Abstract

Spatio-temporal data mining, defined as identification of interesting, useful and non-trivial patterns from large spatio-temporal databases, has huge application in various domains, including environmental management, transportation, epidemiology, climatology and so on. Several models, taking care of the underlying physics in such domain specific problems, have been proposed till date. However, the two major limitations in such physics-driven models are firstly, these assume that all the physical systems are well understood, which is not true in reality; secondly, these models are computationally inefficient, requiring lots of computational power. Therefore, recently, the *data-driven approaches* have emerged as a new paradigm in this regard, with an aim to extensively and efficiently analyze historical data for generating insights, and utilize those in further studies.

The focus of our research is to explore *data-driven modeling* for *spatio-temporal* (ST) *prediction* of *time series data*. In this regard, we have proposed two separate prediction frameworks to deal with: i) *spatial time series data, abundant in time*, and ii) *spatial time series data, abundant in space*, respectively. The objective is to extend various *computational intelligence* (CI) techniques to plug into these frameworks for better modeling of the spatio-temporal dependency, and thereby deriving various data-driven models to deal with different contexts of ST prediction. Our aim is to make use of wealth of observed as well as simulated data, and thereby to enhance data-driven modeling as a complement for the physics-driven approaches.

The major contributions in this thesis are as follows: (1) proposing data-driven model, based on *Bayesian network with residual correction* (BNRC), for ST prediction under scarcity of influencing variables; (2) proposing data-driven model, based on *spatial Bayesian network* (SpaBN), for ST prediction under profusion of influencing variables; (3) proposing spatially explicit data-driven model, based on a *semantic Bayesian network* (semBnet), for ST prediction with incorporated domain knowledge; (4) proposing Deep-STEP as a data-driven model for ST prediction of *large-scale raster time series data*; and (5) proposing NFBN as a *new variant* of *fuzzy Bayesian network* to reduce parameter uncertainty in discrete Bayesian network learning. Comparative study with the benchmark and state-of-the-art prediction techniques has demonstrated the efficacy and superiority of our proposed data-driven models in ST prediction. The overall study has been made considering the *spatial time series data* from the domains of *climatology/meteorology*, *hydrology*, and *remote sensing*. However, the proposed approaches are generic enough to be applied for predicting spatial time series from other domains as well. In future, we like to explore data-driven modeling for spatio-temporal change pattern mining and large-scale data analysis. We have also a plan to combine both the physical and the data-driven approaches, for developing theory-guided data-driven models for ST prediction.

Keywords. *Spatio-temporal prediction, Data-driven approach, Spatial time series, Computational Intelligence, Bayesian network, Fuzzy set, Deep learning*