Chapter XX
Rough Sets and Granular Computing in Geospatial Information

Iftikhar U. Sikder
Cleveland State University, USA

ABSTRACT

The representation of geographic entities is characterized by inherent granularity due to scale and resolution specific observations. This article discusses the various aspects of rough set-based approximation modeling of spatial and conceptual granularity. It outlines the context and applications of rough set theory in representing objects with intermediate boundaries, spatial reasoning and knowledge discovery.

GRANULAR COMPUTING

Granular computing is an emerging paradigm in the computing world. It deals with representation of information in the form of some aggregates and their processing (Pedrycz, 2001; Y. Y. Yao, 2001). According to Zadeh's definition (Zadeh, 1996) "- *information granulation involves partitioning a class of objects (points) into granules, with a granule being a clump of objects (points) which are drawn together by indistinguishability, similarity, or functionality.*" Instead of focusing on machine-centric approach to information, granular computing emphasizes human-centric perception in terms of formation of abstract concept in representing real world. By selectively focusing on individual granules of information at different level of representations one can gain greater understanding of the inherent knowledge structure and problem solving in implementation of intelligent systems. The theoretical foundation of granular computing involves formal approach to interval mathematics including fuzzy sets, rough sets, shadowed set, and random sets. Typical application areas of granular computing include soft computing systems such as fuzzy systems (Pedrycz & Vukovich, 1999) quotient space theory (Zhang & Zhang, 2004), machine learning from example (Yao & Yao, 2002), neural networks (Zhang *et al*., 2000), rough logic and rough set theory (Liu & Jiang, 2002).

GRANULATION IN SPATIAL INFORMATION

In geographic data information granulation is a central aspect of observation or abstract representation of geographic entities. The representation of space and spatial relationships in terms of idealized abstraction of geometric primitives (e.g., polygons, lines and points or pixels) and topological constraints inherently presupposes granulation of the space-time continuum. Spatial categories essentially represent hierarchies of abstractions. In spatial classification, the thematic classification systems are fixed by a well-defined crisp boundary or abstract concept in the context in which the data set was generated at a particular hierarchy. By establishing a coarse view of the world, it is possible to deal with large information granules such as states, districts, and counties. For example, in raster data model a regular grid is used to cover space where each grid cell uniformly represents spatial characteristics of local phenomena. Since the spatial variations of the local features within a grid are indistinguishable, a granular representation is implicit in a raster data model. The functional aspects of information processing also vary with the degrees of granularity. At finer details, the image processing tasks involve image segmentation, edge detection, and noise removal. While at higher end abstractions, computing algorithms are concerned with automated feature recognition, semantic interpretation, etc. The effect of granularity is also reflected in geostatistical prediction, which is classically known as the *modifiable areal unit problem* - a geographic manifestation of ecological fallacy causing spurious conclusion due to spatial aggregation (Openshaw, 1984).

In principle, granularity in an observation makes objects *indiscernible,* i.e., indistinguishable from each other. The indiscernibility relation can lead to the equivalence *relation*, which implies that objects within an equivalence class are *reflexive* (i.e., indistinguishable from itself), *symmetric* (i.e., if object *x* is indiscernible from *y*, then y is indiscernible from *x*), and transitive (if *x* is indiscernible from *y* and *y* is indiscernible from *z* then *x* is indiscernible from *z*). The notion of indiscernibility and the resulting equivalence class allows expressing concepts in terms of approximation space. However, the transitivity constraint may not always be maintained in spatial relations, for example, when the discernibility relation is a distance measure. Hence in those cases one may have to use the *similarity or tolerance relation* (i.e., objects are reflexive and symmetric).

ROUGH SET THEORY

While granularity leads to partitioning of the universe, it is obvious that some concept may not be well defined by the partition. Rough set theory (Pawlak, 1991) was introduced to characterize a concept in terms of elementary sets in an approximation space. Information granulation or the concept of indiscernibility is at the heart of rough set theory. A finer granulation means a more definable concept. Spatial categories can be represented in the form $U, C \cup \{d\}$), where $d \notin C$ is the decision attribute or the thematic feature such as "forest" in a given location and U is the closed universe which consists of a non-empty finite set of objects (a number of spatial categories) and C is a non-empty finite set of attributes that characterizes a spatial category i.e., "forest" such that $c : U \to V_c$ for every $c \in C$, V_c is a value of attribute c. For $P \subseteq C$ the granule of knowledge about a forest with respect to indiscernibility relation can be represented as: