Neighborhood Rough-Sets-Based Spatial Data Analytics

D

Sharmila Banu K
VIT University, India

B. K. Tripathy
VIT University, India

INTRODUCTION

With the huge amount of data generated every day, governments, corporates and scientists look at mining useful information from them. Considering the storage and organization costs of these data, the useful trade-off is to discover useful patterns in them. Transactional, telecommunication, spatial, satellite, remote sensing, medical informatics etc., are some of the domains dealing with *Big Data*. Discovering patterns and inferring predictions are vital to the well-being of man-kind at large. Analysis of spatial data usually includes construction of information system, dimensionality reduction, decision rule extraction based on a computational model and error analysis. This chapter focuses on the construction decision systems, proposes a hybrid method by substantiating the advantages and challenges involved with spatial data.

Spatial Data upholds an important perspective that it can be analysed only with respect to specific spatial reference (or a geographic area).Spatial data include spatial attributes like temperature, rainfall, humidity, slope, land cover etc., These features will be with respect to a spatial reference and encompass spatial auto correlation. Spatial auto correlation is the property of spatial data where the values of a feature will tend to be similar in a neighborhood and vary with increasing distance from the neighborhood. Figure 1 illustrates a sample region identified with latitudes and longitudes. A number of spatial features can be measured and some of them are shown in Table 1.

BACKGROUND

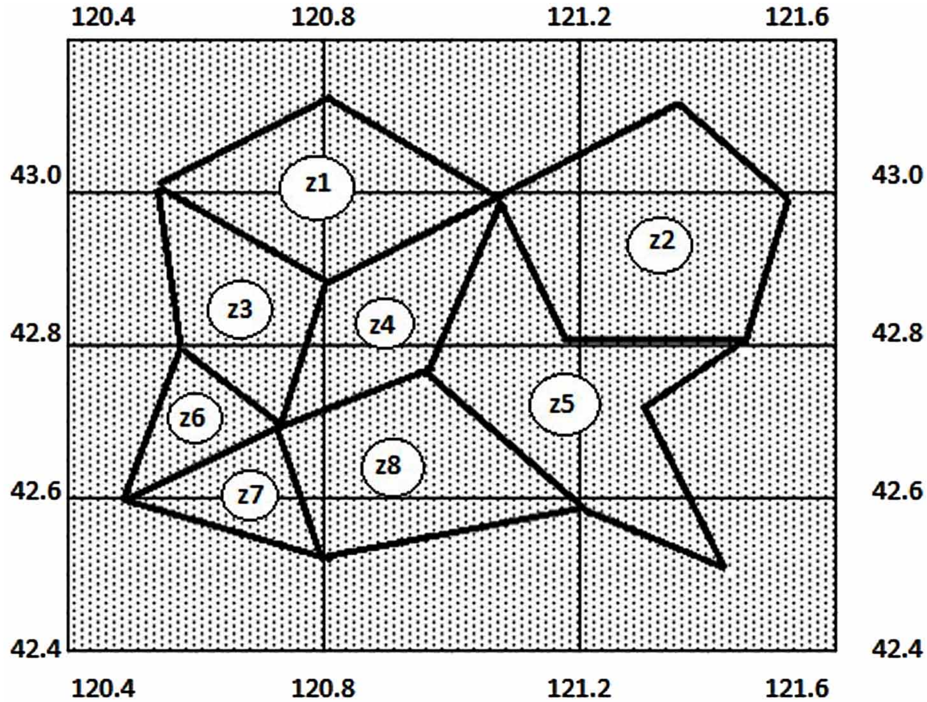
Mining spatial data is useful in fields like weather forecasting, natural calamity prediction, crime management, transmission and spread of infectious diseases and others. This calls for expertise in these areas and the nature of spatial data. For example, representing topology in spatial data modelling is inherent to dealing with uncertainties. And, Rough Sets have been used to deal with uncertainty in spatial data mining. Pawlak's (1982) Rough Set Theory (RST) has been used to model spatial regions with unclear boundaries. Beaubouef and Petry (1994) have demonstrated the use of rough sets have been used to query crisp data in relational databases. The Region Connection Calculus (RCC) proposed by Randell & Cohn (1992) and Egg-Yolk models by Cohn and Gotts (1996) have been blended with the ap-

Table 1. Sample Attribute Data of the region in Figure 1

Zones	Rainfall	Temperature	Slope
z1	21	18	1.6
z2	21	19	1.8
z3	22	19	2.1
z4	26	22	2.4
z5	16	28	1.4
z6	17	26	1.3
z7	19	27	1.7
z8	20	27	1.9

DOI: 10.4018/978-1-5225-2255-3.ch160

Figure 1. A sample geographic region



proximation concepts of RST to identify vague region boundaries. Rough sets have been used by Bai et. al (2010) to identify villages with birth defects, Ahlqvist (2005) for spatial classification and analysis, Leung et. al. (2007) for discovering classification rules in remote sensor data, Øhrn A (1999) for disease diagnosis and outcome prediction and Thangavel and Pethalakshmi (2006) for dimensionality reduction.

Spatial Data Analysis presents a myriad of challenges in modelling due to the uncertainty intrinsic to it. Spatial auto correlation is yet another feature which presents the reference with core and periphery. And clearly RST can handle this with its approximation concepts.

Spatial Data: Challenges

RST uses equivalence relation to partition the universe which needs a discrete value space. Spatial Data involves continuous values and Jensen & Shen (2004) state that there may be loss of information upon discretization of continuous information. Lin (2000) proposed Neighborhood Rough Sets and Dubois & Prade (2009), Fuzzy Rough Sets to solve this situation. Further, the granule of information supported in RST is based on partitions which yield single layer granulation structures. Neighborhood Rough Sets associate each element of universe with a family of Neighborhood granules which leads to multi-layered granulation.

Dimensionality reduction is also accomplished by rough sets based on indiscernibility. A subset of attributes that preserve the indiscernibility and henceforth induce the same partitioning of the universe are identified and called reducts. When applied on a spatial decision system, Bai & Ge (2009) say it is possible that some spatially relevant attributes may be removed while determining the reducts. It is comprehensive to treat spatial and non-spatial attributes separately. While most of the spatial attributes like temperature, rainfall, humidity, altitude etc., are continuous, using Neighborhood Rough Sets to identify knowledge granules will be accomplishing without the loss of information due to discretization.

Related Content