Preface

Data mining has been an active area of research since the early 1990s. Data mining has evolved from relational data to include time and space constraints. More recently, advances in positioning technology and location-based services have led to a growth of spatio-temporal databases that require advanced data mining capabilities.

Temporal data mining is complex due to the existence of time which imposes an ordering on the temporal events. However, temporal data mining is able to discover the behavioral aspects of objects as opposed to simple mining rules that describe their states at a point in time. Spatio-temporal mining has several challenges due to the complexity of geographical domains, the mapping of data in spatial and temporal frameworks, and the spatial and temporal correlation. Mining in spatio-temporal databases needs to consider the multi-states of the data, and integrate spatial and temporal information together to find meaningful spatio-temporal patterns.

The target audience of this book is students, researchers, and professionals who are interested to work in the area of data mining, data analysis, machine learning, and knowledge discovery in spatial and/or temporal databases. More specifically, the book will be a useful companion for graduate students studying the issues of data mining in spatio-temporal databases, and for instructors who can use the book as a reference for advanced topics in spatio-temporal databases.
To the Data Mining Researchers

This book is designed to give an in-depth discussion on the issues and challenges facing pattern discovery in temporal and spatio-temporal databases. The book consists of two parts: the first, comprising of chapters II to V, is focused on temporal data mining. The second part, comprising of chapters VI to XI, concentrates on spatial-temporal mining. Each chapter is self-contained so that you will be able to focus directly on those chapters that interest you.

To the Graduate Students

This book will provide the necessary background information and the latest research in the area of temporal and spatio-temporal data mining. The materials are presented in a clear and easy to understand manner. The emphasis of the book is on the design and implementation of novel algorithms and techniques used for patterns discovery in temporal and spatio-temporal databases. To enable you to better appreciate these algorithms and techniques, two chapters, namely Chapter II and Chapter VI, are included to introduce the necessary background information and related work in time series mining and spatio-temporal mining respectively.

Organization of the Book

The book is organized as follows. Chapter I provides an overview of the various types of patterns that could be discovered in temporal and spatio-temporal databases. Chapters II to XI can be divided into two parts. The first part, comprising of chapters II to V, is focused on temporal data mining. The second part, comprising of chapters VI to XI, concentrates on spatial-temporal mining. Chapter XII summarizes the book and suggests possible future research directions for researchers and graduate students who would like to explore into this exciting field.

In Chapter II, two types of temporal data are dominant in the development of temporal data mining. They are time-series data and sequence data. Interesting temporal patterns that could be discovered from these two types
of data include periodic patterns and sequential patterns, respectively. We will first discuss several issues in time series mining such as similarity measures, dimension reduction, and data discretization. Then we will examine the existing techniques for mining periodic patterns and sequential patterns in temporal databases.

Periodic patterns are patterns that occur regularly over time in a time series database. Periodic patterns could be full, partial, asynchronous, and surprising. A full periodic pattern such as season cycles for every year requires that every event in the database contributes to part of the pattern. Discovering periodic patterns from an entire time series may not be suitable in applications where the periodic patterns occur only within small segments of the time series. Chapter III presents a class of periodic patterns called dense periodic patterns, and describes an algorithm that efficiently discovers short period patterns that may exist in part of a time series.

Sequence patterns aim to show precedence relationships and ordered associations among events in the sequence. Sequential pattern mining has been a well-studied problem and has become an important in applications such as market and customer analysis. Many techniques have been developed for general sequence pattern mining, constraint-based sequence pattern mining, frequent episode mining, sequence pattern mining in noisy environment, and closed sequence pattern mining. Chapter IV analyzes the I/O cost of one of the state-of-the-art sequence mining techniques called GSP and discusses an I/O conscious algorithm called MFS as well as the incremental versions of GSP and MFS for mining sequence patterns in evolving databases.

We observe that many real-world objects have states that change over time. By tracking the state sequences of these objects, we can study their behavior and take preventive measures before they reach some undesirable states. Chapter V introduces a new kind of pattern called progressive confident rules. This class of pattern captures sequences of states with an increasing confidence that leads to a particular end state. In order to reduce the search space, we devise new pruning strategies and employ the concise set analysis of rules in the mining of progressive confident rules. We also demonstrate how progressive confident rules can be incorporated in existing classifiers to improve their accuracy.

Besides techniques to discover time-varying patterns, there has been increasing interest to extract patterns from spatio-temporal databases. While there has been much research on association rule mining on transactional, spatial, and temporal data, there is little literature on finding interesting associations in spatio-temporal data. Chapter VI first examines the early works in
spatio-temporal mining which have been concentrated on finding evolution patterns of natural phenomena, frequent movements of objects over time, and space-time clusters. Then we give a review of the traditional association rules mining algorithms and their variants on transactional data, temporal data, and spatial data.

Chapters VII, VIII, and IX describe new classes of patterns that can be found in spatial-temporal data and algorithms to efficiently mine these patterns. We investigate the discovery of interesting spatio-temporal patterns from two aspects. First, we impose temporal constraints on the mining of spatial collocation patterns to discover topological patterns. Such patterns can reveal the intra-relationships of events in a time period. These events may relate to each other in a star, clique, or star-clique manner within the given time window. Chapter VII deals with topological patterns discovery.

Next, we search for spatial sequence patterns by incorporating spatial information into the process for mining sequence patterns. We will introduce two classes of spatial sequence patterns, called flow patterns and generalized spatio-temporal patterns. These two classes of spatial sequence patterns are useful to the understanding of many real-life applications. Such patterns can disclose the inter-relationships of events in different time windows, for example, how the observation of one event in some location implies the occurrence of another event in a second location, or how changes of events in one location can affect the events in another location. The discovery of spatial sequence patterns is challenging because of the potentially large search space and the large number of candidates. Chapters VIII and IX will describe efficient methods to discover flow patterns and generalized spatio-temporal patterns, respectively.

With appropriate models, a spatio-temporal database can be transformed into a tree or graph database, and the problem of mining frequent spatio-temporal patterns then becomes the problem of finding frequent trees or subgraphs. Chapter X and XI present efficient and scalable methods for mining arbitrary complex spatio-temporal association patterns, in the form of tree patterns and graph patterns, respectively. We note that changes in spatio-temporal databases can cause changes to the graph structures that model the relationships in the spatio-temporal data. Re-execution of the mining algorithm each time the graphs are updated is costly, and may result in an explosion in the demand for computational and I/O resources. Chapter XI also describes an incremental algorithm to handle such situations.

Finally, we conclude in Chapter XII with a summary of the contributions in each chapter and provide directions for future research.