Computers and Electronics in Agriculture 70 (2010) 276–284

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

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

An intelligent system employing an enhanced fuzzy *c*-means clustering model: Application in the case of forest fires

L.S. Iliadis [a,*], M. Vangeloudh [a], S. Spartalis [b]

[a] Democritus University of Thrace, Department of Forestry and Management of the Environment, and Natural Resources, Orestiada, Greece
[b] Democritus University of Thrace, Department of Production Management Engineering, Xanthi, Greece

ARTICLE INFO

Article history:
Received 30 November 2008
Received in revised form 1 July 2009
Accepted 14 July 2009

Keywords:
Extended fuzzy *c*-means clustering
Innovative fuzzy termination criterion
Forest fires
Forest fire risk clustering

ABSTRACT

Fuzzy *c*-means is a well-established clustering algorithm. According to this approach instead of having each data point $Dp_i = (X,Y)$ belonging only to a specific cluster in a crisp manner, each Dp_i belongs to all of the determined clusters with a different degree of membership. In this way cluster overlapping is allowed. This research effort enhances the fuzzy *c*-means model in an intelligent manner, employing a flexible fuzzy termination criterion. The enhanced fuzzy *c*-means clustering algorithm performs several iterations before the proper centers of the clusters "more or less" stabilize, which means that their coordinates remain "almost equal" to the previous ones. In this way the algorithm is expanded to perform in a more flexible and human like intelligent way, avoiding the chance of infinite loops and the performance of unnecessary iterations. A corresponding software system has been developed in C++ programming language applying the extended model. The system has been applied for the clustering of the Greek forest departments according to their forest fire risk. Two risk factors were taken into consideration, namely the number of forest fires and the annual burned forested areas. The design and the development of the innovative model-system and the results of its application are presented and discussed in this research paper.

1. Introduction

This research effort focuses in the development of an intelligent extended version of the fuzzy *c*-means clustering algorithm (FCC) by utilizing a flexible termination criterion of the iteration cycles. It also describes the development of an information tool that automates the whole process and its application in forest fire risk clustering. The developed software system has been named *FUCEMFUTEC* (fuzzy *c*-means with fuzzy termination criterion) and it has been developed in the C++ programming language.

Artificial neural networks, support vector machines (that constitute the mathematical models of learning) plus fuzzy logic systems (that enable flexible knowledge representation) are the basic parts of a whole branch of Artificial Intelligence called Soft Computing.

The problem of forest fires has been faced under various perspectives by several scientists worldwide in the literature. Successful statistical approaches like Regression analysis, Probabilistic analysis and also Artificial Intelligence approaches (or mixtures of both) have been proposed and applied (Brillinger et al., 1986; Cheney, 1991; Iliadis et al., 2002; Iliadis, 2005; Nute et al., 2005; González et al., 2007). Statistical analysis is very powerful, however it is a fact (generally acceptable) that the mathematical tools that we have in our disposal are not always adequate and that we have to try to develop new methodologies all the time. This paper aims in proposing and applying new modeling techniques in a real world environmental problem.

This research effort does not aim in solving the problem of forest fire forecasting by estimating a curve. Its aim is the clustering of the Greek forest departments (GFD) based on two main historical characteristics (the number of forest fires and the burned areas). Moreover, the most important thing is the fact that it performs the clustering in a fuzzy sense with an innovative "*smart*" termination criterion. This means that each forest department is assigned a degree of membership to the "Very Risky" cluster and another degree of membership to the "Risky" cluster.

It is common practice in the risk analysis literature to use fuzzy logic. Most researchers accept and apply fuzzy sets and fuzzy algebra for the risk screening and prioritization at plant and regional levels (Mock and Gheorghe, 1999).

Application of the *FUCEMFUTEC* has been performed for the clustering of the GFD based on the annual number of forest fires and on the hectares of burned areas per year. This is a very useful practical application that determines the forest fire risk cluster—degree of each GFD for a period of 15 years from 1983 till 1997, whereas there are no complete and accurate surveys available for 1998. This is due to the fact that in 1998 the protection

* Corresponding author. Tel.: +30 25520 41135; fax: +30 25520 41135.
E-mail address: liliadis@fmenr.duth.gr (L.S. Iliadis).

0168-1699/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.compag.2009.07.008

of the forests was transferred from the GFD to the Greek fire stations. Though available data involve a quite long period, this can be considered as a pilot application that determines the validity of the fuzzy clustering methods towards forest fire protection.

Of course there are also available forest fire data from 1999 till today, but they are related to prefectures (which are much wider areas) and thus they cannot be combined to the data of past years. An indicative application of the *FUCEMFUTEC* has been performed towards forest fire risk clustering on prefectures for four years, namely 1999, 2000, 2001, 2002. Due to the small number of data records, this application has been performed only to demonstrate the new model and its potential use. However this is a first step before the actual employment of the system in a real situation.

This paper deals with a rather unexplored Soft Computing subject (such as fuzzy termination criterion) and it offers an enhanced version of the fuzzy *c*-means clustering algorithm, by applying it on a very crucial environmental issue (maybe for the first time). This is a positive aspect of the paper.

A future effort will involve the clustering of prefectures based on data from year 1999 till 2008 and also the employment of new techniques that offer reliable solutions when applied in cases with missing data.

The performed clustering has in many cases revealed the most risky GFD and it has shown that it is a useful tool towards the enhancement of a proper prevention and protection policy.

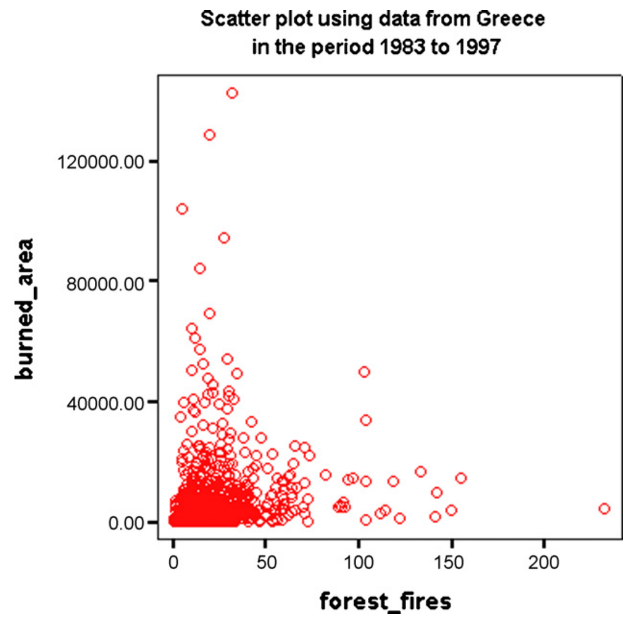

Fig. 1. Scatter plot of burned areas for each forest department over forest fire frequencies.

1.1. The problem of forest fires

1.1.1. The situation in Greece

Forests are invaluable renewable natural resources producing wealth, goods and services towards financial development and improvement of quality of life. However they are often exposed in serious risks due to both natural and human causes. Forest fires are the main threat for all Mediterranean forest ecosystems and their consequences are very serious and disastrous every year. Fauna and flora are destroyed, torrential phenomena are caused, agricultural production is significantly reduced and even human lives are lost.

According to Iliadis (1998) the problem is intense not only in the Mediterranean countries but also in the USA, in Australia and in Canada. Data reveal that Greece has the highest problem between the European Union countries according to the average burned area per forest fire incident. It has been estimated that almost 39.4 ha are burned per forest fire in Greece, whereas in Spain 28.4, in Italy 19.74 and in Portugal 15.29 (Iliadis, 1998; Dimitrakopoulos, 1994). Fig. 1 uses forest fire data of Greece from 1983 to 1997 and it shows the distribution of burned areas over forest fire frequencies and also some extreme outliers.

During the last 30 years the number of forest fires in Greece has been increased by 300% due to intense human intervention and activities, due to the use of forest land for building and construction and also due to the increase in available biomass as a consequence of the population reduction in the countryside (Kailidis, 1990; Markalas and Kailidis, 1993).

The Greeks do not hear about the forest fire problem from the news any more but they face it in their everyday life, near their houses, near their cottages, near their towns and villages. Characteristic cases that support this point of view are the disastrous forest fires of Pendelh mountain (close to Athens) in 1995, 1998 and 2000, the forest fires of the Urban forest “Seich-Sou” of Thessaloniki in 1997 with severe long-term consequences in the levels of the city atmospheric pollution (Damialis et al., 2007), the successive forest fire destructions of Immitos area (close to Athens) in 1996, 1998 and 2005, the very serious fire incidents of Parnitha mountain in 2007 and the vast destruction of the Peloponnesus area (Southern Greece) in August 2007 where 84 people lost their lives (World Wide Fund for Nature www.panda.org, 2007).

1.1.2. Literature review on forest fire risk estimation

There is no widely accepted definition of risk. Generally we can say that risk involves possibility of loss or injury; someone or something that creates or suggests a Hazard. It is clear that the core of all definitions of risk is the same: risk exists when loss is possible and its financial, social or environmental impact is significant.

In this study, the risk related to forest fires is a pure number in the closed interval [0,1] and it is determined by the use of historical data, such as the number of forest fire incidents and the hectares of burned area.

The degree of risk for all examined cases (forest departments) depends on their characteristic cluster and also on their degree of membership to this cluster. Of course both of them (the cluster and the DOM) depend on both of the historical data (number of forest fires and burned area). The number of forest fires is related to the ignition component (both man- and lightning-caused). ‘Hazard’ refers to the state of the fuel, exclusive of weather or the environs in which the fuel is found and it is related to the burned area. In the context of technical risk assessment, the term “risk” considers not only the possibility of an event, but also includes values and expected losses (Hardy, 2005).

Early forest fire risk estimation can be used for the design of a proper resources allocation strategy and towards the determination of effective protection and prevention measures.

Several efforts have been made towards this direction on a global scale during the last 30 years. Deeming et al. (1978) have developed National Fire Danger Rating System (NFDRS) in USA which is still used under several improvements. It estimates the daily Ignition Component and the daily Man Caused Risk indices.

Probabilistic approaches measuring forest fire risk have been carried out in USA, Australia and New Zealand (Brillinger et al., 1986) and also in Spain (Ramon and Olabarria, 2006).

According to Iliadis (1998) Garcia and Soriano have developed a local forest fire risk estimation model in 1993, which is suitable only for the area of Galicia in Northwestern Spain.

An international System is the EFFIS (European Forest Fire Information System) which has been developed by the Institute for Environmental and Sustainability Land Management located in Italy (Ayanz et al., 2003; Lopez et al., 2001). In fact the EFFIS is a common effort of the European countries to develop a common