Big Data Management

The big data paradigm presents a number of challenges for university curricula on big data or data science related topics. On the one hand, new research, tools and technologies are currently being developed to harness the increasingly large quantities of data being generated within our society. On the other, big data curricula at universities are still based on the computer science knowledge systems established in the 1960s and 70s. The gap between the theories and applications is becoming larger, as a result of which current education programs cannot meet the industry's demands for big data talents.

This series aims to refresh and complement the theory and knowledge framework for data management and analytics, reflect the latest research and applications in big data, and highlight key computational tools and techniques currently in development. Its goal is to publish a broad range of textbooks, research monographs, and edited volumes that will:

- Present a systematic and comprehensive knowledge structure for big data and data science research and education
- Supply lectures on big data and data science education with timely and practical reference materials to be used in courses
- Provide introductory and advanced instructional and reference material for students and professionals in computational science and big data
- Familiarize researchers with the latest discoveries and resources they need to advance the field
- Offer assistance to interdisciplinary researchers and practitioners seeking to learn more about big data

The scope of the series includes, but is not limited to, titles in the areas of database management, data mining, data analytics, search engines, data integration, NLP, knowledge graphs, information retrieval, social networks, etc. Other relevant topics will also be considered.

More information about this series at https://link.springer.com/bookseries/15869

Lizhen Wang • Yuan Fang • Lihua Zhou

Preference-based Spatial Co-location Pattern Mining

Science Press
Beijing

Springer

Lizhen Wang
School of Information Science and Engineering
Yunnan University
Kunming, China

Yuan Fang
School of Information Science and Engineering
Yunnan University
Kunming, China

Lihua Zhou
School of Information Science and Engineering
Yunnan University
Kunming, China

ISSN 2522-0179 ISSN 2522-0187 (electronic)
Big Data Management
ISBN 978-981-16-7565-2 ISBN 978-981-16-7566-9 (eBook)
https://doi.org/10.1007/978-981-16-7566-9

Jointly published with Science Press
The print edition is not for sale in China (Mainland). Customers from China (Mainland) please order the print book from Science Press.

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Foreword

With the continuous improvement in Global Positioning System (GPS) accuracy, our location-based services have already gone deeply into all aspects of people's life, and the data containing spatial location information is increasing. In the face of the continuous growth of massive spatial data, people are finding it difficult to fully understand the data without knowledge assistance, and so spatial data mining technology has emerged. Spatial co-location pattern mining is an important branch of the field of spatial data mining. By mining the spatial co-location patterns, we can find interesting relationships between spatial features and play a positive guiding role in various location-based application domains.

Massive spatial data brings new challenges to spatial co-location pattern mining. How can we find the compressed or condensed representation of spatial patterns from a large number of mining results? How can we design appropriate preference constraints for users in spatial co-location pattern mining? How do we derive some pruning strategies from the preference constraints to improve the efficiency of the preference-based mining algorithms? This monograph answers these questions quite well.

Professor Wang's research team has systematically and continuously researched spatial co-location pattern mining since 2008. In particular, fruitful and pioneering research on preference-based spatial co-location pattern mining has been conducted. Many of their results have been published in internationally renowned journals and conferences and have been accepted by the scholars in the world, of course, including myself.

This monograph studies a series of preference-based pattern mining techniques, including: maximal prevalent co-location pattern mining, maximal sub-prevalent co-location pattern mining, SPI-closed co-location pattern mining, top-k prevalent co-location pattern mining, dominant co-location pattern mining, non-redundant co-location pattern mining, high-utility co-location pattern mining, interactive mining, the similarity measurement between spatial co-location patterns, etc. The research results are systematic and promising.

Throughout the monograph, the author gives a clear explanation of the motivation for the research, the definition of the problem, the key research, the statement and analysis of the method, and the application evaluation, etc. It is believed that this monograph may serve researchers and application developers in spatial data mining technology and related fields and may help them explore this exciting field and develop new methods and applications. It may also provide the current status of this promising latest research theme, both for graduate students and other interested readers.

I find the monograph is enjoyable to read and fully recommend this monograph to you.

Renmin University of China, Beijing, China
July, 2021

Xiaofeng Meng

Preface

The development of information technology has enabled many different technologies to collect large amounts of spatial data every day. It is of very great significance to discover implicit, non-trivial, and potentially valuable information from this spatial data. Spatial co-location patterns expose the distribution rules of spatial features, and their discovery can be of great value to application users. This book is intended to provide commercial software developers with proven and effective algorithms for detecting and filtering these implicit patterns, and easily-implemented pseudocode is provided for all our algorithms. The book is also intended to provide a base for further research in such a promising field. We take the demand from user applications and provide a mathematical and systematic study of preference-based spatial co-location pattern mining.

Spatial co-location pattern mining has broad prospects for spatial data owners. Potential economic value has been found in mining the data from Earth science, public safety, biological information processing, location-based personalized recommendation, geo-information system (GIS), and military strategy planning. We know it has aroused strong interest amongst researchers, both at research institutions (Google, Microsoft, IBM, etc.) and at universities (Stanford University, University of Washington, University of Minnesota, Hong Kong University of Science and Technology, etc.). High-level research results have emerged in authoritative international data engineering journals such as "IEEE Transactions on Knowledge and Data Engineering (TKDE)" and at top academic conferences such as "ACM SIGSPATIAL, ACM SIGKDD, and IEEE ICDE."

The authors' research team began conducting research on spatial pattern mining in 2008. Since then it has been involved with 5 National Natural Science Fund projects on spatial co-location pattern mining, has trained 7 doctoral students and over 80 master students, and has published more than 90 relevant academic papers. We have been given prizes from the Yunnan Science and Technology Award. The research team led by the first author is called the innovation team of the "Yunnan Province Spatial Big Data Mining and Decision Support Research Group," the first provincial innovation team of our college.

Similar to frequent item set mining in transaction databases, spatial co-location pattern mining often generates a huge number of prevalent co-location patterns, but only a few of them satisfy user interests. User preferences are often subjective, and a pattern preferred by one user may not be favored by another, and so cannot be measured by objective-oriented *prevalence* measures. Therefore, the following challenges and/or issues will be answered in this book:

- What is a natural criterion for ranking patterns and presenting the "best" patterns?
- How to find a condensed representation of spatial patterns from a huge number of mined results.
- How to enable a user to have proper constraints/preferences in spatial co-location pattern mining.
- How to derive pruning properties from the constraints which improve the efficiency of the corresponding preference-based mining.

It is essential to study the theories and algorithms of preference-based co-location pattern mining in order to solve these challenges and issues. Preference-based co-location pattern mining refers to mining *constrained* or *condensed* co-location patterns instead of mining *all* prevalent co-location patterns. Specifically, this book includes problems such as *maximal* co-location pattern mining, *closed* co-location pattern mining, *top-k* co-location pattern mining, *non-redundant* co-location pattern mining, *dominant* co-location pattern mining, *high utility* co-location pattern mining, and *user-preferred* co-location pattern mining.

For the above problem areas, this book details the relevant research, the basic concepts, the resulting algorithms and their analysis, with experimental evaluation of each algorithm. These techniques come from the latest results of our research in recent years. For the convenience of readers, the chapters of this book are as integrated as possible, so the reading order of each chapter is quite flexible. You can find the corresponding chapter reading according to your interest. Of course, if you read the book from beginning to end, you will find a small amount of repetitive information intended to guarantee the relative independence of each chapter, but absolutely not information redundancy, as we describe the same content in different forms if possible. Indeed, readers are encouraged to read and study this book in order.

This book can be used both as a textbook for learners and as a good reference for professionals.

Although many people have contributed to this book, we first express our gratitude to our families. Without their encouragement and support, it would have been impossible to finish this book, and so this book is dedicated to them.

Secondly, we should sincerely thank Roger England, a colleague who helped correct the English of the book and gave a lot of valuable comments. We would also particularly like to thank Professor Xiaofeng Meng of Renmin University of China for his guidance and help. After reading the first draft of the book, he not only gave specific comments, but also gladly provided a foreword for the book; we would also like to thank Ms. Xin Li of Science Press of China and Mr. Wei Zhu of Springer. Their efforts have facilitated the smooth publication of the book.

Thanks must also go to the National Natural Science Foundation Committee and the Yunnan Provincial Department of Science and Technology for their long-term project funding (Nos.: 61966036, 61472346, 61272126, 61662086, 61762090, 2018HC019). Without their funding for specific research objectives, it would have been difficult to construct the systematic and forward-looking research results which this book conveys.

This book involves an academic discipline frontier, so numerous references have been given in the various chapters, but here we would like to particularly thank S. Shekhar, Y. Huang, and J.S. Yoo for their pioneering work in the relevant fields.

In the research work, although the authors invested a lot of effort, and in the writing of the book each chapter and every sentence has been carefully checked, although limited to our research depth and knowledge level, errors in the book are probably inevitable and we welcome the reader's criticisms and corrections.

Kunming, China
July 2021

Lizhen Wang
Yuan Fang
Lihua Zhou

Contents