机器学习开源库

Bob @ 2014 年 2 月 25 日星期二

目 录

1. Shogun 将军机器学习库

    SHOGUN 是一个关于大规模机器学习的工具包，它尤其关注支持向量机（SVM）。它提供 一个 SVM 的通用对象来支持不同的 SVM 实现，其中包括 LibSVM，SVMLight，SVMLin 及 GPDT。 每一种 SVM 都可支持多种核函数。SHOGUN 除了实现绝大部分常见的核函数，如线性核函数、 多项式核函数、高斯核函数及 Sigmoid 核函数等，同时还实现一些字符串核函数，如 Locality Improved、 Fischer、TOP、Spectrum、Weighted Degree 等。对于字符串核函数，还实现了 LINADD 优化。同时 SHOGUN 也支持自定义核函数。它的一个重要功能就是可以将许多个 子核函数作线性加权，并且这些子核函数不必作用于相同的定义域。子核函数的权重可 以通过多重核函数学习（Multiple Kernel Learning）获得。目前可以支持 SVM 处理二类分类问题和回归问题。同时 SHOGUN 还实现了其它线性方法，如线性判别分析（LDA）， 线性规划学习器 （LPM），（核函数）感知器和训练 HMM 模型的特征算法等。输入可以是 密集矩阵、称疏矩阵或者字符串，数据类型可是 int/short/double/char，并且可以将其中一种类型转换为 另外一种。此外，还可以使用预处理器对输入 feature 进行预处理（如减去均值）。

    Website：http://www.shogun-toolbox.org/page/home/

    官网也比较分析了机器学习库，具体请参考：http://www.shogun-toolbox.org/page/features/

2. Dlib 库

Dlib is a general purpose cross-platform C++ library designed using contract programming and modern C++ techniques. It is open source software and licensed under the Boost Software License. Theintroduction contains everything you need to know to get started using the library. However, if you have any questions, comments, or complaints feel free to email me or post in the sourceforge Forums.

Dlib 库有 cross-validation 方法，详细的实现请参考：

http://dlib.net/model_selection_ex.cpp.html

事实上就是源代码 model_selection_ex.cpp 文件的实现。

Dlib 包括了很多库，是一个很强大的库。主要特征如下表所示：

Major Features

- **Documentation**
  - Unlike a lot of open source projects, this one provides complete and precise documentation for every class and function. There are also debugging modes that check the documented preconditions for functions. When this is enabled it will catch the vast majority of bugs caused by calling functions incorrectly or using objects in an incorrect manner.
  - Lots of example programs are provided
  - *I consider the documentation to be the most important part of the library*. So if you find anything that isn't documented, isn't clear, or has out of date documentation, tell me and I will fix it.
- **High Quality Portable Code**
  - Good unit test coverage. The ratio of unit test lines of code to library lines of code is about 1 to 4.
  - The library is tested regularly on MS Windows, Linux, and Mac OS X systems. However, it should work on any POSIX system and has been used on Solaris, HPUX, and the BSDs.
  - No other packages are required to use the library. Only APIs that are provided by an out of the box OS are needed.
  - There is no installation or configure step needed before you can use the library. See the How to compile page for details.
  - All operating system specific code is isolated inside the OS abstraction layers which are kept as small as possible. The rest of the library is either layered on top of the OS abstraction layers or is pure ISO standard C++.
- **Threading**
  - The library provides a portable and simple threading API
  - A message passing pipe for inter-thread and inter-process communication
  - A timer object capable of generating events that are regularly spaced in time
  - Threaded objects
  - Threaded functions
  - Parallel for loops
  - A thread_pool with support for futures
- **Networking**
  - The library provides a portable and simple TCP sockets API
  - An object to help you make TCP based servers
  - iostream and streambuf objects that enables TCP sockets to interoperate with the C++ iostreams library
  - A simple HTTP server object you can use to embed a web server into your applications
  - A message passing pipe for inter-thread and inter-process communication
  - A tool used to implement algorithms using the Bulk Synchronous Parallel (BSP) computing model
- **Graphical User Interfaces**
  - The library provides a portable and simple core GUI API
  - Implemented on top of the core GUI API are numerous widgets
  - Unlike many other GUI toolkits, the entire dlib GUI toolkit is threadsafe
- **Numerical Algorithms**
  - A fast matrix object implemented using the expression templates technique and capable of using

- BLAS and LAPACK libraries when available.
  - Numerous linear algebra and mathematical operations are defined for the matrix object such as the singular value decomposition, transpose, trig functions, etc.
  - General purpose unconstrained non-linear optimization algorithms using the conjugate gradient, BFGS, and L-BFGS techniques
  - Levenberg-Marquardt for solving non-linear least squares problems
  - Box-constrained derivative-free optimization via the BOBYQA algorithm
  - An implementation of the Optimized Cutting Plane Algorithm
  - Several quadratic program solvers
  - Combinatorial optimization tools for solving optimal assignment and min cut/max flowproblems as well as the CKY algorithm for finding the most probable parse tree
  - A big integer object
  - A random number object
- **Machine Learning Algorithms**
  - Conventional SMO based Support Vector Machines for classification and regression
  - Reduced-rank methods for large-scale classification and regression
  - Relevance vector machines for classification and regression
  - General purpose multiclass classification tools
  - A Multiclass SVM
  - A tool for solving the optimization problem associated with structural support vector machines.
  - Structural SVM tools for sequence labeling
  - Structural SVM tools for solving assignment problems
  - Structural SVM tools for object detection in images
  - Structural SVM tools for labeling nodes in graphs
  - A large-scale SVM-Rank implementation
  - An online kernel RLS regression algorithm
  - An online SVM classification algorithm
  - Semidefinite Metric Learning
  - An online kernelized centroid estimator/novelty detector and offline support vector one-class classification
  - Clustering algorithms: linear or kernel k-means, Chinese Whispers, and Newman clustering.
  - Radial Basis Function Networks
  - Multi layer perceptrons
- **Graphical Model Inference Algorithms**
  - Join tree algorithm for exact inference in a Bayesian network.
  - Gibbs sampler markov chain monte carlo algorithm for approximate inference in a Bayesian network.
  - Routines for performing MAP inference in chain-structured, Potts, or general factor graphs.
- **Image Processing**
  - Routines for reading and writing common image formats.
  - Automatic color space conversion between various pixel types
  - Common image operations such as edge finding and morphological operations
  - Implementations of the SURF, HOG, and FHOG feature extraction algorithms.
  - Tools for detecting objects in images including a frontal face detector.

- **Data Compression and Integrity Algorithms**
    - o A CRC 32 object
    - o MD5 functions
    - o Various abstracted objects representing parts of data compression algorithms. Many forms of the PPM algorithm are included.
- **Testing**
    - o A thread safe logger object styled after the popular Java logger log4j
    - o A modular unit testing framework
    - o Various assert macros useful for testing preconditions
- **General Utilities**
    - o A type-safe object to convert between big and little endian byte orderings
    - o A command line parser with the ability to parse and validate command lines with various types of arguments and options
    - o An XML parser
    - o An object that can perform base64 conversions
    - o Many container classes
    - o Serialization support
    - o Many memory manager objects that implement different memory pooling strategies

3. MLDemo

MLDemos is an open-source visualization tool for machine learning algorithms created to help studying and understanding how several algorithms function and how their parameters affect and modify the results in problems of classification, regression, clustering, dimensionality reduction, dynamical systems and reward maximization. MLDemos is open-source and free for personal and academic use.

Website: http://mldemos.b4silio.com/

这是一个可视化的机器学习库，通过 qt 实现算法结果可视化

4. Libol 在线学习开源库

LIBOL is an open-source library that consists of a family of state-of-the-art online learning algorithms for machine learning and data mining research. It includes two categories of online learning methods: regular linear online learning algorithms and kernel-based online learning algorithms.

Linear methods

LIBOL consists of a large family of state-of-the-art linear onliine learning algorithms in two major categories: (i) first-order online learning, and (ii) second-order online learning. The list of linear algoirthms is summarzed below:

First-order Alogirhtms

Perceptron

Passive-Aggressive (PA) variants (PA, PA-I, PA-II)

Online Gradient Descent (OGD)

Second-order Alogirhtms

Second-order Perceptron (SOP)

Confidence-weighted learning (CW)

AROW: Adaptive Regularization Of Weights

SCW: Soft Confidence Weighted learning

Kernel methods

The list of kernel-based algorithms is summarized below:

Unbounded Algorithms

Kernel Perceptron

Kernel Passive-Aggressive (PA) variants (PA, PA-I, PA-II)

Kernel Online Gradient Descent (OGD)

Bounded Algorithms

Forgetron

RBP: Randomized Budget Perceptron

Projectron vairants (Projectron and Projectron++)

Bounded Online Gradient Descent (BOGD and BOGD++)

5. Weka

Weka 的全名是怀卡托智能分析环境（Waikato Environment for Knowledge Analysis），是一款免费的，非商业化的，基于 JAVA 环境下开源的机器学习以及数据挖掘软件。它和它的源代码可在其官方网站下载。有趣的是，该软件的缩写 WEKA 也是 New Zealand 独有的一种鸟名，而 Weka 的主要开发者同时恰好来自新西兰的 the University of Waikato。（来自百度百科）。

Weka 提供的功能有数据处理，特征选择、分类、回归、聚类、关联规则、可视化等。本文将对 Weka 的使用做一个简单的介绍，并通过简单的示例，使大家了解使用 weka 的流程。本文将仅对图形界面的操作做介绍，不涉及命令行和代码层面的东西。

IBM 文档库有一个序列介绍 Weka 的使用，分为三个部分，链接如下：

http://www.ibm.com/developerworks/cn/opensource/os-weka1/index.html

http://www.ibm.com/developerworks/cn/opensource/os-weka2/index.html

http://www.ibm.com/developerworks/cn/opensource/os-weka3/index.html

Weka 官网：http://www.cs.waikato.ac.nz/ml/weka/

Machine Learning Group at the University of Waikato 和 weka 一起出版了书《Data Mining:Practical Machine Learning Tools and Techniques》，已下载。

6. Orange

Open source data visualization and analysis for novice and experts. Data mining through visual programming or Python scripting. Components for machine learning. Add-ons for bioinformatics and text

mining. Packed with features for data analytics.

　　主要特征是可视化。而且有点类似于 matlab 的 simulink，图形化的编程模式可以快速的进行机器学习算法的实现和可视化展现，而且提供了许多脚本语言接口，方便编程使用。以后写文章可视化方面要仔细用这个工具。

　　官网：http://orange.biolab.si/features/

7. Tapkee

Tapkee is a C++ template library for dimensionality reduction with some bias on spectral methods. The Tapkee origins from the code developed during GSoC 2011 as the part of theShogun machine learning toolbox. The project aim is to provide efficient and flexible standalone library for dimensionality reduction which can be easily integrated to existing codebases. Tapkee leverages capabilities of effective Eigen3 linear algebra library and optionally makes use of the ARPACK eigensolver. The library uses CoverTree and VP-tree data structures to compute nearest neighbors. To achieve greater flexibility we provide a callback interface which decouples dimension reduction algorithms from the data representation and storage schemes.

　　Tapkee 是一个专门实现维数降解的开源库，主要是基于 c++模板进行的。

8. Bob：free signal-processing and machine learning toolbox

Bob 是一个开源的信号处理和机器学习库，基于 python 和 c++的混合编程方式，同时支持用户基于 python 和 c++开发组件，也很容易集成已有的第三方 c++库和 python 库，最后整个算法验证测试统一基于 python 环境，方便，不需要编译、链接等常规测试过程。Bob 的组成如下图：

Bob 网址：http://www.idiap.ch/software/bob/docs/releases/last/sphinx/html/

9. LibSVM
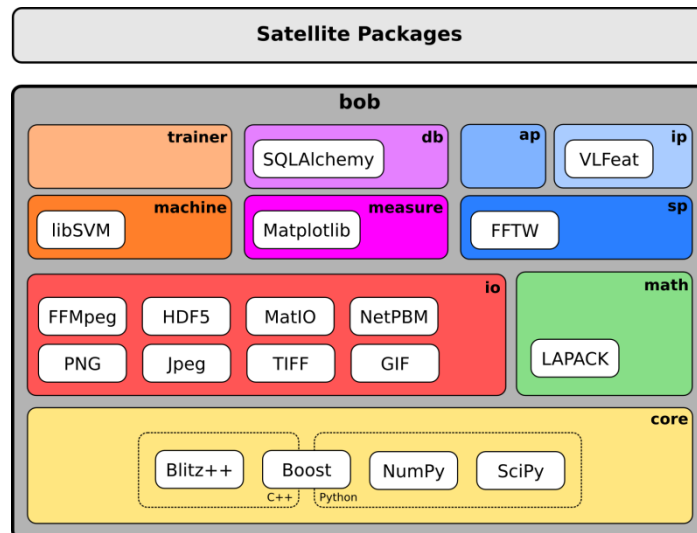
http://www.csie.ntu.edu.tw/~cjlin/libsvm/，还需要依赖 GNUPlot

10. a Multi-class Support Vector Machine package

MSVMpack is an open source package dedicated to multi-class support vector machines: SVMs which can handle classification problems with more than two classes without relying on decomposition methods. The aim is to provide a unified framework and implementation for all the different M-SVM models in a single package.

网页：http://www.loria.fr/~lauer/MSVMpack/MSVMpack.html

11. Fast Artificial Neural Network Library

Fast Artificial Neural Network Library is a free open source neural network library, which implements multilayer artificial neural networks in C with support for both fully connected and sparsely connected networks. Cross-platform execution in both fixed and floating point are supported. It includes a framework for easy handling of training data sets. It is easy to use, versatile, well documented, and fast. Bindings to more than 20 programming languages are available. An easy to read introduction article and a reference manual accompanies the library with examples and recommendations on how to use the library. Several graphical user interfaces are also available for the library.

神经网络的 C 代码实现，这个库引用较多，而且在一直更新。

网页：http://leenissen.dk/fann/wp/

12. Scikit-learn

Python 的机器学习库，文档较全，维护较好。网页：http://scikit-learn.org/stable/

13. Machine Learning Matlab Toolbox (MLT)

机器学习 matlab 工具箱，

http://mirlab.org/jang/matlab/toolbox/machineLearning/

14.网上资源（待整理）

To support the open source software movement, JMLR MLOSS publishes contributions related to implementations of non-trivial machine learning algorithms, toolboxes or even languages for scientific computing. Submission instructions are available here.

A Library for Locally Weighted Projection Regression

Stefan Klanke, Sethu Vijayakumar, Stefan Schaal; 9(Apr):623--626, 2008.

[abs][pdf]   [code][mloss.org]

Shark

Christian Igel, Verena Heidrich-Meisner, Tobias Glasmachers; 9(Jun):993--996, 2008.

[abs][pdf]   [code][mloss.org]

LIBLINEAR: A Library for Large Linear Classification

Rong-En Fan, Kai-Wei Chang, Cho-Jui Hsieh, Xiang-Rui Wang, Chih-Jen Lin; 9(Aug):1871--1874, 2008.

[abs][pdf]   [code][mloss.org]

JNCC2: The Java Implementation Of Naive Credal Classifier 2

Giorgio Corani, Marco Zaffalon; 9(Dec):2695--2698, 2008.

[abs][pdf]   [code][mloss.org]

Python Environment for Bayesian Learning: Inferring the Structure of Bayesian Networks from Knowledge and Data

Abhik Shah, Peter Woolf; 10(Feb):159--162, 2009.

[abs][pdf]   [code][mloss.org]

Nieme: Large-Scale Energy-Based Models

Francis Maes; 10(Mar):743--746, 2009.

[abs][pdf]   [code][mloss.org]

Java-ML: A Machine Learning Library

Thomas Abeel, Yves Van de Peer, Yvan Saeys; 10(Apr):931--934, 2009.

[abs][pdf]   [code][mloss.org]

Model Monitor (M^2): Evaluating, Comparing, and Monitoring Models

Troy Raeder, Nitesh V. Chawla; 10(Jul):1387--1390, 2009.

[abs][pdf]   [code][mloss.org]

Dlib-ml: A Machine Learning Toolkit

Davis E. King; 10(Jul):1755--1758, 2009.

[abs][pdf]   [code][mloss.org]

RL-Glue: Language-Independent Software for Reinforcement-Learning Experiments

Brian Tanner, Adam White; 10(Sep):2133--2136, 2009.

[abs][pdf]   [code][mloss.org]

DL-Learner: Learning Concepts in Description Logics

Jens Lehmann; 10(Nov):2639−2642, 2009.

[abs][pdf]   [code][mloss.org]

Error-Correcting Output Codes Library

Sergio Escalera, Oriol Pujol, Petia Radeva; 11(Feb):661−664, 2010.

[abs][pdf]   [code][mloss.org]

PyBrain

Tom Schaul, Justin Bayer, Daan Wierstra, Yi Sun, Martin Felder, Frank Sehnke, Thomas Rückstieß, Jürgen Schmidhuber; 11(Feb):743−746, 2010.

[abs][pdf]   [code][mloss.org]

Continuous Time Bayesian Network Reasoning and Learning Engine

Christian R. Shelton, Yu Fan, William Lam, Joon Lee, Jing Xu; 11(Mar):1137−1140, 2010.

[abs][pdf]   [code][mloss.org]

SFO: A Toolbox for Submodular Function Optimization

Andreas Krause; 11(Mar):1141−1144, 2010.

[abs][pdf]   [code][mloss.org]

MOA: Massive Online Analysis

Albert Bifet, Geoff Holmes, Richard Kirkby, Bernhard Pfahringer; 11(May):1601−1604, 2010.

[abs][pdf]   [code][mloss.org]

FastInf: An Efficient Approximate Inference Library

Ariel Jaimovich, Ofer Meshi, Ian McGraw, Gal Elidan; 11(May):1733−1736, 2010.

[abs][pdf]   [code][mloss.org]

The SHOGUN Machine Learning Toolbox

Sören Sonnenburg, Gunnar Rätsch, Sebastian Henschel, Christian Widmer, Jonas Behr, Alexander Zien, Fabio de Bona, Alexander Binder, Christian Gehl, Vojtěch Franc; 11(Jun):1799−1802, 2010.

[abs][pdf]   [code][mloss.org]

A Surrogate Modeling and Adaptive Sampling Toolbox for Computer Based Design

Dirk Gorissen, Ivo Couckuyt, Piet Demeester, Tom Dhaene, Karel Crombecq; 11(Jul):2051−2055, 2010.

[abs][pdf]   [code][mloss.org]

Model-based Boosting 2.0

Torsten Hothorn, Peter Bühlmann, Thomas Kneib, Matthias Schmid, Benjamin Hofner; 11(Aug):2109−2113, 2010.

[abs][pdf]   [code][mloss.org]

libDAI: A Free and Open Source C++ Library for Discrete Approximate Inference in Graphical Models

Joris M. Mooij; 11(Aug):2169−2173, 2010.

[abs][pdf]   [code][mloss.org]

Gaussian Processes for Machine Learning (GPML) Toolbox

Carl Edward Rasmussen, Hannes Nickisch; 11(Nov):3011−3015, 2010.

[abs][pdf]   [code][mloss.org]

CARP: Software for Fishing Out Good Clustering Algorithms

Volodymyr Melnykov, Ranjan Maitra; 12(Jan):69−73, 2011.

[abs][pdf]   [code][mloss.org]

The arules R-Package Ecosystem: Analyzing Interesting Patterns from Large Transaction Data Sets

Michael Hahsler, Sudheer Chelluboina, Kurt Hornik, Christian Buchta; 12(Jun):2021−2025, 2011.

[abs][pdf]   [code][mloss.org]

MSVMpack: A Multi-Class Support Vector Machine Package

Fabien Lauer, Yann Guermeur; 12(Jul):2293−2296, 2011.

[abs][pdf]   [code][mloss.org]

Waffles: A Machine Learning Toolkit

Michael Gashler; 12(Jul):2383−2387, 2011.

[abs][pdf]   [code][mloss.org]

MULAN: A Java Library for Multi-Label Learning

Grigorios   Tsoumakas,   Eleftherios   Spyromitros-Xioufis,   Jozef   Vilcek,   Ioannis   Vlahavas;
12(Jul):2411−2414, 2011.

[abs][pdf]   [code][mloss.org]

LPmade: Link Prediction Made Easy

Ryan N. Lichtenwalter, Nitesh V. Chawla; 12(Aug):2489−2492, 2011.

[abs][pdf]   [code][mloss.org]

Scikit-learn: Machine Learning in Python

*Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier
Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre
Passos,   David   Cournapeau,   Matthieu   Brucher,   Matthieu   Perrot,   Édouard   Duchesnay*;
12(Oct):2825−2830, 2011.

[abs][pdf]   [code][mloss.org]

The Stationary Subspace Analysis Toolbox

Jan Saputra Müller, Paul von Bünau, Frank C. Meinecke, Franz J. Király, Klaus-Robert Müller;
12(Oct):3065−3069, 2011.

[abs][pdf]   [code][mloss.org]

MULTIBOOST: A Multi-purpose Boosting Package

Djalel Benbouzid, Róbert Busa-Fekete, Norman Casagrande, François-David Collin, Balázs Kégl; 13(Mar):549−553, 2012.

[abs][pdf]   [code][mloss.org]

ML-Flex: A Flexible Toolbox for Performing Classification Analyses In Parallel

Stephen R. Piccolo, Lewis J. Frey; 13(Mar):555−559, 2012.

[abs][pdf]   [code][sourceforge.net]

GPLP: A Local and Parallel Computation Toolbox for Gaussian Process Regression

Chiwoo Park, Jianhua Z. Huang, Yu Ding; 13(Mar):775−779, 2012.

[abs][pdf]   [code][mloss.org]

NIMFA : A Python Library for Nonnegative Matrix Factorization

Marinka Žitnik, Blaž Zupan; 13(Mar):849−853, 2012.

[abs][pdf]   [code][mloss.org]

The huge Package for High-dimensional Undirected Graph Estimation in R

Tuo Zhao, Han Liu, Kathryn Roeder, John Lafferty, Larry Wasserman; 13(Apr):1059−1062, 2012.

[abs][pdf]   [code][cran.r-project.org]

glm-ie: Generalised Linear Models Inference & Estimation Toolbox

Hannes Nickisch; 13(May):1699−1703, 2012.

[abs][pdf]   [code][mloss.org]

Jstacs: A Java Framework for Statistical Analysis and Classification of Biological Sequences

Jan Grau, Jens Keilwagen, André Gohr, Berit Haldemann, Stefan Posch, Ivo Grosse; 13(Jun):1967−1971, 2012.

[abs][pdf]   [code][mloss.org]

Pattern for Python

Tom De Smedt, Walter Daelemans; 13(Jun):2063−2067, 2012.

[abs][pdf]   [code][mloss.org]

DEAP: Evolutionary Algorithms Made Easy

Félix-Antoine Fortin, François-Michel De Rainville, Marc-André Gardner, Marc Parizeau, Christian Gagné; 13(Jul):2171−2175, 2012.

[abs][pdf]   [code][deap.gel.ulaval.ca]

A Topic Modeling Toolbox Using Belief Propagation

Jia Zeng; 13(Jul):2233−2236, 2012.

[abs][pdf]   [code][mloss.org]

PREA: Personalized Recommendation Algorithms Toolkit

Joonseok Lee, Mingxuan Sun, Guy Lebanon; 13(Sep):2699−2703, 2012.

[abs][pdf]   [code][mloss.org]

Oger: Modular Learning Architectures For Large-Scale Sequential Processing

David Verstraeten, Benjamin Schrauwen, Sander Dieleman, Philemon Brakel, Pieter Buteneers, Dejan Pecevski; 13(Oct):2995−2998, 2012.

[abs][pdf]   [code][mloss.org]

Sally: A Tool for Embedding Strings in Vector Spaces

Konrad Rieck, Christian Wressnegger, Alexander Bikadorov; 13(Nov):3247−3251, 2012.

[abs][pdf]   [code][mloss.org]

DARWIN: A Framework for Machine Learning and Computer Vision Research and Development

Stephen Gould; 13(Dec):3533−3537, 2012.

[abs][pdf]   [code][mloss.org]

SVDFeature: A Toolkit for Feature-based Collaborative Filtering

Tianqi Chen, Weinan Zhang, Qiuxia Lu, Kailong Chen, Zhao Zheng, Yong Yu; 13(Dec):3619−3622, 2012.

[abs][pdf]   [code][mloss.org]

A C++ Template-Based Reinforcement Learning Library: Fitting the Code to the Mathematics

Hervé Frezza-Buet, Matthieu Geist; 14(Feb):625−628, 2013.

[abs][pdf]   [code][malis.metz.supelec.fr]

MLPACK: A Scalable C++ Machine Learning Library

Ryan R. Curtin, James R. Cline, N. P. Slagle, William B. March, Parikshit Ram, Nishant A. Mehta, Alexander G. Gray; 14(Mar):801−805, 2013.

[abs][pdf]   [code][mloss.org]

GPstuff: Bayesian Modeling with Gaussian Processes

Jarno Vanhatalo, Jaakko Riihimäki, Jouni Hartikainen, Pasi Jylänki, Ville Tolvanen, Aki Vehtari; 14(Apr):1175−1179, 2013.

[abs][pdf]   [code][mloss.org]

JKernelMachines: A Simple Framework for Kernel Machines

David Picard, Nicolas Thome, Matthieu Cord; 14(May):1417−1421, 2013.

[abs][pdf][bib]   [code][mloss.org]

Orange: Data Mining Toolbox in Python

Janez Demšar, Tomaž Curk, Aleš Erjavec, Črt Gorup, Tomaž Hočevar, Mitar Milutinovič, Martin Možina, Matija Polajnar, Marko Toplak, Anže Starič, Miha Štajdohar, Lan Umek, Lan Žagar, Jure Žbontar, Marinka Žitnik, Blaž Zupan; 14(Aug):2349−2353, 2013.

[abs][pdf][bib]   [code][mloss.org]

Tapkee: An Efficient Dimension Reduction Library

Sergey Lisitsyn, Christian Widmer, Fernando J. Iglesias Garcia; 14(Aug):2355−2359, 2013.

[abs][pdf][bib]　[code][mloss.org]

The CAM Software for Nonnegative Blind Source Separation in R-Java

Niya Wang, Fan Meng, Li Chen, Subha Madhavan, Robert Clarke, Eric P. Hoffman, Jianhua Xuan, Yue Wang; 14(Sep):2899−2903, 2013.

[abs][pdf][bib]　[code][mloss.org]

QuantMiner for Mining Quantitative Association Rules

Ansaf Salleb-Aouissi, Christel Vrain, Cyril Nortet, Xiangrong Kong, Vivek Rathod, Daniel Cassard; 14(Oct):3153−3157, 2013.

[abs][pdf][bib]　[code][github.com]

Divvy: Fast and Intuitive Exploratory Data Analysis

Joshua M. Lewis, Virginia R. de Sa, Laurens van der Maaten; 14(Oct):3159−3163, 2013.

[abs][pdf][bib]　[code][mloss.org]

GURLS: A Least Squares Library for Supervised Learning

Andrea Tacchetti, Pavan K. Mallapragada, Matteo Santoro, Lorenzo Rosasco; 14(Oct):3201−3205, 2013.

[abs][pdf][bib]　[code][github.com]

网上资源：

1）mlpack is a C++ machine learning library.

2）PLearn is a C++ library aimed at research and development in the field of

statistical machine learning algorithms. Its originality is to allow to easily express, directly in C++ in a straightforward manner, complex non-linear functions to be optimized.

3）Waffles- C++ Machine Learning。

4）Torch7 provides a Matlab-like environment for state-of-the-art machine learning

algorithms. It is easy to use and provides a very efficient implementation

5）SHARK is a modular C++ library for the design and optimization of adaptive

systems. It provides methods for linear and nonlinear optimization, in particular evolutionary and gradient-based algorithms, kernel-based learning algorithms and neural networks, and various other machine learning techniques. SHARK serves as a toolbox to support real world applications as well as research in different domains of computational intelligence and machine learning. The sources are compatible with the following platforms: Windows, Solaris, MacOS X, and Linux.

6）Dlib-ml is an open source library, targetedat both engineers and research

scientists, which aims to provide a similarly rich environment fordeveloping machine learning software in the C++ language.

7) Eblearn is an object-oriented C++ library that implements various machine learning models, including energy-based learning, gradient-based learning for machine composed of multiple heterogeneous modules. In particular, the library provides a complete set of tools for building, training, and running convolutional networks.

8)   Machine Learning Open Source Software :Journal of Machine Learning Research: http://jmlr.csail.mit.edu/mloss/.

9) search in google: c++ site:jmlr.csail.mit.edu filetype:pdf   , Machine Learning Toolkit

10) SIGMA: Large-Scale and Parallel Machine-Learning Tool Kit

11) http://sourceforge.net/directory/science-engineering/ai/machinelearning/os:windows/freshness:recently-updated/

-------------   2012.9.12   ---------

12) ELF: ensemble learning framework。特点：c++，监督学习，使用了 intel 的 IPP

和 MKL，training speed 和 accuracy 是主要目标。http://elf-project.sourceforge.net/

------------- 2012.11.03   ---------

13) http://mloss.org/software/ machine learning open sources software。算是一个索引网站吧。

14) http://drwn.anu.edu.au/index.html

------------- 2013.4.09   ---------