On the Evaluation of Similarity for Time Series

Silvia Maria Ojeda[1], Juan Carlos Bellassai Gauto[2] and Marcos Alejandro Landi[2,3]

[1] FAMAF-Universidad Nacional de Córdoba, Argentina
[2] CIEM-CONICET, Córdoba, Argentina
[3] Instituto de Altos Estudios Espaciales Mario Gulich- CONAE, Córdoba, Argentina
juancbellassai@gmail.com

Abstract. The search and detection of similarities is a central problem in the analysis and processing of time series databases. The issue is relevant, for example, in problems of classification of time series and in situations in which a predictive process must be evaluated, or when it is necessary to compare two or more prediction methods. Many of the works oriented to the evaluation of similarity in time series have focused on the notion of dynamic distortion, with good results in the quantification of similarity, but with a high computational cost. As a result, the interest in the development of new similarity indexes and the improvement of existing similarity measures remains in force; even more considering the remarkable increase and availability of time series databases and the urgency that applications demand daily. The expectation about the new proposals is that they are able to quantify quickly and not only effectively the similarity between time series, in response to different application problems. Therefore, an interesting alternative is to investigate about simple mathematical formulation measures, which have proven useful for measuring the similarity in two-dimensional scenarios and assess their adaptation to measure similarity be-tween time series. One of the proposals to measure similarity between two-dimensional scenarios is the SSIM similarity index, defined to quantify similarity between digital images. The development was presented by Wang et al. in 2004 and has shown excellent results to evaluate the similarity between two digital images. SSIM has the advantage over other proposals, its simple mathematical formulation. In effect, this index is calculated from the product of three factors: the luminance, the contrast and the correlation between the images to be compared. These factors represent, respectively, simple relations between the means, the contrast and the correlation between the images. In this work, we adapted the SSIM index for images to the problem of evaluating the similarity in time series, obtaining a temporal similarity index called SSIMT. The results presented here showed that although the SSIM index was developed to measure similarity between images, it can be used as an index of similarity between time series (in this case called SSIMT). SSIMT and the two robust versions of the SSIMT proposed (SSIMM and SSIMR), showed better results than the D index developed by Chouakria and Nagabhushan [7], which is an index with a high performance [9], [16].

Keywords: Time Series, Classification, Clustering.