

- [首页](#)
- [开源项目](#)
 - [国产开源项目](#)
 - [项目分类](#)
 - [最新收录项目](#)
 -
 - [Java 开源软件](#)
 - [C# 开源软件](#)
 - [PHP 开源软件](#)
 - [C/C++ 开源软件](#)
 - [Ruby 开源软件](#)
 - [Python 开源软件](#)
 - [Go开源软件](#)
 - [JS开源软件](#)
- [问答](#)
 - [技术问答 »](#)
 - [技术分享 »](#)
 - [IT大杂烩 »](#)
 - [职业生涯 »](#)
 - [站务/建议 »](#)
 - [支付宝专区 »](#)
 - [MoPaaS专区 »](#)
 - [开源硬件专区 »](#)
- [代码](#)
- [博客](#)
- [翻译](#)
- [资讯](#)
- [专题](#)
 - [源创会 视频](#)
 - [高手问答 访谈](#)
 - [周刊 乱弹](#)
 - [公司开源导航页](#)
 -
 - [Android开发专区](#)
 - [iOS开发专区](#)
 - [iOS代码库](#)
 - [Windows Phone](#)
- [城市圈](#)

当前访客身份：游客 [[登录](#) | [加入开源中国](#)]
[开源中国](#)

开源资讯

当前位置：
[资讯首页](#) » [综合资讯](#) » 显示全文 ， [投递新闻»](#)

资讯、软件、分享、代码、

搜 索

推荐！国外程序员整理的机器学习资源大全

[oschina](#) 发布于：2014年07月22日 ([56评](#))

分享到：

收藏 +750

本列表选编了一些机器学习领域牛B的框架、库以及软件（按编程语言排序）。

C++

计算机视觉

- [CCV](#) —基于C语言/提供缓存/核心的机器视觉库，新颖的机器视觉库
- [OpenCV](#)—它提供C++, C, Python, Java 以及 MATLAB接口，并支持Windows, Linux, Android and Mac OS操作系统。

通用机器学习

- [MLPack](#)
- [DLib](#)
- [ecogg](#)
- [shark](#)

Closure

通用机器学习

- [Closure Toolbox](#)—Clojure语言库与工具的分类目录

Go

自然语言处理

- [go-porterstemmer](#)—一个Porter词干提取算法的原生Go语言净室实现
- [paicehusk](#)—Paice/Husk词干提取算法的Go语言实现
- [snowball](#)—Go语言版的Snowball词干提取器

通用机器学习

- [Go Learn](#)— Go语言机器学习库
- [go-pr](#) —Go语言机器学习包.
- [bayesian](#)—Go语言朴素贝叶斯分类库。
- [go-galib](#)—Go语言遗传算法库。

数据分析/数据可视化

- [go-graph](#)—Go语言图形库。
- [SVGo](#)—Go语言的SVG生成库。

Java

自然语言处理

- [CoreNLP](#)—斯坦福大学的CoreNLP提供一系列的自然语言处理工具，输入原始英语文本，可以给出单词的基本形式（下面Stanford开头的几个工具都包含其中）。
- [Stanford Parser](#)—一个自然语言解析器。
- [Stanford POS Tagger](#) —一个词性分类器。
- [Stanford Name Entity Recognizer](#)—Java实现的名词识别器
- [Stanford Word Segmenter](#)—分词器，很多NLP工作中都要用到的标准预处理步骤。
- [Tregex, Tsurgeon and Semgrex](#) —用来在树状数据结构中进行模式匹配，基于树关系以及节点匹配的正则表达式（名字是“tree regular expressions”的缩写）。
- [Stanford Phrasal](#):最新的基于统计短语的机器翻译系统，_java编写
- [Stanford Tokens Regex](#)—用以定义文本模式的框架。
- [Stanford Temporal Tagger](#)—SUTime是一个识别并标准化时间表达式的库。
- [Stanford SPIED](#)—在种子集上使用模式，以迭代方式从无标签文本中学习字符实体
- [Stanford Topic Modeling Toolbox](#) —为社会科学及其他希望分析数据集的人员提供的主题建模工具。
- [Twitter Text Java](#)—Java实现的推特文本处理库
- [MALLET](#) --基于Java的统计自然语言处理、文档分类、聚类、主题建模、信息提取以及其他机器学习文本应用包。
- [OpenNLP](#)—处理自然语言文本的机器学习工具包。
- [LingPipe](#) —使用计算机语言学处理文本的工具包。

通用机器学习

- [MLlib in Apache Spark](#)—Spark中的分布式机器学习程序库

- [Mahout](#) —分布式的机器学习库
- [Stanford Classifier](#) —斯坦福大学的分类器
- [Weka](#)—Weka是数据挖掘方面的机器学习算法集。
- [ORYX](#)—提供一个简单的大规模实时机器学习/预测分析基础架构。

数据分析/数据可视化

- [Hadoop](#)—大数据分析平台
- [Spark](#)—快速通用的大规模数据处理引擎。
- [Impala](#) —为Hadoop实现实时查询

Javascript

自然语言处理

- [Twitter-text-js](#) —JavaScript实现的推特文本处理库
- [NLP.js](#) —javascript及coffeescript编写的NLP工具
- [natural](#)—Node下的通用NLP工具
- [Knwl.js](#)—JS编写的自然语言处理器

数据分析/数据可视化

- [D3.js](#)
- [High Charts](#)
- [NVD3.js](#)
- [dc.js](#)
- [chartjs](#)
- [dimple](#)
- [amCharts](#)

通用机器学习

- [Convnet.js](#)—训练深度学习模型的JavaScript库。
- [Clustering.js](#)—用JavaScript实现的聚类算法，供Node.js及浏览器使用。
- [Decision Trees](#)—Node.js实现的决策树，使用ID3算法。
- [Node-fann](#) —Node.js下的快速人工神经网络库。
- [Kmeans.js](#)—k-means算法的简单Javascript实现，供Node.js及浏览器使用。
- [LDA.js](#) —供Node.js用的LDA主题建模工具。
- [Learning.js](#)—逻辑回归/c4.5决策树的JavaScript实现
- [Machine Learning](#)—Node.js的机器学习库。
- [Node-SVM](#)—Node.js的支持向量机
- [Brain](#) —JavaScript实现的神经网络
- [Bayesian-Bandit](#) —贝叶斯强盗算法的实现，供Node.js及浏览器使用。

Julia

通用机器学习

- [PGM](#)—Julia实现的概率图模型框架。
- [DA](#)—Julia实现的正则化判别分析包。
- [Regression](#)—回归分析算法包（如线性回归和逻辑回归）。
- [Local Regression](#) —局部回归，非常平滑！
- [Naive Bayes](#) —朴素贝叶斯的简单Julia实现
- [Mixed Models](#) —（统计）混合效应模型的Julia包
- [Simple MCMC](#) —Julia实现的基本mcmc采样器

- [Distance](#)—Julia实现的距离评估模块
- [Decision Tree](#) —决策树分类器及回归分析器
- [Neural](#) —Julia实现的神经网络
- [MCMC](#) —Julia下的MCMC工具
- [GLM](#) —Julia写的广义线性模型包
- [Online Learning](#)
- [GLMNet](#) —GMLNet的Julia包装版，适合套索/弹性网模型。
- [Clustering](#)—数据聚类的基本函数：k-means, dp-means等。
- [SVM](#)—Julia下的支持向量机。
- [Kernal Density](#)—Julia下的核密度估计器
- [Dimensionality Reduction](#)—降维算法
- [NMF](#) —Julia下的非负矩阵分解包
- [ANN](#)—Julia实现的神经网络

自然语言处理

- [Topic Models](#) —Julia下的主题建模
- [Text Analysis](#)—Julia下的文本分析包

数据分析/数据可视化

- [Graph Layout](#) —纯Julia实现的图布局算法。
- [Data Frames Meta](#) —DataFrames的元编程工具。
- [Julia Data](#)—处理表格数据的Julia库
- [Data Read](#)—从Stata、SAS、SPSS读取文件
- [Hypothesis Tests](#)—Julia中的假设检验包
- [Gladfly](#) —Julia编写的灵巧的统计绘图系统。
- [Stats](#)—Julia编写的统计测试函数包
- [RDataSets](#) —读取R语言中众多可用的数据集的Julia函数包。
- [DataFrames](#) —处理表格数据的Julia库。
- [Distributions](#)—概率分布及相关函数的Julia包。
- [Data Arrays](#) —元素值可以为空的数据结构。
- [Time Series](#)—Julia的时间序列数据工具包。
- [Sampling](#)—Julia的基本采样算法包

杂项/演示文稿

- [DSP](#) —数字信号处理
- [JuliaCon Presentations](#)—Julia大会上的演示文稿
- [SignalProcessing](#)—Julia的信号处理工具
- [Images](#)—Julia的图片库

Lua

通用机器学习

- [Torch7](#)
 - [cephes](#) —Cephes数学函数库，包装成Torch可用形式。提供并包装了超过180个特殊的数学函数，由Stephen L. Moshier开发，是SciPy的核心，应用于很多场合。
 - [graph](#) —供Torch使用的图形包。
 - [randomkit](#)—从Numpy提取的随机数生成包，包装成Torch可用形式。
 - [signal](#) —Torch-7可用的信号处理工具包，可进行FFT, DCT, Hilbert, cepstrums, stft等变换。

- [nn](#) —Torch可用的神经网络包。
- [nngraph](#) —为nn库提供图形计算能力。
- [nnx](#)—一个不稳定实验性的包，扩展Torch内置的nn库。
- [optim](#)—Torch可用的优化算法库，包括 SGD, Adagrad, 共轭梯度算法, LBFGS, RProp等算法。
- [unsup](#)—Torch下的非监督学习包。提供的模块与nn(LinearPsd, ConvPsd, AutoEncoder, ...)及独立算法 (k-means, PCA)等兼容。
- [manifold](#)—操作流形的包。
- [svm](#)—Torch的支持向量机库。
- [lbfgs](#)—将liblbfgs包装为FFI接口。
- [vowpalwabbit](#) —老版的vowpalwabbit对torch的接口。
- [OpenGM](#)—OpenGM是C++编写的图形建模及推断库，该binding可以用Lua以简单的方式描述图形，然后用OpenGM优化。
- [sphagetti](#) —MichaelMathieu为torch7编写的稀疏线性模块。
- [LuaSHKit](#) —将局部敏感哈希库SHKit包装成lua可用形式。
- [kernel smoothing](#) —KNN、核权平均以及局部线性回归平滑器
- [cutorch](#)—torch的CUDA后端实现
- [cunn](#) —torch的CUDA神经网络实现。
- [imggraph](#)—torch的图像/图形库，提供从图像创建图形、分割、建立树、又转化回图像的例程
- [videograph](#)—torch的视频/图形库，提供从视频创建图形、分割、建立树、又转化回视频的例程
- [saliency](#) —积分图像的代码和工具，用来从快速积分直方图中寻找兴趣点。
- [stitch](#) —使用hugin拼合图像并将其生成视频序列。
- [sfm](#)—运动场景束调整/结构包
- [fex](#) —torch的特征提取包，提供SIFT和dSIFT模块。
- [OverFeat](#)—当前最高水准的通用密度特征提取器。
- [Numeric Lua](#)
- [Lunatic Python](#)
- [SciLua](#)
- [Lua - Numerical Algorithms](#)
- [Lunum](#)

演示及脚本

- [Core torch7 demos repository](#).核心torch7演示程序库
 - 线性回归、逻辑回归
 - 人脸检测（训练和检测是独立的演示）
 - 基于mst的断词器
 - train-a-digit-classifier
 - train-autoencoder
 - optical flow demo
 - train-on-housenumbers
 - train-on-cifar
 - tracking with deep nets
 - kinect demo
 - 滤波可视化
 - saliency-networks
- [Training a Convnet for the Galaxy-Zoo Kaggle challenge\(CUDA demo\)](#)
- [Music Tagging](#)—torch7下的音乐标签脚本

- [torch-datasets](#) 读取几个流行的数据集的脚本，包括：
 - BSR 500
 - CIFAR-10
 - COIL
 - Street View House Numbers
 - MNIST
 - NORB
- [Atari2600](#) —在Arcade Learning Environment模拟器中用静态帧生成数据集的脚本。

Matlab

计算机视觉

- [Contourlets](#) —实现轮廓波变换及其使用函数的MATLAB源代码
- [Shearlets](#)—剪切波变换的MATLAB源码
- [Curvelets](#)—Curvelet变换的MATLAB源码（Curvelet变换是对小波变换向更高维的推广，用来在不同尺度角度表示图像。）
- [Bandlets](#)—Bandlets变换的MATLAB源码

自然语言处理

- [NLP](#) —一个Matlab的NLP库

通用机器学习

- [Training a deep autoencoder or a classifier on MNIST digits](#)—在MNIST字符数据集上训练一个深度的autoencoder或分类器[深度学习]。
- [t-Distributed Stochastic Neighbor Embedding](#) —获奖的降维技术，特别适合于高维数据集的可视化
- [Spider](#)—Matlab机器学习的完整面向对象环境。
- [LibSVM](#) —支持向量程序库
- [LibLinear](#) —大型线性分类程序库
- [Machine Learning Module](#) —M. A .Girolami教授的机器学习课程，包括PDF，讲义及代码。
- [Caffe](#)—考虑了代码清洁、可读性及速度的深度学习框架
- [Pattern Recognition Toolbox](#) —Matlab中的模式识别工具包，完全面向对象

数据分析/数据可视化

- [matlab_gbl](#)—处理图像的Matlab包
- [gamic](#)—图像算法纯Matlab高效实现，对MatlabBGL的mex函数是个补充。

.NET

计算机视觉

- [OpenCVDotNet](#) —包装器，使.NET程序能使用OpenCV代码
- [Emgu CV](#)—跨平台的包装器，能在Windows, Linus, Mac OS X, iOS, 和Android上编译。

自然语言处理

- [Stanford.NLP for .NET](#) —斯坦福大学NLP包在.NET上的完全移植，还可作为NuGet包进行预编译。

通用机器学习

- [Accord.MachineLearning](#) —支持向量机、决策树、朴素贝叶斯模型、K-means、高斯混合模型和机器学习应用的通用算法，例如：随机抽样一致性算法、交叉验证、网格搜索。这个包是Accord.NET框架的一部分。
- [Vulpes](#)—F#语言实现的Deep belief和深度学习包，它在Alea.cuBase下利用CUDA GPU来执行。
- [Encog](#) —先进的神经网络和机器学习框架，包括用来创建多种网络的类，也支持神经网络需要的数据规则化及处理的类。它的训练采用多线程弹性传播。它也能使用GPU加快处理时间。提供了图形化界面来帮助建模和训练神经网络。

- [Neural Network Designer](#) —这是一个数据库管理系统和神经网络设计器。设计器用WPF开发，也是一个UI，你可以设计你的神经网络、查询网络、创建并配置聊天机器人，它能问问题，并从你的反馈中学习。这些机器人甚至可以从网络搜集信息用来输出，或是用来学习。

数据分析/数据可视化

- [numl](#) —numl这个机器学习库，目标就是简化预测和聚类的标准建模技术。
- [Math.NET Numerics](#) — Math.NET项目的数值计算基础，着眼提供科学、工程以及日常数值计算的方法和算法。支持 Windows, Linux 和 Mac上的 .Net 4.0, .Net 3.5 和 Mono , Silverlight 5, WindowsPhone/SL 8, WindowsPhone 8.1 以及装有 PCL Portable Profiles 47 及 344的 Windows 8 , 装有 Xamarin的Android/iOS 。
- [Sho](#) — Sho是数据分析和科学计算的交互式环境，可以让你将脚本（IronPython语言）和编译的代码（.NET）无缝连接，以快速灵活的建立原型。这个环境包括强大高效的库，如线性代数、数据可视化，可供任何.NET语言使用，还为快速开发提供了功能丰富的交互式shell。

Python

计算机视觉

- [SimpleCV](#)—开源的计算机视觉框架，可以访问如OpenCV等高性能计算机视觉库。使用Python编写，可以在Mac、Windows以及Ubuntu上运行。

自然语言处理

- [NLTK](#) —一个领先的平台，用来编写处理人类语言数据的Python程序
- [Pattern](#)—Python可用的web挖掘模块，包括自然语言处理、机器学习等工具。
- [TextBlob](#)—为普通自然语言处理任务提供一致的API，以NLTK和Pattern为基础，并和两者都能很好兼容。
- [jieba](#)—中文断词工具。
- [SnowNLP](#) —中文文本处理库。
- [loso](#)—另一个中文断词库。
- [genius](#) —基于条件随机域的中文断词库。
- [nut](#) —自然语言理解工具包。

通用机器学习

- [Bayesian Methods for Hackers](#) —Python语言概率规划的电子书
- [MLlib in Apache Spark](#)—Spark下的分布式机器学习库。
- [scikit-learn](#)—基于SciPy的机器学习模块
- [graphlab-create](#) —包含多种机器学习模块的库（回归，聚类，推荐系统，图分析等），基于可以磁盘存储的DataFrame。
- [BigML](#)—连接外部服务器的库。
- [pattern](#)—Python的web挖掘模块
- [NuPIC](#)—Numenta公司的智能计算平台。
- [Pylearn2](#)—基于Theano的机器学习库。
- [hebel](#) —Python编写的使用GPU加速的深度学习库。
- [gensim](#)—主题建模工具。
- [PyBrain](#)—另一个机器学习库。
- [Crab](#) —可扩展的、快速推荐引擎。
- [python-recsys](#) —Python实现的推荐系统。
- [thinking bayes](#)—关于贝叶斯分析的书籍
- [Restricted Boltzmann Machines](#) —Python实现的受限波尔兹曼机。[深度学习]。
- [Bolt](#) —在线学习工具箱。
- [CoverTree](#) —cover tree的Python实现，scipy.spatial.kdtree便捷的替代。
- [nilearn](#)—Python实现的神经影像学机器学习库。
- [Shogun](#)—机器学习工具箱。
- [Pyevolve](#) —遗传算法框架。
- [Caffe](#) —考虑了代码清洁、可读性及速度的深度学习框架

- [breze](#)—深度及递归神经网络的程序库，基于Theano。

数据分析/数据可视化

- [SciPy](#) —基于Python的数学、科学、工程开源软件生态系统。
- [NumPy](#)—Python科学计算基础包。
- [Numba](#) —Python的低级虚拟机JIT编译器，Cython and NumPy的开发者编写，供科学计算使用
- [NetworkX](#) —为复杂网络使用的高效软件。
- [Pandas](#)—这个库提供了高性能、易用的数据结构及数据分析工具。
- [Open Mining](#)—Python中的商业智能工具（Pandas web接口）。
- [PyMC](#) —MCMC采样工具包。
- [zipline](#)—Python的算法交易库。
- [PyDy](#)—全名Python Dynamics，协助基于NumPy, SciPy, IPython以及 matplotlib的动态建模工作流。
- [SymPy](#) —符号数学Python库。
- [statsmodels](#)—Python的统计建模及计量经济学库。
- [astropy](#) —Python天文学程序库，社区协作编写
- [matplotlib](#) —Python的2D绘图库。
- [bokeh](#)—Python的交互式Web绘图库。
- [plotly](#) —Python and matplotlib的协作web绘图库。
- [vincent](#)—将Python数据结构转换为Vega可视化语法。
- [d3py](#)—Python的绘图库，基于D3.js。
- [ggplot](#) —和R语言里的ggplot2提供同样的API。
- [Kartograph.py](#)—Python中渲染SVG图的库，效果漂亮。
- [pygal](#)—Python下的SVG图表生成器。
- [pycascading](#)

杂项脚本/iPython笔记/代码库

- [pattern_classification](#)
- [thinking stats 2](#)
- [hyperopt](#)
- [numpic](#)
- [2012-paper-diginorm](#)
- [ipython-notebooks](#)
- [decision-weights](#)
- [Sarah Palin LDA](#) —Sarah Palin关于主题建模的电邮。
- [Diffusion Segmentation](#) —基于扩散方法的图像分割算法集合。
- [Scipy Tutorials](#) —SciPy教程，已过时，请查看scipy-lecture-notes
- [Crab](#)—Python的推荐引擎库。
- [BayesPy](#)—Python中的贝叶斯推断工具。
- [scikit-learn tutorials](#)—scikit-learn学习笔记系列
- [sentiment-analyzer](#) —推特情绪分析器
- [group-lasso](#)—坐标下降算法实验，应用于（稀疏）群套索模型。
- [mne-python-notebooks](#)—使用 mne-python进行EEG/MEG数据处理的IPython笔记
- [pandas cookbook](#)—使用Python pandas库的方法书。
- [climin](#)—机器学习的优化程序库，用Python实现了梯度下降、LBFGS、rmsprop、adadelta 等算法。

Kaggle竞赛源代码

- [wiki challenge](#) —Kaggle上一个维基预测挑战赛 Dell Zhang解法的实现。

- [kaggle insults](#)—Kaggle上“从社交媒体评论中检测辱骂”竞赛提交的代码
- [kaggle_acquire-valued-shoppers-challenge](#)—Kaggle预测回头客挑战赛的代码
- [kaggle-cifar](#) —Kaggle上CIFAR-10 竞赛的代码，使用cuda-convnet
- [kaggle-blackbox](#) —Kaggle上blackbox赛代码，关于深度学习。
- [kaggle-accelerometer](#) —Kaggle上加加速度计数据识别用户竞赛的代码
- [kaggle-advertised-salaries](#) —Kaggle上用广告预测工资竞赛的代码
- [kaggle amazon](#) —Kaggle上给定员工角色预测其访问需求竞赛的代码
- [kaggle-bestbuy_big](#)—Kaggle上根据bestbuy用户查询预测点击商品竞赛的代码（大数据版）
- [kaggle-bestbuy_small](#)—Kaggle上根据bestbuy用户查询预测点击商品竞赛的代码（小数据版）
- [Kaggle Dogs vs. Cats](#) —Kaggle上从图片中识别猫和狗竞赛的代码
- [Kaggle Galaxy Challenge](#) —Kaggle上遥远星系形态分类竞赛的优胜代码
- [Kaggle Gender](#) —Kaggle竞赛：从笔迹区分性别
- [Kaggle Merck](#)—Kaggle上预测药物分子活性竞赛的代码（默克制药赞助）
- [Kaggle Stackoverflow](#)—Kaggle上 预测Stack Overflow网站问题是否会被关闭竞赛的代码
- [wine-quality](#) —预测红酒质量。

Ruby

自然语言处理

- [Treat](#)—文本检索与注释工具包，Ruby上我见过的最全面的工具包。
- [Ruby Linguistics](#)—这个框架可以用任何语言为Ruby对象构建语言学工具。包括一个语言无关的通用前端，一个将语言代码映射到语言名的模块，和一个含有很有英文语言工具的模块。
- [Stemmer](#)—使得Ruby可用 libstemmer_c中的接口。
- [Ruby Wordnet](#) —WordNet的Ruby接口库。
- [Raspel](#) —aspell绑定到Ruby的接口
- [UEA Stemmer](#)—UEALite Stemmer的Ruby移植版，供搜索和检索用的保守的词干分析器
- [Twitter-text-rb](#)—该程序库可以将推特中的用户名、列表和话题标签自动连接并提取出来。

通用机器学习

- [Ruby Machine Learning](#) —Ruby实现的一些机器学习算法。
- [Machine Learning Ruby](#)
- [jRuby Mahout](#) —精华！在JRuby世界中释放了Apache Mahout的威力。
- [CardMagic-Classifer](#)—可用贝叶斯及其他分类法的通用分类器模块。
- [Neural Networks and Deep Learning](#)—《神经网络和深度学习》一书的示例代码。

数据分析/数据可视化

- [rsruby](#) - Ruby - R bridge
- [data-visualization-ruby](#)—关于数据可视化的Ruby Manor演示的源代码和支持内容
- [ruby-plot](#) —将gnuplot包装为Ruby形式，特别适合将ROC曲线转化为svg文件。
- [plot-rb](#)—基于Vega和D3的ruby绘图库
- [scruffy](#) —Ruby下出色的图形工具包
- [SciRuby](#)
- [Glean](#)—数据管理工具
- [Bioruby](#)
- [Arel](#)

Misc 杂项

- [Big Data For Chimps](#)—大数据处理严肃而有趣的指南书

R

通用机器学习

- [Clever Algorithms For Machine Learning](#)
- [Machine Learning For Hackers](#)
- [Machine Learning Task View on CRAN](#)—R语言机器学习包列表，按算法类型分组。
- [caret](#)—R语言150个机器学习算法的统一接口
- [SuperLearner](#) and [subsemble](#)—该包集合了多种机器学习算法
- [Introduction to Statistical Learning](#)

数据分析/数据可视化

- [Learning Statistics Using R](#)
- [ggplot2](#)—基于图形语法的数据可视化包。

Scala

自然语言处理

- [ScalaNLP](#)—机器学习和数值计算库的套装
- [Breeze](#) —Scala用的数值处理库
- [Chalk](#)—自然语言处理库。
- [FACTORIE](#)—可部署的概率建模工具包，用Scala实现的软件库。为用户提供简洁的语言来创建关系因素图，评估参数并进行推断。

数据分析/数据可视化

- [MLlib in Apache Spark](#)—Spark下的分布式机器学习库
- [Scalding](#) —CAscading的Scala接口
- [Summing Bird](#)—用Scalding 和 Storm进行Streaming MapReduce
- [Algebird](#) —Scala的抽象代数工具
- [xerial](#) —Scala的数据管理工具
- [simmer](#) —化简你的数据，进行代数聚合的unix过滤器
- [PredictionIO](#) —供软件开发者和数据工程师用的机器学习服务器。
- [BIDMat](#)—支持大规模探索性数据分析的CPU和GPU加速矩阵库。

通用机器学习

- [Conjecture](#)—Scalding下可扩展的机器学习框架
- [brushfire](#)—scalding下的决策树工具。
- [ganitha](#) —基于scalding的机器学习程序库
- [adam](#)—使用Apache Avro, Apache Spark 和 Parquet的基因组处理引擎，有专用的文件格式，Apache 2软件许可。
- [bioscala](#) —Scala语言可用的生物信息学程序库
- [BIDMach](#)—机器学习CPU和GPU加速库。

原文链接：[awesome-machine-learning](#) 翻译：[伯乐在线 - toolate](#)
译文链接：<http://blog.jobbole.com/73806/>

相关链接

想通过手机客户端（支持 Android、iPhone 和 Windows Phone）访问开源中国：[请点击这里](#)

本站文章除注明转载外，均为本站原创或编译

欢迎任何形式的转载，但请务必注明出处，尊重他人劳动共创开源社区

转载请注明：文章转载自：开源中国社区 [<http://www.oschina.net>]

本文标题：推荐！国外程序员整理的机器学习资源大全

本文地址：<http://www.oschina.net/news/53818/awesome-machine-learning>

旧一篇：[Percona Monitoring Plugins 1.1.4 发布](#) 2年前

新一篇：[Apache Commons DbUtils 1.6 发布](#) 2年前

相关资讯

- [好程序员 + 烂技术 = 痛苦](#) 2天前
- [Codelf 搜索开源代码帮程序员命名](#) 1周前
- [程序员如何谋划出月薪 3 万](#) 2周前
- [OSChina 周日乱弹 —— 程序员的鄙视链](#) 3周前
- [35 岁程序员的独家面试经历](#) 1个月前
- [CmlPHP V2.5.1 发布，为程序员而生的开发框架](#) 1个月前
- [CmlPHP V2.5 发布，为程序员而生的开发框架](#) 1个月前
- [程序员必知的前端演讲史](#) 2个月前
- [Python 程序员都会喜欢的 6 个库](#) 2个月前
- [\[译\] 1000 赞的程序员高薪职业建议](#) 2个月前

网友评论，共 56 条

[发表评论](#) [回到顶部](#)



1楼：**mingshun** 发表于 2014-07-22 07:15 [回复此评论](#)
marked



2楼：**opmetic** 发表于 2014-07-22 08:31 [回复此评论](#)
我也marked



3楼：**Disappeared** 来自 [Android](#) 发表于 2014-07-22 08:37 [回复此评论](#)
我也也marked



4楼：**JonasBollack** 发表于 2014-07-22 08:51 [回复此评论](#)
我也也也marked



5楼：**dclink** 发表于 2014-07-22 08:52 [回复此评论](#)
我也也也marked



6楼：**黑炭** 发表于 2014-07-22 08:59 [回复此评论](#)
学海无涯啊！



7楼：**邪恶八稚女** 发表于 2014-07-22 09:04 [回复此评论](#)
mark



8楼：**Broly** 发表于 2014-07-22 09:05 [回复此评论](#)
marked



9楼：**jrchens** 发表于 2014-07-22 09:06 [回复此评论](#)
我想知道这是一个人多久能学习完的？



10楼：**iamxi** 发表于 2014-07-22 09:06 [回复此评论](#)
mark



11楼：**chenqiyl** 发表于 2014-07-22 09:06 [回复此评论](#)

[引用来自“黑炭”的评论](#)

学海无涯啊！
回头...



12楼：**star_qu** 发表于 2014-07-22 09:07 [回复此评论](#)
学习！！！！



13楼：**Raydeer** 发表于 2014-07-22 09:14 [回复此评论](#)
很强大的样子



14楼：**lnyu** 发表于 2014-07-22 09:19 [回复此评论](#)
回头



15楼：**mickelfeng** 发表于 2014-07-22 09:19 [回复此评论](#)
没有我php。



16楼：**XzhiF** 发表于 2014-07-22 09:20 [回复此评论](#)
marked



17楼：吐槽的达达仔 发表于 2014-07-22 09:22 [回复此评论](#)
没啥大用。。



18楼：在路上逛开源 发表于 2014-07-22 09:24 [回复此评论](#)
太高大上...



19楼：oschinaee 发表于 2014-07-22 09:29 [回复此评论](#)
marked



20楼：cobbage 发表于 2014-07-22 09:35 [回复此评论](#)
收藏了

[1](#)
[2](#)
[3](#)
[≥](#)

网名： (必填)

邮箱： (必填，不公开)

网址：

验证码： [换一个](#)

与内容无关的评论将被删除，严重者禁用帐号

[回到顶部](#) [回到评论列表](#)



阿里云存储新品秒杀
¥198/1T1年
[立即抢购](#)

本周热点资讯

- [Chair：支付宝前端团队推出的 N...](#) 7天前
- [前端程序员：月薪 5K 到 5 万，...](#) 4天前
- [jQuery 3.0 Beta 发布...](#) 7天前
- [2016 最佳 Linux 发行版排行 榜...](#) 4天前
- [好程序员 + 烂技术 = 痛苦](#) 3天前
- [核心开发者宣称比特币失败 清 货...](#) 3天前
- [透过微信应用号，看HTML5与 Nat...](#) 2天前
- [百度开源其关键人工智能系统 WA...](#) 7天前
- [ASP.NET 5 已终结，迎来 ASPNE...](#) 1天前

- 德国慕尼黑政府花巨资用 Linux ... 12小时前

本站最新资讯

- 糖果CMS V1.1 发布，内容管理系统... 8小时前
- Dropwizard 0.9.2 发布，RESTful API... 10小时前
- Dat 8.0.1 发布，数据集版本控制... 10小时前
- MongoDB 3.0.9 RC0 发布... 10小时前
- 德国慕尼黑政府花巨资用 Linux ... 12小时前
- GParted 0.25.0 和 GParted Live 12.0 发布... 13小时前
- Google Chrome v48.0.2564.82 正式发布... 14小时前
- Node v5.5.0 (稳定版) 发布... 14小时前
- 华为推出云容器服务 Cloud Container Engine... 15小时前
- JDK 8u72 发布，Java 开发工具包... 15小时前



你也许会喜欢 [X](#)

- 好程序员 + 烂技术 = 痛苦 2天前
- Codef 搜索开源代码帮程序员命名 1周前
- 程序员如何谋划出月薪 3 万 2周前