International Journal of Innovative
Computing, Information and Control
Volume **5**, Number **6**, June **2009**

ICIC International ©2009 ISSN 1349-4198
pp. 1475–1483

FUZZY REGRESSION MODELS ON ENTROPY BASED BLOCKING STRUCTURES

Mika Sato-Ilic

Faculty of Systems and Information Engineering
University of Tsukuba
Tsukuba, 305-8573, Japan
mika@sk.tsukuba.ac.jp

Received February 2008; revised August 2008

Abstract. *An entropy based fuzzy blocking regression model is proposed in terms of entropy for the distribution of degree of belongingness of an object to clusters. Our approach is based on using fuzzy clustering as a sufficiently explainable method to describe the distribution of an object over the classification structure. We extend the conventional blocking regression model by using two types of degrees. One is the degree of belongingness of an object to a fuzzy cluster and the other is the degree of uncertainty of the object to the classification structure. This extension has the following merits: (1) the fitness is improved due to the increase in the degree of freedom of the clustering result. (2) we can simultaneously capture the relationship between the fuzzy clusters and a dependent variable along with the relationship between the independent variables and a dependent variable. (3) objectively obtained clusters can be used, in spite of this, conventional blocking regression requires external information to identify the clusters. (4) by considering not only the degree of belonging to a single cluster, but also the relatively related degree of clusters for a fixed object, we can avoid the noise of the data. Several numerical examples show improved performance in terms of the fitness for the proposed model.*
Keywords: Fuzzy regression, Fuzzy cluster, Blocking regression, Entropy

1. **Introduction.** The efficiency and usefulness of linear regression analysis is well known in the area of multivariate data analysis and widely used in many areas. The primary objective of this research is to capture the homogeneous linearity of data. The restriction of homogeneous linearity of the linear regression model is not adaptable, when it pertains to the implementation of an explicit "spatial" perspective in data. This provides our motivation for including spatial effects in regression models, both from a theory-driven as well as from a data-driven perspective. In order to carry out the inclusion of the spatial effects, representation of the heterogeneous nonlinear structure in the data is necessary. Therefore, the key problem is how to represent the heterogeneous nonlinearity. Sometimes, such heterogeneous nonlinearity is described by a locally distributed classification structure. A typical case of such data is when the data is distributed geographically and each specific area has a different linear structure. Otherwise we can consider a case in which the individual source of the data is locally separated in the population and the local structure is assumed to be captured as a linear structure. In both these cases, areas or sources are treated as clusters (or blocks) in the data. Therefore, by including the spatial distribution over the clusters and by revealing relations to the local distribution of the data, the use of a classification structure for the regression model is introduced. In these models, the heterogeneous nonlinearity is represented by the classification structure. The fuzzy c-regression model by [8], geographically weighted regression model by [3], and fuzzy cluster loading models by [12] are typical examples. Using a tree structure, bagging