IMPROVING ICA ALGORITHMS APPLIED TO PREDICTING STOCK RETURNS

J. M. Górriz
University of Cádiz
Avda Ramón Puyol E 11202 s/n Algeciras

C. G. Puntonet
University of Granada
Daniel Saucedo E 18071 s/n

R. Martín-Clemente
University of Seville
Avda. Descubrimientos s/n

Keywords: Independent Component Analysis, On-line Support Vector Machines, Regularization Theory, Preprocessing Techniques.

Abstract: In this paper we improve a well known signal processing technique such as independent component analysis (ICA) or blind source separation applied to predicting multivariate financial such as portfolio of stock returns using the Vapnik-Chervonenkis theory. The key idea in ICA algorithms is to linearly map the input space series (stock returns) into a new space which contains statistically independent components. There´s a wide class of ICA algorithms however they usually fail due to their high convergence rates or their limited ability of local search, as the number of observed signals increases.

1 INTRODUCTION

ICA algorithms have been applied successfully to several fields such as biomedicine, speech, sonar and radar, signal processing, etc. and more recently also to time series forecasting (Górriz et al., 2003), i.e. using stock data (Back and Weigend, 1997), and they have faced the problem of blind source separation (BSS) (Bell and Sejnowski, 1995). In the latter application the mixing process of multiple sensors is based on linear transformation making the following assumptions:

1. the original (unobservable) sources are statistically independent which are related to social-economic events.
2. the number of sensors (stock series) is equal to that of sources.
3. the Darmois-Skitovick conditions are satisfied (Cao and Liu, 1996).

In this work we apply Genetic Algorithms to ICA in the search of the separation matrix, in order to improve the performance of endogenous learning support vector machines (SVMs) in real time series forecasting speeding up convergence rates (scenarios with the BSS problem in higher dimension). Thus we are using ICA as preprocessing technique in order to extract interesting structure in the stock.

SVMs are learning algorithms based on the structural risk minimization principle (Vapnik, 1999) (SRM) characterized by the use of the expansion of SV "admissible" kernels and the sparsity of the solution. They have been proposed as a technique in time series forecasting (Muller et al., 1999) and have faced the overfitting problem, presented in classical neural networks, thanks to their high capacity for generalization. The solution for SVM prediction is achieved solving the constrained quadratic programming problem thus SV machines are nonparametric techniques, i.e. the number of basis functions are unknown before hand. The solution of this complex problem in *real-time applications* can be extremely uncomfortable because of high computational time demand. Moreover the low rates of ICA algorithms can cause a design fault in real time applications.

We organize the essay as follows. SV-ICA algorithm for financial time series prediction will be presented in section 3. In section we show a guided GA to improve convergence rates in ICA algorithms. Finally we diplay results and state some conclusions in section 5.

M. Górriz J., G. Puntonet C. and Martín-Clemente R. (2004).
IMPROVING ICA ALGORITHMS APPLIED TO PREDICTING STOCK RETURNS.
In *Proceedings of the First International Conference on E-Business and Telecommunication Networks*, pages 351-355
DOI: 10.5220/0001399803510355

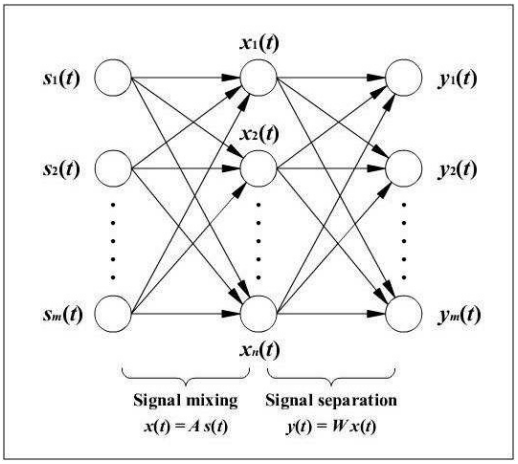


Figure 1: Independent Component Analysis and BSS problem. We suppose in financial ICA that the observed signal are linearly generated by some underlying independent factors.

2 INTRODUCTION TO ICA

ICA has been used as a solution of the blind source separation problem (Bell and Sejnowski, 1995) denoting the process of taking a set of measured signal in a vector, $\mathbf{x}$, and extracting from them a new set of statistically independent components (ICs) in a vector $\mathbf{y}$. In the basic ICA each component of the vector $\mathbf{x}$ is a linear instantaneous mixture of independent source signals in a vector $\mathbf{s}$ with some unknown deterministic mixing coefficients:

$$x_i = \sum_{i=1}^{N} a_{ij} s_j \tag{1}$$

Due to the nature of the mixing model we are able to estimate the original sources $\tilde{s}_i$ and the de-mixing weights b_{ij} applying i.e. ICA algorithms based on higher order statistics like cumulants.

$$\tilde{s}_i = \sum_{i=1}^{N} b_{ij} x_j \tag{2}$$

Using vector-matrix notation and defining a time series vector $\mathbf{x} = (x_1, \ldots, x_n)^T$, $\mathbf{s}$, $\tilde{\mathbf{s}}$ and the matrix $\mathbf{A} = \{a_{ij}\}$ and $\mathbf{B} = \{b_{ij}\}$ we can write the overall process as:

$$\tilde{\mathbf{s}} = \mathbf{Bx} = \mathbf{BAs} = \mathbf{Gs} \tag{3}$$

where we define $\mathbf{G}$ as the overall transfer matrix. The estimated original sources will be, under some conditions included in Darmois-Skitovich theorem (Cao and Liu, 1996), a permuted and scaled version of the original ones. Thus, in general, it is only possible to find $\mathbf{G}$ such that $\mathbf{G} = \mathbf{PD}$ where $\mathbf{P}$ is a permutation matrix and $\mathbf{D}$ is a diagonal scaling matrix.

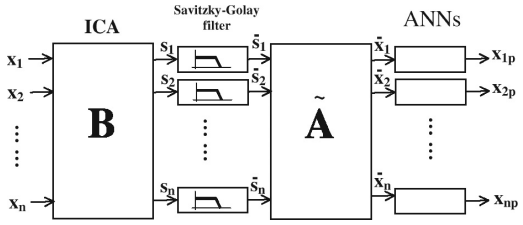


Figure 2: Schematic representation of the ICA preprocessing step.

3 SVM-ICA ALGORITHM

SVM are essentially Regularization Networks (RN) with the kernels being Green´s function of the corresponding regularization operators (Smola et al.,). Using this connection, with a clever choice of regularization operator (based on SVM philosophy), we should obtain a parametric model being very resistant to the overfitting problem. Our parametric model is a *Resource allocating Network* (Platt, 1991) characterized by the control of neural resources and by the use of matrix decompositions, i.e. Singular Value Decomposition (SVD) and QR Decomposition to input selection and neural pruning (Górriz et al., 2004).

The on-line RN based on "Resource Allocating Network" algorithms (RAN) [1] (Platt, 1991) which consist of a network using RBFs, a strategy for

- Allocating new units (RBFs), using two part novelty condition (Platt, 1991)
- Input space selection and neural pruning using matrix decompositions such as SVD and QR with pivoting (Salmerón-Campos, 2001).

and a learning rule based on the Structural Risk Minimization principle as discuss in (Górriz et al., 2004).

4 GA APPLIED TO THE SEPARATION MATRIX IN ICA PROCEDURE.

A GA can be modelled by means of a *time inhomogeneous Markov* chain obtaining interesting properties related with weak and strong ergodicity, convergence and the distribution probability of the process. A canonical GA is constituted by operations of parameter encoding, population initialization, crossover

[1] The principal feature of these algorithms is sequential adaptation of neural resources.

, mutation, mate selection, population replacement, fitness scaling, etc. proving that with these simple operators a GA does not converge to a population containing only optimal members. However, there are GAs that converge to the optimum, *The Elitist GA* and those which introduce *Reduction Operators* (Eiben et al., 1991). Our GA combines the two balancing goals: exploiting the blindly search like a canonical GA and using statistical properties like a standard ICA algorithm. In order to include statistical information into the algorithm (it would be a nonsense to ignore it!) we define the hybrid statistical genetic operator based on reduction operators as follows:

$$\langle q, \mathbf{M}_{\mathbf{G}}^{\mathbf{n}} p \rangle = \frac{1}{\aleph(T_n)} \exp\left(\frac{||q - \mathbf{S}^{\mathbf{n}} \cdot p||^2}{T_n}\right); \qquad p, q \in \wp_N \tag{4}$$

where $\aleph(T_n)$ is the normalization constant depending on temperature T_n, n is the iteration and $\mathbf{S}^{\mathbf{n}}$ is the step matrix which contains statistical properties, i.e based on cumulants it can be expressed using quasi-Newton algorithms as (Hyvärinen and Oja,):

$$\mathbf{S}^{\mathbf{n}} = (\mathbf{I} - \mu^n(\mathbf{C}_{y,y}^{1,\beta}\mathbf{S}_y^{\beta} - \mathbf{I})); \qquad p_i \in \mathbf{C} \tag{5}$$

where $\mathbf{C}_{y,y}^{1,\beta}$ is the cross-cumulant matrix whose elements are $[\mathbf{C}_{y,y}^{\alpha,\beta}]_{ij} = Cum(\underbrace{y_i, \ldots, y_i}_{\alpha}, \underbrace{y_j, \ldots, y_j}_{\beta})$ and $\mathbf{S}_y^{\beta}$ is the sign matrix of the output cumulants.

Finally the guided GA (GGA) is modelled, at each step, as the stochastic matrix product acting on probability distributions over populations:

$$\mathbf{G}^{\mathbf{n}} = \mathbf{P}_R^n \cdot \mathbf{F}_{\mathbf{n}} \cdot \mathbf{C}_{\mathbf{P_c^n}}^{\mathbf{k}} \cdot \mathbf{M}_{(\mathbf{P_m}, \mathbf{G})^{\mathbf{n}}} \tag{6}$$

where $\mathbf{F}_{\mathbf{n}}$ is the selection operator, $\mathbf{P}_R^n$ a reduction operator, $\mathbf{C}_{\mathbf{P_c^n}}^{\mathbf{k}}$ is the cross-over operator and $\mathbf{M}_{(\mathbf{P_m}, \mathbf{G})^{\mathbf{n}}}$ are the mutation and guided operators.

The GA used applies local search (using the selected mutation and crossover operators) around the values (or individuals) found to be optimal (elite) the last time. The computational time depends on the encoding length, number of individuals and genes. Because of the probabilistic nature of the GA-based method, the proposed method almost converges to a global optimal solution on average. In our simulation nonconvergent case was found.

5 SIMULATIONS AND CONCLUSIONS

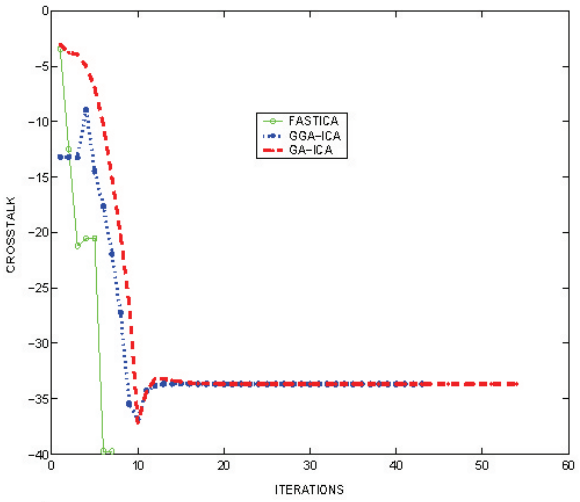
To check the performance of the proposed hybrid algorithm, 50 computer simulations were conducted to test the GGA vs. the GA method (without guide) and the most relevant ICA algorithm to date, FastICA (Hyvärinen and Oja,). In this paper we neglect the evaluation of the computational complexity of the current methods, described in detail in several references such as (Tan and Wang, 2001). The main reason lies in the fact that we are using a 8 nodes Cluster Pentium II 332MHz 512Kb Cache, thus the computational requirements of the algorithms (fitness functions, encoding, etc.) are generally negligible compared with the cluster capacity. Logically GA-based BSS approaches suffer from a higher computational complexity.

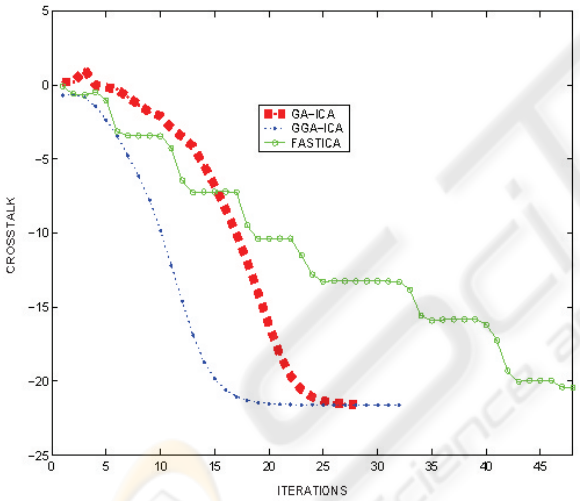
Consider the mixing cases from 2 to 20 independent random super-gaussian input signals. We focuss our attention on the evolution of the crosstalk vs. the number of iterations using a mixing matrix randomly chosen in the interval $[-1, +1]$. The number of individuals chosen in the GA methods were $N_p = 30$ in the 50 (randomly mixing matrices) simulations for a number of input sources from 2 (standard BSS problem) to 20 (BSS in biomedicine or finances). The standard deviation of the parameters of the separation over the 50 runs never exceeded 1% of their mean values while using the FASTICA method we found large deviations from different mixing matrices due to its limited capacity of local search as dimension increases. The results for the crosstalk are displayed in Table 5. It can be seen from the simulation results that the FASTICA convergence rate decreases as dimension increases whereas GA approaches work efficiently.

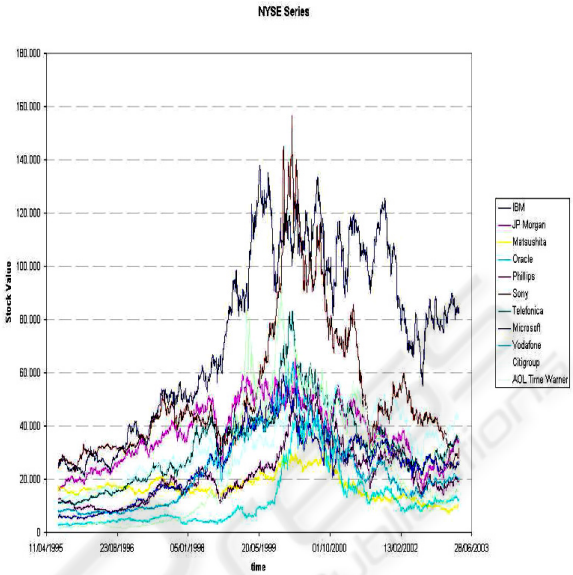
A GGA-based BSS method has been developed to solve BSS problem from the linear mixtures of independent sources. The proposed method obtain a good performance overcoming the local minima problem over multidimensional domains (see table 5). Extensive simulation results prove the ability of the proposed method. This is particular useful in some medical applications where input space dimension increases and in real time applications where reaching fast convergence rates is the major objective.

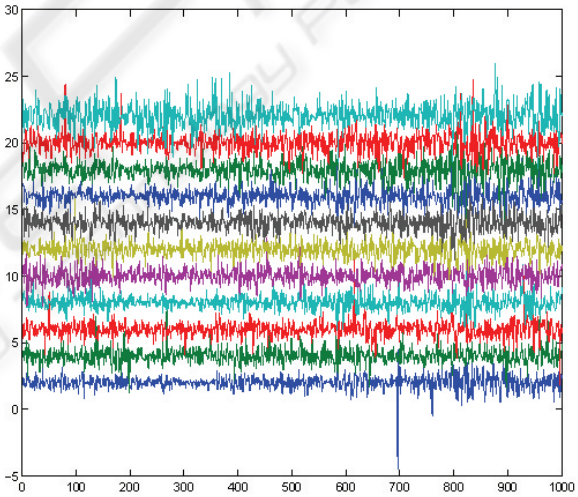
Minimizing the regularizated risk functional, using an operator the enforce flatness in feature space, we build a hybrid model that achieves high prediction performance (Górriz et al., 2003), comparing with the previous on-line algorithms for time series forecasting. This performance is similar to the one achieve by SVM but with lower computational time demand, essential feature in real-time systems. The benefits of SVM for regression choice consist in solving a -uniquely solvable- quadratic optimization problem, unlike the general RBF networks, which requires suitable non-linear optimization with danger of getting stuck in local minima. Nevertheless the RBF networks used in this paper, with the help of various techniques obtain high performance, even under extremely volatile conditions, since the level of noise and the change of delay operation mode applied to the chaotic dynamics was rather high.

Table 1: Figures: 1) Mean Crosstalk (50 runs) vs. iterations to reach the convergence for num. sources equal to 2 using Ga approach and FASTICA (Hyvärinen and Oja,). 2) Mean Crosstalk (50 runs) vs. iterations to reach the convergence for num. sources equal to 20 3) Set of Series used. 4) Set of ICs.









REFERENCES

Back, A. and Weigend, A. (1997). A first application of independent component analysis to extracting structure from stock returns. *International Journal of Neural Systems*, vol. 8(5).

Bell, A. and Sejnowski, T. (1995). An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, vol. 7:1129–1159.

Cao, X. and Liu, W. (1996). General approach to blind source separation. In *IEEE Transactions on signal Processing*, volume vol. 44,3, pages 562–571.

Eiben, A., Aarts, E., and Hee, K. V. (1991). Global convergence of genetic algorithms:a markov chain analysis. *Parallel Problem Solving from Nature, Lecture Notes in Computer Science*, vol. 496:4–12.

Górriz, J. M., Puntonet, C. G., Salmerón, M., and Ortega, J. (June 2003). New method for filtered ica signals applied to volatile time series. *Lecture Notes in Computer Science, LNCS*, vol. 2687:433–440,ISSN 0302–9743.

Górriz, J., Puntonet, C., and Salmerón, M. (2004). On line algorithm for time series prediction based on support vector machine philosophy. In *The International Conference on Computational Science. Lecture notes on Computer Science. ICCS 2004*, Krakow, Poland USA.

Hyvärinen, A. and Oja, E. A fast fixed point algorithm for independent component analysis. *Neural Computation*, 9:1483–1492.

Muller, K., Smola, A., Ratsch, G., Scholkopf, B., and Kohlmorgen, J. (1999). Using support vector machines for time series prediction. *Advances in kernel Methods-Support Vector Learning, MIT Press*, pages 243–254.

Platt, J. (1991). A resource-allocating network for function interpolation. *Neural Computation*, 3:213–225.

Salmerón-Campos, M. (2001). *Predicción de Series Temporales con Rede Neuronales de Funciones Radiales y Técnicas de Descomposición Matricial*. Phd thesis, University of Granada, Departamento de Arquitectura y Tecnología de Computadores.

Smola, A., Scholkopf, B., and Muller, K. The connection between regularization operators and support vector kernels. *Neural Networks*, 11:637–649.

Tan, Y. and Wang, J. (2001). Nonlinear blind source separation using higher order statistics and a genetic algorithm. *IEEE Transactions on Evolutionary Computation*, vol. 5(6).

Vapnik, V. (1999). *The Nature of Statistical Learning Theory*. Springer, 2nd edition edition.