

M A S S I V E C O M P U T I N G

DATA MINING FOR  
SCIENTIFIC AND  
ENGINEERING  
APPLICATIONS

Robert L. Grossman, Chandrika Kamath,  
Philip Kegelmeyer, Vipin Kumar  
and Raju R. Namburu (Eds.)

KLUWER ACADEMIC PUBLISHERS

# Table of Contents

<b>Foreword</b> .....	ix
<b>List of Contributors</b> .....	xi
<b>List of Reviewers</b> .....	xvii
<b>Preface</b> .....	xix
<b>1 ON MINING SCIENTIFIC DATASETS</b> Chandrika Kamath.....	1
<b>2 UNDERSTANDING HIGH DIMENSIONAL AND LARGE DATA SETS: SOME MATHEMATICAL CHALLENGES AND OPPORTUNITIES</b> Jagdish Chandra.....	23
<b>3 DATA MINING AT THE INTERFACE OF COMPUTER SCIENCE AND STATISTICS</b> Padhraic Smyth.....	35
<b>4 MINING LARGE IMAGE COLLECTIONS</b> Michael C. Burl.....	63
<b>5 MINING ASTRONOMICAL DATABASES</b> Roberta M. Humphreys, Juan E. Cabanela, and Jeffrey Kriessler.....	85
<b>6 SEARCHING FOR BENT-DOUBLE GALAXIES IN THE FIRST SURVEY</b> Chandrika Kamath, Erick Cantú-Paz, Imola K. Fodor, and Nu Ai Tang.....	95
<b>7 A DATASPACE INFRASTRUCTURE FOR ASTRONOMICAL DATA</b>	



	Robert Grossman, Emory Creel, Marco Mazzucco, and Roy Williams.....	115
<b>8</b>	<b>DATA MINING APPLICATIONS IN BIOINFORMATICS</b> Naren Ramakrishnan and Ananth Y. Grama.....	125
<b>9</b>	<b>MINING RESIDUE CONTACTS IN PROTEINS</b> Mohammed J. Zaki and Chris Bystroff.....	141
<b>10</b>	<b>KDD SERVICES AT THE GODDARD EARTH SCIENCES DISTRIBUTED ARCHIVE CENTER</b> Christopher Lynnes and Robert Mack.....	165
<b>11</b>	<b>DATA MINING IN INTEGRATED DATA ACCESS AND DATA ANALYSIS SYSTEMS</b> Ruixin Yang, Menas Kafatos, Kwang-Su Yang, and X. Sean Wang.....	183
<b>12</b>	<b>SPATIAL DATA MINING FOR CLASSIFICATION, VISUALISATION AND INTERPRETATION WITH ARTMAP NEURAL NETWORK</b> Weiguo Liu, Sucharita Gopal, and Curtis Woodcock.....	201
<b>13</b>	<b>REAL TIME FEATURE EXTRACTION FOR THE ANALYSIS OF TURBULENT FLOWS</b> I. Marusic, G.V. Candler, V. Interrante, P.K. Subbareddy, and A. Moss.....	223
<b>14</b>	<b>DATA MINING FOR TURBULENT FLOWS</b> Eui-Hong (Sam) Han, George Karypis, and Vipin Kumar.....	239
<b>15</b>	<b>EVITA-EFFICIENT VISUALIZATION AND INTERROGATION OF TERA-SCALE DATA</b> Raghu Machiraju, James E. Fowler, David Thompson, Bharat Soni, and Will Schroeder.....	257
<b>16</b>	<b>TOWARDS UBIQUITOUS MINING OF</b>	

	<b>DISTRIBUTED DATA</b>	
	Hillol Kargupta, Krishnamoorthy Sivakumar, Weiyun Huang, Rajeev Ayyagari, Rong Chen, Byung-Hoon Park, and Erik Johnson.....	281
<b>17</b>	<b>DECOMPOSABLE ALGORITHMS FOR DATA MINING</b>	
	Raj Bhatnagar.....	307
<b>18</b>	<b>HDDI<sup>TM</sup>: HIERARCHICAL DISTRIBUTED DYNAMIC INDEXING</b>	
	William M. Pottenger, Yong-Bin Kim, and Daryl D. Meling.....	319
<b>19</b>	<b>PARALLEL ALGORITHMS FOR CLUSTERING HIGH-DIMENSIONAL LARGE-SCALE DATASETS</b>	
	Harsha Nagesh, Sanjay Goil, and Alok Choudhary.....	335
<b>20</b>	<b>EFFICIENT CLUSTERING OF VERY LARGE DOCUMENT COLLECTIONS</b>	
	Inderjit S. Dhillon, James Fan, and Yuqiang Guan.....	357
<b>21</b>	<b>A SCALABLE HIERARCHICAL ALGORITHM FOR UNSUPERVISED CLUSTERING</b>	
	Daniel Boley.....	383
<b>22</b>	<b>HIGH-PERFORMANCE SINGULAR VALUE DECOMPOSITION</b>	
	David B. Skillicorn and Xiaolan Yang.....	401
<b>23</b>	<b>MINING HIGH-DIMENSIONAL SCIENTIFIC DATA SETS USING SINGULAR VALUE DECOMPOSITION</b>	
	Ekaterina Maltseva, Clara Pizzuti, and Domenico Talia.....	425
<b>24</b>	<b>SPATIAL DEPENDENCE IN DATA MINING</b>	
	James P. LeSage and R. Kelley Pace.....	439

<b>25</b>	<b>SPARC: SPATIAL ASSOCIATION RULE-BASED CLASSIFICATION</b>	461
	Jiawei Han, Anthony K. H. Tung, and Jing He.....	
<b>26</b>	<b>WHAT'S SPATIAL ABOUT SPATIAL DATA MINING: THREE CASE STUDIES</b>	487
	Shashi Shekhar, Yan Huang, Weili Wu, C.T. Lu, and S. Chawla.....	
<b>27</b>	<b>PREDICTING FAILURES IN EVENT SEQUENCES</b>	515
	Mohammed J. Zaki, Neal Lesh, and Mitsunori Ogihara.....	
<b>28</b>	<b>EFFICIENT ALGORITHMS FOR MINING LONG PATTERNS IN SCIENTIFIC DATA SETS</b>	541
	Ramesh C. Agarwal and Charu C. Aggarwal.....	
<b>29</b>	<b>PROBABILISTIC ESTIMATION IN DATA MINING</b>	567
	Edwin P. D. Pednault and Chidanand Apte.....	
<b>30</b>	<b>CLASSIFICATION USING ASSOCIATION RULES: WEAKNESSES AND ENHANCEMENTS</b>	591
	Bing Liu, Yiming Ma, and Ching-Kian Wong.....	