An Application of One-Class Support Vector Machine for Currency Crises Discrimination

Claudio M. Rocco S.

Universidad Central Venezuela, Facultad de Ingeniería,
Apartado 47937, Caracas 1040A
e-mail: crocco@reacciun.ve

Abstract

In this paper, we investigate the utilization of one-class Support Vector Machine (OC-SVM) in assessing the occurrence of currency crises. OC-SVM faces the problem that there is only information available regarding one class. So the main idea is to develop an estimation algorithm, by training a model using examples only on a non-crises data set, that can be trained to distinguish two classes. The proposed approach is illustrated to model currency crises in Venezuela. Results are encouraging: whereas a standard (2-classes) SVM model previously used by the author scored a Noise/Signal Ratio (NSR) of 0.143, the one-class approach increases the specificity to produce a better NSR of 0.119.

Keywords: One-Class Support Vector Machine, Currency crises.

1. INTRODUCTION

In this paper, we investigate the utilization of 'one-class classifiers' [1,2] or 'novelty detectors' [3] to develop a model that helps policy makers to identify situations in which currency crises may happen.

Different methodologies have been used for early currency crisis detection. The first approach is based on multivariate logit or probit model. The second approach (the "signals" approach) has been to compare the behavior of selected variables in the period preceding crises with their behavior in a control group [4]. In [5], the authors employ empirical models built by training a Support Vector Machine (SVM).

In this two-class classification problem (crisis and non-crisis), the main difficulty is the presence of a high imbalance between classes. For example, in the example to be presented regarding currency crises in Venezuela, only 15 out of 232 cases (6.47 %) corresponds to crisis events. The class imbalance is then investigated using one-class Support Vector Machine (OC-SVM).

OC-SVM provides a different approach for conventional binary classification (crisis or non-crisis). The technique consists in simply ignoring one of the two classes and learning from a single class (in our cases, the non-crisis events or "normal events"). The crisis events are then identified as abnormal or outlier cases, that is cases that deviate from the normal profile [6].

OC-SVM has been used in engineering problems (for example condition monitoring of machinery [7]) where most of the data available is sampled from healthy systems. To our best knowledge, OC-SVM has not been yet used in currency crisis evaluation.

Section 2 contains an overview of Crisis Definition. The OC-SVM approach is presented in Section 3. Finally, section 4 presents the proposed approach and the results of an example.

2. CRISIS DEFINITION

A currency crisis is defined as a situation in which an attack on the currency leads to a sharp depreciation of the currency, a large decline in international reserves or a combination [4]. According with most of the literature on this topic, a currency crisis is identified by the behavior of an index of "Speculative Pressure" (ISP), defined as [8]:

$$ISP_t = \Delta_{\%}\text{ Exchange rate} + \Delta_{\%}\text{ Interest rates} - \Delta_{\%}\text{International reserve}$$

where all the variables (expressed in monthly percentage changes) are standardized to have mean zero and unit variance. An increase in the index due to variation on these variables, for example a loss of international reserves, reflects stronger selling pressure on the domestic currency [4]

A crisis is defined as a period with an unusual "pressure"[8: $ISP_t > \mu + k\ \sigma$, where μ is the sample mean, σ the standard deviation of the ISP series and $k \geq 1$. As suggested in [9] k=1.5 is selected to detect a crisis event while k=1.0 is used to detect a financially fragile event. Thus, we define the binary variable $Crisis_t$ as:

$$Crisis_t = \begin{cases} 1 & \text{if } ISP_t > \mu + k\sigma \\ -1 & \text{otherwise} \end{cases}$$

There is a wide set of variables that can be used to build a model to explain a crisis. In general, the choice of variables is dictated by theoretical considerations and by the availability of information on a monthly basis [4]. In the example to be presented later we use the following variables:

- Real Domestic Credit
- M2/International Reserves
- Inflation
- Oil Prices (Brent)
- An index of equity prices
- Exchange rate
- Exchange rate overvaluation (using the Hodrick-Prescott descomposition approach [10])

3. SUPPORT VECTOR MACHINES

3.1. Basic SVM (Two-classes)

Suppose $\mathbf{X}_t$ is a set of variable and y_t is the corresponding crisis evaluation. Suppose we have N training data points $\{(\mathbf{X}_1,y_1), \ldots (\mathbf{X}_N,y_N)\}$. The main idea is to obtain a hyperplane that separates crisis from non-crisis in this space that is to construct the hyperplane H: $y = \mathbf{w} \cdot \mathbf{X} - b = 0$, and two hyperplanes parallel to it:

$$H_1: y = \mathbf{w} \cdot \mathbf{X} - b = +1 \text{ and}$$
$$H_2: y = \mathbf{w} \cdot \mathbf{X} - b = -1$$

with the condition, that there are no data points between H_1 and H_2, and the distance between H_1 and H_2 (the margin) is maximized. Figure 1 shows the situation [11].

The quantities $\mathbf{w}$ and b are the parameters that control the function and are referred as the weight vector and bias [11].

The problem can be formulated as:

$$\min_{\mathbf{w},b} \tfrac{1}{2}\mathbf{w}^T\mathbf{w}$$
$$\text{s.t} \quad y_i(\mathbf{w} \cdot \mathbf{X} - b) \geq 1$$

This is a convex, quadratic programming problem in ($\mathbf{w}$, b), in a convex set. It can be equivalently solved by searching for the values of the Lagrange multipliers α_i in the Wolfe dual problem.

Once we have trained a SVM, we simply determine on which side of the decision boundary a given test pattern $\mathbf{X}^*$ lies and assign the corresponding class label, using sgn $(\mathbf{w} \cdot \mathbf{X}^* + b)$.

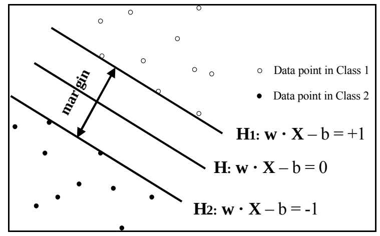

Figure 1: Decision hyperplanes generated by a linear SVM [11]

When the maximal margin hyperplane is found, only those points which lie closest to the hyperplane have $\alpha_i > 0$ and these points are the support vectors, that is, the critical elements of the training set. All other points have $\alpha_i = 0$. This means that if all other training points were removed and training was repeated, the same separating hyperplane would be found [12]. In figure 2, the points a, b, c, d and e are examples of support vectors [11].

If the surface separating the two classes is not linear, we can transform the data points to another high dimensional feature space such the data points will be linearly separable. Figure 3 is an example of such transformation [13].

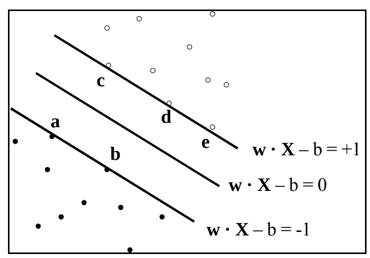

Figure 2: Example of support vectors [11]

The algorithm that finds a separating hyperplane in the feature space can be obtained in terms of vector in input space and a transformation function $\Phi(\cdot)$. We need not be explicit about the transformation $\Phi(\cdot)$ as long as we know that a *kernel function* $\mathbf{K}(\mathbf{X_i},\mathbf{X_j})$ is equivalent to the dot product of some other high dimensional feature space [12,13].

There are many kernel functions that can be used this way, for example [12,13]:

$\mathbf{K}(\mathbf{X_i},\mathbf{X_j}) = e^{-\|\mathbf{X}_i-\mathbf{X}_j\|^2/2\sigma^2}$ the Gaussian radial basis function kernel

$\mathbf{K}(\mathbf{X_i},\mathbf{X_j}) = (\mathbf{X_i} \cdot \mathbf{X_j} + m)^p$ the polynomial kernel

The Mercer's condition can be used to determine if a function can be used as kernel function [14].

With a suitable kernel, SVM can separate in the feature space the data that in the original input space was non-separable. This property means that we can obtain nonlinear algorithms by using proven methods to handle linearly separable data sets [14].

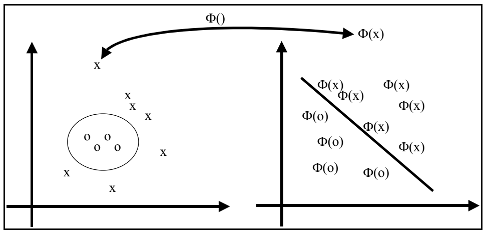

Figure 3: A non-linear separating region transformed in to a linear one [13]

The need of properly choosing the kernel is a limitation of the support vector approach. In general, the SVM with lower complexity should be selected.

3.2. One-Class SVM

Support Vector Machines provide a valid approach to the two-category classification problem (crisis/non-crisis). An adaptation of the SVM methodology in order to handle classification problems using data from only one class has been proposed by [15,16]. This adapted method, termed one-class SVM, is only trained by the data of one class and never uses the other class data. It must estimate the boundary that separates those two classes based only on data that lies on one side of it. The problem therefore is to define this boundary in order to minimize misclassifications.

After mapping input data space into a high-dimensional feature space via a kernel, OC-SVM treats the origin as the only member of the second class. Then using "relaxation parameters", the method separates the image of the one class cases from origin and then the two-class classification SVM technique is employed. Schölkopf et al. [15] formulate the one-class SVM approach as follows:

Consider that a data set has a probability distribution *P* in the feature space. Consider a subset *S* of the feature space. We want to estimate the probability that a test point $\mathbf{X}^*$ drawn from *P* lies outside of *S* is bounded by some a priori specified value $\nu \in (0,1)$. The solution for this problem is obtained by estimating a function f defined as:

$$f(X^*) = \begin{cases} +1 & \text{if } X^* \in S \\ -1 & \text{if } X^* \in \overline{S} \end{cases}$$

which is positive on S and negative on the complement of S. In other words, function f takes the value +1 in a "small" region where most of the data lies, and -1 elsewhere.

The main idea is that the algorithm maps the data into a feature space H using an appropriate kernel function, and then attempts to find the hyperplane that separate the mapped vectors from the origin with maximum margin.

To separate the data set from the origin, we need to solve the following quadratic programming problem:

$$\min\left(\frac{1}{2}\|w\|^2 + \frac{1}{\nu n}\sum_{i=1}^{n}\xi_i - \mathrm{b}\right)$$

s.t. $y_i(w \bullet \Phi(x_i)) \geq \mathrm{b} - \xi_i,\ \xi_i \geq 0, i = 1,..,n$

where $\nu \in (0,1)$ is a parameter that controls the trade off between maximizing the distance from the origin and containing most of the data in the region created by the hyper-plane, ξ_i are slack variables that penalize the objective function. Then the decision function f ($\mathbf{X}^*$) = sgn ($\mathbf{w} \cdot \mathbf{X}^*$ + b) will be positive for most examples $\mathbf{X}^*_i$ contained in the training set.

3.3. Performance of a classifier

The performance of a binary classifier is measured using sensitivity, specificity and accuracy [17]:

$$\text{sensitivity} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}}$$

$$\text{specificity} = \frac{\mathrm{TN}}{\mathrm{TN}+\mathrm{FP}}$$

$$\text{accuracy} = \frac{\mathrm{TP}+\mathrm{TN}}{\mathrm{TP}+\mathrm{TN}+\mathrm{FP}+\mathrm{FN}}$$

where:

TP=Number of True Positive classified cases (SVM correctly classifies)

TN=Number of True Negative classified cases (SVM correctly classifies)

FP=Number of False Positive classified cases (SVM labels a case as positive while it is a negative

FN=Number of False Negative classified cases (SVM labels a case as negative while it is a positive

For crisis evaluation, sensitivity gives the percentage of correctly classified non-crisis event and the specificity the percentage of correctly classified crisis event. In order to select the best model, we used the Noise/Signal Ratio (NSR), as suggested in [4,8]:

$$\mathrm{NSR} = \frac{1\text{-sensitivity}}{\text{specificity}}$$

This index measures the false signals as a ratio of the good signals issued. The selection rule is to choose the model that minimizes the NSR [8].

4. EXAMPLE

We applied the OC-SVM approach to evaluate currency crises in Venezuela, during the period January 1980-May 1999. We used a monthly database with 232 observations for each variable mentioned in section 2 [18]. During this period 15 crises were detected, using the IPE, with k=1.5. In our research we used the LIBSVM [19], a software that implements OC-SVM as well other SVM formulations.

For the OC-SVM, we trained the SVM with 203 non-crisis observations, using a Gaussian kernel. In order to select the parameters ν and σ, we use a validation data set with 14 non-crisis events and to test the novelty detector we use a data set with 15 crisis events. We tried several values for ν and σ, and the best NSR obtained was 0.119. This NSR is better than the NSR of 0.143 reported in [5], using a two-classes SVM. Basically the specificity obtained in the OC-SVM (60.0 %) is higher than the specificity reported in [5] (46.5 %).

5. CONCLUSIONS

This paper has presented an approach to evaluate currency crises based on one-class SVM. OC-SVM is able to cope with class imbalance, an important characteristic when dealing with currency crisis events, where the number of non-crisis events is much greater than the crisis events. The Noise/Signal Ratio value calculated using OC-SVM is better than the obtained with a classical SVM, suggesting that OC-SVM is a promising approach for this type of problems. Additional works are required to compare the proposed approach with other traditional financial crisis discrimination or detection tools.

6. REFERENCES

[1] Tax, D.M.J., Duin, R.P.W.: Support vector domain description. Pattern Recognition Letters. 20(11-13) (1999) 1191–1199.

[2] Tax, D.M.J.: One-class classification. Technische Universiteit Delft Ph.D.Thesis, 2001.

[3] Hayton,P., Schölkopf, B., Tarassenko, L., Anuzis, P.: Support Vector Novelty Detection Applied to Jet Engine Vibration Spectra. Advances in Neural Information Processing Systems 13, eds Leen, Todd K. and Dietterich, Thomas G. and Tresp, Volker, MIT Press, (2001) 946 - 952.

[4] Kaminsky G., Lizondo S., Reinhart C.: "Leading Indicators of Currency Crisis", IMF Staff Paper, No. 45, 1998

[5] Rocco S., C.M., Moreno J.A.:"A Support Vector Machine Model for Currency Crises Discrimination", in Computational Intelligence in Economics and Finance", Ed Shu-Heng Chen, Paul P. Wang, Springer, Berlin, 2004

[6] Cohen G., Hilario M., Sax H., Hogonnet S.: "An application of One-Class Support Vector Machine to Nonsocomial Infection Detection", Medinfo 2004

[7] Hayton P., Schölkopf B., Tarassenko L., Anuzis P.: Support Vector Novelty Detection Applied to Jet Engine Vibration Spectra, NIPS 2000, 946-952

[8] Herrera S., García C.: "A User's Guide to an Early Warning System of Macroeconomic Vulnerability for LAC Countries", XVII Latin American Meeting of Econometric Society, Mexico 1999

[9] Arreaza A., Fernández M.A., Mirabal M.J., Alvarez F.: "Fragilidad Financiera en Venezuela: Determinantes e indicadores", Revista BCV, Vol. XV, No.1, 2001

[10] Goldfajn I., Valdéz R.:The aftermath of Appreciations. Quaterly Journal of Economics, 1999, 114(1)

[11] http://www.ics.uci.edu/~xge/svm/

[12] Burges C.: "A tutorial on Support Vector Machines for Patter Recognition", http://www.kernel-machines.org/

[13] Cristianini N., Shawe-Taylor J.: "*An introduction to Support Vector Machines*", Cambridge University Press, 2000

[14] Campbell C.: "An Introduction to Kernel Methods", In R.J. Howlett and L.C. Jain, editors, *Radial Basis Function Networks: Design and Applications*, page 31. Springer Verlag, Berlin, 2000

[15] Schölkopf B., Williamson RC, Smola AJ, Shawe-Taylor J, Platt J.:"Support vector method for novelty detection, in: Adv. In Neural Information Proccessing Systems 12; 2000; MIT Press 2000, pp 582-588

[16] Schölkopf B., Platt J, Shawe-Taylor J, Smola AJ, Williamson RC: Estimating the support of a high-dimensional distribution, Neural Computing 1999, 13:1443-1471

[17] Veropoulos K., Campbell C., Cristianini N.: "Controlling the Sensitivity of Support Vector Machines", *Proceedings of the International Joint Conference on Artificial Intelligence, Stockholm, Sweden, 1999 (IJCAI99)*, Workshop ML3, p. 55-60.

[18] Alvarez F., Aponte P., Zambrano P.: "Logit y Probit: modelos para el tratamiento de variables polítomas", internal report, Facltad de Ingeniería, Universidad Central de Venezuela, 2000

[19] Chih-Chung Chang and Chih-Jen Lin. LIBSVM: a library for support vector machines, 2001.Software available at http: //www.csie.ntu.edu.tw/~cjlin/libsvm.