SpringerBriefs in Computer Science

Guangtao Xue · Yi-Chao Chen · Feng Lyu · Minglu Li

Robust Network Compressive Sensing

Springer

SpringerBriefs in Computer Science

Series Editors
Stan Zdonik, Brown University, Providence, RI, USA
Shashi Shekhar, University of Minnesota, Minneapolis, MN, USA
Xindong Wu, University of Vermont, Burlington, VT, USA
Lakhmi C. Jain, University of South Australia, Adelaide, SA, Australia
David Padua, University of Illinois Urbana-Champaign, Urbana, IL, USA
Xuemin Sherman Shen, University of Waterloo, Waterloo, ON, Canada
Borko Furht, Florida Atlantic University, Boca Raton, FL, USA
V. S. Subrahmanian, University of Maryland, College Park, MD, USA
Martial Hebert, Carnegie Mellon University, Pittsburgh, PA, USA
Katsushi Ikeuchi, University of Tokyo, Tokyo, Japan
Bruno Siciliano, Università di Napoli Federico II, Napoli, Italy
Sushil Jajodia, George Mason University, Fairfax, VA, USA
Newton Lee, Institute for Education, Research and Scholarships, Los Angeles, CA, USA

SpringerBriefs present concise summaries of cutting-edge research and practical applications across a wide spectrum of fields. Featuring compact volumes of 50 to 125 pages, the series covers a range of content from professional to academic.

Typical topics might include:

- A timely report of state-of-the art analytical techniques
- A bridge between new research results, as published in journal articles, and a contextual literature review
- A snapshot of a hot or emerging topic
- An in-depth case study or clinical example
- A presentation of core concepts that students must understand in order to make independent contributions

Briefs allow authors to present their ideas and readers to absorb them with minimal time investment. Briefs will be published as part of Springer's eBook collection, with millions of users worldwide. In addition, Briefs will be available for individual print and electronic purchase. Briefs are characterized by fast, global electronic dissemination, standard publishing contracts, easy-to-use manuscript preparation and formatting guidelines, and expedited production schedules. We aim for publication 8–12 weeks after acceptance. Both solicited and unsolicited manuscripts are considered for publication in this series.

**Indexing: This series is indexed in Scopus, Ei-Compendex, and zbMATH **

Guangtao Xue • Yi-Chao Chen • Feng Lyu •
Minglu Li

Robust Network Compressive Sensing

Springer

Guangtao Xue
Shanghai, China

Yi-Chao Chen
Shanghai, China

Feng Lyu
Hunan, China

Minglu Li
Jinhua, China

ISSN 2191-5768 ISSN 2191-5776 (electronic)
SpringerBriefs in Computer Science
ISBN 978-3-031-16828-4 ISBN 978-3-031-16829-1 (eBook)
https://doi.org/10.1007/978-3-031-16829-1

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022
This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

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

Preface

The networks are constantly generating a wealth of rich and diverse information. This information creates exciting opportunities for network analysis and provides insight into the complex interactions between network entities. However, network analysis often faces the problems of (1) under-constrained, where there is too little data due to feasibility and cost issues in collecting data, or (2) over-constrained, where there is too much data, so the analysis becomes unscalable. Compressive sensing is an effective technique to solve both problems. It utilizes the underlying data structure for analysis. Specifically, to solve the under-constrained problem, compressive sensing technologies can be applied to reconstruct the missing elements or predict the future data; to solve the over-constraint problem, compressive sensing technologies can be applied to identify significant elements.

To well support compressive sensing in network data analysis, a robust and general framework is needed to support diverse applications, yet this is challenging for real-world data where noise, anomalies, and lack of synchronization are common. First, the number of unknowns for network analysis can be much larger than the number of measurements we have. For example, traffic engineering requires knowing the complete traffic matrix between all source and destination pairs in order to properly configure traffic and avoid congestion. However, measuring the flow between all source and destination pairs is very expensive or even infeasible. Reconstructing data from a small number of measurements is an under-constrained problem. In addition, real-world data are complex and heterogeneous and often violate the low-level assumptions required by existing compressive sensing techniques. Such violations significantly reduce the applicability and effectiveness of existing compressive sensing methods. Third, synchronization of network data reduces the data ranks and increases spatial locality. However, periodic time series exhibit not only misalignment but also different frequencies, which makes it difficult to synchronize data in the time and frequency domains.

In this monograph, we investigate compressive sensing techniques to provide a robust and general framework for network data analytics. Our goal is to build a compressive sensing framework for missing data interpolation, anomaly detection, data segmentation, and activity recognition, and to demonstrate its benefits. In

Chap. 1, we introduce compressive sensing, including its definition, limitation, and how it supports different network analysis applications. In Chap. 2, to demonstrate the feasibility of compressive sensing in network analytics, we apply it to detect anomalies in the customer care call dataset from a Tier 1 ISP in the United States. A regression-based model is applied to find the relationship between calls and events. We show that compressive sensing is effective in identifying important factors and can leverage the low-rank structure and temporal stability to improve the detection accuracy. In Chap. 3, we show that there are several challenges in applying compressive sensing to real-world data. Understanding the reasons behind the challenges is important for designing methods and mitigating their impact. Therefore, we analyze a wide range of real-world traces, and our analysis shows that there are different factors that contribute to the violation of the low-rank property in real data. In particular, we find that (1) noise, errors, and anomalies, and (2) asynchrony in the time and frequency domains lead to network-induced ambiguity and can easily cause low-rank matrices to become higher-ranked. To address the problem of noise, errors, and anomalies, in Chap. 4, we propose a robust compressive sensing technique. It explicitly accounts for anomalies by decomposing real-world data represented in matrix form into a low-rank matrix, a sparse anomaly matrix, an error term, and a small noise matrix. In Chap. 5, to address the problem of lack of synchronization, we propose a data-driven synchronization algorithm. It can eliminate misalignment while taking into account the heterogeneity of real-world data in both time and frequency domains. The data-driven synchronization can be applied to any compressive sensing technique and is general to any real-world data. We show that the combination of the two techniques can reduce the ranks of real-world data, improve the effectiveness of compressive sensing, and have a wide range of applications.

Shanghai, China — Guangtao Xue
Shanghai, China — Yi-Chao Chen
Hunan, China — Feng Lyu
Zhejiang, China — Minglu Li

Contents

1 Introduction 1
- 1.1 Apply Compressive Sensing to Real-World Data 2
- 1.2 Understand the Limitation 3
- 1.3 Improve Compressive Sensing for Real-World Data 3
- 1.4 Aim of the Monograph 4

2 Event Detection System 7
- 2.1 Problem Formulation 7
- 2.2 Our Approach 12
  - 2.2.1 Problem Formulation 12
  - 2.2.2 Our Regression 13
  - 2.2.3 Reducing Categories 14
  - 2.2.4 Combining Multiple Classifiers 17
- 2.3 Evaluation 18
- 2.4 External Data: Social Media 21
- 2.5 Conclusion 23

References 23

3 Limitation of Compressive Sensing 25
- 3.1 Datasets 25
- 3.2 Analysis 28
  - 3.2.1 Anomalies Increase Ranks 28
  - 3.2.2 Lack of Synchronization Increases Ranks 31
  - 3.2.3 Temporal Stability Varies Across Traces 32
- 3.3 Conclusion 33

References 33

4 Robust Compressive Sensing 35
- 4.1 LENS Decomposition 35
  - 4.1.1 LENS Decomposition Framework 35
  - 4.1.2 Optimization Algorithm 39
  - 4.1.3 Setting Parameters 41

4.2 Evaluation ... 43
4.2.1 Evaluation Methodology ... 43
4.2.2 Performance Results ... 45
4.3 Conclusion ... 55
References ... 59

5 Data-Driven Synchronization ... 61
5.1 Motivation ... 61
5.2 Data-Driven Synchronization ... 61
5.3 Applications ... 65
5.3.1 Interpolation ... 65
5.3.2 Segmentation ... 67
5.3.3 Activity Recognition ... 69
5.4 Evaluation ... 70
5.4.1 Evaluation of Synchronization ... 70
5.4.2 Evaluation of Interpolation Accuracy ... 73
5.4.3 Evaluation of Segmentation ... 75
5.4.4 Evaluation of Activity Recognition ... 82
5.4.5 Summary of Evaluation Results ... 82
5.5 Existing Sychronization Approaches ... 83
5.6 Conclusion ... 83
References ... 83

6 Conclusion and Future Research Direction ... 85

A Alternating Direction Method ... 87
A.1 Algorithm ... 87