Knowledge Discovery in Spatial Data

Bearbeitet von
Yee Leung

1. Auflage 2012. Taschenbuch. xxix, 360 S. Paperback
ISBN 978 3 642 26170 1
Format (B x L): 15,5 x 23,5 cm
Gewicht: 599 g

Preface

When I first came across the term data mining and knowledge discovery in databases, I was excited and curious to find out what it was all about. I was excited because the term tends to convey a new field that is in the making. I was curious because I wondered what it was doing that the other fields of research, such as statistics and the broad field of artificial intelligence, were not doing. After reading up on the literature, I have come to realize that it is not much different from conventional data analysis. The commonly used definition of knowledge discovery in databases: "the non-trivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data" is actually in line with the core mission of conventional data analysis. The process employed by conventional data analysis is by no means trivial, and the patterns in data to be unraveled have, of course, to be valid, novel, useful and understandable. Therefore, what is the commotion all about? Careful scrutiny of the main lines of research in data mining and knowledge discovery again told me that they are not much different from that of conventional data analysis. Putting aside data warehousing and database management aspects, again a main area of research in conventional database research, the rest of the tasks in data mining are largely the main concerns of conventional data analysis. Model identification, model construction, and discovery of plausible hypotheses, for example, are not unique to data mining. They are, in addition to model estimation and hypothesis testing, in the agenda of conventional data analysis, such as statistics, also. Searching for clusters, looking for separating surfaces or rules for classification, mining for association rules and relationships, and detecting temporal trends or processes in data constitute the core of the knowledge discovery process that is not unique to data mining. They form the backbone of research in conventional data analysis. From this perspective, there is very little novelty in data mining and knowledge discovery.

On the other hand, if we look at the environment within which data mining and knowledge discovery is taking place, there is something genuine that is worthy of our attention. Though we have traditionally looked for patterns in data by performing clustering, classification, relational analysis, and trend or process analysis, the kinds of data that we are dealing with nowadays are quite different from

that targeted by conventional data analysis methods. The sheer volume and complexity of the data that we need to handle nowadays are substantially different from that of the past. Effective discovery of knowledge hidden in data requires novel methods for accomplishing the old tasks. Therefore, it is from this perspective that the mission of data mining and knowledge discovery is justified. This field of research can actually be treated as the continuation of the mission of conventional data analysis into the information and knowledge age. And, our main objective is simply to discover knowledge in data as we have always been doing. Nevertheless, I have no problem in using the term data mining and knowledge discovery adopted by the research community as long as we know exactly what we are doing.

Following up on the literature, I also encountered the term spatial data mining and knowledge discovery. A natural question again is what it is all about, and how it is different from data mining and knowledge discovery in general. An examination of the research activities in this area again tells me that in principle it is more or less similar to that of the general field. The major difference is that data in spatial data mining are mostly geo-referenced and much more complex. And the knowledge to be discovered is often location specific and takes on geometric shapes. Space and time are the two main dimensions along which knowledge discovery is performed. Thus, there is something unique in spatial data mining and knowledge discovery that is worth looking into.

Our main goal is then to discover knowledge in spatial data. It is again in line with conventional spatial data analysis but with special emphasis placed on the nature of spatial data. The idea is to develop novel methods for spatial knowledge discovery. Whether we should call such process spatial data mining and knowledge discovery, or just simply discovery of knowledge in spatial and temporal data is just a matter of terminology. It all involves the discovery of spatial structures, processes, and relationships from spatial and temporal data. As data mining and knowledge discovery have become a commonly employed collective term for such activities, it is used indiscriminately throughout this book. I would not painstakingly point out whether a method should be called a data mining and knowledge discovery method, or just a data analysis method targeting the unraveling of structures, processes and relationships in voluminous and complex spatial and temporal databases.

As a good number of text books and research monographs have been written on data mining and knowledge discovery, one needs a good justification to write another book on the topic. Given the unique features of knowledge discovery in spatial data and the burgeoning growth of research interest in this area, it is an opportune time to make a critical analysis of the field and explore directions for further research. Instead of repeating what has been written in many current books on data mining and knowledge discovery, I would like to write it from the perspective of my own research in this area. So, it is not a text book on data mining and knowledge discovery. It is not a book, like many others, that discusses all aspects of the knowledge discovery process. So there are no discussions on topics such as data warehousing, on-line analytical processing (OLAP), data query, and data mining software. There is no intention to give a comprehensive survey of the

literature of the field, although state-of-the-art reviews under relevant topics are made in the book.

This book is intended to be a research monograph on methods and algorithms, conventionally called data mining methods, for the discovery of knowledge in spatial and temporal data. The majority of the methods discussed are based on our own research. So, when I discuss topics such as clustering, classification, relationships and temporal processes, algorithms in the literature are not discussed in detail. Emphasis is placed on the development of our own methods. Nevertheless, it is not difficult to see that some of our methods can, more or less, fit into the family of research methodologies on the same topics. They are developed on the foundation of mathematics, statistics, and artificial intelligence. In brief, the present monograph is not a text book for spatial data mining and knowledge discovery. It is a book for researchers and advanced graduate students who are interested or might have an interest in the methodologies for the discovery of knowledge in spatial and temporal data. The view is more personal, but it fits in with the overall picture of research in the field.

Yee Leung

http://www.springer.com/978-3-642-02663-8

Knowledge Discovery in Spatial Data

Leung, Y.

2009, XXIX, 360 p. 113 illus., Hardcover
ISBN: 978-3-642-02663-8