


# 机器学习最牛B的框架、库和软件！吐血整理！



来自: **李苗苗**(专注大数据BI众包。dashengzb.cn) 2016-08-08 15:47:42


这次又准备了什么必备收藏？

全是机器学习最牛B的框架、库和软件！

吐血整理，堪称史上最全！



- 一、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和Mac OS操作系统
- 通用机器学习
- MLPack
- DLib
- ecogg
- shark
- 二、Closure
- 通用机器学习
- Closure Toolbox—Clojure语言库与工具的分类目录
- 三、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




斯坦福大学机器学习导论

2916 人聚集在这个小组

加入小组

- 最新话题 (更多)
- 关于资讯机器人的技术与运用解读 (智搜Giiso)
- 斯坦福大学新生群 (懒懒的Sally)
- 【实力见证】Giiso智搜获颁2017高交会初创展最具投... (智搜Giiso)
- AI淘汰80万工人，却创造了350万个新岗位！ (智搜Giiso)
- 深度学习培训 (mamaxin)



2018豆瓣电影日历，你抢到了吗？

豆瓣市集

广告

豆 下载豆瓣App

TRAVELZOO旅游族

只要¥9.9！

日本3年多次往返签证 限时惊喜价

立即出发

广告

#### 四、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 Sengrex—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

#### 五、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

##### 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

##### 数据分析/数据可视化

D3.js

High Charts

NVD3.js

在这里发现跟你一样特别的人，并为之交流...

注册

登录

第三方登录：



dimple  
amCharts

## 六、Julia

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

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

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

## 七、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Black Box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

[在这里发现跟你一样特别的人，并与之交流...](#)

注册

登录

第三方登录： 



kaggle-bestbuy\_big—Kaggle上根据Best Buy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Best Buy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

## 八、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—Torch7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

[在这里发现跟你一样特别的人，并与之交流...](#)

注册

登录

第三方登录： 



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

16人喜欢

喜欢

回应推荐喜欢

查看所有回应



李苗苗 (专注大数据BI众包。dashengzb.cn) 2016-08-09 10:46:57

第一次回帖，竟然八小时之前的帖子，还是一楼。。。 [内cool超人](#)

沙发是你的

赞 回应

你的回应

回应请先 [登录](#)，或 [注册](#)

☐ 推荐到广播

加上去