Pattern Recognition of Multivariate Time Series using Wavelet Features

Maharaj, Elizabeth Ann
Monash University, Department of Econometrics and Business Statistics
900 Dandenong Road, Caulfield East
Melbourne (3145), Australia
E-mail: ann.maharaj@monash.edu

D'Urso, Pierpaolo
Sapienza Universita di Roma, Dipartimento di Analisi Economiche e Sociali
P.za Aldo Moro,
Rome (00185), Italy
E-mail: pierpaolo.durso@imoroma1.it

Introduction

The pattern recognition of multivariate time series is of special interest in fields such as geology, environmental studies and medical research. Several authors have proposed such pattern recognition methods, amongst them being, Kakizawa et al. (1998) who considered extensions of spectral measures for multivariate time series, Maharaj (1999) who fitted vector autoregressive (VAR) models to multivariate time series and used an algorithm based on *p-values* of hypothesis tests to cluster the time series, and D'Urso (2005) who in a three-way framework, proposed cross sectional and longitudinal fuzzy clustering models for classifying multivariate time series.

We propose the use of wavelets features associated with the multivariate time series incorporated into crisp and fuzzy clustering methods to achieve pattern recognition outcomes. This approach is based on combination of the wavelet variance of each individual component series, and the wavelet correlation between every pair of component series in the multivariate set. Apart from Maharaj (1999), none of the other authors take into account the relationship between the components of the multivariate time series. However, this existing approach by Maharaj (1999) first involves modeling each multivariate time series. Our approach also takes into account the relationship between the components of the multivariate time series, but it does not require modeling the time series, hence overcoming the uncertainty associated with modeling.

We decompose each component series of the multivariate set into wavelet series on a number of scales and obtain the wavelet variances at each scale. We then obtain the wavelet correlations at each scale for every pair of component series of the multivariate set, and concatenate the wavelet variances and correlations into a single vector to represent the multivariate time series. The crisp clustering methods that are applied are the hierarchical procedures of single linkage, complete linkage, average linkage and Ward's method, and the k-means and k-medoids non-hierarchical procedures. The fuzzy clustering methods that are applied are the fuzzy k-means and the fuzzy relational procedures (see Kaufman and Rousseeuw, 1990, for more details on all these methods).

Simulation Studies

We conducted two sets of simulation studies; one using stationary bivariate time series and the other using bivariate time series that were stationary in the mean but nonstationary in the variance. In both cases we used crisp and fuzzy clustering procedures.

For crisp clustering we considered the scenario of three well-separated clusters consisting of four time series each. Four pairs of bivariate time series of lengths T=64, 256, 1024, 4096 were generated from each of three types of vector autoregressive moving average (VARMA(p, q)) model.

VARMA(1,0) or VAR(1): $X_t = A_t X_{t-1} + \epsilon_t$ where $X_t = [X_{1t}\ ,\ X_{2t},]$,

$A_t = \begin{bmatrix} A_{11t} & A_{12t} \\ A_{21t} & A_{22t} \end{bmatrix} = \begin{bmatrix} 0.5 & 0.1 \\ 0.7 & 0.5 \end{bmatrix}$, is the matrix of autoregressive coefficients and $\epsilon_t = [\epsilon_{1t}\ ,\ \epsilon_{2t},]$ is a bivariate white noise process.

VARMA(0,1) or VMA(1): $Y_t = B_t \epsilon_{t-1} + \epsilon_t$ where $Y = [Y_{1t}\ ,\ Y_{2t},]$, $B_t = \begin{bmatrix} B_{11t} & B_{12t} \\ B_{21t} & B_{22t} \end{bmatrix} = \begin{bmatrix} 0.3 & 0.7 \\ 0.1 & 0.3 \end{bmatrix}$, the matrix of moving average coefficients.

VARMA(1,1): $Z_t = A_t Z_{t-1} + B_t \epsilon_{t-1} + \epsilon_t$ where $Z_t = [Z_{1t}\ ,\ Z_{2t},]$,

$A_t = \begin{bmatrix} A_{11t} & A_{12t} \\ A_{21t} & A_{22t} \end{bmatrix} = \begin{bmatrix} 0.5 & 0.1 \\ 0.7 & 0.5 \end{bmatrix}$, $B_t = \begin{bmatrix} B_{11t} & B_{12t} \\ B_{21t} & B_{22t} \end{bmatrix} = \begin{bmatrix} -0.2 & 0.1 \\ 0.3 & 0.1 \end{bmatrix}$,

While one of the matrices of coefficients, namely, A_t in the VARMA(1,1) process is the same as that of the VAR(1) process, the other matrix of coefficients B_t is different from that in the VMA(1) process. Hence the expectation is that series generated from these three processes will group into three different clusters.

For fuzzy clustering, we considered the scenario of two-well separated clusters consisting of four time series each generated from the VAR(1) and VMA(1) models described above and a single switching series generated from a VARMA(1,1) model with parameters defined in A_t and B_t for the VAR(1) and VMA(1) respectively. Since this VARMA(1,1) model is a combination of the VAR(1) and VMA(1) models, the expectation is that the bivariate series generated from it, will display switching behaviour between the clusters of VAR(1) and VMA(1) series.

For both crisp and fuzzy clustering, we considered four correlation structures between the pairs of processes in the bivariate white noise process which were used to generating the bivariate time series from each process. These are given in Table 1.

Wavelet filters of lengths 2, 4, 6 and 8 of the Daubechies family (DB2, DB4, DB6, DB8), of length 8 of the Symletts family (SYM8), and of length 6 of the Coiflets family (CF6) were used to generate the MODWT wavelet coefficients and hence the MODWT wavelet variances and correlations (see Percival and Walden, 2000, for more details). Table 2 shows the maximum allowable number of scales for each of the filters for each series length. This is to ensure that the boundary coefficients that have an effect on the estimated scale-by-scale variance and correlation coefficients are excluded.

In each case, we examined the performance of the clustering methods for the maximum allowable number of scales down to two scales, for example, for T= 1024 with the SYM8 filter we used 7, 6, 5, 4, 3 and 2 scales.

Table 1: Correlation Structures

	1 Independent noise	2 Correlated noise	3 Correlated noise	4 Correlated noise
VAR(1)	0	0.5	0.25	0.75
VMA(1)	0	0.5	0.5	0.5
VARMA(1,1)	0	0.5	0.75	0.5

Table 2: Maximum allowable number of scales

Wavelet filter	Maximum allowable number of scales			
	T=64	*T*=256	*T*=1024	*T*=4096
DB2	6	8	10	12
DB4	4	6	8	10
DB6	3	5	7	9
DB8	3	5	7	9
SYM8	3	5	7	9
CF6	3	5	7	9

For both crisp and fuzzy clustering, we considered three scenarios for input variables for each time series into the clustering procedure. The total number of input variables for the first scenario were, only wavelet variances of each series in the bivariate pair; for the second, wavelet variances and the scale-by-scale wavelet correlations; for the third scenario only scale-by-scale wavelet correlations.

One hundred simulations were carried out each time. For crisp clustering we assessed the accuracy of the algorithms by calculating a similarity measure, namely, the adjusted *Rand Index* as proposed by Hubert and Arabie (1985) for each simulation, and then averaging them over the 100 simulations. The adjusted *Rand Index* takes on values in the interval [0, 1] with values closer to 1, indicating a higher degree of cluster accuracy.

We evaluated the fuzzy clustering approaches according to the frequency with which the four bivariate series generated from the VAR(1) process would group together in one cluster and the four bivariate series generated from the VMA(1) would group together in another cluster, but the single switching series generated from the VARMA(1,1) process that has characteristics of each of other processes would belong simultaneously to both clusters, to a substantial degree, i.e., its membership degrees would be between 0.3 and 0.7. On the other hand, if the membership degrees in one cluster is greater than 0.7, the time series would be considered to be a much more likely to be member of this cluster. The rationale for selecting the specific membership degree constraints above have been discussed by Maharaj and D'Urso (2009)

The same scenarios as those above were carried out when we introduced nonstationarity in the variance in the VAR(1), VMA(1) and VARMA(1,1) processes from which the time series were generated. Namely the generation process for the bivariate series were $W_t = C_t \times V_t$, $C_t = exp(-(t-500)^2/2 \times 200^2)$, where V_t is defined respectively as X_t , Y_t and Z_t above. Note that we did not consider time series with trend in this study because the wavelet filters have the property of making such time series stationary in the mean.

As indicated, as far as we are aware, of the other methods used to cluster multivariate time series, just the approach by Maharaj (1999) takes into account the relationship between the components of the multivariate time series. Given that our approach takes into account this relationship, we compare the output of clustering for the features used by Maharaj (1999) using the same scenarios described above. Maharaj (1999) first fitted multivariate autoregressive (VAR) models to the time series before using the *p-value* approach to cluster the time series. However, with the results of the *p-value* approach, direct comparisons cannot be made with our method because the *Rand Index* cannot be determined with this approach. This is because the correct number of clusters is not known in advance even in the simulation studies. Hence we used the estimates of the fitted VAR models from five to ten lags to cluster the time series.

Summary of Simulation Study Results

- For the shorter time series the performances of all crisp clustering methods are inconsistent across the different error correlation structures and are mostly poor for nonstationary time series.
- For the longer time series, most crisp clustering methods perform reasonably well to very well
- The performances of all crisp clustering method are generally better when both wavelet variances and wavelet correlations are input together as the clustering variables.
- Overall, with a few exceptions, there is little difference in the performance of the various crisp clustering methods.
- The performances of fuzzy clustering methods are generally poor for the shorter time series.
- In most cases for the longer time series, the fuzzy relational method with both wavelet variances and wavelet correlations together as inputs generally performs very well and better than the fuzzy k-means method.
- For longer time series, it appears that using both wavelet variance and wavelet correlations together as clustering variables more often leads to better performance than when only wavelet variance or only wavelet correlations are used as inputs.
- In all cases, for both fuzzy and crisp methods similar observations were made for all wavelet filters, namely DB2, DB4, DB6, DB8, SYM8 and CF6.
- The results of both crisp and fuzzy clustering using the wavelet features was in all cases superior to when the time series where clustered using the VAR model estimates.

Application

In Rome, a monitoring network consisting of twelve stations at various locations is used to assess air quality. Information from the municipality of Rome reveals that the monitoring stations are classified into four classes. Two classes denoted ‘B’ and ‘C’ can be considered as the urban classes. They are distinguished according to the traffic density of the areas involved. In particular, class ‘B’ with stations Arenula (AR), Cinecittà (CN), Magnagrecia (MG) and Preneste (PR), refers to residential areas and class ‘C’ with stations Francia (FR), Fermi (FE), Libia (LI), Montezemolo (MO) and Tiburtina (TI), refers to high traffic areas. One monitoring station Ada (AD) belongs to class ‘A’. This station is located near a park and therefore will have lower pollution. It should provide information about the lowest level of pollution in Rome. However, we can observe that meteorological phenomena can lead to air pollution also in this area. Two monitoring stations, Castel di Guido and Tenuta del Cavaliere belong to class ‘D’. They are areas indirectly exposed to vehicular pollution.

We consider hourly levels of air pollution in Rome[1] over a three month (October to December 2000) period. In particular we investigate the levels of carbon monoxide CO, nitrogen oxide NO and nitrogen dioxide NO2 recorded at ten monitoring station, i.e., Ada (group A), Arenula (B), Cinecittà (B), Fermi (C), Francia (C), Libia (C), Magnagrecia (B), Montezemolo (C), Preneste (B), Tiburtina (C). The other two monitoring stations were not included because of lack of data for the analysis period.

Different levels of vehicular pollution that contributes to the levels of CO, NO and NO2 characterise the locations of the stations. Note that due to large amounts of missing data records for the other pollutants, we based our analysis on just these three chemicals. The goal of the application is to show the usefulness of the clustering approaches in identifying locations with similar hourly levels of CO, NO and NO2. Thus, this analysis together with the analyses of these and other pollutants levels for subsequent years could lead to suggestions concerning locations where the monitoring occurs.

The maximum average silhouette width (see Kaufman and Rousseeuw, 1990 for more details) was used as a means of selecting the optimal number of clusters, and for most of the scales under consideration, the two-cluster solution appeared to be the optimal solution. The silhouette coefficients also indicated that one of the stations could not be clearly identified as belonging to one cluster or the other. This was confirmed by the fuzzy clustering procedures. The plot of the membership degrees of the stations in Cluster 1 against those in Cluster 2 are plotted in Figure 1.

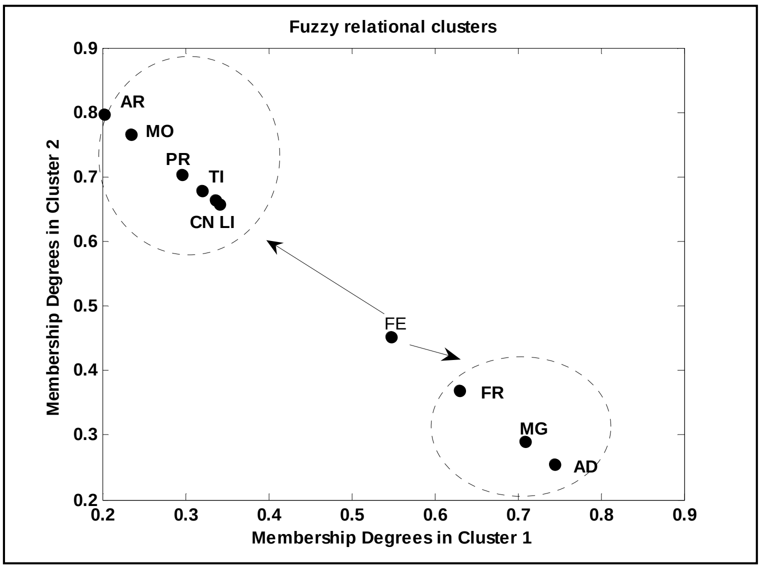

Figure 1: Fuzzy relational clusters for 5 scales

In order to validate the 2-cluster solution, we applied the crisp and fuzzy cluster methods to five subsets of the data set with the starting point randomly selected so that the time series consisted of 1024 observations. This was to ensure that up to 7 scales could be considered so that comparisons can be made with the solution from the full data set that consists of 1976 observations. The results reveal that the corresponding hard membership of the fuzzy relational method displays complete consistency across the full time series and the subset time series on all scales. In the case of the other methods, there was a slight inconsistency in that the

---

1 Source: Italian Environmental Protection Agency.

monitoring station FERMI has hard membership in either Cluster 1 or 2. Overall, we can conclude that clustering results from the four methods have been validated reasonably well.

The cluster consisting of the stations (ADA, FRANCIA, MGRECIA) is quite heterogeneous in relation to the assumed levels of pollutions they are located in, i.e. Classes A, B and C. However, the cluster (ARENULA, CINECITTA, LIBIA, PRENESTE, MONTEZEMOLO, TIBURTINA) consists of four stations in Class B and remaining two stations in Class C. The station FERMI is in Class C and from our a nalysis has fairly fuzzy membership in both clusters. Since our analysis was based on just three (CO, NO, NO2) of the six pollutants recorded (CO, SO2, O3, NO, NO2, NOX, PM10), it is quite possible that our results may be different if recordings for the other chemicals were available for the study period without huge amounts of missing data.

Conclusion

The simulation studies show the wavelet variances and correlations combined together as inputs into the crisp clustering methods that have been considered, have very good discriminatory power especially for longer time series. Likewise, for longer time series, wavelet variances and correlations combined together as inputs into the fuzzy relational clustering method are well able to detect fuzzy characteristics of multivariate time series if they exist, and successfully detect common patterns. The application to multivariate greenhouse gases time series shows that the methods considered are well validated

REFERENCES

D'Urso, P. (2005) Fuzzy clustering for data time arrays with inlier and outlier time trajectories, *IEEE Transactions on Fuzzy Systems*, 13, 5, 583-604.

D'Urso, P., and Maharaj, E.A. (2009), Autocorrelation-based fuzzy clustering of time series, *Fuzzy Sets and Systems* 160 3565–3589.

Hubert, L.J., and Arabie, P. (1985), Comparing partitions, *Journal of Classification*, 2, 193-218.

Kakizawa, Y., Shumway, H., Taniguchi M. (1998) Discriminant and clustering for multivariate time series, Journal of the American Statistical Association, 93, 441, 328-340.

Kaufman, L., and Rousseeuw, P.J. (1990) *Finding groups in data*: *An introduction to cluster analysis*, New York, J. Wiley and Sons.

Maharaj E.A. (1999) The comparison and classification of stationary multivariate time series, *Pattern Recognition* 32, 7, 1129-1138.

Percival, D.B., and Walden, A.T. (2000) *Wavelet methods for time series analysis*. Cambridge University Press Cambridge.

ABSTRACT

We propose the use of wavelets features associated with the multivariate time series incorporated into crisp and fuzzy clustering methods to achieve pattern recognition outcomes. This approach is based on combination of the wavelet variance of each individual component series, and the wavelet correlation between every pair of component series in the multivariate set.