Chapter 31
Neighborhood Rough-Sets-Based Spatial Data Analytics

Sharmila Banu K
VIT University, India

B. K. Tripathy
VIT University, India

ABSTRACT

Rough set theory partitions a universe using single-layered granulation. The equivalence classes induced by rough sets are based on discretized values. Considering the fact that the spatial data are continuous at large, discretizing them may cause loss of data. Neighborhood approximations can lead to closely related coverings using continuous values. Besides, the spatial attributes also need to be given due consideration and should be handled unlike non-spatial attributes in the process of dimensionality reduction. This chapter analyzes the use of neighborhood rough sets for continuous data and handling spatially correlated attributes using rough sets.

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

DOI: 10.4018/978-1-5225-7598-6.ch031

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 approximation 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.

Figure 1. A sample geographic region