SpringerBriefs in Computer Science

Series Editors
Stan Zdonik
Peng Ning
Shashi Shekhar
Jonathan Katz
Xindong Wu
Lakhmi C Jain
David Padua
Xuemin Shen
Borko Furht
VS Subrahmanian

For further volumes:
http://www.springer.com/series/10028

Baris M. Kazar • Mete Celik

Spatial AutoRegression (SAR) Model

Parameter Estimation Techniques

Springer

Baris M. Kazar
Oracle America Inc.
Nashua, NH, USA
baris.kazar@Oracle.com

Mete Celik
Erciyes University
Kayseri, Turkey
mcelik@erciyes.edu.tr

ISSN 2191-5768 e-ISSN 2191-5776
ISBN 978-1-4614-1841-2 e-ISBN 978-1-4614-1842-9
DOI 10.1007/978-1-4614-1842-9
Springer New York Dordrecht Heidelberg London

Library of Congress Control Number: 2012931935

© The Author(s) 2012
All rights reserved. This work may not be translated or copied in whole or in part without the written permission of the publisher (Springer Science+Business Media, LLC, 233 Spring Street, New York, NY 10013, USA), except for brief excerpts in connection with reviews or scholarly analysis. Use in connection with any form of information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed is forbidden.
The use in this publication of trade names, trademarks, service marks, and similar terms, even if they are not identified as such, is not to be taken as an expression of opinion as to whether or not they are subject to proprietary rights.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Dedicated

to our advisors Prof. Shashi Shekhar and Prof. David Lilja.

Baris Kazar & Mete Celik

To my wife Neslihan, my teacher parents Mualla Hatice and Kemal, my brothers Ozgur and Caglar, my parent in-laws Fadime and Muzaffer, my brother in-law Emrullah, my sister in-law Emriye, and my whole family.

Baris Kazar

To my wife Filiz, my son Cagri Fatih, my daughter Elif Huma, my parents Fatma and Metin, and my whole family.

Mete Celik

Acknowledgements

We would like to thank our advisors Professor Shashi Shekhar and Professor David J. Lilja. Without their invaluable help, guidance, and endless support we would never be able to make this study a reality.

We also would like to thank Minnesota Supercomputing Institute (MSI), University of Minnesota Digital Technology Center (DTC) and Army High Performance and Computing Center (AHPCRC), Department of Electrical and Computer Engineering, and Department of Computer Science and Engineering at the University of Minnesota and Department of Computer Engineering at Erciyes University for their financial support and permissions to use their facilities and computing resources.

We are very grateful to the all members of Spatial Database and Spatial Data Mining Group and ARCTiC Labs (Laboratory for Advanced Research in Computing Technology and Compilers) at the University of Minnesota and Birali Runesha and Dr. Shuxia Zhang at the Minnesota Supercomputing Institute for their valuable discussions, critiques, and contributions.

This research also benefited from discussions with many colleagues and friends. We would like to thank Professor Daniel Boley, Professor Pen Chung Yew, Professor Vipin Kumar, Dr. Siva Ravada, Dr. Joshua Yi, Chris Hescott, Dr. Sreekumar Kodakara, Dr. Ying Chen, Dr. Keqiang Wu, Dr. Alex Zhang, Professor Hui Xiong, Professor Sanjay Chawla, Dr. Betsy George, Professor Jin Soung Yoo, Dr. Sangho Kim, Professor Kelley Pace, Professor James P. LeSage, Professor Giannakis Georgios, Professor Resit Sendag, Cetin Yavuz, Professor Vladimir Cherkassky, Professor Paul Schrater, Professor Filiz Dadaser-Celik, Neslihan Erdogan Kazar (MBA), Dr. Shuxia Zhang, and Dr. Birali Runesha for participating in white board discussions, proof reading papers, discussing spatial autoregressive model (SAR) on the phone or on our presentations for hours with great patience, and helping us debugging serial and parallel codes.

Contents

1 Introduction 1
- 1.1 Background Information 2
- 1.2 Spatial Data Mining 2
- 1.3 Related Work 3
- 1.4 Contributions 4
- 1.5 Outline and Scope 5

2 Theory behind the SAR Model 7
- 2.1 Example Neighborhood Matrix (**W**) on Regular (Uniform) Grid Space. 8
- 2.2 Illustration of the Neighborhood Matrix Formation on a 4-by-4 Regular Grid Space. 9
- 2.3 Constructing the Neighborhood Matrix **W** on Irregular Grid Space . 11
- 2.4 Derivation of the ML (Concentrated Log-likelihood) Function 12
  - 2.4.1 The Effect of SAR Autoregression Parameter ρ 16
- 2.5 Types of Optimization 17

3 Parallel Exact SAR Model Solutions 19
- 3.1 Problem Statement 21
- 3.2 The Serial Implementation 21
- 3.3 Proposed Parallel Formulation 23
  - 3.3.1 Stage $\mathscr{A}$: Computing Eigenvalues 24
  - 3.3.2 Stage $\mathscr{B}$: The Golden Section Search 24
  - 3.3.3 Stage $\mathscr{C}$: Least Squares 25
- 3.4 Algebraic Cost Model 25
- 3.5 Experimental Work and Discussion 29
  - 3.5.1 Which load-balancing method provides the best speedup? . . 30
  - 3.5.2 How does problem size impact speedup? 31
  - 3.5.3 How does chunk size affect speedup? 32
  - 3.5.4 How does number of processors affect speedup? 32
- 3.6 Summary 33

4 Comparing Exact and Approximate SAR Model Solutions 35
4.1 Problem Statement 36
4.2 Approximation by Taylor's Series Expansion 36
4.3 Why is Taylor's Series Approximation valid? 37
4.4 Approximation by Chebyshev Polynomials 38
4.5 Why is Chebyshev Polynomial Approximation valid? 41
4.6 Experiment Design 42
4.7 Experimental Results 42
4.8 Summary 45

5 Parallel Implementations of Approximate SAR Model Solutions 47
5.1 Problem Statement 47
5.2 Related Work 48
5.3 Operation Cost Analysis 48
5.4 Experimental Design 49
5.5 Experimental Results 50
5.6 Summary 50

6 A New Approximation: Gauss-Lanczos Approximated SAR Model Solution 51
6.1 Problem Statement 52
6.2 A New Approximation: Gauss-Lanczos Method 52
6.3 Algebraic Error Ranking 55
6.4 Experimental Design and System Setup 55
6.5 Summary 58

7 Conclusions and Future Work 59

8 Supplementary Materials 61
8.1 Moran's I Index: Quantifying the Auto-correlation in Datasets 61
8.2 Simple Overview of Log-likelihood Theory 61
8.3 Derivation of Log-likelihood Function for SARMA Model 62
8.4 Proof for Eigen-values of Markov Matrix which are bounded in $[-1,+1]$ and occur in $\pm$ pairs 65
8.5 Basic Linear Algebra Facts 67
8.6 Proof of symmetry of $(\mathbf{I}-\rho\mathbf{W})^T(\mathbf{I}-\mathbf{M})(\mathbf{I}-\rho\mathbf{W})$ 68
8.7 Proof of $(\mathbf{I}-\rho\mathbf{W})^T(\mathbf{I}-\mathbf{M})(\mathbf{I}-\rho\mathbf{W}) \geq 0$ 68
8.8 Single Variable Optimization: The Golden Section Search 69
8.9 Multi-variable Search 70
References 71