

[BT](#)x [InfoQ调研吧!](#)

- [投稿](#)
- [关于我们](#)
- [合作伙伴](#)

- 欢迎关注我们的:



InfoQ - 促进软件开发领域知识与创新的传播

[登录](#)

- [En](#)
- [中文](#)
- [日本](#)
- [Fr](#)
- [Br](#)

482,381 十一月 独立访问用户

- [语言 & 开发](#)
 - [Java](#)
 - [Clojure](#)
 - [Scala](#)
 - [.Net](#)
 - [移动](#)
 - [Android](#)
 - [iOS](#)
 - [HTML 5](#)
 - [JavaScript](#)
 - [函数式编程](#)
 - [Web API](#)

特别专题 语言 & 开发

[专访豌豆荚搜索技术负责人：应用内搜索——移动时代的搜索技术](#)

[浏览所有 语言 & 开发](#)

- [架构 & 设计](#)

- [架构](#)
- [企业架构](#)
- [性能和可伸缩性](#)
- [设计](#)
- [案例分析](#)
- [设计模式](#)
- [安全](#)

特别专题 架构 & 设计

[专访豌豆荚搜索技术负责人：应用内搜索——移动时代的搜索技术](#)



[浏览所有 架构 & 设计](#)

- [数据科学](#)
 - [大数据](#)
 - [NoSQL](#)
 - [关系型数据库](#)

特别专题 数据科学

[利用外部数据封装器打破数据筒仓](#)



[作者Lenlev Hensarling探讨了Postgre数据库外部数据封装器（FDW）。FDW提供了一个SQL](#)

[接口，用于访问远程数据存储中的数据对象。借助该接口，可以从像NoSQL数据库这样的、完全不同的数据源中集成数据，将它们存入一种公用模型。](#)

[浏览所有 数据科学](#)

- [文化 & 方法](#)
 - [Agile](#)
 - [领导能力](#)
 - [团队协作](#)
 - [测试](#)
 - [用户体验](#)
 - [Scrum](#)
 - [精益](#)

特别专题

[浏览所有](#)

- [DevOps](#)
 - [持续交付](#)
 - [自动化操作](#)
 - [云计算](#)

特别专题 DevOps

[专访WebRTC标准之父Daniel C. Burnett](#)



[WebRTC 是一个由 Google、Mozilla 和 Opera 主导的开源项目，2015年11月11日，由声网（Agora）与美国 TMC 联合主办的亚太区首届（WebRTC）网络实时通信大会在北京中关村皇冠假日酒店举行。会上，InfoQ 对 WebRTC 之父 Daniel C. Burnett 进行了专访，声网（Agora）北美运营总监王骅补充了部分问题，以下是专访实录。](#)



[StuQ教育](#)
[EGO组织](#)
[移动](#)
[Docker](#)
[开源](#)
[云计算](#)
[大数据](#)
[架构师](#)
[阿里百川](#)
[QCon](#)
[ArchSummit](#)
[UnitedStack](#)

[全部话题](#)

您目前处于: [InfoQ首页](#) [新闻](#) 机器学习的11个开源项目

机器学习的11个开源项目

作者 [张天雷](#) 发布于 2014年12月18日 | [讨论](#)

[+](#) 分享到: [微博](#) [微信](#) [Facebook](#) [Twitter](#) [有道云笔记](#) [邮件分享](#)

- [“稍后阅读”](#)
- [“我的阅读清单”](#)

机器学习是目前数据分析领域的一个热点内容，在平时的学习和生活中经常会用到各种各样的机器学习算法。实际上，基于Python、Java等的很多机器学习算法基本都被前人实现过很多次了。这些算法在网上可以找到很多，然而往往存在很多“脏”或者“乱”的开源代码。

在这样的背景下，[InfoWorld](#)近日公布了机器学习领域11个最受欢迎的开源项目，这11个开源项目大多与垃圾邮件过滤、人脸识别、推荐引擎相关。它们大多数基于现今最流行的语言以及平台，推广以及扩展了机器学习领域的很多重要算法。从中，用户不但可以找到LDA等主题模型，也可以找到HMM等隐马尔科夫模型。这些模型都是应用领域的热点，也是研究者们最需要的。

1. [Scikit-learn](#)

Scikit-learn是一个非常强大的Python机器学习工具包。它通过在现有Python的基础上构建了NumPy和Matplotlib，提供了非常便利的数学工具。这个工具包包括了很多简单且高效的工具，很适合用于数据挖掘和数据分析。

在主页中，可以看到User Guide，这是整个机器学习的索引，其中用户可以学到各种有效的方法。在Reference里，用户可以找到各个类具体的用法索引。

2. [Shogun](#)

Shogun是一个基于C++的最古老的机器学习开源库，它创建于1999年。作为一个SWIG库，Shogun可以轻松地嵌入Java、Python、C#等主流处理语言中。它的重点在于大尺度上的内核方法，特别是“支持向量机”的学习工具箱。其中，它包括了大量的线性方法，如LDA、LPM、HMM等等。

3. [Accord Framework/AForge.net](#)

Accord是AForge.net的扩展，是一个基于.Net的机器学习与信号处理框架。它包括了一系列的对图像和音频的机器学习算法，如人脸检测、SIFT拼接等等。同时，Accord支持移动对象的实时跟踪等功能。它提供了一个从神经网络到决策树系统的机器学习库。

4. [Mahout](#)

Mahout是一个广为人知的开源项目，它是Apache Software旗下的一个开源项目，提供了众多的机器学习经典算法的实现，旨在帮助开发人员更加方便快捷地创建智能应用程序。Mahout内包含了聚类、分类、推荐等很多经典算法，并且提供了很方便的云服务的接口。

5. [MLlib](#)

MLlib是Apache自己的Spark和Hadoop机器学习库，它被设计用于大规模高速度地执行MLlib所包含的大部分常见机器学习算法。MLlib是基于Java开发的项目，同时可以方便地与Python等语言对接。用户可以自己设计针对MLlib编写代码，这是很具有个性化的设计。

6. [H2O](#)

H2O是0xdata的旗舰产品，是一款核心数据分析平台。它的一部分是由R语言编写的，另一部分是由Java和Python语言编写的。用户可以部署H2O的R程序安装包，之后就可以在R语言环境下运行了。H2P的算法是面向业务欺诈活着趋势预测的，目前正在新一轮的融资中。

7. [Cloudera Oryx](#)

Oryx也是由Hadoop所设计的机器学习开源项目，由Cloudera Hadoop Distribution的创造者所提供。Oryx能够让机器学习的模型使用在实时的数据流上，如垃圾邮件过滤等。

8. [GoLearn](#)

GoLearn是谷歌所构建的Go语言的一体化机器学习库，目标是简单并且可定制。Go语言是谷歌的主打语言，目前使用已经越来越广泛。GoLearn的简单在于数据在库内被加载和处理，因此能够可定制地扩展数据结构以源码。

9. [Weka](#)

>Weka是使用Java开发的用于数据挖掘的开源项目。Weka作为一个公开的数据挖掘工作平台，集合了大量能够承担数据挖掘任务的机器学习算法，包括了对数据进行预处理、分类、回归、聚类等等。同时，Weka实现了对大数据的可视化，通过Java设计的新式交互界面上，实现人与程序的交互。

10. [CUDA-Convnet](#)

CUDA是我们众所周知的GPU加速套件。而CUDA-Convnet是一个基于GPU加速的神经网络应用程序机器学习库。它使用C++编写，并且使用了Nvidia的CUDA GPU处理技术。

目前，这个项目已经被重组成为CUDA-Convnet2，支持多个GPU和Kepler-generation GPUs。Vuples项目与之类似，使用F#语言编写，并且适用于.Net平台上。

11. [ConvNetJS](#)

ConvNetJS是一款基于JavaScript的在线深度学习库，它提供了在线的深度学习训练方式。它能够帮助深度学习的初学者更快、更加直观的理解算法，通过一些简单的Demo给用户最直观的解释。

感谢[郭蕾](#)对本文的审校。

给InfoQ中文站投稿或者参与内容翻译工作，请邮件至editors@cn.infoq.com。也欢迎大家通过新浪微博（@InfoQ）或者腾讯微博（@InfoQ）关注我们，并与我们的编辑和其他读者朋友交流。

【ArchSummit北京2015】全球架构师峰会精彩升级！手淘从无到有如何建设起了如今的Hybrid App框架？携程如何通过工程解耦及开发框架优化轻松应对业务和新技术的快速发展？苏宁如何应对风险识别系统建设的挑战？爱奇艺在面对其视频转码对IaaS平台弹性计算的高要求又是如何应对呢？近百精英压轴年度技术大戏！[精彩内容，期待您的参与](#)。

- [领域](#)
- [架构 & 设计](#)
- [语言 & 开发](#)
- [专栏](#)
- [开放源代码](#)
- [开源项目发布](#)

- [机器学习](#)
- [开源软件](#)
- [开源](#)

相关内容

[《利用开源项目，学习软件开发》内容回顾与作者采访](#)

[王为：开源本身是一件好事](#)

[小米对开源的使用与参与之道](#)

[如何打造一个成功的开源项目](#)

[360的开源软件使用以及开源文化构建经验](#)

相关厂商内容

[手机百度APS深度剖析](#)

[手淘如何从无到有的Hybrid App框架创建历程。](#)

[万人在线直播教室如何搭建？](#)

[公司高速发展，研发团队如何调优](#)

[支付宝红包瞬时支付量挑战双十一零点峰值！](#)

相关赞助商

全球架构师峰会，12月18-19日，北京·国际会议中心，[精彩内容](#)邀您参与！



您好，朋友！

您需要 [注册一个InfoQ账号](#) 或者 [登录](#) 才能进行评论。在您完成注册后还需要进行一些设置。

获得来自InfoQ的更多体验。

告诉我们您的想法

请输入主题	信息
-------	----

允许的HTML标签：a, b, br, blockquote, i, li, pre, u, ul, p

☐ 当有人回复此评论时请E-mail通知我

[发送信息](#)

社区评论 [Watch Thread](#)

[关闭](#)

by

发布于

- [查看](#)
- [回复](#)
- [回到顶部](#)

2015/12/2

机器学习的11个开源项目

[关闭](#)

主题您的回复

允许的HTML标签: a, b, br, blockquote, i, li, pre, u, ul, p

☐ 当有人回复此评论时请E-mail通知我

发送信息

取消

[关闭](#)

主题您的回复

允许的HTML标签: a, b, br, blockquote, i, li, pre, u, ul, p

☐ 当有人回复此评论时请E-mail通知我

取消

[关闭](#)

OK

相关内容

- [Quora是如何应用机器学习的?](#) 2015年12月1日
- [专访YYKit作者郭曜源: 开源大牛是怎样炼成的](#) 2015年11月26日



- [开源启示录·第二季](#) 2015年10月12日
- [机器学习和大数据:是先有鸡还是先有蛋](#) 2015年11月9日



- [OpenStack开源实践](#) 2015年11月28日

- [《利用开源项目, 学习软件开发》内容回顾与作者采访](#) 2015年11月20日
- [Pivotal开源基于PostgreSQL的数据库Greenplum](#) 2015年11月4日

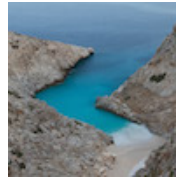


- [专访GitHub CIO Scott Chacon: 开源文化与开放API经济](#) 2015年11月18日
- [开源等于安全吗?](#) 2015年11月3日





相关内容



- [Diffy: Twitter的开源自动化测试工具](#) 2015年10月29日
- [开源文化对360天擎架构演进的影响](#) 2015年9月29日



- [腾讯代码管理平台架构和开源实践](#) 2015年9月25日



- [杨锦涛：主流开源存储方案孰优孰劣](#) 2015年8月28日



- [王为：开源本身是一件好事](#) 2015年8月10日
- 语言 & 开发



[专访豌豆荚搜索技术负责人：应用内搜索——移动时代的搜索技术](#)

[三个非常出色的开源硬件项目](#)

- [小米对开源的使用与参与之道](#) 2015年7月27日
- [物联网技术周报第22期：Raspberry Pi 推出轻量级 Pi Zero 只要5 美元](#)



架构 & 设计

[专访豌豆荚搜索技术负责人：应用内搜索——移动时代的搜索技术](#)

[三个非常出色的开源硬件项目](#)

北京·国际会议中心-203B

[物联网技术周报第22期：Raspberry Pi 推出轻量级 Pi Zero 只要5 美元](#)

免费参加

文化 & 方法

[专访豌豆荚搜索技术负责人：应用内搜索——移动时代的搜索技术](#)

[Real World Kanban作者访谈](#)

[赞助商链接](#)

阿里百川无线开放平台，致力于为移动开发者与App创业者提供后端服务、商业组件以及资本支持，可快速搭建应用并带来卓越用户体验，详情请见 [阿里百川专区](#)>>

[MongoDB 3.2版本即将发布，向企业市场进军](#)

[百度移动App测试平台，可提供全面兼容、深度性能、安全漏洞扫描以及手游兼容等多种测试套餐](#)
[AdaptiFi发布移动App测试数据实时分析软件](#)
应用挑战。详情>>

[AWS Summit中国站活动将于12月17日在上海全面开幕，六大专场涵盖运维、物联网、机器学习等领域，而报名并出席本次活动还将享有现场专属福利。立即报名>>](#)

[DevOps](#)

[JavaOne 2015概览](#)

[Apache Maven JDeps插件3.0.0版本发布](#)

[Verizon的ThingSpace平台简化了新的物联网应用的开发](#)

- [首页](#)
- [全部话题](#)
- [QCon全球软件开发大会](#)
- [关于我们](#)
- [投稿](#)
- [创建账号](#)
- [登录](#)

- [全球QCon](#)
- [伦敦 2016年3月7-11日](#)
- [圣保罗 2016年3月28日-4月1日](#)

- [北京 2016年4月](#)
- [东京 2016年4月](#)
- [纽约 2016年6月13日-6月17日](#)
- [里约热内卢 2016年10月](#)
- [Shanghai, Oct 2016](#)
- [旧金山 2016年11月7-11日](#)

InfoQ每周精要

通过个性化定制的新闻邮件、RSS Feeds和InfoQ业界邮件通知，保持您对感兴趣的社区内容的时刻关注。

[点击这里
查看样刊](#)



- [属于您的个性化RSS](#)
- [InfoQ官方微博](#)
- [InfoQ官方微信](#)
- [社区新闻和热点](#)

特别专题

- [UnitedStack专区](#)
- [月刊：《架构师》](#)
- [AWS专区](#)
- [百度技术沙龙专区](#)
- [阿里百川专区](#)
- [七牛专区](#)
- [融云](#)
- [EGO超级极客邦](#)
- [StuQ提升你技能](#)
- [全球容器技术大会](#)
- [信息无障碍参考文档](#)

定制您感兴趣的技术领域

- ☒ 语言 & 开发
- ☒ 架构 & 设计
- ☒ 数据科学
- ☒ 文化 & 方法
- ☒ DevOps

这会影响您在主页和RSS订阅中看到的内容。点击“偏好设置”可选择更多精彩定制内容。

InfoQ.com
及所有内
容，版权所
有 © 2006-
2015
C4Media
Inc.

[提供反馈](#)[错误报告](#)[商务合作](#)[内容合作](#)[Marketing](#)feedback@cn.infoq.combugs@cn.infoq.comsales@cn.infoq.comeditors@cn.infoq.commarketing@infoq.com

InfoQ.com
服务器由
Contegix提供, 我们最
信赖的ISP
伙伴。
北京创新网
媒广告有限
公司 京ICP
备09022563
号-7 [隐私
政策](#)

[BT](#)