The Data Collection For Machine Learning

ABS-TEMP-2015A-No.002-TJx

汤吉(Ji TANG)
ID: ABS-TEMP-2015A, E-mail: tangji08@hotmail.com
Date: 2015.11.16

November 16, 2015

Contents

Chapter 1

Machine Learning

1.1 Journals

1. Journal of Machine Learning Research, IF = 3.420
   http://www.jmlr.org

2. Machine Learning, IF = 1.467
   http://www.springer.com/computer/ai/journal/10994

3. IEEE Transactions on Knowledge and Data Engineering, IF = 2.067
   http://ieeexplore.ieee.org/xpl/RecentIssue.jsp?punumber=69

4. Neural Computation, IF = 2.207
   http://www.mitpressjournals.org/loi/neco

5. Journal of Artificial Intelligence Research, IF = 1.257
   http://www.jair.org

6. Artificial Intelligence, IF = 3.371
   http://www.journals.elsevier.com/artificial-intelligence/

7. IEEE Transactions on Pattern Analysis and Machine Intelligence, IF = 5.781
   http://www.computer.org/web/tpami/index

8. Pattern Recognition, IF = 3.096
   http://www.journals.elsevier.com/pattern-recognition/

9. Neural Networks, IF = 2.708
   http://www.journals.elsevier.com/neural-networks

1.2 Person

There are some important professors or person whom we may encounter frequently when doing researches.

1. Edoardo M. Airoldi, Harvard University, USA
   http://www.people.fas.harvard.edu/ăiroldi/

2. Peter Auer, University of Leoben, Austria
   http://portal.uni-freiburg.de/sdd/personen/auer/index.html/startseite

3. Francis Bach, INRIA, France
   http://www.di.ens.fr/ fbach/

4. Andrew Bagnell, Carnegie Mellon University, USA
   http://www.ri.cmu.edu/person.html?person_id=689

5. David Barber, University College London, UK
   http://web4.cs.ucl.ac.uk/staff/D.Barber/pmwiki/pmwiki.php?n=Brml.HomePage

6. Mikhail Belkin, Ohio State University, USA
   http://soscholar.net/author?author_id=42e8023a-8f4e-4ed8-908b-2bacea0034b4

7. Yoshua Bengio, Université de Montréal, Canada
   http://www.iro.umontreal.ca/ bengioy/yoshua_en/

8. Samy Bengio, Google Research, USA
   http://bengio.abracadoudou.com

9. Jeff Bilmes, University of Washington, USA
   http://melodi.ee.washington.edu/ bilmes/pgs/index.html

10. Karsten Borgwardt, MPI For Intelligent systems, Germany
    https://www.bsse.ethz.ch/mlcb?page=employee&employee=Karsten

11. Lawrence Carin, Duke University, USA
    http://people.ee.duke.edu/ lcarin/

12. Zhihua Zhang, Shanghai Jiao Tong University, China
    http://bcmi.sjtu.edu.cn/ zhzhang/

1.3 Directions

1. Improving classification accuracy by learning ensembles of classifiers.
2. Methods for scaling up supervised learning algorithms.
3. Reinforcement learning.
4. Learning complex stochastic models.

1.4 Open Source Software

1. Shark
2. A Library for Locally Weighted Projection Regression
3. LIBLINEAR: A Library for Large Linear Classification
4. JNCC2: The Java Implementation Of Naive Credal Classifier 2
5. Python Environment for Bayesian Learning: Inferring the Structure of Bayesian Networks from Knowledge and Data
6. Nieme: Large-Scale Energy-Based Models
7. Java-ML: A Machine Learning Library
8. Model Monitor: Evaluating, Comparing, and Monitoring Models
9. Dlib-ml: A Machine Learning Toolkit

10. RL-Glue: Language-Independent Software for Reinforcement-Learning Experiments
11. DL-Learner: Learning Concepts in Description Logics
12. Error-Correcting Output Codes Library
13. PyBrain
14. Continuous Time Bayesian Network Reasoning and Learning Engine
15. SFO: A Toolbox for Submodular Function Optimization
16. MOA: Massive Online Analysis
17. FastInf: An Efficient Approximate Inference Library
18. The SHOGUN Machine Learning Toolbox
19. A Surrogate Modeling and Adaptive Sampling Toolbox for Computer Based Design
20. Model-based Boosting 2.0
21. libDAI: A Free and Open Source C++ Library for Discrete Approximate Inference in Graphical Models
22. Gaussian Processes for Machine Learning (GPML) Toolbox
23. CARP: Software for Fishing Out Good Clustering Algorithms
24. The arules R-Package Ecosystem: Analyzing Interesting Patterns from Large Transaction Data Sets
25. MSVMpack: A Multi-Class Support Vector Machine Package
26. Waffles: A Machine Learning Toolkit
27. MULAN: A Java Library for Multi-Label Learning
28. LPmade: Link Prediction Made Easy
29. Scikit-learn: Machine Learning in Python
30. The Stationary Subspace Analysis Toolbox
31. MULTIBOOST: A Multi-purpose Boosting Package
32. ML-Flex: A Flexible Toolbox for Performing Classification Analyses In Parallel

33. GPLP: A Local and Parallel Computation Toolbox for Gaussian Process Regression
34. NIMFA : A Python Library for Nonnegative Matrix Factorization
35. The huge Package for High-dimensional Undirected Graph Estimation in R
36. glm-ie: Generalised Linear Models Inference & Estimation Toolbox
37. Jstacs: A Java Framework for Statistical Analysis and Classification of Biological Sequences
38. Pattern for Python
39. DEAP: Evolutionary Algorithms Made Easy
40. A Topic Modeling Toolbox Using Belief Propagation
41. PREA: Personalized Recommendation Algorithms Toolkit
42. Oger: Modular Learning Architectures For Large-Scale Sequential Processing
43. Sally: A Tool for Embedding Strings in Vector Spaces
44. DARWIN: A Framework for Machine Learning and Computer Vision Research and Development
45. SVDFeature: A Toolkit for Feature-based Collaborative Filtering
46. A C++ Template-Based Reinforcement Learning Library: Fitting the Code to the Mathematics
47. MLPACK: A Scalable C++ Machine Learning Library
48. GPstuff: Bayesian Modeling with Gaussian Processes
49. JKernelMachines: A Simple Framework for Kernel Machines
50. Orange: Data Mining Toolbox in Python
51. Tapkee: An Efficient Dimension Reduction Library
52. The CAM Software for Nonnegative Blind Source Separation in R-Java
53. QuantMiner for Mining Quantitative Association Rules
54. Divvy: Fast and Intuitive Exploratory Data Analysis

55. GURLS: A Least Squares Library for Supervised Learning
56. BudgetedSVM: A Toolbox for Scalable SVM Approximations
57. EnsembleSVM: A Library for Ensemble Learning Using Support Vector Machines
58. Information Theoretical Estimators Toolbox
59. The FASTCLIME Package for Linear Programming and Large-Scale Precision Matrix Estimation in R
60. LIBOL: A Library for Online Learning Algorithms
61. Conditional Random Field with High-order Dependencies for Sequence Labeling and Segmentation
62. Manopt, a Matlab Toolbox for Optimization on Manifolds
63. pystruct - Learning Structured Prediction in Python
64. ooDACE Toolbox: A Flexible Object-Oriented Kriging Implementation
65. The Gesture Recognition Toolkit
66. SPMF: A Java Open-Source Pattern Mining Library
67. BayesOpt: A Bayesian Optimization Library for Nonlinear Optimization, Experimental Design and Bandits
68. SAMOA: Scalable Advanced Massive Online Analysis
69. The flare Package for High Dimensional Linear Regression and Precision Matrix Estimation in R
70. Introducing CURRENNT: The Munich Open-Source CUDA RecurREnt Neural Network Toolkit
71. A Classification Module for Genetic Programming Algorithms in JCLEC
72. Encog: Library of Interchangeable Machine Learning Models for Java and C#
73. RLPy: A Value-Function-Based Reinforcement Learning Framework for Education and Research

To get more information and download for using, click http://www.jmlr.org/mloss/

1.5 Classic Review

1. Learning internal representations by error propagation - Rumelhart, Hinton, et al. - 1986 (Show Context)
2. Support-vector networks - Cortes, Vapnik - 1995
3. A training algorithm for optimal margin classifiers - Boser, Guyon, et al. - 1992 (Show Context)
4. Spline Models for Observational Data - Wahba - 1990
5. Nonlinear component analysis as a kernel eigenvalue problem - Scholkopf, Smola, et al. - 1998
6. A Probabilistic Theory of Pattern Recognition - Devroye, Györfi, et al. - 1996 (Show Context)
7. On the uniform convergence of relative frequencies of events to their probabilities. Theory of Probability and its Applications - Vapnik, Chervonenkis - 1971
8. V.N.: Estimation of Dependences Based on Empirical Data - Vapnik - 1982 (Show Context)
9. Learnability and the Vapnik-Chervonenkis dimension - Blumer, Ehrenfeucht, et al. - 1989
10. Regularization theory and neural networks architectures - Girosi, Jones, et al. - 1995
11. An equivalence between sparse approximation and support vector machines - Girosi - 1997
12. Scale-sensitive dimensions, uniform convergence, and learnability - ALON, BEN-DAVID, et al. - 1993
13. Simpli support vector decision rules - Burges - 1996
14. Generalization performance of support vector machines and other pattern classifiers - Bartlett, Shawe-Taylor - 1999
15. A Theory of Learning and Generalization - Vidyasagar - 1997
16. Fat-shattering and the learnability of real-valued functions - Bartlett, Long, et al. - 1996

17. Y.: Learning process in an asymmetric threshold network. In: Disordered systems and biological organization - Cun - 1986 (Show Context)

18. A framework for structural risk minimization - Shawe-Taylor, Bartlett, et al. - 1996

19. The Glivenko-Cantelli problem, ten years later - Talagrand - 1996

20. sufficient conditions for the uniform convergence of means to their expectations," Theory Probab - "Necessary - 1981 (Show Context)

21. the uniform convergence of relative frequencies of events to their probabilities," Theory Probab - "On - 1971 (Show Context)

22. On the annealed vc entropy for margin classifiers: A statistical mechanics study - Opper - 1998

23. invariance in kernel-based methods - "Geometry - 1999

24. connection between regularization operators and support vector kernels - "The - 1998

25. necessary and sufficient conditions for consistency of the method of empirical risk minimization," Yearbook of the Academy of Sciences of the USSR - "The - 1989 (Show Context)

26. support vector kernels," in - Williamson, Smola, et al. - 1999

1.6 Conferences

1. International Conference on Machine Learning
   http://icml.cc/2013/

2. International Joint Conference on Artificial Intelligence
   http://www.ijcai.org

3. Pacific Rim International Conference on Artificial Intelligence
   http://ktw.mimos.my/pricai2012/

4. International Conference on Pattern Recognition
   http://www.icpr2014.org

5. International Conference on Document Analysis and Recognition
   http://www.icdar2013.org

6. International Conference on Automatic Face and Gesture Recognition
   http://fg2013.cse.sc.edu

7. International Conference on Artificial Neural Networks
   https://www.waset.org/Conferences

Chapter 2

Big Data

Big data is a broad term for data sets so large or complex that traditional data processing applications are inadequate. Challenges include analysis, capture, data curation, search, sharing, storage, transfer, visualization, and information privacy. The term often refers simply to the use of predictive analytics or other certain advanced methods to extract value from data, and seldom to a particular size of data set. Accuracy in big data may lead to more confident decision making. And better decisions can mean greater operational efficiency, cost reduction and reduced risk.

Analysis of data sets can find new correlations, to ”spot business trends, prevent diseases, combat crime and so on.” Scientists, business executives, practitioners of media, and advertising and governments alike regularly meet difficulties with large data sets in areas including Internet search, finance and business informatics. Scientists encounter limitations in e-Science work, including meteorology, genomics,connectomics, complex physics simulations,and biological and environmental research.[1]

2.1 Person

1. Wikipedia contributors. Vladimir Vapnik. Wikipedia, The Free Encyclopedia. https://en.wikipedia.org/w/index.php?title=Vladimir_Vapnik&oldid=690001724.

2. Wikipedia contributors. Lawrence Rabiner. Wikipedia, The Free Encyclopedia. https://en.wikipedia.org/w/index.php?title=Lawrence_Rabiner&oldid=660725634.

3. R. Tyrrell Rockafellar,University of Washington http://www.math.washington.edu/ rtr/mypage.html

4. Sergey Brin,Google Co-founder
   http://www.ted.com/speakers/sergey_brin

5. Jitendra Malik,University of California at Berkeley
   http://www.cs.berkeley.edu/ malik/

6. Jianbo Shi, University of Pennsylvania
   http://www.cis.upenn.edu/ jshi/

7. Jiawei Han,Abel Bliss Professor, Department of Computer Science
   http://web.engr.illinois.edu/ hanj/

8. Henning Schulzrinne,Professor in the Dept. of Computer Science
   http://www.cs.columbia.edu/ hgs/

2.2 Journals

2.3 Directions

1. Cloud/Grid/Stream Computing for Big Data
2. High Performance/Parallel Computing Platforms for Big Data
3. Autonomic Computing and Cyber-infrastructure, System Architectures, Design and Deployment
4. Energy-efficient Computing for Big Data
5. Programming Models and Environments for Cluster, Cloud, and Grid Computing to Support Big Data
6. Software Techniques andArchitectures in Cloud/Grid/Stream Computing
7. Big Data Open Platforms
8. New Programming Models for Big Data beyond Hadoop/MapReduce, STORM
9. Software Systems to Support Big Data Computing

2.4 Open Source Software

2.5 Classic Review

1. 2011. “Six Provocations for Big Data - Boyd, Crawford
2. Super crunchers, Bantam - Ayres - 2007
3. Survey sampling - Kish - 1965 (Show Context)
4. Big data: The next frontier for innovation, competition, and productivity, McKinsey Global Institute - Manyika, Chui - 2011 (Show Context)
5. Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data - Zikopoulos, Eaton - 2011
6. Practical text mining and statistical analysis for non-structured text data applications, 1st ed - Miner, Elder, et al. - 2012
7. Expand your digital horizon with Big Data. Forrester Research - Hopkins, Evelson - 2011
8. MapReduce: simplified data processing on large clusters - Dean, Ghemawat - 2008
9. k-anonymity: a model for protecting privacy - Sweeney - 2002
10. White paper, —Big Data Meets Big Data Analytics - Troester, SAS
11. Oracle Whitepaper (August, 2012), Oracle Information Architecture: An Architect’s Guide to Big Data - Sun, Heller
12. Ebook, —Strategic Guide to Big Data Analytics - Carr, Jackson - 2012

2.6 Conferences

Bibliography

[1] Wikipedia. International conference on machine learning — wikipedia, the free encyclopedia, 2015. [Online; accessed 18-November-2015].