DOCTORAL THESIS

Learning Complex Spatio-Temporal Dependency: Model Design, Information-Theoretic Analysis, and Systematic Validation

TAN, Qi

Date of Award:
2021

Abstract

Learning the intrinsic dependency among heterogeneous data is one of the most challenging problems in predictive spatio-temporal analytics (PSTA) tasks because the dependency to be learned generally manifests, interacts, and integrates in a complex way at varying scales within and/or across diverse data sources, which renders characterization of such underlying dependency extremely difficult. To address this challenge, a key question must be answered: Given a specific learning task and the corresponding dataset, how should one design an appropriate learning model and theoretically determine its desired configuration by analyzing its learning behavior and quantifying its learning capacity so that the useful information contained in the data can be effectively extracted to yield outstanding predictive performance?

Existing methods, including classical time series models, representative tensor-based learning models, and state-of-the-art deep neural network models, have shown their effectiveness in capturing specific types of dependency, such as short-range temporal dependency or homogeneous/aligned data dependency. However, these methods do not explicitly address the critical issue of multi-scale dependency modeling for heterogeneous data, and their performance thus presents certain limitations. More importantly, even though the existing deep learning models have achieved state-of-the-art performance in certain learning tasks, and earlier studies have examined the behaviors of learning models from the perspective of optimization or representation learning, they are incomplete. Specifically, they failed to provide an in-depth understanding of why a specific deep learning model works well on a given dataset and of the relationship between the learning model's configura-

tion and its corresponding model capacity on the given task and dataset. Without such a clear understanding, it remains a mystery how to determine the desired configurations of a certain deep model with respect to the given learning data sets, thus making it difficult to sustain the success of deep learning.

To address these unsolved yet challenging issues in a theoretically sound and practically feasible way, in this thesis, we aim to systematically answer four research questions:

1. How can we design a learning model to characterize the complex multi-scale dependency of heterogeneous data in PSTA?

2. How can we analyze and quantify the designed model's capacity to capture the multiscale dependency of spatiotemporal data?

3. How can we validate the learning behavior and performance of the designed model in various scenarios of multi-scale spatio-temporal dependency?

4. How can we make the designed model practically applicable?

To answer the first question, we develop and demonstrate a novel interactively and integratively connected deep recurrent neural network (I^2DRNN) model to capture the multi-scale dependency in heterogeneous spatio-temporal data. The proposed I^2DRNN comprises three key modules: (i) an Input module that integrates data from heterogeneous sources; (ii) a Hidden module that captures information on various scales while allowing the information to flow between layers in an interactive manner; and (iii) an Output module that models the integrative effects of information from various hidden layers to generate the output predictions. By integrating these modules, the I^2DRNN can model the integrative effects of various scales of spatio-temporal data within and/or across diverse factors, as observed from heterogeneous sources.

To answer the second question, we propose an information-theoretic framework. This framework enables us to theoretically analyze the learning behavior of I^2DRNN, quantitatively characterize the information-based model capacity (i-CAP) of each of its compo-

nents in terms of capturing the multi-scale spatiotemporal information, and appropriately determine its necessary and sufficient configurations with respect to a given dataset. With these information-theoretic guarantees, the developed framework serves as rigorous and explainable guidance in the design of a desirable deep architecture for a given task.

To answer the third question, we conduct a series of experiments with both synthetic datasets and real-world PSTA tasks to validate the I^2DRNN model and to confirm its information-theoretically described behavior. The experimental results show that the I^2DRNN model outperforms both classical and state-of-the-art models on all datasets and PSTA tasks. Furthermore, as readily validated, the proposed model captures multi-scale spatio-temporal dependency, which is meaningful in the real-world context. More importantly, the model configuration that corresponds to the best performance on a given dataset always falls within the range between the necessary and sufficient configurations, as described by the information-theoretic analysis.

To answer the fourth question, we tackle two practical issues in PSTA: incomplete data imputation before multi-scale dependency learning and hidden interaction mining after multi-scale dependency learning. To infer the missing data for PSTA, we propose a heterogeneous neural metric learning method to restore the data integrity by referring to heterogeneous data sources. We examine the proposed method in a real-world spatio-temporal analytics task, malaria risk mapping, and show that the proposed method produces accurate and reliable risk inference. To characterize the hidden interaction among the spatio-temporal data, we develop two structure-aware methods to uncover the primitive motif prior and the mesoscale connection structure of the underlying interaction network, respectively. Our results with various datasets validate the effectiveness of the proposed methods in capturing the hidden interactions for PSTA.

Keywords: Multi-Scale Spatio-Temporal Dependency Learning, Predictive Spatio-Temporal Analytics (PSTA), Interactively- and Integratively-Connected Deep Recurrent Neural Network (I^2DRNN), Information-Theoretic Analysis, Incomplete Data Imputation, Hidden Interaction Characterization

Table of Contents

List of Tables

List of Figures