

Classification of Countries based on Macro-Economic Variables using Fuzzy Support Vector Machine

Divya

Indian Institute of Information Technology
Allahabad, India

Sonali Agarwal

Lecturer, Indian Institute of Information
Technology, Allahabad, India

ABSTRACT

During recent years, socio economic statuses of countries have been reviewed by the statistical researcher to find out interconnection between macro-economic variables. In this paper a cross country data set have been taken including various macro-economic variables. Data mining techniques have been applied to classify the countries using Control of Corruption, Human Development Index, Economic Freedom Index, Political Stability etc. The outcome of this research work can benefit the countries involved, in its regulation and monitoring processes, investors and business parties involved and also to maintain stability in fast changing economic scenario in the era of globalization. This paper suggests a different approach to classify countries using Fuzzy Support Vector Machine (FSVM).

Keywords

Macro-Economic variables, Classification, Fuzzy Support Vector Machine.

1. INTRODUCTION

Macro-economic variables deal with the performance structure and decision- making about the entire economy. They are helpful to determine the economy of a country. These variables are used by government and large corporation to evaluate the economic policies and business strategies. In this paper various macro-economic variables of different countries have been considered for classification and countries are classified into three groups: Developed, Developing and under development country. A fuzzy c-means clustering (FCM) has been performed to group the countries based on similar class level and then SVM Multi-classification has been applied to classify the countries. The one of most important methodology for classification problems is Support Vector Machine (SVM). SVM works well for both linear and non-linear classification problem [1][2]. To solve non-linear classification problem, SVM first maps the input data points into a high-dimensional space that is also known as feature space. It then constructs a separating hyper-plane in order to maximize the margin between two classes in this space. The mapping of data points into higher dimensional feature space is done using kernel function. Only few points are responsible for constructing the hyper-plane, those are known as Support Vector [3][4]. One of the main drawbacks associated with real world classification problems is noise and outlier. The working of standard SVM classifier is influence in the presence of outlier and noise. Training process of SVM is sensitive to the outliers or noises in the training dataset due to over-fitting [5]. Uncertainty in the result of training data points affect the decision making process. So in

this research paper work we have presented SVM using FCM Clustering approach. To deal with classification problems with outliers or noises a FCM clustering-based fuzzy SVM algorithm is developed. Jair and Xiaou Li have proposed fuzzy SVM that is suitable for handling large dataset. This paper presents a new approach of SVM Classification that is based on fuzzy clustering [6]. Another paper work that are presented by Huang and Liu [7] and Lin and Wang [8], have proposed fuzzy SVMs independently that is based on unilateral weights and standard SVM model for classification problems with outliers or noises. The paper is organized as follows: In the II section we discuss about the fuzzy c-means clustering and section III deals with SVM Multi-classification techniques. Section IV is about the results and experiment and after that in section V we conclude the paper.

2. FUZZY C-MEANS CLUSTERING

Clustering is unsupervised classification techniques that classifies or groups the objects into different-2 subset, in each subset all the object have same type of properties. The objects that belong to one cluster are similar to each other than the objects that are belonging to another cluster. The process of Clustering is splitting a set of patterns and obtains a number of classes that are homogeneous based on a suitable similarity measure in such away that the patterns that are belonging to one of the clusters are similar and the patterns that are from different clusters are as dissimilar as possible. Consider a finite set $X = \{x_1, x_2, \dots, x_n\}$, such as $x_j \in R^d$, where d is the dimension of dataset in Euclidean space and $j=1, 2, \dots, n$. The problem is to partitioning of these data into k -fuzzy sets with respect to a given criterion. The main aim is to optimize an objective function. We express the result of fuzzy clustering by a partitioning matrix U such that $U = [u_{ij}]$, $i=1, \dots, k$, $j=1, \dots, n$. u_{ij} is a numeric value in $[0,1]$. There are two constraints on the value of u_{ij} :

1. Total memberships of the element $x_j \in X$ in all classes are equal to 1.
2. Every constructed cluster is non-empty set that is different from the entire set.

$$\sum_{i=1}^k u_{ij} = 1 \text{ for all } j = 1, 2, \dots, n \quad (1)$$

$$0 < \sum_{j=1}^n u_{ij} < n, \text{ for all } i = 1, 2, \dots, k$$

A general of objective function is given below: $J(u_{ij}, v_k) = \sum_{i=1}^k \sum_{j=1}^n \sum_{l=1}^k g[w(x_i), u_{ij}] d(x_j, v_l) \quad (2)$

Where $w(x_i)$: prior weight for each x_i

$d(x_j, v_i)$: Degree of dissimilarity between the data x_j and supplemental element v_i

v_i : Central vector of i^{th} cluster.

Degree of dissimilarity is defined as: $d(x_j, v_i) \geq 0$

$$d(x_j, v_i) = d(x_i, v_j)$$

We can formulate fuzzy clustering into an optimization problem:

$$\min J(u_{ij}, v_i), i = 1, 2, \dots, k, j = 1, 2, \dots, n \quad \text{Subject to (1)}$$

The objective function is:

$$J(u_{ij}, v_i) = \sum_{i=1}^k \sum_{j=1}^n u_{ij}^m \|x_i, v_i\|^2, m > 1 \quad (3)$$

m =exponent weight that influences degree of fuzziness of membership function.

To optimize objective function differentiate the equation (3) with respect to cluster center and membership function.

$$v_i = \frac{1}{\sum_{j=1}^n u_{ij}^m} \sum_{j=1}^n u_{ij}^m \cdot x_j \quad i = 1, \dots, k \quad (4)$$

$$u_{ij} = \frac{(1/\|x_j, v_i\|^2)^{\frac{1}{m-1}}}{\sum_{l=1}^k (1/\|x_j, v_l\|^2)^{\frac{1}{m-1}}} \quad (5)$$

Where $i=1, \dots, k$, and $j=1, \dots, n$

3. SUPPORT VECTOR MACHINE

The fundamental theory of SVM is originated from statistical learning theory given by Vapnik and others [3]. Support Vector Machines (SVMs) are based on supervised learning methods which analyze the data and identify the patterns. Support Vector Machine is based on the construction of single or multiple hyper-planes in a very high dimensional space for the utilization of classification and regression techniques [9]. To solve classification problem, training dataset is given that contain 'n' observation (x_i, y_i) for $i=1$ to n . Here $x_i \in R^D$ and $y_i \in \{1, \dots, k\}$ refers the class labels. The main purpose of SVM classification technique is to learn a classification rule, $\varphi(x): R^D \rightarrow \{1, \dots, k\}$, that closely matches attributes, x_i , to the class label, y_i .

Equation of hyper plane is given as:

$$w \cdot x + b = 0$$

where w and b are two equation parameters of the hyper plane.

From this we build up a hyper plane given by:

$$w \cdot x + b > 0, \Rightarrow f = 1$$

$$w \cdot x + b < 0, \Rightarrow f = -1$$

SVM is initially intended to perform binary classification later it was designed to solve multi- classification problem by three methods such as "One-Against-Many", "Multi-category SVM" and "One-Against-One" [10]. In conventional way, "one-against-one" takes two classes into consideration at the same

time. Final decision of the class to which data point belongs is based on voting strategy. The class that have maximum number of vote for a particular data point is become final class and data point is classified into that. One-Against-Many is one of SVM multi-classification method in which a number of SVM classifiers are constructed. In this method each class is tested against rest of other class. The final class to which data point belongs based upon the value of decision function. The data point assigned to that class that has maximum value from the decision function.

For non-linear mapping of input data samples to feature space require kernel function $K(x_i, x_j)$. In this paper we have performed experiment using two kernel functions that includes:

(a) Polynomial kernel function:

$$K(x_i, x_j) = (1 + x_i^T x_j)^2$$

(b) Radial basis kernel function:

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2) \text{ for } \gamma > 0$$

3.1 One-Against-Many SVM

The first method for solving multiclass problem is called the one-against-many method. One-against-many constructs k SVM models where k is denoted the number of classes [10]. One-against-method suggests the training of i^{th} support vector. Training includes example with positive labels in i^{th} class and rest example with negative label. Therefore, training data for given $l(x_1, y_1), (x_2, y_2), \dots, (x_l, y_l)$ where $x_i \in R^n$, $i=1, \dots, l$ and $y_i \in \{1, \dots, k\}$ is the x_i class, following problem showed by the i^{th} SVM:

$$\min_{w_i, b_i, \xi_j} \frac{1}{2} (w_i)^T w_i + C \sum_{j=1}^l \xi_j^i (w_i)^T$$

$$(w_i)^T \varphi(x_j) + b_i \geq 1 - \xi_j^i, \quad \text{if } y_j = i$$

$$(w_i)^T \varphi(x_j) + b_i \leq -1 + \xi_j^i, \quad \text{if } y_j \neq i$$

$$\xi_j^i \geq 0, \quad j = 1, \dots, l \quad (6)$$

where mapping of training data x_i into higher dimensional feature space using function φ and C is represented as the penalty parameter. To obtain maximal margin hyper-plane try to minimize $1/2 (w_i)^T w_i$ which means that we would like to maximize $2/\|w\|$. But when data are not linear separable, then for this to reduce the number of training errors penalty term $C \sum_{j=1}^l \xi_j^i$ is added. The balance between the training errors and the regularization term $1/2 (w_i)^T w_i$ is achieved by SVM. After solving equation (6), we obtain k decision functions:

$$(w_1)^T \varphi(x) + b^1 \dots (w_k)^T \varphi(x) + b^k$$

Classification of sample data ‘X’ is based on the argument of decision function. If argument of decision function for any class is maximum; X is classified into that class.

$$\text{class of } x \equiv \arg \max_{i=1, \dots, k} ((w_i)^T \phi(x) + b_i) \quad (7)$$

3.2 One-against-one Method

In this method k (k-1)/2 classifiers are constructed and each one classifier is trained on data from two classes. Consider two classes ‘i’ and ‘j’ and training data that are from the ith and the jth classes, following binary classification problem is solve as:

$$\min_{w_{ij}, b_{ij}, \xi_{ij}^t} \frac{1}{2} (w_{ij})^T w_{ij} + C \sum_t \xi_{ij}^t (w_{ij})^T$$

$$(w_{ij})^T \phi(x^t) + b_{ij} \geq 1 - \xi_{ij}^t \text{ if } y^t = i$$

$$(w_{ij})^T \phi(x^t) + b_{ij} \leq -1 + \xi_{ij}^t \text{ if } y^t = j$$

$$\xi_{ij}^t \geq 0 \quad (8)$$

After constructed all k (k-1)/2 classifiers, there are different methods for doing the future testing. After some tests, we decide to use the following voting strategy: take the sign of decision function into consideration. If $\text{sign}((w_{ij})^T \phi(x) + b_{ij})$ is positive then data point ‘x’ is in the ith class. If any data point belongs to ith class then the vote for this class is increased by one. Otherwise, if data point belongs to second class, the jth is increased by one. The data point ‘x’ is belongs to the class with the largest vote. The above voting procedure is also called the “Winner-takes-all” strategy. If there is case of similar number of vote, we simply select the one with the smaller index.

4. RESULT AND EXPERIMENTS

A cross country dataset of 70 countries has been taken with following macro-economic variables: Press Freedom index (PFI), Economic Freedom Index (EFI), Human Development Index (HDI), Control of Corruption(COC), Political Stability(PS), Government Effectiveness (GE), Voice and Accountability(VA), Combined Gross Enrollment Ratio for Primary, Secondary and Tertiary schools (GER), GDP per Capita, Net Foreign Direct Investment inflows (NFDI), Female Economic Activity Rate (FEAR), Religious Freedom (RF), Consumer Price Index (CPI), GDP Growth and Population. A fuzzy SVM (FSVM) classification has been implemented using Matlab 7.0. For experiment we used two kernel functions

Polynomial and Gaussian kernel function. With the help of fuzzy clustering, class labels have been derived on the basis of RF, EFI, HDI, GDP, COC and PS. The outputs of fuzzy clustering which are class labels have been further taken as input in SVM multi-classification to classify the countries. The result indicates that the countries can be classified in three different groups on the basis of above mentioned macro-economic variables. It states that United Kingdom, United States, and Japan etc are having high HDI, good PS and better COC can be considered as developed countries. Similarly countries like India, Sri Lanka, China, and Nepal etc can be placed in developing countries because they have less HDI, PS, and COC as compared to developed countries. Afghanistan, Sierra Leone with poor stability, poor in handle the corruption, low GDP are considered in the category of under development countries.

Table 1. Comparison of accuracy of different Fuzzy-SVM Classifiers

Kernel Function	Fuzzy SVM classifier	
	One-against-one	One-against-many
Polynomial	85.34%	89.87%
Gaussian	87.12%	90.91%

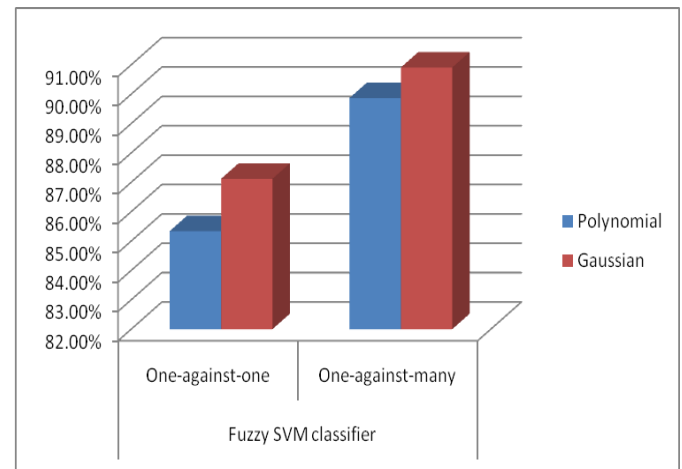


Fig 1. Accuracy comparison of different Fuzzy-SVM Classifier

5. CONCLUSION

Macro-Economic variables contribute toward the development of any country. They are helpful to determine the economic policies and business strategies of a country. On the basis of these variables we can classify the economy of different countries. In this paper we have taken 6 macro-economic variable into consideration and on the basis of these variables, we classify the country into 3 groups. We have performed experiment using Fuzzy Support Vector Machine Multi-classification with Gaussian and Polynomial kernel function. For classification purpose we have used one-against-one and one-against-many SVM multi-classification approach. The result of experiment shows that one-against-many with Gaussian Kernel function have better accuracy.

6. REFERENCES

- [1] Cortes C. and Vapnik V. 1995 Support-Vector networks Mach.,Learn.,vol.20,no.3,pp.273-297.
- [2] Lin C.T.,Yeh C.M., Liang S.F., Chung J. F. and Kumar N. 2006 Support-vector-based fuzzy neural network for pattern classification .IEEE Trans. Fuzzy Syst., vol. 14, no.1pp 31-41.
- [3] Vapnik V. N. 1998 *Statistical Learning Theory*. New York: Wiley.
- [4] Taylor J.S. and Cristianini N. 2000 Support Vector Machines and other kernel-based learning methods,” Cambridge University Press.
- [5] Zhang X. G. 1999 Using class-center vectors to build support vector machines. in Proc. IEEE Signal Process.Soc.Workshop.New York: IEEE Press,pp.3-11.
- [6] Cervantes J., Li X. and Yu W. 2006 Support Vector Machine classification Based on Fuzzy clustering for Large data sets. Springer-Verlag Berlin Heidelberg.
- [7] Huang H.P. and Liu Y.H. 2002 Fuzzy support vector Machines for pattern recognition and data mining. Int.J.Fuzzy Syst., vol. 4, no. 3, pp. 826-835.
- [8] Lin C.F. and Wang S.D. 2002 Fuzzy support vector machines. IEEE Trans.Neural Netw., vol. 13 , no. 2 , pp. 464-471(Mar. 2002).
- [9] Hsu C.W. and Lin C.J. 2002 A comparison of methods for multi-class support vector machines .IEEE Transactions on Neural Networks, 13:415-425.
- [10] Weston J. and Watkins C. Multi-Class Support Vector Machines.
- [11] Liu Y., Zheng Y.F. 2005. One-against-all multi-class SVM classification using reliability measures. IEEE International Joint Conference on, vol.2, no., pp. 849- 854 vol. 2, 31 (Aug.2005).