Chapter 58
Rough Fuzzy Set Theory and Neighbourhood Approximation Based Modelling for Spatial Epidemiology

Balakrushna Tripathy
VIT University, India

Sharmila Banu K.
VIT University, India

ABSTRACT

Modern epidemiological studies involve understanding individual and social level inferences and their role in the transmission and distribution of disease instances. The geographic relevance in epidemiology has been analysed in concurrence with these inferences. The substantial amount of data involved in an epidemiological study is usually very large and intuitively involves missing values and uncertainty. Rough Set Theory (RST) has been used in medical informatics for 'outcome prediction' and 'feature selection'. It can be used to construct the decision system involving spatial, medical and demographic data effectively. This chapter proposes the use of rough sets in conjunction with parallel techniques like Fuzzy sets, Intuitionistic systems and Granular (Neighborhood Approximation) computing for the classic problem of data representation, dimensionality reduction, generation and harvest of minimal rules. RST handles missing values and uncertainty more specific to spatial and medical features of data.

BACKGROUND

Recent and past literatures have documented the relationship between locations, individuals and diseases. Geographic Information Systems (GIS) have been widely used to study problems involving public health. Spatial analysis with respect to epidemiology has been addressed in recent researches. Transmission and distribution of SARS - severe acute respiratory syndrome was studied and analysed by Meng et al. (2002), Wang (2006) documented risk exposure pattern, Ulegtekin et al. (2007) analysed distribution

DOI: 10.4018/978-1-5225-1674-3.ch058

of measles in Turkey, Slowinski et al. (1996) predicted *pancreatitis* using Rough Set Theory, Rowland et al. (1998) predicted ambulation after spinal cord injuries, Vinterbo and Øhrn (1999) built a rough set based predictor for myocardial infarction, Bai et al. (2010) used RST to uncover spatial decision rules in neural-tube birth defect. Spatial analysis employing statistical models and spatial regression methods to study population dynamics is reported in Chi and Zhu (2008) and the use of weighted centroid method to predict outbreak of *Escherichia Coli* in Buscema et al (2013). The results have depended on specific features of dataset like configuration, distribution, spatial heterogeneity and autocorrelation. Bai et al. (2010) substantiate that being discernibility based, ability to handle inconsistent data, applicability to any number of outcomes, dimensionality reduction, suitability for spatial data are some of the features that make Rough Sets very conducive to epidemiological study.

To better express the multifaceted nature of the real world and address the limitation of knowledge and uncertainty of factual data, *fuzziness* can be used to represent some attributes of data. It has been used to represent the classification of land-cover types in Shi (2005) and effect of environmental factors on birth defects in Bai et al. (2010). A geographic phenomenon may tend to be closely related and distant related entities based on the distance. This is spatial auto correlation and upheld by Tobler's *law of geography* as in Miller (2004). In RST, an object tends to have *roughness* where the object is a subset of universe with some property states Pawlak (1984). Lower and Upper approximations are used to define an object. The roughness of an object can be précised upon collecting more attributes about the object. Bai et al. (2014) affirm that *roughness* is not a *fuzzy* concept by nature and so fuzzy sets cannot be used to represent roughness *Rough Fuzzy Sets* which is an extension of rough sets can be used to construct the decision system for spatial analytics. Combining *Intuitionistic* approach along with rough fuzzy sets will tend to better accuracy of results leading to crisp conditions and probability based fuzzy decisions.

Dimensionality reduction which is also addressed by RST needs an extra step on dealing with spatial and non-spatial attributes of the decision system. Spatial attributes which are *continuous* in nature will have to be discretised for RST to construct equivalence classes. Jensen and Shen (2004) approve that the discretization may sometimes lead to loss of information. Liao (2012) substantiated the use of *Neighborhood Rough Set approximation* to work with continuous attributes without discretising them.

Neighbourhoods are defined using nearest neighbour methods or by distance from central point to boundary. Using the this method, for a given set of continuous attributes in space, a neighbourhood is defined for every object in *U,* as $\delta\left(x\right) = \{x \mid x \in U, d(x, x1 \leq \delta)$ where δ>0 and δ(*x*) is δ neighbourhood information granule of *x*. Neighborhood approximation is applied to identify the positive and boundary region. The set of objects in the positive region are identified into decision classes without ambiguity. Non-spatial attributes can be reduced using RST reducts. The minimal reduct induced can be combined with the attributes determined using Neighbourhood Rough Sets from which the rules can be inferred. Treating the spatial and non-spatial attributes separately and integrating them later will uphold the inherent spatial features critical to the semantics of rules.

This chapter will discuss Rough Fuzzy Intuitionistic Decision System for identifying the Spatial Distribution of Disease Instances involving demographic, medical, continuous and auto correlated spatial attributes for spatial data based on Neighborhood Rough Sets. However, the temporal nature of medical data is to be accounted for. Any errors due to missing or inconsistent data should be addressed with the construction of error matrix. The rules generated should be verified for being minimal, correctness and accuracy.