Spatial and Spatial Temporal Data Mining: Accomplishments and Challenges

Shashi Shekhar
University of Minnesota
Latifur Khan and Bhavani Thuraisingham
University of Texas at Dallas

ABSTRACT

The importance of spatial and spatio-temporal data mining is growing with the increasing availability and awareness of large geo-spatial datasets such as maps, repositories of remote-sensing images, and the decennial census. Applications include location-based services, environmental studies (e.g. global change, climatological effects of El Nino, land-use classification from satellite imagery), public health (predicting and controlling spread of disease), public safety (finding crime hot spots), transportation (vehicle navigation, site selection, fuel efficient routes), etc.

However, classical data mining techniques often perform poorly when applied to spatial and spatio-temporal data sets for many reasons. First, spatial data is embedded in a continuous space, whereas classical datasets discrete notions like transactions. Second, a common assumption about independence of data samples in classical statistical analysis is generally false because spatial data tends to be highly auto-correlated. Ignoring autocorrelation in spatial and spatio-temporal data may hypotheses or models that are inaccurate or exhibit residues with auto-correlation. Third, these datasets have a multi-jurisdiction multi-temporal (MJMT) nature and thus exhibit a high degree of heterogeneity.

Other challenges include edge effects, categorization of spatial and spatio-temporal patterns, interest measures to quantify those, and design of computationally efficient and scalable algorithms to mine their instances.

Our talk will present a survey of the new spatial and spatio-temporal data mining methods including those for discovering spatial co-locations, detecting spatial outliers and location prediction. It also presents a gap analysis towards identify hard open computational problems, which represent critical barriers to the progress of geographic information science application domains like environment criminology.