ARTICLE IN PRESS
FSS:7328

Available online at www.sciencedirect.com

ScienceDirect

Fuzzy Sets and Systems ••• (••••) •••–•••

FUZZY sets and systems

www.elsevier.com/locate/fss

Spatiotemporal extended fuzzy C-means clustering algorithm for hotspots detection and prediction

Ferdinando Di Martino [a], Witold Pedrycz [b], Salvatore Sessa [a,c,*]

[a] Università degli Studi di Napoli Federico II, Dipartimento di Architettura, Via Toledo 402, 80134 Napoli, Italy
[b] University of Alberta, Department of Electrical and Computer Engineering, Edmonton, AB, Canada
[c] Università degli Studi di Napoli Federico II, Centro Interdipartimentale di Ricerca "Alberto Calza Bini", Via Toledo 402, 80134 Napoli, Italy

Received 6 January 2016; received in revised form 12 November 2017; accepted 13 November 2017

Abstract

We present a spatiotemporal clustering method, namely SEFCM, which is a generalization of the extended fuzzy C-Means (EFCM) method for detecting hotspots in spatial analysis. Each pattern is formed by three features: the geographical coordinates and the period in which a certain event is occurred. This method is applied to a spatial dataset (formed by earthquake epicenters occurred in Southern Italy since 2001 till to 2014) for prediction of the hotspots obtained in a given year. Comparisons of the prediction results are also made with the ones obtained by applying the known ST-DBSCAN algorithm.

Keywords: EFCM; SEFCM; GIS; Hotspot

1. Introduction

In spatial analysis an hotspot is generally defined as an area containing dense clusters of localized events (e.g., criminal incidents, presence of fire, localization of strains of disease, etc.). Identification of hotspots on the map is usually made by geo-referencing as points the events happened in a certain period. One of the most known ways to detect hotspots is to use cluster techniques, which are an effective way for determining areas exhibiting high concentrations of localized events. Many clustering techniques were used for hotspot detection and events distribution and time-evolution in spatial analysis: K-means algorithm [18] were applied to detection of hotspots with high number of fires [30] and road accidents [1], fuzzy C-means algorithms (FCM) [3,4] were used for detecting hotspots in crime analysis [7,8,11,19,26,28]. The National Institute of Justice at Washington DC (USA) has developed a spatial statistical tool, called CrimeSTAT (see http://www.icpsr.umich.edu/CrimeStat/about.html), for GIS (Geographical Information Systems) analysis of crime incident locations. Kernel density estimation methods [8,10,22,26] were used

* Corresponding author at: Università degli Studi di Napoli Federico II, Dipartimento di Architettura, Via Toledo 402, 80134 Napoli, Italy. Fax: 390812538909.

E-mail addresses: fdimarti@unina.it (F. Di Martino), wpedrycz@ualberta.ca (W. Pedrycz), sessa@unina.it (S. Sessa).

https://doi.org/10.1016/j.fss.2017.11.011
0165-0114/

Please cite this article in press as: F. Di Martino et al., Spatiotemporal extended fuzzy C-means clustering algorithm for hotspots detection and prediction, Fuzzy Sets Syst. (2017), https://doi.org/10.1016/j.fss.2017.11.011

in crime analysis [6,7,28] and road accident analysis [1]. In [9] a hybrid approach was presented in which integrated are FCM and kernel density estimation methods. In [31] two spatiotemporal approaches are based on density-based clustering methods. A spatiotemporal extension of the density based algorithm DBSCAN [9] is used to detect seismic events. Density-based clustering algorithms detect cluster prototypes as polygons on the map. They produce accurate results with a high computational complexity. In [5] a density clustering based on DBSCAN is proposed, namely ST-DBSCAN, applied on spatiotemporal datasets. The computational complexity of this algorithm is $O(n \cdot \log n)$ as in DBSCAN.

Other categories of techniques are applied for spatiotemporal hotspots detection as STScan algorithms [17,23–25, 29,31]. They scan an entire space, hence they are computationally expensive and unsuited to be applied on large data sets.

On the other hand, the FCM-based algorithms, having a linear computational complexity and hence attractive from this point of view, exhibit some drawbacks such as to fix a priori the number of clusters and less robustness to the presence of noise and outliers.

In [20,21] a generalization of the FCM algorithm, called Extended Fuzzy C-means (EFCM), was presented: the cluster prototypes are extended to hyper-spheres whose volume is determined automatically from the data. By using a function of similarity among the clusters, a heuristic supporting cluster merging was introduced: the user starts the iteration by setting an excessively large number of clusters and in the sequel similar clusters are merged in order to obtain a suitable partition during the iteration process. The merging process is guided by monitoring values of an index of fuzzy inclusion between two clusters.

With respect to the FCM algorithm, the EFCM algorithm mainly exhibits two advantages: one obtains recursively the optimal number of clusters and it is robust to noise.

In [2,12–15] a new hotspot detection method based on the EFCM algorithm was proposed. The EFCM algorithm is implemented inside a GIS tool and the patterns are given by localized events occurred in the area of study. The features are specified by presenting two geographic coordinates and the cluster prototypes are represented as circles.

In addition to these methods above described, the evident advantages of the application of the EFCM with respect to the classical FCM method is that the cluster prototypes are circular areas that can be displayed on the map and hence they can approximate the hotspots. Having this in mind, we use geo-processing and spatial analysis operators for analyzing the distribution and evolution of the hotspots and their impact on the area of study.

In [2,12,15] the spatiotemporal evolution of the detected hotspots was studied, comparing the hotspots obtained in two consecutive years and by analyzing their intersection on the map. In this way, it was possible to follow the evolution of a particular phenomenon through the time. In [16] the extension of the Gustafson–Kessel algorithm is useful for detecting ellipsoidal hotspots on the map and analyzing their spatiotemporal evolution.

In [2,12,15] the EFCM algorithm gives efficient way to analyze the spatiotemporal evolution of the hotspots: the events dataset is segmented into periods and for each subset of patterns occurring in a certain period, the EFCM algorithm is applied for displaying the hotspots detected as circles on the map. By measuring the spatial relations between hotspots detected in consecutive periods, it is possible to analyze how the phenomenon evolves spatially and temporally.

In this paper, we present a variation of the EFCM algorithm, namely the spatiotemporal EFCM algorithm (SEFCM) for detecting spatiotemporal hotspots. It adopts the distance used in the augmented FCM algorithm [27] where both the spatial and temporal features of the patterns were considered. Each pattern is formed by r spatial features and q temporal features. The temporal part of the distance is expressed by a multiplicative parameter (weight) λ that allows to control the impact of temporal features of the patterns on the formation of the clusters.

We consider event patterns having three features: the latitude and longitude coordinates and the temporal period in which the event happened. We adopt the distance defined in [27] in which the contribution of the temporal period is augmented by the multiplicative factor λ. We use the reconstruction criterion defined in [27] to obtain the optimal value of λ. This criterion uses the resulting centroid prototypes of the clusters and the partition matrix for reconstructing the patterns. If the error obtained measuring the Euclidean distance between the patterns and the corresponding reconstructed patterns becomes minimal, this means that the quality of the reconstruction is maximal.

The hotspots corresponding to a specific period are visualized as circles on the map setting the third coordinate in the resulting prototype clusters which are spheres. We can study the spatiotemporal evolution of the hotspots by analyzing the spatial relations on the map of hotspots obtained in two consecutive time periods.

Please cite this article in press as: F. Di Martino et al., Spatiotemporal extended fuzzy C-means clustering algorithm for hotspots detection and prediction, Fuzzy Sets Syst. (2017), https://doi.org/10.1016/j.fss.2017.11.011