

Cambridge University Press

978-0-521-19224-8 - Scaling Up Machine Learning: Parallel and Distributed Approaches

Edited by Ron Bekkerman, Mikhail Bilenko and John Langford

Table of Contents

[More information](#)

Contents

<i>Contributors</i>	xi
<i>Preface</i>	xv
1 Scaling Up Machine Learning: Introduction	1
<i>Ron Bekkerman, Mikhail Bilenko, and John Langford</i>	
1.1 Machine Learning Basics	2
1.2 Reasons for Scaling Up Machine Learning	3
1.3 Key Concepts in Parallel and Distributed Computing	6
1.4 Platform Choices and Trade-Offs	7
1.5 Thinking about Performance	9
1.6 Organization of the Book	10
1.7 Bibliographic Notes	17
References	19
Part One Frameworks for Scaling Up Machine Learning	
2 MapReduce and Its Application to Massively Parallel Learning of Decision Tree Ensembles	23
<i>Biswanath Panda, Joshua S. Herbach, Sugato Basu, and Roberto J. Bayardo</i>	
2.1 Preliminaries	24
2.2 Example of PLANET	30
2.3 Technical Details	33
2.4 Learning Ensembles	38
2.5 Engineering Issues	39
2.6 Experiments	41
2.7 Related Work	44
2.8 Conclusions	46
Acknowledgments	47
References	47

Cambridge University Press

978-0-521-19224-8 - Scaling Up Machine Learning: Parallel and Distributed Approaches

Edited by Ron Bekkerman, Mikhail Bilenko and John Langford

Table of Contents

[More information](#)

vi

CONTENTS

3 Large-Scale Machine Learning Using DryadLINQ	49
<i>Mihai Budiu, Dennis Fetterly, Michael Isard, Frank McSherry, and Yuan Yu</i>	
3.1 Manipulating Datasets with LINQ	49
3.2 k -Means in LINQ	52
3.3 Running LINQ on a Cluster with DryadLINQ	53
3.4 Lessons Learned	65
References	67
4 IBM Parallel Machine Learning Toolbox	69
<i>Edwin Pednault, Elad Yom-Tov, and Amol Ghoting</i>	
4.1 Data-Parallel Associative-Commutative Computation	70
4.2 API and Control Layer	71
4.3 API Extensions for Distributed-State Algorithms	76
4.4 Control Layer Implementation and Optimizations	77
4.5 Parallel Kernel k -Means	79
4.6 Parallel Decision Tree	80
4.7 Parallel Frequent Pattern Mining	83
4.8 Summary	86
References	87
5 Uniformly Fine-Grained Data-Parallel Computing for Machine Learning Algorithms	89
<i>Meichun Hsu, Ren Wu, and Bin Zhang</i>	
5.1 Overview of a GP-GPU	91
5.2 Uniformly Fine-Grained Data-Parallel Computing on a GPU	93
5.3 The k -Means Clustering Algorithm	97
5.4 The k -Means Regression Clustering Algorithm	99
5.5 Implementations and Performance Comparisons	102
5.6 Conclusions	105
References	105
Part Two Supervised and Unsupervised Learning Algorithms	
6 PSVM: Parallel Support Vector Machines with Incomplete Cholesky Factorization	109
<i>Edward Y. Chang, Hongjie Bai, Kaihua Zhu, Hao Wang, Jian Li, and Zhihuan Qiu</i>	
6.1 Interior Point Method with Incomplete Cholesky Factorization	112
6.2 PSVM Algorithm	114
6.3 Experiments	121
6.4 Conclusion	125
Acknowledgments	125
References	125
7 Massive SVM Parallelization Using Hardware Accelerators	127
<i>Igor Durdanovic, Eric Cosatto, Hans Peter Graf, Srihari Cadambi, Venkata Jakkula, Srimat Chakradhar, and Abhinandan Majumdar</i>	
7.1 Problem Formulation	128
7.2 Implementation of the SMO Algorithm	131

Cambridge University Press

978-0-521-19224-8 - Scaling Up Machine Learning: Parallel and Distributed Approaches

Edited by Ron Bekkerman, Mikhail Bilenko and John Langford

Table of Contents

[More information](#)

CONTENTS	vii
7.3 Micro Parallelization: Related Work	132
7.4 Previous Parallelizations on Multicore Systems	133
7.5 Micro Parallelization: Revisited	136
7.6 Massively Parallel Hardware Accelerator	137
7.7 Results	145
7.8 Conclusion	146
References	146
8 Large-Scale Learning to Rank Using Boosted Decision Trees	148
<i>Krysta M. Svore and Christopher J. C. Burges</i>	
8.1 Related Work	149
8.2 LambdaMART	151
8.3 Approaches to Distributing LambdaMART	153
8.4 Experiments	158
8.5 Conclusions and Future Work	168
8.6 Acknowledgments	169
References	169
9 The Transform Regression Algorithm	170
<i>Ramesh Natarajan and Edwin Pednault</i>	
9.1 Classification, Regression, and Loss Functions	171
9.2 Background	172
9.3 Motivation and Algorithm Description	173
9.4 TReg Expansion: Initialization and Termination	177
9.5 Model Accuracy Results	184
9.6 Parallel Performance Results	186
9.7 Summary	188
References	189
10 Parallel Belief Propagation in Factor Graphs	190
<i>Joseph Gonzalez, Yucheng Low, and Carlos Guestrin</i>	
10.1 Belief Propagation in Factor Graphs	191
10.2 Shared Memory Parallel Belief Propagation	195
10.3 Multicore Performance Comparison	209
10.4 Parallel Belief Propagation on Clusters	210
10.5 Conclusion	214
Acknowledgments	214
References	214
11 Distributed Gibbs Sampling for Latent Variable Models	217
<i>Arthur Asuncion, Padhraic Smyth, Max Welling, David Newman, Ian Porteous, and Scott Triglia</i>	
11.1 Latent Variable Models	217
11.2 Distributed Inference Algorithms	220
11.3 Experimental Analysis of Distributed Topic Modeling	224
11.4 Practical Guidelines for Implementation	229
11.5 A Foray into Distributed Inference for Bayesian Networks	231
11.6 Conclusion	236
Acknowledgments	237
References	237

12 Large-Scale Spectral Clustering with MapReduce and MPI	240
<i>Wen-Yen Chen, Yangqiu Song, Hongjie Bai, Chih-Jen Lin, and Edward Y. Chang</i>	
12.1 Spectral Clustering	241
12.2 Spectral Clustering Using a Sparse Similarity Matrix	243
12.3 Parallel Spectral Clustering (PSC) Using a Sparse Similarity Matrix	245
12.4 Experiments	251
12.5 Conclusions	258
References	259
13 Parallelizing Information-Theoretic Clustering Methods	262
<i>Ron Bekkerman and Martin Scholz</i>	
13.1 Information-Theoretic Clustering	264
13.2 Parallel Clustering	266
13.3 Sequential Co-clustering	269
13.4 The DataLoom Algorithm	270
13.5 Implementation and Experimentation	274
13.6 Conclusion	277
References	278
Part Three Alternative Learning Settings	
14 Parallel Online Learning	283
<i>Daniel Hsu, Nikos Karampatziakis, John Langford, and Alex J. Smola</i>	
14.1 Limits Due to Bandwidth and Latency	285
14.2 Parallelization Strategies	286
14.3 Delayed Update Analysis	288
14.4 Parallel Learning Algorithms	290
14.5 Global Update Rules	298
14.6 Experiments	302
14.7 Conclusion	303
References	305
15 Parallel Graph-Based Semi-Supervised Learning	307
<i>Jeff Bilmes and Amarnag Subramanya</i>	
15.1 Scaling SSL to Large Datasets	309
15.2 Graph-Based SSL	310
15.3 Dataset: A 120-Million-Node Graph	317
15.4 Large-Scale Parallel Processing	319
15.5 Discussion	327
References	328
16 Distributed Transfer Learning via Cooperative Matrix Factorization	331
<i>Evan Xiang, Nathan Liu, and Qiang Yang</i>	
16.1 Distributed Coalitional Learning	333
16.2 Extension of DisCo to Classification Tasks	343

Cambridge University Press

978-0-521-19224-8 - Scaling Up Machine Learning: Parallel and Distributed Approaches

Edited by Ron Bekkerman, Mikhail Bilenko and John Langford

Table of Contents

[More information](#)

CONTENTS	ix
16.3 Conclusion	350
References	350
17 Parallel Large-Scale Feature Selection	352
<i>Jeremy Kubica, Sameer Singh, and Daria Sorokina</i>	
17.1 Logistic Regression	353
17.2 Feature Selection	354
17.3 Parallelizing Feature Selection Algorithms	358
17.4 Experimental Results	363
17.5 Conclusions	368
References	368
Part Four Applications	
18 Large-Scale Learning for Vision with GPUs	373
<i>Adam Coates, Rajat Raina, and Andrew Y. Ng</i>	
18.1 A Standard Pipeline	374
18.2 Introduction to GPUs	377
18.3 A Standard Approach Scaled Up	380
18.4 Feature Learning with Deep Belief Networks	388
18.5 Conclusion	395
References	395
19 Large-Scale FPGA-Based Convolutional Networks	399
<i>Clément Farabet, Yann LeCun, Koray Kavukcuoglu, Berin Martini, Polina Akselrod, Selcuk Talay, and Eugenio Culurciello</i>	
19.1 Learning Internal Representations	400
19.2 A Dedicated Digital Hardware Architecture	405
19.3 Summary	416
References	417
20 Mining Tree-Structured Data on Multicore Systems	420
<i>Shirish Tatikonda and Srinivasan Parthasarathy</i>	
20.1 The Multicore Challenge	422
20.2 Background	423
20.3 Memory Optimizations	427
20.4 Adaptive Parallelization	431
20.5 Empirical Evaluation	437
20.6 Discussion	442
Acknowledgments	443
References	443
21 Scalable Parallelization of Automatic Speech Recognition	446
<i>Jike Chong, Ekaterina Gonina, Kisun You, and Kurt Keutzer</i>	
21.1 Concurrency Identification	450
21.2 Software Architecture and Implementation Challenges	452
21.3 Multicore and Manycore Parallel Platforms	454
21.4 Multicore Infrastructure and Mapping	455

Cambridge University Press

978-0-521-19224-8 - Scaling Up Machine Learning: Parallel and Distributed Approaches

Edited by Ron Bekkerman, Mikhail Bilenko and John Langford

Table of Contents

[More information](#)

x

CONTENTS

21.5 The Manycore Implementation	459
21.6 Implementation Profiling and Sensitivity Analysis	462
21.7 Application-Level Optimization	464
21.8 Conclusion and Key Lessons	467
References	468
<i>Subject Index</i>	471