

Chapter 3.6

Spatio–Temporal Prediction Using Data Mining Tools

Margaret H. Dunham

Southern Methodist University, Texas, USA

Nathaniel Ayewah

Southern Methodist University, Texas, USA

Zhigang Li

Southern Methodist University, Texas, USA

Kathryn Bean

University of Texas at Dallas, USA

Jie Huang

University of Texas Southwestern Medical Center, USA

ABSTRACT

The spatio-temporal prediction problem requires that one or more future values be predicted for time series input data obtained from sensors at multiple physical locations. Examples of this type of problem include weather prediction, flood prediction, network traffic flow, and so forth. In this chapter we provide an overview of this problem, highlighting the principles and issues

that come to play in spatio-temporal prediction problems. We describe some recent work in the area of flood prediction to illustrate the use of sophisticated data mining techniques that have been examined as possible solutions. We argue the need for further data mining research to attack this difficult problem. This chapter is directed toward professionals and researchers who may wish to engage in spatio-temporal prediction.

INTRODUCTION

Forecasting future values for systems that contain both spatial and temporal features (spatio-temporal) is extremely complex. As an example, consider the problem of predicting precipitation at one location. The amount of previous rainfall in areas close to the target certainly affects this forecast. However, there are many other factors (temperature, time of day, wind direction, wind speed, and so forth) that impact the rainfall prediction. The area of spatio-temporal prediction has been the focus of much research in recent years (Deutsch, & Ramos, 1986; Dougherty, Corne, & Openshaw, 1997; Jothityangkoon, Sivapalan, & Viney, 2000; Kelly, Clapp, & Rodriguez, 1998; Pokrajac, & Obradovic, 2001; Roddick, Hornsby, & Spiliopoulou, 2000; Singh, Chaplain, & McLachlan, 1999). Due to the extreme complexity of predicting these future values, common practice is to utilize domain experts with extensive experience in both forecasting and the problem domain itself. For example, for flood prediction, the *National Oceanic and Atmospheric Administration (NOAA)* actually employs specialists whose job is to understand the history and specifics of predicting floods on one river. A different domain expert may be hired for a different river. Due to the extensive use of domain experts, spatio-temporal prediction is extremely expensive, and due to the complexity of the nature of the problems, prediction accuracy is often low.

In this chapter we argue for more data mining research into the development of sophisticated modeling, machine learning and prediction tools that can assist domain experts in solving spatio-temporal prediction problems. The interesting challenge is to see if these models can use all the data available without the need of expert intervention. This could be useful, since experts in the particular domain of interest – flood prediction, networks – often may not be experts in these data mining tools, which are drawn from statistics, genetic modeling, algorithmic heuristics and

much more. But creating models general enough to be understood by non-experts has proven to be a difficult balancing act, especially when these complex models use many parameters and when there are many types of data to consider.

In this chapter we first introduce the problem and some previous research and solutions. Finally we summarize and make recommendations for future work.

PROBLEM DEFINITION

Spatio-temporal prediction is widely used in many diverse applications, such as environmental protection, flood control, atmosphere surveillance, waste disposal management, traffic management and so on. The general discrete time spatio-temporal prediction problem can be defined as:

Given a collection of spatial locations where time series data is collected at discrete points in time, the spatio-temporal prediction problem is to predict a future value at one of the locations.

Thus, we can state the spatio-temporal prediction problem as one of predicting one or more future values of time series data. Although we could examine a continuous time spatio-temporal prediction, most work has been done looking at a discrete time system where data is collected at regular intervals. For example, the Minnesota Department of Transportation collects freeway data at monitoring stations in the Twin Cities metropolitan area every 30 seconds. (www.dot.state.mn.us/tmc/trafficinfo/data). Example 1 is based on the Tao Dataset of sea surface temperature readings in the Pacific (http://ingrid.ldgo.columbia.edu/SOURCES/IGOSS/TOGA-TAO.cdf/dataset_documentation.html).

Example 1. There are n temperature sensor stations located at buoys at various sites in the ocean. Each station measures the temperature on the surface of the ocean every day at noon. The spatio-temporal prediction problem here would

14 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/chapter/spatio-temporal-prediction-using-data/7705

Related Content

Text Mining Methods for Hierarchical Document Indexing

Han-Joon Kim (2005). *Encyclopedia of Data Warehousing and Mining* (pp. 1113-1119).

www.irma-international.org/chapter/text-mining-methods-hierarchical-document/10763

Data Mining in Practice

Sherry Y. Chen and Xiaohui Liu (2008). *Data Warehousing and Mining: Concepts, Methodologies, Tools, and Applications* (pp. 2273-2280).

www.irma-international.org/chapter/data-mining-practice/7760

Why General Outlier Detection Techniques Do Not Suffice for Wireless Sensor Networks

Yang Zhang, Nirvana Meratnia and Paul Havinga (2010). *Intelligent Techniques for Warehousing and Mining Sensor Network Data* (pp. 136-158).

www.irma-international.org/chapter/general-outlier-detection-techniques-not/39544

An Approach to Mining Crime Patterns

Sikha Bagui (2008). *Data Warehousing and Mining: Concepts, Methodologies, Tools, and Applications* (pp. 2523-2550).

www.irma-international.org/chapter/approach-mining-crime-patterns/7781

Drawing Representative Samples from Large Databases

Wen-Chi Hou, Hong Guo, Feng Yan and Qiang Zhu (2005). *Encyclopedia of Data Warehousing and Mining* (pp. 413-420).

www.irma-international.org/chapter/drawing-representative-samples-large-databases/10633