimensions of data. The objective of the GA kernel is to identify patterns that are capable to creating profitable trading rules. These are then tested using real data from S&P500, where all transaction costs are considered in order to simulate a real market investment system.

2. RELATED WORK

Many of the approaches to multi-dimensional time series analysis are based on dimensional reduction methods, followed by the application of a one-dimensional pattern discovery method. *Tanaka et al.* [5] uses Principal Component Analysis (PCA) to reduce multi-dimensional data to one-dimensional data, to which then a Minimum Description Length (MDL) algorithm is applied to identify patterns. One of the main drawbacks of these type of approaches, where different dimension are compressed to just only one, is that some important information and signals are loss.

In *McGovern et al.* [4] a search for patterns in multi-dimensional data, without dimensional reduction, is made. McGovern's work, which also uses SAX to represent the data, the implemented search for patterns is a two step process, first the SAX sequences populates a *trie* [3], a tree-like structure, where the patterns are classified. Using the classification it is possible to extract the most relevant patterns for each dimension. The final step is an exhaustive search of temporal ordering of the most important patterns discovered for detecting occurrences among dimensions. The SAX form of representation is very adequate to combine with the GA [1,2], since every symbol can simply be fitted to a gene. Making the adopted combination of SAX-GA a natural choice for the presented work.

3. MULTI-DIMENSIONAL SAX-GA

The method described in this paper is based in the SAX-GA approach described in detail in [1,2] which combine the SAX representation method with a GA kernel. In the previous reference works, the SAX-GA was feed with the close stock price time series, in order to discover meaningful patterns, in this work four more time series are analyzed. Usually financial time series are composed of a 5-dimensional data, *Open*, *Maximum*, *Minimum*, *Close* prices and *Volume*. In this work, these dimensions are converted to SAX and then the GA kernel will discover a combination set of patterns among several dimensions and the respective trading rules. The *Volume* time series will be used with the important role of confirming any patterns discovered in the other four dimensions of the financial data, since usually a large increase / decrease is associated with possible inversions on market trend. In order to reduce the risk that could result from long time periods in the market, some safety measures are implemented to make the algorithm close the position and exit the market. One of these safety measures is the use of the maximum drawdown (MDD) in order to close the trade if some pre-defined value is reached.

Multi-dimensional SAX-GA will only consider possible trading situations when on more than half of the dimensions a pattern is detected, like a voting system, and being mandatory the presence of a pattern in the *Volume* time series. The presence of a pattern in each of the dimension is signal by the zero distance between the pattern generated by the GA and the discretized SAX time series. In the case that three or more patterns are discovered the algorithm will enter the market, assuming a strategy accordingly to the type of position defined by the chromosome, *Long* or *Short*. To leave the market the algorithm will search for exit patterns, in this case any pattern on any of the dimensions, except *Volume*, is sufficient to make the algorithm leave the market, this will increase the exit response time. Combining the SAX parameters with the trading strategy described above, the chromosome structure that composes the population in the GA is presented in Figure 1.

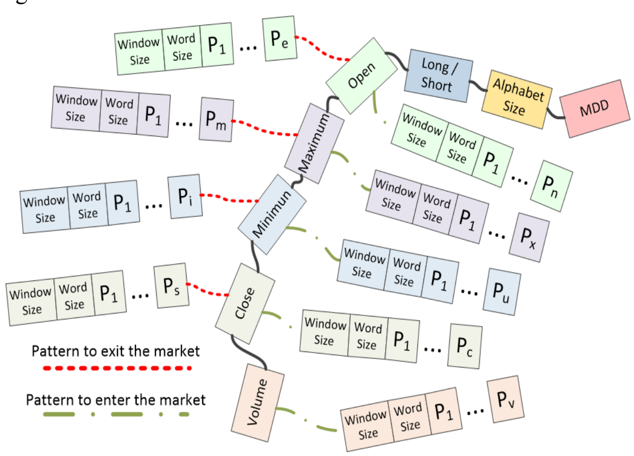

Figure 1. Chromosome structure

4. EXPERIMENTS AND RESULTS

In this section the test results for the Multi-Dimensional SAX-GA are presented. These tests were made considering all the transaction cost, which were 0.15% of the transaction with a minimum of 25 USD. The investment strategies begin the simulation with an initial capital of 10.000 USD and always invest the total available capital when decides to enter the market.

Data was retrieved from 2003 to 2010 from 5 stocks of the S&P500 index. The Multi-Dimensional SAX-GA uses a sliding window strategy for training and testing the methodology. Each training window has the size of 300 days and for testing a window of 60 days was used. The results are compared with the classic Buy & Hold (B&H) strategy and to the Multi-Chromosome SAX-GA (M-Chrom. SAX-GA) [1]. The metric used to evaluate the methodologies performance is the return on investment (ROI). Because evolutionary methods start with different initial populations, usually leading to different solutions, the consistency of the obtained results were validate over 10 runs for each stock. In Table 1, the average final results for the 5 stocks and all the 10 runs are presented.

Table 1. Multi-Dimensional SAX-GA average 10 runs results

Stock Symbol	M-Dim. SAX-GA		M-Chrom. SAX-GA		Buy & Hold
	ROI	Time in Market	ROI	Time in Market	
AMD	84%	27%	67%	91%	-54%
AYE	73%	37%	27%	92%	19%
BBT	27%	25%	27%	87%	9%
BSX	15%	30%	-47%	84%	-78%
C	12%	28%	-61%	92%	-88%

The average combined investment on these 5 stocks over the test period is shown in Figure 2, where it is possible to verify an almost linear characteristic of the earnings, this type of response indicates that the algorithm is capable of maintain earnings in all kinds of markets, *Bull* or *Bear*.

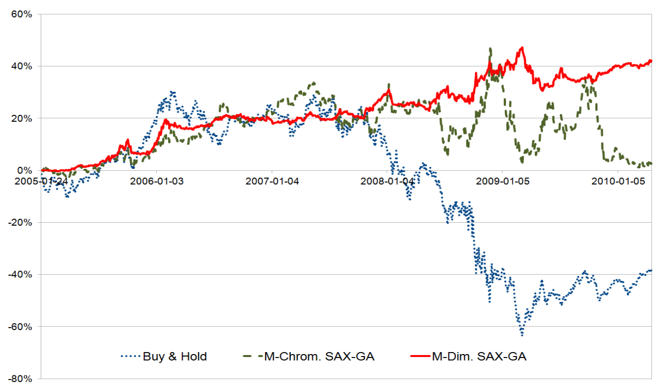

Figure 2. Average results M-Dim. SAX-GA vs. M-Chrom. SAX-GA and Buy & Hold

5. CONCLUSIONS

The new Multi-Dimensional SAX-GA approach to pattern discovery using genetic algorithms (GA), presented in this paper, is applied and validated with multi-dimensional financial time series. The proposed approach not only beats B&H and the reference one-dimensional SAX-GA in terms of ROI but also reduces risk.

6. REFERENCES

[1] A. Canelas, R. Neves, and N. Horta. A new SAX-GA methodology applied to investment strategies optimization. *Proc. 14th Inter. conf. on Genetic Evol. Comp. (GECCO'12)*, pages 1055-1062. July 2012

[2] A. Canelas, R. Neves, and N. Horta. A SAX-GA approach to evolve investment strategies on financial markets based on pattern discovery techniques. *Expert Syst. With Appl.* 40 (5):1579-1590. April 2013.

[3] E. Keogh, J. Lin, and A. Fu. HOT SAX: Efficiently Finding the Most Unusual Time Series Subsequence. *Proc. 5th IEEE Inter. Conf. Data Min.* (ICDM '05), pages 226-233. November 2005.

[4] A. Mcgovern, D.H. Rosendahl, R.A. Brown, and K.K. Droegemeier. Identifying predictive multi-dimensional time series motifs: an application to severe weather prediction. *Data Min. Knowl. Discov.* 22(1-2):232-258. January 2011.

[5] Y. Tanaka, K. Iwamoto, and K. Uehara. Discovery of Time-Series Motif from Multi-Dimensional Data Based on MDL Principle. *Mach. Learn.* 58(2-3):269-300. February 2005.