Synthesis Lectures on Computer Science

SYNTHESIS
COLLECTION OF TECHNOLOGY

Lei Zhu · Jingjing Li ·
Zheng Zhang

Dynamic Graph Learning for Dimension Reduction and Data Clustering

Springer

Synthesis Lectures on Computer Science

The series publishes short books on general computer science topics that will appeal to advanced students, researchers, and practitioners in a variety of areas within computer science.

Lei Zhu · Jingjing Li · Zheng Zhang

Dynamic Graph Learning for Dimension Reduction and Data Clustering

Springer

Lei Zhu
Shandong Normal University
Jinan, Shandong, China

Jingjing Li
University of Electronic Science
and Technology of China
Chengdu, Sichuan, China

Zheng Zhang
Harbin Institute of Technology
Shenzhen, Guangdong, China

ISSN 1932-1228 ISSN 1932-1686 (electronic)
Synthesis Lectures on Computer Science
ISBN 978-3-031-42312-3 ISBN 978-3-031-42313-0 (eBook)
https://doi.org/10.1007/978-3-031-42313-0

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Paper in this product is recyclable.

We dedicate this book to the passionate researchers who devote their efforts to the fields of dimension reduction and data clustering.

Preface I

In the present era, the exponential growth of diverse datasets has underscored the increasing significance of data mining and analysis. However, as data dimensions expand and data annotation becomes more challenging, the complexity of data mining and analysis also intensifies. In response to this issue, researchers have proposed numerous methods for dimension reduction and data clustering. Among these methods, graph-based dimension reduction has emerged as a particularly effective approach due to its enhanced adaptability to data distributions. By constructing similarity relation graphs between data points, graph-based methods can achieve dimension reduction and clustering, thereby showcasing promising performance while retaining the structural information of the original data.

This groundbreaking book represents the first comprehensive exploration of dynamic graph learning. It delves into the modeling of data correlations through adaptive learning of similarity graphs and the extraction of low-dimensional feature vectors that preserve the underlying data structure via spectral embedding and eigendecomposition. Moreover, this book encompasses various subsequent learning tasks within this transformed feature space. Serving as a systematic introduction to dynamic graph learning for dimension reduction and data clustering, it provides an in-depth survey of current advancements and the state-of-the-art in this burgeoning research field.

Notably, this book not only presents the core concepts and methodologies of dynamic graph learning but also elucidates its practical applications in dimension reduction and data clustering. By summarizing recent developments and offering introductory studies on dynamic graph learning for dimension reduction and clustering, this book equips researchers with a comprehensive understanding of crucial issues and serves as an excellent entry point for further exploration in this area of research.

The authors of this remarkable book have conducted extensive research on dynamic graph learning for dimension reduction, data clustering, and related topics. Their expertise and contributions have significantly advanced the field, making this book an invaluable resource for researchers and practitioners seeking to explore the frontiers of dynamic graph learning.

Jinan, China Lei Zhu

Preface II

In our rapidly evolving world, the proliferation of the Internet and information technology has generated an enormous amount of data in various domains. However, the high dimensionality, diversity, and complexity of these data make it challenging to directly utilize or extract meaningful information from them. Furthermore, the labor-intensive and resource-demanding process of data annotation often results in the absence of label information or only partial labeling, further complicating data processing and analysis. To unlock valuable insights for subsequent learning tasks, it becomes imperative to employ dimension reduction and clustering techniques to process and analyze these data. Dimension reduction and clustering methodologies have already found extensive applications in fields such as bioinformatics, medical image analysis, computer vision, and recommendation systems, showcasing promising performance.

Drawing inspiration from the ability of graphs to model correlations between data points, researchers have introduced graph theory into machine learning, particularly in the realm of unsupervised learning tasks. By leveraging graph theory, the similarity matrix of sample data is utilized to decompose features, projecting the original high-dimensional data into a structured and expressive feature vector space. This graph-based feature representation enables effective support for dimensionality reduction and data clustering tasks. While graph learning has made significant strides in the past few decades, challenges remain in the areas of dimension reduction and data clustering. Some of these challenges include:

(1) Joint optimization. Existing methods often separate the graph construction process from subsequent learning tasks, limiting the ability to learn similarity graphs tailored to specific learning objectives and leading to sub-optimal results. Developing effective joint learning strategies is crucial to improving performance.
(2) Inaccurate data correlation modeling. Graph-based methods are sensitive to the quality of learned graphs. Enhancing the accuracy and quality of graphs is of paramount importance.

(3) Multi-view fusion. Current methods treat each view equally when dealing with learning tasks involving multiple views, disregarding the unique characteristics inherent in each view. A more comprehensive approach that considers the distinctive attributes of each view is needed.
(4) Out-of-sample extension. Many existing graph clustering methods are ill-equipped to handle the challenge of clustering out-of-sample data. They often require re-running the entire algorithm to obtain clustering labels for previously unseen data, which poses significant computational burden in practical applications.

In this book, we address the aforementioned research challenges by presenting several state-of-the-art dynamic graph learning methods for dimension reduction and data clustering. These methods have been extensively tested and validated through rigorous experiments. Specifically, we begin by introducing a dynamic graph learning method for feature projection. This method enables simultaneous learning of a feature projection matrix and a dynamic graph. We then explore two dynamic graph methods for feature selection. One method incorporates dynamic graph learning and binary hashing into the unsupervised feature selection process, presenting a unified framework for single-view and multi-view settings. Another method proposes an adaptive collaborative similarity learning approach for unsupervised multi-view feature selection and extends it to a more efficient variant.

Additionally, we delve into two dynamic graph frameworks for data clustering. The first framework focuses on adaptively learning structured graphs and directly generating discrete cluster labels without incurring information loss. The second framework introduces a flexible and self-adaptive multi-view spectral clustering method, which performs adaptive multi-view graph fusion, learns structured graphs, and supports flexible out-of-sample extension simultaneously.

In conclusion, this book provides a comprehensive exploration of dynamic graph learning. It not only addresses critical research challenges but also offers practical methodologies for dimension reduction and data clustering, supported by extensive experimental validation. The authors hope that this book will contribute to the advancement of dynamic graph learning and inspire further research in this rapidly evolving field.

Jinan, China — Lei Zhu
Chengdu, China — Jingjing Li
Shenzhen, China — Zheng Zhang
June 2023

Acknowledgments

We would like to express our heartfelt gratitude to all the individuals who have contributed to the completion of this book. Without their support and assistance, this project would not have been possible, or at least not in its current form. In particular, we extend our sincere appreciation to our colleagues from the Big Media Data Computing Lab at Shandong Normal University, whose invaluable support has greatly influenced the outcome of this time-consuming endeavor. Their contributions have provided essential ingredients for insightful discussions throughout the writing process, and we are truly grateful for their involvement.

First and foremost, we would like to acknowledge Ms. Dan Shi from Shandong Normal University, Mr. Yudong Han and Dr. Xuemeng Song from Shandong University, Mr. Xiao Dong from Sun Yat-Sen University, Prof. Xiaojun Chang from the University of Technology Sydney, Dr. Zhiyong Cheng and Dr. Zhihui Li from Shandong Artificial Intelligence Institute, Xiaobai Liu from San Diego State University, as well as Dr. Huaxiang Zhang from Shandong Normal University. We have consulted with them on specific technical chapters, and they have made significant contributions to several chapters. Their valuable feedback and comments at various stages of the book's development have played a crucial role in shaping its content. We would also like to express our deep gratitude to Prof. Heng Tao Shen from the University of Electronic Science and Technology of China, whose unwavering support, advice, and invaluable experience have been instrumental whenever we reached out to him.

We would also like to extend our appreciation to the editors for their outstanding efforts in bringing this book to fruition. Their dedication and assistance have contributed to the smooth publication process and made the journey enjoyable. Last but certainly not least, we would like to express our profound gratitude to our beloved families for their selfless consideration, unwavering love, and unconditional support throughout this entire undertaking.

We would like to acknowledge the financial support provided by the National Natural Science Foundation of China under Grant 62172263, the Natural Science Foundation of Shandong, China, under Grant ZR2020YQ47 and Grant ZR2019QF002, and the Youth Innovation Project of Shandong Universities, China, under Grant 2019KJN040.

June 2023

Lei Zhu
Jingjing Li
Zheng Zhang

Contents