Spatial Statistics 21 (2017) 209–225

Contents lists available at ScienceDirect

Spatial Statistics

journal homepage: www.elsevier.com/locate/spasta

Copula-based fuzzy clustering of spatial time series

Marta Disegna [a,*], Pierpaolo D'Urso [b], Fabrizio Durante [c]

[a] *Accounting, Finance & Economics Department, Faculty of Management, Bournemouth University, 89 Holdenhurst Road, Bournemouth, BH8 8EB, United Kingdom*
[b] *Department of Social Sciences and Economics, Sapienza University of Roma, P.le Aldo Moro 5, 00185 Roma, Italy*
[c] *Dipartimento di Scienze dell'Economia, Università del Salento, via Monteroni 165, 73100 Lecce, Italy*

HIGHLIGHTS

- A copula-based clustering algorithm for spatial-time series is suggested.
- The dissimilarity measure combines the dependence and the spatial information.
- The clustering is performed around medoids time series.
- The case study shows the usefulness of the proposed algorithm in regional economics.
- Different simulation studies have been presented.

ARTICLE INFO

Article history:
Received 28 February 2017
Accepted 10 July 2017
Available online 21 July 2017

Keywords:
Copula
Fuzzy clustering
Partitioning around medoids
Spatial statistics
Time series
Tourism economics

ABSTRACT

This paper contributes to the existing literature on the analysis of spatial time series presenting a new clustering algorithm called COFUST, i.e. COpula-based FUzzy clustering algorithm for Spatial Time series. The underlying idea of this algorithm is to perform a fuzzy Partitioning Around Medoids (PAM) clustering using copula-based approach to interpret comovements of time series. This generalisation allows both to extend usual clustering methods for time series based on Pearson's correlation and to capture the uncertainty that arises assigning units to clusters. Furthermore, its flexibility permits to include directly in the algorithm the spatial information. Our approach is presented and discussed using both simulated and real data, highlighting its main advantages.

* Corresponding author.
E-mail addresses: disegnam@bournemouth.ac.uk (M. Disegna), pierpaolo.durso@uniroma1.it (P. D'Urso), fabrizio.durante@unisalento.it (F. Durante).

http://dx.doi.org/10.1016/j.spasta.2017.07.002
2211-6753/

1. Introduction

Clustering of time series aims to identify similarities in patterns across time. As such, several methods have been developed according to different concepts of *similarity* that can be based on values, functional shapes, autocorrelation structure, approximation by prototype objects, etc.

Following Caiado et al., (2015) time series clustering methods can be classified into three methodological approaches (for more details, see also Warren Liao, 2005; Caiado et al., 2015; D'Urso et al., 2016a):

1. Observation-based clustering approach: in this case, the methods are based on the observed time series or suitable transformations thereof (see, e.g., Coppi and D'Urso, 2002, 2003, 2006; D'Urso, 2005; Coppi et al., 2010 and references therein).
2. Feature-based clustering approach: it contains methods that exploit specific features of the time series. For instance, these methods are based on:
    - time domain features such as autocorrelation function (ACF) (Alonso and Maharaj, 2006; Caiado et al., 2006, 2009; D'Urso and Maharaj, 2009), partial autocorrelation function (PACF) and inverse autocorrelation function (IACF) (Caiado et al., 2006), quantile autocovariance function (QAF) (Lafuente-Rego and Vilar, 2016; Vilar et al., 2017);
    - frequency domain features such as periodogram and its transformations (Caiado et al., 2009), coherence (Maharaj and D'Urso, 2010) and cepstral (Maharaj and D'Urso, 2011);
    - wavelet features such as wavelet decomposition (D'Urso and Maharaj, 2012; D'Urso et al., 2014).
3. Model-based clustering approach: the methods belonging to this class assume the existence of a stochastic mechanism generating the time series. Moreover, they are based on the fact that a set of time series generated from the same model would most likely have similar patterns. In general, here the time series are clustered by means of the parameter estimates or exploiting the residuals of the fitted models (Caiado et al., 2015). In this class, one can include, among others, methods based on:
    - ARMA or ARIMA models (see, e.g., Piccolo, 1990; Maharaj, 1996; Kalpakis et al., 2001; D'Urso et al., 2013b);
    - GARCH representation (see, e.g., Caiado and Crato, 2010; Otranto, 2010; D'Urso et al., 2013a, 2016a);
    - density function and forecast density (Alonso and Maharaj, 2006; D'Urso et al., 2017);
    - functional approach (see, e.g., James and Sugar, 2003);
    - splines (see, e.g., Garcia-Escudero and Gordaliza, 1999).
    - copulas, measures of association, and tail dependence (see, e.g., De Luca and Zuccolotto, 2011; Durante et al., 2014b; De Luca and Zuccolotto, 2015; Durante et al., 2015; Di Lascio and Giannerini, 2016).

Inside the class of model-based methods, here we focus on the copula-based approach, as recently reviewed in Di Lascio et al. (2017). Copulas are probability distribution functions with uniform marginals, which can be also seen as aggregation functions with special properties (see Durante and Sempi, 2016; Grabisch et al., 2009). They have been extensively used for modelling uncertainty of different types, from probabilistic methods (see Joe, 2015; Nelsen, 2006) to imprecise probabilities and decision theory (see Yager, 2013; Klement et al., 2014; Montes et al., 2015). Nowadays, copula-based models are also frequently used in many problems from spatial statistics; (see, e.g., Bárdossy and Li, 2008; Durante and Salvadori, 2010; Kazianka and Pilz, 2010; Guthke and Bárdossy, 2017).

In stochastic models, copulas are employed in order to represent a joint probability distribution function of a random vector in terms of its marginal distributions. A copula-based model for time series assume that (1) each time series is a realisation of a suitable univariate model (like ARMA, GARCH, ARIMA, etc.) and (2) the innovations (ε_{it}) of the individual time series are jointly coupled by means of a time-invariant copula C (see Patton, 2012).