DE GRUYTER OPEN

International Conference KNOWLEDGE-BASED ORGANIZATION
Vol. XXII No 3 2016

KERNEL METHODS FOR DATA CLASSIFICATION

Raluca-Mariana ŞTEFAN, Măriuţa ŞERBAN, Iulian-Ion HURLOIU, Bianca-Florentina RUSU

"Spiru Haret" University, Bucharest, Romania
ralucastefan77@yahoo.com, mariuta_serban@yahoo.com, iulian_hurloiu@yahoo.com, bianca_preda1974@yahoo.com

Abstract*: In the past decades, the exponential evolution of data collection for macroeconomic databases in digital format caused a huge increase in their volume. As a consequence, the automatic organization and the classification of macroeconomic data show a significant practical value. Various techniques for categorizing data are used to classify numerous macroeconomic data according to the classes they belong to. Since the manual construction of some of the classifiers is difficult and time consuming, are preferred classifiers that learn from action examples, a process which forms the supervised classification type. A variant of solving the problem of data classification is the one of using the kernel type methods. These methods represent a class of algorithms used in the automatic analysis and classification of information. Most algorithms of this section focus on solving convex optimization problems and calculating their own values. They are efficient in terms of computation time and are very stable statistically. Shaw-Taylor, J. and Cristianini, N. have demonstrated that this type of approach to data classification is robust and efficient in terms of detection of existing stable patterns in a finite array of data. Thus,in a modular manner data will be incorporated into a space where it can cause certain linear relationship.*

Keywords: data, macroeconomic, classifiers, kernel methods

1. Introduction

The classification defines an algorithmic procedure that assigns an object received as the input value (*input*), a category of a set of given categories. An example is the membership of a country in one of two classes to which it may belong on the basis of macroeconomic indicators that characterize it: "prosperous country" - "non-prosperous country" or will lead to the optimum economic decision based on the characteristics observed at the macroeconomic indicators level.

The most known and widely used algorithm in this category is the SVM classifier (*support vector machine*). SVM are classifiers that have the greatest development, with excellent results both theoretically and empirically.

An algorithm that implements a classification problem is called a classifier. Also, classifier is a term used to describe a mathematical function implemented by a classification algorithm that converts/maps the input data (inputs) to a certain class. The entry data set consists of input instances/ objects which are grouped in categories/ lasses. An instance is described by a vector of characteristics of the item/the object. Most times it happens that nominal and ordinal data are grouped together; as in the case of integers and real values. Moreover, there are algorithms that work with nominal data and require the discretization of the actual integer /real values in groups. The classification usually refers to a supervised procedure which is a procedure that classifies new instances,

DOI: 10.1515/kbo-2016-0098

learning from a training set of instances which have been correctly categorized.

The unsupervised corresponding procedure called clustering (clustering) consists of grouping data into classes using a measure of similarity, most often using the calculation of distances between instances/objects, the latter being represented as vectors in a multidimensional vectorial space.

Empirical tests were conducted to compare different classifiers between them, in order to determine an optimal classifier for a given problem, but still at the stage where the choice lies with the user and is now a subjective choice.

2. Macroeconomic data classification

Categorization/Classification of macroeconomic data is the automatic action to assign macroeconomic indicators to some predefined classes. This mechanism can provide a conceptual vision on macroeconomic data collections and has important applications in the real world.

The requirement for the classification of macroeconomic indicators can be: 1) supervised classification which contains information on correct classification and the classes to which they belong to and 2) unsupervised classification (*clustering*) in which the classification is made only on the basis of the similarities discovered.

Lately, the number of data characterizing the macroeconomic indicators has increased enormously in volume and continues to grow as macroeconomic professionals need to consider more and more factors represented by economic indicators. Therefore, automatic organization and classification of data that compose databases of macroeconomic indicators show a major practical importance.

Given both the high cost of human resources and/or materials involved for manually organizing macroeconomic indicators or determining the classes to which they belong to, as well that certain classifications of macroeconomic data are effectively impossible to achieve in a certain period of time, explains the growing interest on finding more and more efficient methods in the field of macroeconomic data classification.

Automatic learning theory uses the features of an inductive process that builds an automatic data systematizer through learning from a set of data a priori classified. The advantages posed by this approach are accuracy, considerable reducing effort due to the fact that there is no need for intervention from the outside or from specialists in the field, neither for building the systematizer nor for its adjustment.

3. Kernel functions for data classification

Kernel functions that we can reffer to in order to perform data classification deliver the similarities between objects. Kernel methods are based on working with a matrix, the kernel matrix, containing similarities between the objects of the set of data. Therefore, kernel methods do not work with data directly.

Kernel functions are used for the non-linear extension of linear methods, in the event that if an algorithm is written in terms of scalar products, then it may change the matrix of scalar products with a function or with an arbitrary positive semi-defined matrix. This function or matrix should contain the scalar products of data in the space determined by the characteristics of the data (*feature space*). As a result, a linear extension of the algorithm will be obtained.

Kernel methods aim at learning from data and such methods have the following advantages:

- they have a theoretical and mathematical foundation extremely rigorous for defining both kernel functions and the kernel space by using kernel functions, characterization theorems and theorems regarding the degree of retention of the properties for statistical methods;

- are a tool that can be used in various fields and this versatility is due to the ability to learn which data are represented vectorially or in a manner other than vectorially;
- they can effectively solve problems of data classification used in various fields such as bioinformatics, classification of documents, computer macroeconomic, information retrieval and processing of images;
- have a common feature, a specific feature that defines all methods of this type, namely, the analysis of nonlinear patterns within an array of data (*Shawe-Taylor, Cristianini*, 2004).

According to *Shawe-Taylor, Cristianini* (2004) and to *Hofmann et al.* (2008) both the definitions of kernel methods as well as the main characteristics of data classification methods based on kernel functions can be described as follows:

1. Kernel methods are defined on the basis of two components: a) a function φ that sinks the input space χ in a space of a larger size together with the scalar product F, forming the space of characteristics and b) a classification or regression algorithm used to detect within the characteristics' space F of linear template functions that are represented as scalar products amongst the points of the characteristics' space.

2. The Kernel function k is the function which follows the relation

$k(x,z) = \langle \varphi(x), \varphi(z) \rangle$, no matter what x and $z \in \chi$, **(1)**

and φ is a function defined within the space χ with values from the space of characterics F, with the scalar product of:

$\varphi : \chi \rightarrow \varphi(x) \in F$. **(2)**

For a kernel function to become a proper candidate for solving a problem of classification, it must meet two basic properties (*Shawe-Taylor, Cristianini*, 2004):

- the kernel function should be a measure of similarity appropriate for the problem and the domain where its resolution is needed;
- the evaluation of the kernel function should contain a computational time significantly lower than the corresponding time for the detailed calculation of the scalar products from the vectors defined by φ.

A common feature of kernel methods is the ability to analyze data in a space of characteristics defined by a kernel function of higher complexity compared to the initial data space, starting only from information relating to scalar products resulted from the initial data, provided by the kernel matrix.

Kernel methods have been reintroduced in the 1990's together with SVMs and represent linear functions, which in multidimensional spaces are equivalent to linear functions of the input space. Regarding these methods, the statistical analysis shows that the extended edge can solve the problem of data dimensionality, which has led to using the kernel methods in many other areas, and the algorithms are implemented so that the calculation of scalar products between vectors is made (*Shawe-Taylor*, 2014) .

Using kernel functions ensures the discovery of non-linear connections based on linear algorithms applied in a space of characteristics specifically chosen.

This approach makes designing the algorithm no longer dependent on the properties of the space of the characteristics, and this method produces an increase in flexibility and is suitable both for automatic learning algorithms as well for designing more adaptable kernel functions for data analysis. It results that, regardless of the algorithm used, the theoretical properties of a given kernel functions remain the same.

4. The characterization theorem of kernel functions

The characterization theorem of kernel functions: A function $k: X \times X \rightarrow \mathbb{R}$, that is continuous or has as definition domain a finite array, can be decomposed into a transformation φ of features as $k(x, z) = \langle \varphi(x), \varphi(z) \rangle$, within a Hilbert space, F, applied to both arguments. Followed by the evaluation of the scalar product from F if and only if this transformation satisfies the property of being semi-defined positive finite (*Shawe-Taylor, Cristianini,* 2004).

5. Conclusions

The kernel type method can be combined with a kernel type function thus making it possible to implement and reuse the algorithm in a multidimensional space through modularity properties.

Modularity represents the ability to work with any type of kernel function, resulting in the applicability of an algorithm to any type of data from any field, including the macroeconomic field.

The kernel type approach of the data classification problem leads to the possibility to combine different modules in order to obtain complex data classification systems.

In the specialty literature of statistics and automatic learning, a consequence of continuous study and researches in the field of linear type connections between data, as well as the consequence of the development of robust algorithms, the kernel uses a kernel function, this funtions offers a simplified computational formula which streamlines the representation of linear models in multidimensional spaces. In this way, a high degree of data representability is ensured.

.

References

[1] Shawe-Taylor J., Cristianini N., *Kernel Methods for Pattern Analysis,* Cambridge University Press, 2004.

[2] Shawe-Taylor, J., *Kernel Methods for Pattern Analysis,* Machine Learning Tutorial Imperial College, February 2014.

[3] Ştefan R.M., *A Comparison of Data Classification Methods,* International Conference Emerging Markets Queries in Finance and Business, Tg. Mureş, 24-27 October 2012, Procedia Economics and Finance Elsevier, 2012.

[4] Ştefan R.M., Şerban M., *Linear Discriminant Analysis for Data Classification*, Arad Academic Days 22nd Edition, Applied Computer Science Communications, Universitatea de Vest *Vasile Goldiş*, May 2012.

[5] Ştefan R.M., Şerban M., *Tehnici de instruire SVM pentru rezolvarea problemelor de clasificare a datelor*, Conferinţa Economică Naţională "Echilibre şi dezechilibre ale pieţei româneşti în perioada actuală", 8 mai 2012, Universitatea *Spiru Haret*; 2012.

[6] Ştefan R.M., *Designining an Algorithmic Model for Notional Assertive Computing for Supervised Database with Applications in Macroeconomics*, Publicat în Journal of Economics and Technologies Knowledge, No. 4, April 2015, 2015.

[7] Ştefan R.M., *Configurarea algoritmilor de instruire şi a criteriilor de optimalitate în analiza discriminantă a datelor*, Publicat în Revista pentru Dezvoltarea Bazată pe Cunoaştere, Vol. 1, Nr. 3, Mai 2015, pag. 98-103, 2015.