

机器学习(Machine Learning)&深度学习(Deep Learning)资料(Chapter 1)

注:机器学习资料[篇目一](#)共500条,[篇目二](#)开始更新

希望转载的朋友,你可以不用联系我.但是一定要保留原文链接,因为这个项目还在继续也在不定期更新.希望看到文章的朋友能够学到更多.此外:某些资料在中国访问需要梯子.

- [《Brief History of Machine Learning》](#)
- <http://www.erogol.com/brief-history-machine-learning/>

介绍:这是一篇介绍机器学习历史的文章,介绍很全面,从感知机、神经网络、决策树、SVM、Adaboost到随机森林、Deep Learning.

- [《Deep Learning in Neural Networks: An Overview》](#)
- <http://www.idsia.ch/~juergen/DeepLearning15May2014.pdf>

介绍:这是瑞士人工智能实验室Jurgen Schmidhuber写的最新版本《神经网络与深度学习综述》本综述的特点是以时间排序,从1940年开始讲起,到60-80年代,80-90年代,一直讲到2000年后及最近几年的进展.涵盖了deep learning里各种tricks,引用非常全面.

- [《A Gentle Introduction to Scikit-Learn: A Python Machine Learning Library》](#)
- <http://machinelearningmastery.com/a-gentle-introduction-to-scikit-learn-a-python-machine-learning-library/>

介绍:这是一份python机器学习库,如果您是一位python工程师而且想深入的学习机器学习.那么这篇文章或许能够帮助到你.

- [《How to Layout and Manage Your Machine Learning Project》](#)
- <http://machinelearningmastery.com/how-to-layout-and-manage-your-machine-learning-project/>

介绍:这一篇介绍如果设计和管理属于你自己的机器学习项目的文章,里面提供了管理模版、数据管理与实践方法.

- [《Machine Learning is Fun!》](#)
- <https://medium.com/code-poet/80ea3ec3c471>

介绍:如果你还不知道什么是机器学习,或则是刚刚学习感觉到很枯燥乏味.那么推荐一读.这篇文章已经被翻译成中文,如果有兴趣可以移步<http://blog.jobbole.com/67616/>

- [《R语言参考卡片》](#)
- <http://cran.r-project.org/doc/contrib/Liu-R-refcard.pdf>

介绍:R语言是机器学习的主要语言,有很多的朋友想学习R语言,但是总是忘记一些函数与关键字的含义.那么这篇文章或许能够帮助到你

- [《Choosing a Machine Learning Classifier》](#)
- <http://blog.echen.me/2011/04/27/choosing-a-machine-learning-classifier/>

介绍:我该如何选择机器学习算法,这篇文章比较直观的 compares Naive Bayes, Logistic Regression, SVM, 决策树等方法的优劣,另外讨论了样本大小、Feature与Model权衡等问题.此外还有已经翻译了的版本:<http://www.52ml.net/15063.html>

- [《An Introduction to Deep Learning: From Perceptrons to Deep Networks》](#)
- <http://www.toptal.com/machine-learning/an-introduction-to-deep-learning-from-perceptrons-to-deep-networks>

介绍：深度学习概述：从感知机到深度网络，作者对于例子的选择、理论的介绍都很到位，由浅入深。翻译版本：<http://www.cnblogs.com/xiaowanyer/p/3701944.html>

- 《The LION Way: Machine Learning plus Intelligent Optimization》
- <http://vdisk.weibo.com/s/ayG13we2vxyKl>

介绍:<机器学习与优化>这是一本机器学习的小册子,短短300多页道尽机器学习的方方面面.图文并茂,生动易懂,没有一坨坨公式的烦恼.适合新手入门打基础,也适合老手温故而知新.比起MLAPP/PRML等大部头,也许这本你更需要!具体内容推荐阅读:<http://intelligent-optimization.org/LIONbook/>

- 《深度学习与统计学习理论》
- <http://php-52cs.rhcloud.com/?cat=7>

介绍:作者是来自百度,不过他本人已经在2014年4月份申请离职了。但是这篇文章很不错如果你不知道深度学习与支持向量机/统计学习理论有什么联系?那么应该立即看看这篇文章。

- 《计算机科学中的数学》
- http://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-042j-mathematics-for-computer-science-fall-2010/readings/MIT6_042JF10_notes.pdf

介绍:这本书是由谷歌公司和MIT共同出品的计算机科学中的数学: [Mathematics for Computer Science](#), Eric Lehman et al 2013。分为5大部分: 1) 证明, 归纳。2) 结构, 数论, 图。3) 计数, 求和, 生成函数。4) 概率, 随机行走。5) 递归。等等

- 《信息时代的计算机科学理论(Foundations of Data Science)》
- <http://research.microsoft.com/en-US/people/kannan/book-no-solutions-aug-21-2014.pdf>

介绍: 信息时代的计算机科学理论,目前国内有纸质书购买, [iTunes购买](#)

- 《Data Science with R》
- <http://vdisk.weibo.com/s/ayG13we2vx5qg>

介绍:这是一本由雪城大学新编的第二版《数据科学入门》教材: 偏实用型, 浅显易懂, 适合想学习R语言的同学选读。

- 《Twenty Questions for Donald Knuth》
- <http://www.informit.com/articles/article.aspx?p=2213858>

介绍:这不是一篇文档或书籍。这是篇向图灵奖得主Donald Knuth提问记录稿: 近日, Charles Leiserson, Al Aho, Jon Bentley等大神向Knuth提出了20个问题, 内容包括TAOCP, P/NP问题, 图灵机, 逻辑, 以及为什么大神不用电邮等等。

- 《Automatic Construction and Natural-Language Description of Nonparametric Regression Models》
- <http://arxiv.org/pdf/1402.4304v2.pdf>

介绍: 不会统计怎么办? 不知道如何选择合适的统计模型怎么办? 那这篇文章你的好好读一读!麻省理工Joshua B. Tenenbaum和剑桥Zoubin Ghahramani合作, 写了一篇关于automatic statistician的文章。可以自动选择回归模型类别, 还能自动写报告...

- 《ICLR 2014论文集》
- <http://openreview.net/venue/iclr2014>

介绍:对深度学习和representation learning最新进展有兴趣的同学可以了解一下

- 《Introduction to Information Retrieval》
- <http://www-nlp.stanford.edu/IR-book/>

介绍：这是一本信息检索相关的书籍，是由斯坦福Manning与谷歌副总裁Raghavan等合著的Introduction to Information Retrieval一直是北美最受欢迎的信息检索教材之一。最近作者增加了该课程的幻灯片和作业。IR相关资源：<http://www-nlp.stanford.edu/IR-book/information-retrieval.html>

- 《Machine learning in 10 pictures》
- <http://www.denizyuret.com/2014/02/machine-learning-in-5-pictures.html>

介绍:Deniz Yuret用10张漂亮的图来解释机器学习重要概念：1. Bias/Variance Tradeoff 2. Overfitting 3. Bayesian / Occam's razor 4. Feature combination 5. Irrelevant feature 6. Basis function 7. Discriminative / Generative 8. Loss function 9. Least squares 10. Sparsity.很清晰

- 《雅虎研究院的数据集汇总》
- <http://webscope.sandbox.yahoo.com/catalog.php?datatype=l>

介绍：雅虎研究院的数据集汇总：包括语言类数据，图与社交类数据，评分与分类数据，计算广告学数据，图像数据，竞赛数据，以及系统类的数据。

- 《An Introduction to Statistical Learning with Applications in R》
- <http://www-bcf.usc.edu/~gareth/ISL/>

介绍：这是一本斯坦福统计学著名教授Trevor Hastie和Robert Tibshirani的新书，并且在2014年一月已经开课：<https://class.stanford.edu/courses/HumanitiesScience/StatLearning/Winter2014/about>

- Best Machine Learning Resources for Getting Started
- <http://machinelearningmastery.com/best-machine-learning-resources-for-getting-started/>

介绍：机器学习最佳入门学习资料汇总是专为机器学习初学者推荐的优质学习资源，帮助初学者快速入门。而且这篇文章的介绍已经被翻译成[中文版](#)。如果你不怎么熟悉，那么我建议你先看一看中文的介绍。

- My deep learning reading list
- http://blog.sina.com.cn/s/blog_bda0d2f1010101fpp4.html

介绍:主要是顺着Bengio的PAMI review的文章找出来的。包括几本综述文章，将近100篇论文，各位山头们的Presentation。全部都可以在google上找到。

- Cross-Language Information Retrieval
- <http://www.morganclaypool.com/doi/abs/10.2200/S00266ED1V01Y201005HLT008?journalCode=hit>

介绍：这是一本书籍，主要介绍的是跨语言信息检索方面的知识。理论很多

- 探索推荐引擎内部的秘密，第 1 部分: 推荐引擎初探
- http://www.ibm.com/developerworks/cn/web/1103_zhaoct_recommstudy1/index.html?ca=drs-

介绍:本文共有三个系列，作者是来自IBM的工程师。它主要介绍了推荐引擎相关算法，并帮助读者高效的实现这些算法。探索推荐引擎内部的秘密，第 2 部分: 深度推荐引擎相关算法 - 协同过滤,探索推荐引擎内部的秘密，第 3 部分: 深度推荐引擎相关算法 - 聚类

- 《Advice for students of machine learning》
- <http://mimno.infosci.cornell.edu/b/articles/ml-learn/>

介绍：康奈尔大学信息科学系助理教授David Mimno写的《对机器学习初学者的一点建议》，写的挺实际，强调实践与理论结合，最后还引用了冯·诺依曼的名言：“Young man, in mathematics you don't understand things. You just get used to them.”

- 分布式并行处理的数据
- <http://web.stanford.edu/group/pdplab/pdphandbook/>

介绍：这是一本关于分布式并行处理的数据《Explorations in Parallel Distributed Processing: A Handbook of Models, Programs, and Exercises》,作者是斯坦福的James L. McClelland。着重介绍了各种神级网络算法的分布式实现,做Distributed Deep Learning 的童鞋可以参考下

- 《“机器学习”是什么？》
- <http://blogs.technet.com/b/machinelearning/archive/2014/07/01/what-is-machine-learning.aspx>

介绍：【“机器学习”是什么？】John Platt是微软研究院杰出科学家，17年来他一直在机器学习领域耕耘。近年来机器学习变得炙手可热，Platt和同事们遂决定开设[博客](#)，向公众介绍机器学习的研究进展。机器学习是什么，被应用在哪里？来看Platt的这篇[博文](#)

- 《2014年国际机器学习大会ICML 2014 论文》
- <http://icml.cc/2014/index/article/15.htm>

介绍：2014年国际机器学习大会（ICML）已经于6月21-26日在国家会议中心隆重举行。本次大会由微软亚洲研究院和清华大学联手主办，是这个有着30多年历史并享誉世界的机器学习领域的盛会首次来到中国，已成功吸引海内外1200多位学者的报名参与。干货很多，值得深入学习下

- 《Machine Learning for Industry: A Case Study》
- <http://blogs.technet.com/b/machinelearning/archive/2014/07/11/machine-learning-for-industry-a-case-study.aspx>

介绍：这篇文章主要是以Learning to Rank为例说明企业界机器学习的具体应用，RankNet对NDCG之类不敏感，加入NDCG因素后变成了LambdaRank，同样的思想从神经网络改为应用到Boosted Tree模型就成就了LambdaMART。[Chirs Burges](#)，微软的机器学习大神，Yahoo 2010 Learning to Rank Challenge 第一名得主，排序模型方面有RankNet，LambdaRank，LambdaMART，尤其以LambdaMART最为突出，代表论文为：

[From RankNet to LambdaRank to LambdaMART: An Overview](#)

此外，Burges还有很多有名的代表作，比如：[A Tutorial on Support Vector Machines for Pattern Recognition](#)

[Some Notes on Applied Mathematics for Machine Learning](#)

- 100 Best GitHub: Deep Learning
- <http://meta-guide.com/software-meta-guide/100-best-github-deep-learning/>

介绍:100 Best GitHub: Deep Learning

- 《UFLDL-斯坦福大学Andrew Ng教授“Deep Learning”教程》
- <http://www.52ml.net/12019.html>

介绍:本教程将阐述无监督特征学习和深度学习的主要观点。通过学习，你也将实现多个功能学习/深度学习算法，能看到它们为你工作，并学习如何应用/适应这些想法到新问题上。本教程假定机器学习的基本知识（特别是熟悉的监督学习，逻辑回归，梯度下降的想法），如果你不熟悉这些想法，我们建议你去看[这里机器学习课程](#)，并先完成第II，III，IV章（到逻辑回归）。此外这关于这套教程的源代码在github上面已经有python版本了 [UFLDL Tutorial Code](#)

* 《Deep Learning for Natural Language Processing and Related Applications》

介绍:这份文档来自微软研究院,精髓很多。如果需要完全理解，需要一定的机器学习基础。不过有些地方会让人眼前一亮,茅塞顿开。

- Understanding Convolutions
- <https://colah.github.io/posts/2014-07-Understanding-Convolutions/>

介绍:这是一篇介绍图像卷积运算的文章，讲的已经算比较详细的了

- 《Machine Learning Summer School》
- <http://mlss2014.com/>

介绍：每天请一个大牛来讲座，主要涉及机器学习，大数据分析，并行计算以及人脑研究。<https://www.youtube.com/user/smolix>（需翻墙）

- 《Awesome Machine Learning》
- <https://github.com/josephmisiti/awesome-machine-learning>

介绍：一个超级完整的机器学习开源库总结，如果你认为这个碉堡了，那后面这个列表会更让你惊讶：【Awesome Awesomeness】，国内已经有热心的朋友进行了翻译[中文介绍](#)，[机器学习数据挖掘免费电子书](#)

- [斯坦福《自然语言处理》课程视频](#)
- <http://cs224d.stanford.edu/syllabus.html>

介绍:ACL候任主席、斯坦福大学计算机系Chris Manning教授的《自然语言处理》课程所有视频已经可以在斯坦福公开课网站上观看了（如Chrome不行，可用IE观看）作业与测验也可以下载。

- [《Deep Learning and Shallow Learning》](#)
- <http://freemind.pluskid.org/machine-learning/deep-learning-and-shallow-learning/>

介绍:对比 Deep Learning 和 Shallow Learning 的好文，来着浙大毕业、MIT 读博的 Chiyuan Zhang 的博客。

- [《Recommending music on Spotify with deep learning》](#)
- <http://benanne.github.io/2014/08/05/spotify-cnns.html>

介绍:利用卷积神经网络做音乐推荐。

- [《Neural Networks and Deep Learning》](#)
- <http://neuralnetworksanddeeplearning.com/index.html>

介绍：神经网络的免费在线书，已经写了三章了，还有对应的开源代码：<https://github.com/mnielsen/neural-networks-and-deep-learning> 爱好者的福音。

- [《Java Machine Learning》](#)
- <http://machinelearningmastery.com/java-machine-learning/>

介绍：Java机器学习相关平台和开源的机器学习库，按照大数据、NLP、计算机视觉和Deep Learning分类进行了整理。看起来挺全的，Java爱好者值得收藏。

- [《Machine Learning Theory: An Introductory Primer》](#)
- <http://www.oschina.net/translate/6-tips-for-writing-better-code>

介绍：机器学习最基本的入门文章，适合零基础者

- [《机器学习常见算法分类汇总》](#)
- <http://www.ctocio.com/hotnews/15919.html>

介绍：机器学习的算法很多。很多时候困惑人们都是，很多算法是一类算法，而有些算法又是从其他算法中延伸出来的。这里，我们从两个方面来给大家介绍，第一个方面是学习的方式，第二个方面是算法的类似性。

- [《机器学习经典论文/survey合集》](#)
- <http://suanfazu.com/discussion/68/%E6%9C%BA%E5%99%A8%E5%AD%A6%E4%B9%A0%E7%BB%8F%E5%85%B8%E8%AE%BA%E6%96%87survey>

介绍：看题目你已经知道了是什么内容,没错。里面有很多经典的机器学习论文值得仔细与反复的阅读。

- [《机器学习视频库》](#)
- <http://work.caltech.edu/library/>

介绍：视频由加州理工学院（Caltech）出品。需要英语底子。

- 《机器学习经典书籍》
- <http://suanfazu.com/discussion/109/%E6%9C%BA%E5%99%A8%E5%AD%A6%E4%B9%A0%E7%BB%8F%E5%85%B8%E4%B9%A6%E7%B1%8D>

介绍：总结了机器学习的经典书籍，包括数学基础和算法理论的书籍，可做为入门参考书单。

- 《16 Free eBooks On Machine Learning》
- <http://efytimes.com/e1/fullnews.asp?edid=121516>

介绍:16本机器学习的电子书，可以下载下来在pad，手机上面任意时刻去阅读。不多我建议你看完一本再下载一本。

- 《A Large set of Machine Learning Resources for Beginners to Mavens》
- <http://www.erogol.com/large-set-machine-learning-resources-beginners-mavens/>

介绍:标题很大，从新手到专家。不过看完上面所有资料。肯定是专家了

- 《机器学习最佳入门学习资料汇总》
- <http://article.yeeyan.org/view/22139/410514>

介绍：入门的书真的很多，而且我已经帮你找齐了。

- 《Sibyl》
- <http://users.soe.ucsc.edu/~niejiazhong/slides/chandra.pdf>

介绍：Sibyl 是一个监督式机器学习系统，用来解决预测方面的问题，比如 YouTube 的视频推荐。

- 《Neural Network & Text Mining》
- <http://www.slideshare.net/ssuser9cc1bd/piji-li-dltm>

介绍:关于(Deep) Neural Networks在 NLP 和 Text Mining 方面一些paper的总结

- 《前景目标检测1（总结）》
- <http://www.cnblogs.com/lxy2017/p/3927226.html>

介绍:计算机视觉入门之前景目标检测1（总结）

- 《行人检测》
- <http://www.52ml.net/17004.html>

介绍:计算机视觉入门之行人检测

- 《Deep Learning – important resources for learning and understanding》
- <http://www.kdnuggets.com/2014/08/deep-learning-important-resources-learning-understanding.html>

介绍:Important resources for learning and understanding . Is awesome

- 《Machine Learning Theory: An Introductory Primer》
- <http://www.toptal.com/machine-learning/machine-learning-theory-an-introductory-primer>

介绍:这又是一篇机器学习初学者的入门文章。值得一读

- 《Neural Networks and Deep Learning》
- <http://neuralnetworksanddeeplearning.com/>

介绍:在线Neural Networks and Deep Learning电子书

- 《Python 网页爬虫 & 文本处理 & 科学计算 & 机器学习 & 数据挖掘兵器谱》
- <http://www.52nlp.cn/python-%E7%BD%91%E9%A1%B5%E7%88%AC%E8%99%AB-%E6%96%87%E6%9C%AC%E5%A4%84%E7%90%86-%E7%A7%91%E5%AD%A6%E8%AE%A1%E7%AE%97-%E6%9C%BA%E5%99%A8%E5%AD%A6%E4%B9%A0-%E6%95%B0%E6%8D%AE%E6%8C%96%E6%8E%98>

介绍:python的17个关于机器学习的工具

- 《神奇的伽玛函数(上)》
- <http://www.flickering.cn/%E6%A6%82%E7%8E%87%E7%BB%9F%E8%AE%A1/2014/06/%E7%A5%9E%E5%A5%87%E7%9A%84%E4%BC%BD%E7%8E>

介绍:下集在这里神奇的伽玛函数(下)

- 《分布式机器学习的故事》
- <http://cxwangyi.github.io/notes/2014-01-20-distributed-machine-learning.html>

介绍:作者王益目前是腾讯广告算法总监，王益博士毕业后在google任研究。这篇文章王益博士7年来从谷歌到腾讯对于分布机器学习的所见所闻。值得细读

- 《机器学习提升之道 (Level-Up Your Machine Learning) 》
- <http://metacademy.org/roadmaps/cjrd/level-up-your-ml>

介绍:把机器学习提升的级别分为0~4级，每级需要学习的教材和掌握的知识。这样，给机器学习者提供一个上进的路线图，以免走弯路。另外，整个网站都是关于机器学习的，资源很丰富。

- 《Machine Learning Surveys》
- <http://www.mlsurveys.com/>

介绍:机器学习各个方向综述的网站

- 《Deep Learning Reading list》
- <http://deeplearning.net/reading-list/>

介绍:深度学习阅资源列表

- 《Deep Learning: Methods and Applications》
- http://research.microsoft.com/pubs/219984/DeepLearningBook_RefsByLastFirstNames.pdf

介绍：这是一本来自微的研究员 li Peng和Dong Yu所著的关于深度学习的方法和应用的电子书

- 《Machine Learning Summer School 2014》
- <http://pan.baidu.com/s/1pJ0ok7T>

介绍:2014年七月CMU举办的机器学习夏季课刚刚结束 有近50小时的视频、十多个PDF版幻灯片，覆盖 深度学习，贝叶斯，分布式机器学习，伸缩性 等热点话题。所有13名讲师都是牛人：包括大牛Tom Mitchell（他的[机器学习]是名校的常用教材），还有CMU李沐。（1080P高清哟）

- 《Sibyl: 来自Google的大规模机器学习系统》
- <http://users.soe.ucsc.edu/~niejiazhong/slides/chandra.pdf>

介绍:在今年的IEEE/IFIP可靠系统和网络 (DSN) 国际会议上, Google软件工程师Tushar Chandra做了一个关于Sibyl系统的主题演讲。 Sibyl是一个监督式机器学习系统, 用来解决预测方面的问题, 比如YouTube的视频推荐。详情请阅读[google sibyl](#)

- 《Building a deeper understanding of images》
- <http://googleresearch.blogspot.com/2014/09/building-deeper-understanding-of-images.html>

介绍:谷歌研究院的Christian Szegedy在谷歌研究院的博客上简要地介绍了他们今年参加ImageNet取得好成绩的GoogLeNet系统.是关于图像处理的。

- 《Bayesian network 与python概率编程实战入门》
- <https://github.com/memect/hao/blob/master/awesome/bayesian-network-python.md>

介绍:贝叶斯学习。如果不是很清可看看[概率编程语言与贝叶斯方法实践](#)

- 《AMA: Michael I Jordan》
- http://www.reddit.com/r/MachineLearning/comments/2fxi6v/ama_michael_i_jordan/

介绍:网友问伯克利机器学习大牛、美国双料院士Michael I. Jordan: “如果你有10亿美金, 你怎么花? Jordan: “我会用这10亿美金建造一个NASA级别的自然语言处理研究项目。”

- 《机器学习&数据挖掘笔记_16 (常见面试之机器学习算法思想简单梳理) 》
- <http://www.cnblogs.com/tornadomeet/p/3395593.html>

介绍:常见面试之机器学习算法思想简单梳理,此外作者还有一些其他的[机器学习与数据挖掘文章](#)和[深度学习文章](#),不仅是理论还有源码。

- 《文本与数据挖掘视频汇总》
- <http://www.kdnuggets.com/2014/09/most-viewed-web-mining-lectures-videlectures.html>

介绍: Videolectures上最受欢迎的25个文本与数据挖掘视频汇总

- 《怎么选择深度学习的GPUs》
- <http://timdettmers.wordpress.com/2014/08/14/which-gpu-for-deep-learning/>

介绍:在Kaggle上经常取得不错成绩的Tim Dettmers介绍了他自己是怎么选择深度学习的GPUs, 以及个人如何构建深度学习的GPU集群: <http://t.cn/RhpuD1G>

- 《对话机器学习大神Michael Jordan : 深度模型》
- <http://www.infoq.com/cn/news/2014/09/depth-model>

介绍:对话机器学习大神Michael Jordan

- 《Deep Learning 和 Knowledge Graph 引爆大数据革命》
- http://blog.sina.com.cn/s/blog_46d0a3930101fswl.html

介绍:还有 2 , 3 部分。 http://blog.sina.com.cn/s/blog_46d0a3930101gs5h.html

- 《Deep Learning 教程翻译》
- http://blog.sina.com.cn/s/blog_46d0a3930101h6nf.html

介绍:是Stanford 教授 Andrew Ng 的 Deep Learning 教程, 国内的机器学习爱好者很热心的把这个教程翻译成了中文。如果你英语不好, 可以看看这个

- [《Deep Learning 101》](#)
- <http://markus.com/deep-learning-101/>

介绍:因为近两年来,深度学习在媒体界被炒作很厉害(就像大数据)。其实很多人都还不知道什么是深度学习。这篇文章由浅入深。告诉你深度学习究竟是什么!

- [《UFLDL Tutorial》](#)
- http://ufldl.stanford.edu/wiki/index.php/UFLDL_Tutorial

介绍:这是斯坦福大学做的一门课程(很勉强),这个可以给你在深度学习的路上给你一个学习的思路。里面提到了一些基本的算法。而且告诉你如何去应用到实际环境中。 [中文版](#)

- [《Toronto Deep Learning Demos》](#)
- <http://deeplearning.cs.toronto.edu/>

介绍:这是多伦多大学做的一个深度学习用来识别图片标签/图转文字的demo。是一个实际应用案例。有源码

- [《Deep learning from the bottom up》](#)
- http://metacademy.org/roadmaps/rgrosse/deep_learning

介绍:机器学习模型,阅读这个内容需要有一定的基础。

- [《R工具包的分类汇总》](#)
- <http://cran.r-project.org/web/views/>

介绍: (CRAN Task Views, 34种常见任务,每个任务又各自分类列举若干常用相关工具包) 例如: 机器学习, 自然语言处理, 时间序列分析, 空间信息分析, 多重变量分析, 计量经济学, 心理统计学, 社会学统计, 化学计量学, 环境科学, 药物代谢动力学 等

- [《机器学习常见算法分类汇总》](#)
- <http://www.ctocio.com/hotnews/15919.html>

介绍: 机器学习无疑是当前数据分析领域的一个热点内容。很多人在平时的工作中都或多或少会用到机器学习的算法。本文为您总结一下常见的机器学习算法,以供您在工作和学习中参考。

- [《Deep Learning \(深度学习\) 学习笔记整理系列》](#)
- <http://blog.csdn.net/zouxy09/article/details/8775360>

介绍: 很多干货,而且作者还总结了好几个系列。另外还作者还了一个[文章导航](#).非常的感谢作者总结。

[Deep Learning \(深度学习\) 学习笔记整理系列之\(二\)](#)

[Deep Learning \(深度学习\) 学习笔记整理系列之\(三\)](#)

[Deep Learning \(深度学习\) 学习笔记整理系列之\(四\)](#)

[Deep Learning \(深度学习\) 学习笔记整理系列之\(五\)](#)

[Deep Learning \(深度学习\) 学习笔记整理系列之\(六\)](#)

[Deep Learning \(深度学习\) 学习笔记整理系列之\(七\)](#)

[DeepLearning \(深度学习\) 学习笔记整理系列之\(八\)](#)

- [《Tutorials Session A - Deep Learning for Computer Vision》](#)
- <http://research.microsoft.com/apps/video/default.aspx?id=206976&l=i>

介绍:传送理由: Rob Fergus的用深度学习做计算机视觉的NIPS 2013教程。有mp4, mp3, pdf各种[下载](#) 他是纽约大学教授, 目前也在Facebook工作, 他2014年的8篇[论文](#)

- [《FudanNLP》](#)
- <https://github.com/xpqiu/fnlp/>

介绍:FudanNLP, 这是一个复旦大学计算机学院开发的开源中文自然语言处理(NLP)工具包

Fudan NLP里包含中文分词、关键词抽取、命名实体识别、词性标注、时间词抽取、语法分析等功能, 对搜索引擎 文本分析等极为有价值。

- [《Open Sourcing ml-ease》](#)
- <http://engineering.linkedin.com/large-scale-machine-learning/open-sourcing-ml-ease>

介绍:LinkedIn 开源的机器学习工具包,支持单机, Hadoop cluster, 和 Spark cluster 重点是 logistic regression 算法

- [《机器学习周刊》](#)
- <http://ztl2004.github.io/MachineLearningWeekly/index.html>

介绍:对于英语不好, 但又很想学习机器学习的朋友。是一个大的福利。机器学习周刊目前主要提供中文版, 还是面向广大国内爱好者, 内容涉及机器学习、数据挖掘、并行系统、图像识别、人工智能、机器人等等。谢谢作者

- [《线性代数》](#)
- <http://v.163.com/special/opencourse/daishu.html>

介绍:《线性代数》是《机器学习》的重要数学先导课程。其实《线代》这门课讲得浅显易懂特别不容易, 如果一上来就讲逆序数及行列行列式性质, 很容易让学生失去学习的兴趣。我个人推荐的最佳《线性代数》课程是麻省理工Gilbert Strang教授的课程。 [课程主页](#)

- [《Big-data》](#)
- <http://blog.andreamostosi.name/big-data/>

介绍:大数据数据处理资源、工具不完备列表, 从框架、分布式编程、分布式文件系统、键值数据模型、图数据模型、数据可视化、列存储、机器学习等。很赞的资源汇总。

- [《machine learning for smart dummies》](#)
- <http://yahoolabs.tumblr.com/post/97839313996/machine-learning-for-smart-dummies>

介绍:雅虎邀请了一名来自本古里安大学的访问学者, 制作了一套关于机器学习的系列视频课程。本课程共分为7期, 详细讲解了有关SVM, boosting, nearest neighbors, decision trees 等常规机器学习算法的理论基础知识。

- [《Entanglement-Based Quantum Machine Learning》](#)
- <http://arxiv.org/abs/1409.7770>

介绍:应对大数据时代, 量子机器学习的第一个实验 [paper](#) [下载](#)

- [《How a Math Genius Hacked OkCupid to Find True Love》](#)
- <http://www.wired.com/2014/01/how-to-hack-okcupid/all/>

介绍:Wired杂志报道了UCLA数学博士Chris McKinlay (图1) 通过大数据手段+机器学习方法破解婚恋网站配对算法找到真爱的故事, 通过Python脚本控制着12个账号, 下载了婚恋网站2万女用户的600万问题答案, 对他们进行了统计抽样及聚类分析(图2, 3), 最后终于收获了真爱。科技改变命运!

- [《Underactuated Robotics》](#)
- <https://www.edx.org/course/mitx/mitx-6-832x-underactuated-robotics-3511>

介绍:MIT的Underactuated Robotics于 2014年10月1日开课,该课属于MIT研究生级别的课程,对机器人和非线性动力系统感兴趣的朋友不妨可以挑战一下这门课程!

- [《mllib实践经验\(1\)》](#)
- <http://www.csdn.net/article/2014-12-26/2823330>

介绍:mllib实践经验分享

- [《Google Turns To Deep Learning Classification To Fight Web Spam》](#)
- <http://www.seobythesea.com/2014/09/google-turns-deep-learning-classification-fight-web-spam/>

介绍:Google用Deep Learning做的antispam(反垃圾邮件)

- [《NLP常用信息资源》](#)
- <https://github.com/memect/hao/blob/master/awesome/nlp.md>

介绍:NLP常用信息资源* [《NLP常用信息资源》](#)

- [《机器学习速查表》](#)
- <https://github.com/soulmachine/machine-learning-cheat-sheet>

介绍:机器学习速查表

- [《Best Papers vs. Top Cited Papers in Computer Science》](#)
- <http://arnetminer.org/conferencebestpapers>

介绍:从1996年开始在计算机科学的论文中被引用次数最多的论文

- [《InfiniTAM: 基于深度图像的体数据集集成框架》](#)
- <http://mmcheng.net/zh/itam/>

介绍:把今年的一个ACM Trans. on Graphics (TOG)论文中的代码整理为一个开源的算法框架,共享出来了。欢迎大家使用。可以实时的采集3D数据、重建出三维模型。Online learning, GPU Random forest, GPU CRF也会后续公开。

- [《Hacker's guide to Neural Networks》](#)
- <http://karpathy.github.io/neuralnets/>

介绍:【神经网络黑客指南】现在,最火莫过于深度学习(Deep Learning),怎样更好学习它?可以让你在浏览器中,跑起深度学习效果的超酷开源项目ConvNetJS作者karpathy告诉你,最佳技巧是,当你开始写代码,一切将变得清晰。他刚发布了一本图书,不断在线更新

- [《Building a Production Machine Learning Infrastructure》](#)
- <http://machinelearningmastery.com/building-a-production-machine-learning-infrastructure/>

介绍:前Google广告系统工程师Josh Wills 讲述工业界和学术界机器学习的异同,大实话

- [《Deep Learning Sentiment Analysis for Movie Reviews using Neo4j》](#)
- <http://neo4j.com/blog/deep-learning-sentiment-analysis-movie-reviews-using-neo4j/>

介绍:使用Neo4j 做电影评论的情感分析。

- 《DeepLearning.University – An Annotated Deep Learning Bibliography》
- <http://memkite.com/deep-learning-bibliography/>

介绍：不仅是资料，而且还对有些资料做了注释。

- 《A primer on deeping learning》
- <http://www.datarobot.com/blog/a-primer-on-deep-learning/>

介绍：深度学习入门的初级读本

- 《Machine learning is teaching us the secret to teaching 》
- <https://news.ycombinator.com/item?id=8379571>

介绍：机器学习教会了我们什么？

- 《scikit-learn：用于机器学习的Python模块》
- <http://scikit-learn.org/stable/documentation.html>

介绍：scikit-learn是在SciPy基础上构建的用于机器学习的Python模块。

- 《对话机器学习大神Michael Jordan：解析领域中各类模型》
- <http://www.infoq.com/cn/news/2014/10/interview-michael-jordan>

介绍：乔丹教授（Michael I. Jordan）教授是机器学习领域神经网络的大牛，他对深度学习、神经网络有着很浓厚的兴趣。因此，很多提问的问题中包含了机器学习领域的各类模型，乔丹教授对此——做了解释和展望。

- 《A*搜索算法的可视化短教程》
- <http://www.redblobgames.com/pathfinding/a-star/introduction.html>

介绍：A*搜索是人工智能基本算法，用于高效地搜索图中两点的最佳路径，核心是 $g(n)+h(n)$ ： $g(n)$ 是从起点到顶点 n 的实际代价， $h(n)$ 是顶点 n 到目标顶点的估算代价。合集

- 《基于云的自然语言处理开源项目FudanNLP》
- <http://code.csdn.net/news/2822123>

介绍：本项目利用了Microsoft Azure，可以在几分钟内完成NLP on Azure Website的部署，立即开始对FNLP各种特性的试用，或者以REST API的形式调用FNLP的语言分析功能

- 《吴立德《概率主题模型&数据科学基础》
- http://www.youku.com/playlist_show/id_22935176.html

介绍：现任复旦大学首席教授、计算机软件博士生导师。计算机科学研究所副所长.内部课程

- 《机器学习入门资源不完全汇总》
- <http://ml.memect.com/article/machine-learning-guide.html>

介绍：好东西的干货真的很多

- 《收集从2014年开始深度学习文献》
- <http://memkite.com/deep-learning-bibliography/>

介绍：从硬件、图像到健康、生物、大数据、生物信息再到量子计算等，Amund Tveit等维护了一个DeepLearning.University小项目：收集从2014年开始深度学习文献，相信可以作为深度学习的起点。[github](#)

- 《EMNLP上两篇关于股票趋势的应用论文》
- <http://emnlp2014.org/papers/pdf/EMNLP2014148.pdf>

介绍：EMNLP上两篇关于stock trend 用到了deep model组织特征；Exploiting Social Relations and Sentiment for Stock Prediction用到了stock network。

- 《Bengio组（蒙特利尔大学LISA组）深度学习教程》
- <http://deeplearning.net/tutorial/deeplearning.pdf>

介绍：作者是深度学习一线大牛Bengio组写的教程，算法深入显出，还有实现代码，一步步展开。

- 《学习算法的Neural Turing Machine》
- <http://arxiv.org/pdf/1410.5401v1.pdf>

介绍：许多传统的机器学习任务都是在学习function，不过谷歌目前有开始学习算法的趋势。谷歌另外的这篇学习Python程序的Learning to Execute也有相似之处

- 《Learning to Rank for Information Retrieval and Natural Language Processing》
- <http://www.morganclaypool.com/doi/abs/10.2200/S00607ED2V01Y201410HLT026>

介绍：作者是华为技术有限公司，诺亚方舟实验室，首席科学家的李航博士写的关于信息检索与自然语言处理的文章

- 《Rumor has it: Identifying Misinformation in Microblogs》
- <http://www.aclweb.org/anthology/D11-1147>

介绍：利用机器学习在谣言的判别上的应用,此外还有两个。一个是识别垃圾与虚假信息的paper.还有一个是网络舆情及其分析技术

- 《R机器学习实践》
- <http://study.163.com/course/introduction/854064.htm>

介绍：该课程是网易公开课的收费课程，不贵，超级便宜。主要适合于对利用R语言进行机器学习，数据挖掘感兴趣的人。

- 《大数据分析：机器学习算法实现的演化》
- http://ifeve.com/bigdataanalyticsbeyondhadoop_evolutionofmlrealization/

介绍：本章中作者总结了三代机器学习算法实现的演化：第一代非分布式的，第二代工具如Mahout和Rapidminer实现基于Hadoop的扩展，第三代如Spark和Storm实现了实时和迭代数据处理。BIG DATA ANALYTICS BEYOND HADOOP

- 《图像处理，分析与机器视觉》
- <http://book.douban.com/subject/5921462/>

介绍：讲计算机视觉的四部奇书（应该叫经典吧）之一，另外三本是Hartley的《多图几何》、Gonzalez的《数字图像处理》、Rafael C.Gonzalez / Richard E.Woods 的《数字图像处理》

- 《LinkedIn最新的推荐系统文章Browsemaps》
- <http://pan.baidu.com/s/1sjFeLTN>

介绍：里面基本没涉及到具体算法，但作者介绍了CF在LinkedIn的很多应用，以及他们在做推荐过程中获得的一些经验。最后一条经验是应该监控log数据的质量，因为推荐的质量很依赖数据的质量！

- 《初学者如何查阅自然语言处理（NLP）领域学术资料》
- http://blog.sina.com.cn/s/blog_574a437f01019poo.html

介绍：初学者如何查阅自然语言处理（NLP）领域学术资料

- 《树莓派的人脸识别教程》
- <http://www.open-electronics.org/raspberry-pi-and-the-camera-pi-module-face-recognition-tutorial/>

介绍：用树莓派和相机模块进行人脸识别

- 《利用深度学习与大数据构建对话系统》
- http://www.hangli-hl.com/uploads/3/1/6/8/3168008/short_text_conversation_mla.pdf

介绍：如何利用深度学习与大数据构建对话系统

- 《经典论文Leo Breiman：Statistical Modeling: The Two Cultures》
- http://lear.inrialpes.fr/people/mairal/resources/pdf/review_sparse_arxiv.pdf

介绍：Francis Bach合作的有关稀疏建模的新综述(书)：Sparse Modeling for Image and Vision Processing，内容涉及Sparsity, Dictionary Learning, PCA, Matrix Factorization等理论，以及在图像和视觉上的应用，而且第一部分关于Why does the l1-norm induce sparsity的解释也很不错。

- 《Reproducing Kernel Hilbert Space》
- <http://www.umiacs.umd.edu/~hal/docs/daume04rkhs.pdf>

介绍：RKHS是机器学习中的重要概念，其在large margin分类器上的应用也是广为熟知的。如果没有较好的数学基础，直接理解RKHS可能会不易。本文从基本运算空间讲到Banach和Hilbert空间，深入浅出，一共才12页。

- 《Hacker's guide to Neural Networks》
- <http://karpathy.github.io/neuralnets/>

介绍：许多同学对于机器学习及深度学习的困惑在于，数学方面已经大致理解了，但是动起手来却不知道如何下手写代码。斯坦福深度学习博士Andrej Karpathy写了一篇实战版本的深度学习及机器学习教程，手把手教你用Javascript写神经网络和SVM。

- 《【语料库】语料库资源汇总》
- <http://blog.csdn.net/pandalibaba/article/details/17409395>

介绍：【语料库】语料库资源汇总

- 《机器学习算法之旅》
- <http://blog.jobbole.com/60809/>

介绍：本文会过一遍最流行的机器学习算法，大致了解哪些方法可用，很有帮助。

- 《Reproducible Research in Computational Science》
- <http://www.csee.wvu.edu/~xinl/source.html>

介绍：这个里面有很多关于机器学习、信号处理、计算机视觉、深度学习、神经网络等领域的大量源代码（或可执行代码）及相关论文。科研写论文的好资源

- 《NYU 2014年的深度学习课程资料》
- <http://civl.nyu.edu/doku.php?id=deeplearning:slides:start>

介绍：NYU 2014年的深度学习课程资料，有视频

- 《计算机视觉数据集不完全汇总》
- <https://github.com/memect/hao/blob/master/awesome/computer-vision-dataset.md>

介绍：计算机视觉数据集不完全汇总

- 《Machine Learning Open Source Software》
- <http://mloss.org/software/>

介绍：机器学习开源软件

- 《LIBSVM》
- <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>

介绍：A Library for Support Vector Machines

- 《Support Vector Machines》
- <http://www.support-vector-machines.org/index.html>

介绍：数据挖掘十大经典算法之一

- 《100 Best GitHub: Deep Learning》
- <http://meta-guide.com/software-meta-guide/100-best-github-deep-learning/>

介绍：github上面100个非常棒的项目

- 《加州大学欧文分校(UCI)机器学习数据集仓库》
- <http://archive.ics.uci.edu/ml>

介绍：当前加州大学欧文分校为机器学习社区维护着306个数据集。 [查询数据集](#)

- 《Andrej Karpathy个人主页》
- <http://cs.stanford.edu/people/karpathy/>

介绍：Andrej Karpathy 是斯坦福大学Li Fei-Fei的博士生，使用机器学习在图像、视频语义分析领域取得了科研和工程上的突破，发的文章不多，但每个都很扎实，在每一个问题上都做到了state-of-art.

- 《Andrej Karpathy的深度强化学习演示》
- <http://cs.stanford.edu/people/karpathy/convnetjs/demo/rldemo.html>

介绍：Andrej Karpathy的深度强化学习演示， [论文在这里](#)

- 《CIKM数据挖掘竞赛夺冠算法-陈运文》
- <http://www.52nlp.cn/cikm-competition-topdata>

介绍：CIKM Cup(或者称为CIKM Competition)是ACM CIKM举办的国际数据挖掘竞赛的名称。

- 《Geoffrey E. Hinton》
- <http://www.cs.toronto.edu/~hinton/>

介绍：杰弗里·埃弗里斯特·辛顿 FRS是一位英国出生的计算机学家和心理学家，以其在神经网络方面的贡献闻名。辛顿是反向传播算法和对比散度算法的发明人之一，也是深度学习的积极推动者。

- 《自然语言处理的深度学习理论与实际》
- http://cikm2014.fudan.edu.cn/cikm2014/Tpl/Public/slides/CIKM14_tutorial_slides_6.pdf

介绍：微软研究院深度学习技术中心在CIKM2014 上关于《自然语言处理的深度学习理论与实际》教学讲座的幻灯片

- 《用大数据和机器学习做股票价格预测》
- <http://eugenezhulenev.com/blog/2014/11/14/stock-price-prediction-with-big-data-and-machine-learning/>

介绍：本文基于<支持向量机的高频限价订单的动态建模>采用了 Apache Spark和Spark MLlib从纽约股票交易所的订单日志数据构建价格运动预测模型。(股票有风险，投资谨慎)GitHub源代码托管地址。

- 《关于机器学习的若干理论问题》
- <http://dataunion.org/?p=2011>

介绍：徐宗本 院士将于热爱机器学习的小伙伴一起探讨有关于机器学习的几个理论性问题，并给出一些有意义的结论。最后通过一些实例来说明这些理论问题的物理意义和实际应用价值。

- 《深度学习在自然语言处理的应用》
- http://vdisk.weibo.com/s/D2szyg_bBVM0

介绍：作者还著有《这就是搜索引擎：核心技术详解》一书，主要是介绍应用层的东西

- 《Undergraduate machine learning at UBC》
- <http://www.cs.ubc.ca/~nando/340-2012/index.php>

介绍：机器学习课程

- 《人脸识别必读的N篇文章》
- http://blog.sina.com.cn/s/blog_6ae183910101h4jr.html

介绍：人脸识别必读文章推荐

- 《推荐系统经典论文文献及业界应用》
- <http://semocean.com/%E6%8E%A8%E8%8D%90%E7%B3%BB%E7%BB%9F%E7%BB%8F%E5%85%B8%E8%AE%BA%E6%96%87%E6%96%87%E7%>

介绍：推荐系统经典论文文献

- 《人脸识别必读的N篇文章》
- http://blog.sina.com.cn/s/blog_6ae183910101h4jr.html

介绍：人脸识别必读文章推荐

- 《第十二届中国“机器学习及其应用”研讨会PPT》

- 《空间数据挖掘常用方法》
- <http://blog.csdn.net/u012690204/article/details/41853731>

介绍:空间数据挖掘常用方法

- 《Use Google's Word2Vec for movie reviews》
- <https://www.kaggle.com/c/word2vec-nlp-tutorial>

介绍:Kaggle新比赛 "When bag of words meets bags of popcorn" aka "边学边用word2vec和deep learning做NLP" 里面全套教程教一步一步用python和gensim包的word2vec模型,并在实际比赛里面比调参数和清数据。 如果已装过gensim不要忘升级

- 《PyNLPIR》
- <http://pynlpir.readthedocs.org/en/latest/>

介绍:PyNLPIR提供了NLPIR/ICTCLAS汉语分词的Python接口,此外Zhon提供了常用汉字常量,如CJK字符和偏旁,中文标点,拼音,和汉字正则表达式(如找到文本中的繁体字)

- 《深度卷积神经网络下围棋》
- <http://www.technologyreview.com/view/533496/why-neural-networks-look-set-to-thrash-the-best-human-go-players-for-the-first-time/>

介绍:这篇文章说把最近模型识别上的突破应用到围棋软件上,打16万张职业棋谱训练模型识别功能。想法不错。训练后目前能做到不用计算,只看棋盘就给出下一步,大约10级棋力。但这篇文章太过乐观,说什么人类的最后一块堡垒马上就要跨掉了。话说得太早。不过,如果与别的软件结合应该还有潜力可挖。
@万精油墨绿

- 《NIPS审稿实验》
- <http://mrtz.org/blog/the-nips-experiment/>

介绍:UT Austin教授Eric Price关于今年NIPS审稿实验的详细分析,他表示,根据这次实验的结果,如果今年NIPS重新审稿的话,会有一半的论文被拒。

- 《2014年最佳的大数据,数据科学文章》
- <http://www.kdnuggets.com/2014/12/top-kdnuggets-2014-analytics-big-data-science-stories.html>

介绍:KDNuggets分别总结了2014年14个阅读最多以及分享最多的文章。我们从中可以看到多个主题——深度学习,数据科学家职业,教育和薪酬,学习数据科学的工具比如R和Python以及大众投票的最受欢迎的数据科学和数据挖掘语言

- 《机器学习经典算法详解及Python实现-线性回归 (Linear Regression) 算法》
- <http://blog.csdn.net/suipingsp/article/details/42101139>

介绍:Python实现线性回归,作者还有其他很棒的文章推荐可以看看

- 《2014中国大数据技术大会33位核心专家演讲PDF》
- <http://download.csdn.net/album/detail/1367/1/1>

介绍:2014中国大数据技术大会33位核心专家演讲PDF下载

- 《使用RNN和Paragraph Vector做情感分析》
- <http://arxiv.org/abs/1412.5335>

介绍:这是T. Mikolov & Y. Bengio最新论文Ensemble of Generative and Discriminative Techniques for Sentiment Analysis of Movie Reviews,使用RNN和PV在情感分析效果不错,[项目代码](<https://github.com/mesnilgr/iclr15>)公布在github(目前是空的)。这意味着Paragraph Vector终于揭开面纱了嘛。

- [《NLPIR/ICTCLAS2015分词系统大会上的技术演讲》](#)
- <http://pan.baidu.com/s/1o6l9S18>

介绍:NLPIR/ICTCLAS2015分词系统发布与用户交流大会上的演讲,请更多朋友检阅新版分词吧。我们实验室同学的演讲包括:[孙梦姝-基于评论观点挖掘的商品搜索技术研究](#) [李然-主题模型](#)

- [《Machine Learning is Fun!》](#)
- <https://medium.com/code-poet/80ea3ec3c471>

介绍:Convex Neural Networks 解决维数灾难

- [《CNN的反向求导及练习》](#)
- <http://dataunion.org/?p=5395>

介绍:介绍CNN参数在使用bp算法时该怎么训练,毕竟CNN中有卷积层和下采样层,虽然和MLP的bp算法本质上相同,但形式上还是有些区别的,很显然在完成CNN反向传播前了解bp算法是必须的。此外作者也做了一个[资源集:机器学习,深度学习,视觉,数学等](#)

- [《正则表达式优化成Trie树》](#)
- <https://github.com/cloudflare/ahocorasick>

介绍:如果要在一篇文章中匹配十万个关键词怎么办? [Aho-Corasick](#) 算法利用添加了返回边的Trie树,能够在线性时间内完成匹配。但如果匹配十万个正则表达式呢? 这时候可以用到把多个正则优化成Trie树的方法,如日本人写的 [Regexp::Trie](#)

- [《Deep learning Reading List》](#)
- <http://jmozah.github.io/links/>

介绍:深度学习阅读清单

- [《Caffe》](#)
- <http://caffe.berkeleyvision.org/>

介绍:Caffe是一个开源的深度学习框架,作者目前在google工作,作者主页[Yangqing Jia \(贾扬清\)](#)

- [《GoogLeNet:深度学习模型的Caffe复现》](#)
- https://github.com/BVLC/caffe/blob/master/models/bvlc_googlenet/readme.md

介绍:2014 ImageNet冠军GoogLeNet深度学习模型的Caffe复现模型,[GoogleNet](#)论文.

- [《LambdaNet, Haskell实现的开源人工神经网络库》](#)
- <https://github.com/jbarrow/LambdaNet>

介绍:LambdaNetLambdaNet是由Haskell实现的一个开源的人工神经网络库,它抽象了网络创建、训练并使用了高阶函数。该库还提供了一组预定义函数,用户可以采取多种方式组合这些函数来操作现实世界数据。

- [《百度余凯&张潼机器学习视频》](#)
- <http://wenku.baidu.com/course/view/49e8b8f67c1cfad6195fa705>

介绍:如果你从事互联网搜索,在线广告,用户行为分析,图像识别,自然语言理解,或者生物信息学,智能机器人,金融预测,那么这门核心课程你必须深入了解。

- 《杨强在TEDxNanjing谈智能的起源》
- http://v.youku.com/v_show/id_XODQzNDM4MDg0.html

介绍:“人工智能研究分许多流派。其中之一以IBM为代表,认为只要有高性能计算就可得到智能,他们的‘深蓝’击败了世界象棋冠军;另一流派认为智能来自动物本能;还有个很强的流派认为只要找来专家,把他们的思维用逻辑一条条写下,放到计算机里就行……” 杨强在TEDxNanjing谈智能的起源

- 《深度RNN/LSTM用于结构化学习 0)序列标注Connectionist Temporal ClassificationICML06》
- http://www.machinelearning.org/proceedings/icml2006/047_Connectionist_Tempor.pdf

介绍:1)机器翻译Sequence to Sequence NIPS14 2)成分句法GRAMMAR AS FOREIGN LANGUAGE

- 《Deep Learning实战之word2vec》
- <http://techblog.youdao.com/?p=915>

介绍:网易有道的三位工程师写的word2vec的解析文档,从基本的词向量/统计语言模型->NNLM->Log-Linear/Log-Bilinear->层次化Log-Bilinear,到CBOW和Skip-gram模型,再到word2vec的各种tricks,公式推导与代码,基本上是网上关于word2vec资料的大合集,对word2vec感兴趣的朋友可以看看

- 《Machine learning open source software》
- <http://mloss.org/software/>

介绍:机器学习开源软件,收录了各种机器学习的各种编程语言学术与商业的开源软件. 与此类似的还有很多例如:DMOZ - Computers: Artificial Intelligence: Machine Learning: Software, LIBSVM – A Library for Support Vector Machines, Weka 3: Data Mining Software in Java, scikit-learn:Machine Learning in Python, Natural Language Toolkit:NLTK, Machine Learning for Language Toolkit, Data Mining - Fruitful and Fun, Open Source Computer Vision Library

- 《机器学习入门者学习指南》
- <http://www.guokr.com/post/512037/>

介绍:作者是计算机研二(写文章的时候,现在是2015年了应该快要毕业了),专业方向自然语言处理. 这是一点他的经验之谈. 对于入门的朋友或许会有帮助

- 《A Tour of Machine Learning Algorithms》
- <http://machinelearningmastery.com/a-tour-of-machine-learning-algorithms/>

介绍:这是一篇关于机器学习算法分类的文章,非常好

- 《2014年的《机器学习日报》大合集》
- <http://ml.memect.com/download/2014.zip>

介绍:机器学习日报里面推荐很多内容,在这里有一部分的优秀内容就是来自机器学习日报.

- 《Image classification with deep learning常用模型》
- <http://blog.csdn.net/abcjennifer/article/details/42493493>

介绍:这是一篇关于图像分类在深度学习中的文章

- 《自动语音识别:深度学习方法》
- <http://research.microsoft.com/en-us/people/deng/>

介绍:作者与Bengio的兄弟Samy 09年合编《自动语音识别:核方法》3)李开复1989年《自动语音识别》专著,其博导、94年图灵奖得主Raj Reddy作序

- 《NLP中的中文分词技术》
- <http://blog.csdn.net/heyeshuwu/article/details/42554903>

介绍: 作者是360电商技术组成员,这是一篇NLP在中文分词中的应用

- 《Using convolutional neural nets to detect facial keypoints tutorial》
- <http://danielnouri.org/notes/2014/12/17/using-convolutional-neural-nets-to-detect-facial-keypoints-tutorial/>

介绍: 使用deep learning的人脸关键点检测, 此外还有一篇[AWS部署教程](#)

- 《书籍推荐:Advanced Structured Prediction》
- <http://www.amazon.cn/Advanced-Structured-Prediction-Nowozin-Sebastian/dp/0262028379>

介绍: 由Sebastian Nowozin等人编纂MIT出版的新书《Advanced Structured Prediction》<http://t.cn/RZxipKG>, 汇集了结构化预测领域诸多牛文, 涉及CV、NLP等领域, 值得一读。网上公开的几章草稿:[一](#),[二](#),[三](#),[四](#),[五](#)

- 《An Introduction to Matrix Concentration Inequalities》
- <http://arxiv.org/pdf/1501.01571v1.pdf>

介绍: Tropp把数学家用高深装逼的数学语言写的矩阵概率不等式用初等的方法写出来, 是非常好的手册, 领域内的paper各种证明都在用里面的结果。虽说是初等的, 但还是非常的难

- 《The free big data sources you should know》
- <https://agenda.weforum.org/2014/12/the-free-big-data-sources-you-should-know/>

介绍: 不容错过的免费大数据集, 有些已经是耳熟能详, 有些可能还是第一次听说, 内容跨越文本、数据、多媒体等, 让他们伴你开始数据科学之旅吧, 具体包括: Data.gov、US Census Bureau、European Union Open Data Portal、Data.gov.uk等

- 《A Brief Overview of Deep Learning》
- <http://yyue.blogspot.com/2015/01/a-brief-overview-of-deep-learning.html>

介绍: 谷歌科学家、Hinton亲传弟子Ilya Sutskever的深度学习综述及实际建议

- 《A Deep Dive into Recurrent Neural Nets》
- <http://nikhilbuduma.com/2015/01/11/a-deep-dive-into-recurrent-neural-networks/>

介绍: 非常好的讨论递归神经网络的文章, 覆盖了RNN的概念、原理、训练及优化等各个方面内容, 强烈推荐! 本文作者Nikhil Buduma还有一篇[Deep Learning in a Nutshell](#)值得推荐

- 《机器学习: 学习资源》
- <http://qianjiye.de/2014/11/machine-learning-resources/>

介绍: 里面融合了很多的资源, 例如竞赛, 在线课程, demo, 数据整合等。有分类

- 《Statistical foundations of machine learning》
- <https://www.otexts.org/book/sfml>

介绍: 《机器学习的统计基础》在线版, 该手册希望在理论与实践之间找到平衡点, 各主要内容都伴有实际例子及数据, 书中的例子程序都是用R语言编写的。

- 《A Deep Learning Tutorial: From Perceptrons to Deep Networks》
- <http://www.toptal.com/machine-learning/an-introduction-to-deep-learning-from-perceptrons-to-deep-networks>

介绍: IVAN VASILEV写的深度学习导引: 从浅层感知机到深度网络。高可读

- [《Research priorities for robust and beneficial artificial intelligence》](#)
- http://futureoflife.org/static/data/documents/research_priorities.pdf

介绍:鲁棒及有益的人工智能优先研究计划：一封公开信,目前已经有Stuart Russell, Tom Dietterich, Eric Horvitz, Yann LeCun, Peter Norvig, Tom Mitchell, Geoffrey Hinton, Elon Musk等人签署[The Future of Life Institute \(FLI\)](#).这封信的背景是最近霍金和Elon Musk提醒人们注意AI的潜在威胁。公开信的内容是AI科学家们站在造福社会的角度，展望人工智能的未来发展方向，提出开发AI系统的Verification, Validity, Security, Control四点要求，以及需要注意的社会问题。毕竟当前AI在经济领域，法律，以及道德领域相关研究较少。其实还有一部美剧《[疑犯追踪](#)》，介绍了AI的演进从一开始的自我学习，过滤，图像识别，语音识别等判断危险，到第四季的时候出现了机器通过学习成长之后想控制世界的状态。说到这里推荐收看。

- [《metacademy》](#)
- <http://metacademy.org/>

介绍:里面根据词条提供了许多资源，还有相关知识结构，路线图，用时长短等。号称是“机器学习”搜索引擎

- [《FAIR open sources deep-learning modules for Torch》](#)
- <https://research.facebook.com/blog/879898285375829/fair-open-sources-deep-learning-modules-for-torch/>

介绍:Facebook人工智能研究院（FAIR）开源了一系列软件库，以帮助开发者建立更大、更快的深度学习模型。开放的软件库在 Facebook 被称作模块。用它们替代机器学习领域常用的开发环境 Torch 中的默认模块，可以在更短的时间内训练更大规模的神经网络模型。

- [《浅析人脸检测之Haar分类器方法》](#)
- <http://www.cnblogs.com/ello/archive/2012/04/28/2475419.html>

介绍:本文虽然是写于2012年，但是这篇文章完全是作者的经验之作。

- [《如何成为一位数据科学家》](#)
- <http://www.ituring.com.cn/article/55994>

介绍:本文是对《机器学习实战》作者Peter Harrington做的一个访谈。包含了书中部分的疑问解答和一点个人学习建议

- [《Deep learning from the bottom up》](#)
- http://www.metacademy.org/roadmaps/rgrosse/deep_learning

介绍:非常好的深度学习概述，对几种流行的深度学习模型都进行了介绍和讨论

- [《Hands-On Data Science with R Text Mining》](#)
- <http://onepager.togaware.com/TextMiningO.pdf>

介绍:主要是讲述了利用R语言进行数据挖掘

- [《Understanding Convolutions》](#)
- <http://colah.github.io/posts/2014-07-Understanding-Convolutions/>

介绍:帮你理解卷积神经网络，讲解很清晰，此外还有两篇[Conv Nets: A Modular Perspective](#)，[Groups & Group Convolutions](#)。作者的其他关于神经网络文章也很棒

- [《Introduction to Deep Learning Algorithms》](#)
- <http://www.iro.umontreal.ca/~pift6266/H10/notes/deepintro.html#introduction-to-deep-learning-algorithms>

介绍:Deep Learning算法介绍，里面介绍了06年3篇让deep learning崛起的论文

- [《Learning Deep Architectures for AI》](#)
- http://www.iro.umontreal.ca/~bengioy/papers/ftml_book.pdf

介绍:一本学习人工智能的书籍，作者是Yoshua Bengio，相关[国内报道](#)

- [《Geoffrey E. Hinton个人主页》](#)
- <http://www.cs.toronto.edu/~hinton/>

介绍:Geoffrey Hinton是Deep Learning的大牛，他的主页放了一些介绍性文章和课件值得学习

- [《PROBABILITY THEORY: THE LOGIC OF SCIENCE》](#)
- <http://omega.albany.edu:8008/JaynesBook.html>

介绍:概率论：数理逻辑书籍

- [《H2O》](#)
- <https://github.com/h2oai/h2o>

介绍:一个用来快速的统计，机器学习并且对于数据量大的数学库

- [《ICLR 2015会议的arXiv稿件合集》](#)
- <http://www.iclr.cc/doku.php?id=iclr2015:main>

介绍:在这里你可以看到最近深度学习有什么新动向。

- [《Introduction to Information Retrieval》](#)
- <http://www-nlp.stanford.edu/IR-book/>

介绍:此书在信息检索领域家喻户晓，除提供该书的免费电子版外，还提供一个[IR资源列表](#)，收录了信息检索、网络信息检索、搜索引擎实现等方面相关的图书、研究中心、相关课程、子领域、会议、期刊等等，堪称全集，值得收藏

- [《Information Geometry and its Applications to Machine Learning》](#)
- <http://yosinski.com/mlss12/MLSS-2012-Amari-Information-Geometry/>

介绍:信息几何学及其在机器学习中的应用

- [《Legal Analytics – Introduction to the Course》](#)
- <http://computationallegalstudies.com/2015/01/legal-analytics-introduction-course-professors-daniel-martin-katz-michael-j-bommarito/>

介绍:课程《法律分析》介绍幻灯片。用机器学习解决法律相关分析和预测问题，相关的法律应用包括预测编码、早期案例评估、案件整体情况的预测，定价和工作人员预测，司法行为预测等。法律领域大家可能都比较陌生，不妨了解下。

- [《文本上的算法》](#)
- https://github.com/yanxiongglu/text_pdf

介绍:文中提到了最优，模型，最大熵等等理论，此外还有应用篇。推荐系统可以说是一本不错的阅读稿，关于模型还推荐一篇[Generative Model 与 Discriminative Model](#)

- [《NeuralTalk》](#)

- <https://github.com/karpathy/neuraltalk>

介绍: NeuralTalk is a Python+numpy project for learning Multimodal Recurrent Neural Networks that describe images with sentences. NeuralTalk是一个Python的从图像生成自然语言描述的工具。它实现了Google (Vinyals等, 卷积神经网络CNN + 长短期记忆LSTM) 和斯坦福 (Karpathy and Fei-Fei, CNN + 递归神经网络RNN)的算法。NeuralTalk自带了一个训练好的动物模型, 你可以拿狮子大象的照片来试试看

- 《Deep Learning on Hadoop 2.0》
- <https://www.paypal-engineering.com/2015/01/12/deep-learning-on-hadoop-2-0-2/>

介绍: 本文主要介绍了在Hadoop2.0上使用深度学习, 文章来自paypal

- 《Practical recommendations for gradient-based training of deep architectures》
- <http://arxiv.org/abs/1206.5533>

介绍: 用基于梯度下降的方法训练深度框架的实践推荐指导, 作者是Yoshua Bengio .感谢@xuewei4d 推荐

- 《Machine Learning With Statistical And Causal Methods》
- <http://machinelearningmastery.com/machine-learning-statistical-causal-methods/>

介绍: 用统计和因果方法做机器学习 (视频报告)

- 《Machine Learning Course 180'》
- <https://www.youtube.com/playlist?list=PLD0F06AA0D2E8FFBA>

介绍: 一个讲机器学习的Youtube视频教程。160集。系统程度跟书可比拟。

- 《回归(regression)、梯度下降(gradient descent)》
- http://www.cnblogs.com/LeftNotEasy/archive/2010/12/05/mathmatic_in_machine_learning_1_regression_and_gradient_descent.html

介绍: 机器学习中的数学, 作者的研究方向是机器学习, 并行计算如果你还想了解一点其他的可以看看他[博客](#)的其他文章

- 《美团推荐算法实践》
- <http://tech.meituan.com/mt-recommend-practice.html>

介绍: 美团推荐算法实践, 从框架, 应用, 策略, 查询等分析

- 《Deep Learning for Answer Sentence Selection》
- <http://arxiv.org/abs/1412.1632>

介绍: 深度学习用于问答系统答案句的选取

- 《Learning Semantic Representations Using Convolutional Neural Networks for Web Search 》
- <http://www.iro.umontreal.ca/~lisa/pointeurs/WWW2014.pdf>

介绍: CNN用于WEB搜索, 深度学习在文本计算中的应用

- 《Awesome Public Datasets》
- <https://github.com/caesar0301/awesome-public-datasets>

介绍: Awesome系列中的公开数据集

- 《Search Engine & Community》
- <http://www.academics.io/>

介绍: 一个学术搜索引擎

- 《spaCy》
- <http://honnibal.github.io/spaCy/>

介绍: 用Python和Cython写的工业级自然语言处理库, 号称是速度最快的NLP库, 快的原因一是用Cython写的, 二是用了个很巧妙的hash技术, 加速系统的瓶颈, NLP中稀疏特征的存取

- 《Collaborative Filtering with Spark》
- <http://fr.slideshare.net/MrChrisJohnson/collaborative-filtering-with-spark>

介绍: Fields是个数学研究中心,上面的这份ppt是来自Fields举办的活动中Russ Salakhutdinov带来的《大规模机器学习》分享

- 《Topic modeling 的经典论文》
- <http://www.7300days.com/index.php/stds/topic/list/id/27/name/Topic%20modeling>

介绍: Topic modeling 的经典论文,标注了关键点

- 《Move Evaluation in Go Using Deep Convolutional Neural Networks》
- <http://arxiv.org/abs/1412.6564>

介绍: 多伦多大学与Google合作的新论文, 深度学习也可以用来下围棋, 据说能达到六段水平

- 《机器学习周刊第二期》
- <http://ztl2004.github.io/MachineLearningWeekly/issue2.html>

介绍: 新闻, paper, 课程, book, system, CES, Robot, 此外还推荐一个深度学习入门与综述资料

- 《Learning more like a human: 18 free eBooks on Machine Learning》
- <http://www.bigdata-madesimple.com/learning-more-like-a-human-18-free-ebooks-on-machine-learning/>

介绍: 18 free eBooks on Machine Learning

- 《Recommend :Hang Li Home》
- <http://www.hangli-hl.com/>

介绍: Chief scientist of Noah's Ark Lab of Huawei Technologies. He worked at the Research Laboratories of NEC Corporation during 1990 and 2001 and Microsoft Research Asia during 2001 and 2012. [Paper](#)

- 《DEEPLARNING.UNIVERSITY – AN ANNOTATED DEEP LEARNING BIBLIOGRAPHY》
- <http://memkite.com/deep-learning-bibliography/>

介绍: DEEPLARNING.UNIVERSITY的论文库已经收录了963篇经过分类的深度学习论文了, 很多经典论文都已经收录

- 《MLMU.cz - Radim Řehůřek - Word2vec & friends (7.1.2015)》
- <https://www.youtube.com/watch?v=wTp3P2UnTfQ&hd=1>

介绍: Radim Řehůřek(Gensim开发者)在一次机器学习聚会上的报告, 关于word2vec及其优化、应用和扩展, 很实用.[国内网盘](#)

- [《Introducing streaming k-means in Spark 1.2》](#)
- <http://databricks.com/blog/2015/01/28/introducing-streaming-k-means-in-spark-1-2.html>

介绍:很多公司都用机器学习来解决问题, 提高用户体验。那么怎么可以让机器学习更实时和有效呢? Spark MLlib 1.2里面的Streaming K-means, 由斑马鱼脑神经研究的Jeremy Freeman脑神经科学家编写, 最初是为了实时处理他们每半小时1TB的研究数据, 现在发布给大家用了。

- [《LDA入门与Java实现》](#)
- <http://www.hankcs.com/nlp/lda-java-introduction-and-implementation.html>

介绍: 这是一篇面向工程师的LDA入门笔记, 并且提供一份开箱即用Java实现。本文只记录基本概念与原理, 并不涉及公式推导。文中的LDA实现核心部分采用了arbylon的LdaGibbsSampler并力所能及地注解了, 在搜狗分类语料库上测试良好, 开源在[GitHub](#)上。

- [《AMiner - Open Science Platform》](#)
- <http://aminer.org/>

介绍: AMiner是一个学术搜索引擎, 从学术网络中挖掘深度知识、面向科技大数据的挖掘。收集近4000万作者信息、8000万论文信息、1亿多引用关系、链接近8百万知识点; 支持专家搜索、机构排名、科研成果评价、会议排名。

- [《What are some interesting Word2Vec results?》](#)
- <https://www.quora.com/What-are-some-interesting-Word2Vec-results>

介绍: Quora上的主题, 讨论Word2Vec的有趣应用, Omer Levy提到了他在CoNLL2014最佳论文里的分析结果和新方法, Daniel Hammack给出了找特异词的小应用并提供了([Python](#))代码

- [《机器学习公开课汇总》](#)
- <http://blog.coursegraph.com/%E6%9C%BA%E5%99%A8%E5%AD%A6%E4%B9%A0%E5%85%AC%E5%BC%80%E8%AF%BE%E6%B1%87%E6%80%E>

介绍: 机器学习公开课汇总,虽然里面的有些课程已经归档过了, 但是还有个别的信息没有。感谢课程图谱的小编

- [《A First Course in Linear Algebra》](#)
- <http://linear.ups.edu/download.html>

介绍: 【A First Course in Linear Algebra】Robert Beezer 有答案 有移动版、打印版 使用GNU自由文档协议 引用了杰弗逊1813年的信

- [《libfacedetection》](#)
- <https://github.com/ShiqiYu/libfacedetection>

介绍:libfacedetection是深圳大学开源的一个人脸图像识别库。包含正面和多视角人脸检测两个算法。优点:速度快(OpenCV haar+adaboost的2-3倍), 准确度高(FDDB非公开类评测排名第二), 能估计人脸角度。

- [《Inverting a Steady-State》](#)
- <http://dl.acm.org/citation.cfm?doid=2684822.2685310>

介绍:WSDM2015最佳论文 把马尔可夫链理论用在了图分析上面, 比一般的propagation model更加深刻一些。通过全局的平稳分布去求解每个节点影响系数模型。假设合理(转移受到相邻的影响系数影响)。可以用来反求每个节点的影响系数

- [《机器学习入门书单》](#)
- <http://pan.baidu.com/s/1pJogO7x>

介绍:机器学习入门书籍, [具体介绍](#)

- 《The Trouble with SVMs》
- <http://v1v3kn.tumblr.com/post/47193952400/the-trouble-with-svms>

介绍: 非常棒的强调特征选择对分类器重要性的文章。情感分类中, 根据互信息对复杂高维特征降维再使用朴素贝叶斯分类器, 取得了比SVM更理想的效果, 训练和分类时间也大大降低——更重要的是, 不必花大量时间在学习和优化SVM上——特征也一样no free lunch

- 《Rise of the Machines》
- <http://www.stat.cmu.edu/~larry/Wasserman.pdf>

介绍: CMU的统计系和计算机系知名教授Larry Wasserman 在《机器崛起》, 对比了统计和机器学习的差异

- 《实例详解机器学习如何解决问题》
- <http://tech.meituan.com/mt-mlinaction-how-to-ml.html>

介绍: 随着大数据时代的到来, 机器学习成为解决问题的一种重要且关键的工具。不管是工业界还是学术界, 机器学习都是一个炙手可热的方向, 但是学术界和工业界对机器学习的研究各有侧重, 学术界侧重于对机器学习理论的研究, 工业界侧重于如何用机器学习来解决问题。这篇文章是美团的实际环境中的实战篇

- 《Gaussian Processes for Machine Learning》
- <http://www.gaussianprocess.org/gpml/>

介绍: 面向机器学习的高斯过程, 章节概要: 回归、分类、协方差函数、模型选择与超参优化、高斯模型与其他模型关系、大数据集的逼近方法等, [微盘下载](#)

- 《FuzzyWuzzy: Fuzzy String Matching in Python》
- <http://chairnerd.seatgeek.com/fuzzywuzzy-fuzzy-string-matching-in-python/>

介绍: Python下的文本模糊匹配库, 老库新推, 可计算串间ratio(简单相似系数)、partial_ratio(局部相似系数)、token_sort_ratio(词排序相似系数)、token_set_ratio(词集合相似系数)等 [github](#)

- 《Blocks》
- <http://blocks.readthedocs.org/en/latest/>

介绍: Blocks是基于Theano的神经网络搭建框架, 集成相关函数、管道和算法, 帮你更快地创建和管理NN模块.

- 《Introduction to Machine Learning》
- <http://alex.smola.org/teaching/10-701-15/>

介绍: 机器学习大神Alex Smola在CMU新一期的机器学习入门课程"Introduction to Machine Learning"近期刚刚开课, 课程4K高清视频同步到Youtube上, 目前刚刚更新到 2.4 Exponential Families, 课程视频[playlist](#), 感兴趣的同学可以关注, 非常适合入门.

- 《Collaborative Feature Learning from Social Media》
- <http://arxiv.org/abs/1502.01423>

介绍: 用社交用户行为学习图片的协同特征, 可更好地表达图片内容相似性。由于不依赖于人工标签(标注), 可用于大规模图片处理, 难在用户行为数据的获取和清洗; 利用社会化特征的思路值得借鉴.

- 《Introducing practical and robust anomaly detection in a time series》
- <https://blog.twitter.com/2015/introducing-practical-and-robust-anomaly-detection-in-a-time-series>

介绍: Twitter技术团队对前段时间开源的时间序列异常检测算法(S-H-ESD)R包的介绍, 其中对异常的定义和分析很值得参考, 文中也提到——异常是强针对性的, 某个领域开发的异常检测在其他领域直接用可不行.

- 《Empower Your Team to Deal with Data-Quality Issues》
- <http://www.destinationcrm.com/Articles/Web-Exclusives/Viewpoints/Empower-Your-Team-to-Deal-with-Data-Quality-Issues-101308.aspx>

介绍:聚焦数据质量问题的应对,数据质量对各种规模企业的性能和效率都至关重要,文中总结出(不限于)22种典型数据质量问题显现的信号,以及典型的数据质量解决方案(清洗、去重、统一、匹配、权限清理等)

- 《中文分词入门之资源》
- <http://www.52nlp.cn/%E4%B8%AD%E6%96%87%E5%88%86%E8%AF%8D%E5%85%A5%E9%97%A8%E4%B9%8B%E8%B5%84%E6%BA%90>

介绍:中文分词入门之资源.

- 《Deep Learning Summit, San Francisco, 2015》
- <https://www.youtube.com/playlist?list=PLnDbcXCpYZ8ICKExMs8k4PtIbani9ESX3>

介绍:15年旧金山深度学习峰会视频集萃 [国内云盘](#)

- 《Introduction to Conditional Random Fields》
- <http://blog.echen.me/2012/01/03/introduction-to-conditional-random-fields/>

介绍:很好的条件随机场(CRF)介绍文章,作者的学习笔记

- 《A Fast and Accurate Dependency Parser using Neural Networks》
- <http://cs.stanford.edu/~danqi/papers/emnlp2014.pdf>

介绍: 来自Stanford, 用神经网络实现快速准确的依存关系解析器

- 《Which GPU(s) to Get for Deep Learning: My Experience and Advice for Using GPUs in Deep Learning》
- <https://timdettmers.wordpress.com/2014/08/14/which-gpu-for-deep-learning/>

介绍:做深度学习如何选择GPU的建议

- 《Sparse Linear Models》
- <http://new.livestream.com/accounts/10932136/events/3779068>

介绍: Stanford的Trevor Hastie教授在H2O.ai Meet-Up上的报告, 讲稀疏线性模型——面向“宽数据”(特征维数超过样本数)的线性模型, 13年同[主题报告](#)、[讲义](#).

- 《Awesome Computer Vision》
- <https://github.com/jbhuang0604/awesome-computer-vision>

介绍: 分类整理的机器视觉相关资源列表, 秉承Awesome系列风格, 有质有量!作者的更新频率也很频繁

- 《Adam Szeidl》
- http://www.personal.ceu.hu/staff/Adam_Szeidl/

介绍: social networks course

- 《Building and deploying large-scale machine learning pipelines》
- <http://radar.oreilly.com/2015/01/building-and-deploying-large-scale-machine-learning-pipelines.html/>

介绍: 大规模机器学习流程的构建与部署.

- [《人脸识别开发包》](#)
- <http://download.csdn.net/detail/lswtzw/8469997>

介绍: 人脸识别二次开发包, 免费, 可商用, 有演示、范例、说明书.

- [《Understanding Natural Language with Deep Neural Networks Using Torch》](#)
- <http://devblogs.nvidia.com/parallelforall/understanding-natural-language-deep-neural-networks-using-torch/>

介绍: 采用Torch用深度学习网络理解NLP, 来自Facebook 人工智能的文章.

- [《The NLP Engine: A Universal Turing Machine for NLP》](#)
- <http://arxiv.org/pdf/1503.00168.pdf>

介绍: 来自CMU的Ed Hovy和Stanford的Jiwei Li一篇有意思的Arxiv文章, 作者用Shannon Entropy来刻画NLP中各项任务的难度.

- [《The Probabilistic Relevance Framework: BM25 and Beyond》](#)
- http://staff.city.ac.uk/~sb317/papers/foundations_bm25_review.pdf

介绍: 信息检索排序模型BM25(Besting Matching). 1) 从经典概率模型演变而来 2) 捕捉了向量空间模型中三个影响索引项权重的因子: IDF逆文档频率; TF索引项频率; 文档长度归一化. 3) 并且含有集成学习的思想: 组合了BM11和BM15两个模型. 4) 作者是BM25的提出者和Okapi实现者Robertson.

- [《Introduction to ARMA Time Series Models – simplified》](#)
- <http://www.analyticsvidhya.com/blog/2015/03/introduction-auto-regression-moving-average-time-series/>

介绍: 自回归滑动平均(ARMA)时间序列的简单介绍, ARMA是研究时间序列的重要方法, 由自回归模型 (AR模型) 与滑动平均模型 (MA模型) 为基础“混合”构成.

- [《Encoding Source Language with Convolutional Neural Network for Machine Translation》](#)
- <http://arxiv.org/pdf/1503.01838v1.pdf>

介绍: 把来自target的attention signal加入source encoding CNN的输入, 得到了比BBN的模型好的多neural network joint model

- [《Spices form the basis of food pairing in Indian cuisine》](#)
- <http://arxiv.org/abs/1502.03815>

介绍: 揭开印度菜的美味秘诀——通过对大量食谱原料关系的挖掘, 发现印度菜美味的原因之一是其中的味道互相冲突, 很有趣的文本挖掘研究

- [《HMM相关文章索引》](#)
- <http://www.52nlp.cn/hmm%E7%9B%B8%E5%85%B3%E6%96%87%E7%AB%A0%E7%B4%A2%E5%BC%95>

介绍: HMM相关文章, 此外推荐中文分词之HMM模型详解

- [《Zipf's and Heap's law》](#)
- http://www.ccs.neu.edu/home/ekanou/ISU535.09X2/Handouts/Review_Material/zipfslaw.pdf

介绍: 1) 词频与其降序排序的关系, 最著名的是语言学家齐夫(Zipf, 1902-1950) 1949年提出的Zipf's law, 即二者成反比关系. 曼德勃罗(Mandelbrot, 1924- 2010) 引入参数修正了对甚高频和甚低频词的刻画 2) Heaps' law: 词汇表与语料规模的平方根(这是一个参数, 英语0.4-0.6)成正比

- [《I am Jürgen Schmidhuber, AMA》](#)
- http://www.reddit.com/r/MachineLearning/comments/2xcyrl/i_am_j%C3%BCrgen_schmidhuber_ama/

介绍: Jürgen Schmidhuber在Reddit上的AMA(Ask Me Anything)主题, 有不少RNN和AI、ML的干货内容, 关于开源&思想&方法&建议.....耐心阅读, 相信你会受益匪浅。

- [《学术种子网站 : AcademicTorrents》](#)
- <http://academictorrents.com/>

介绍: 成G上T的学术数据, HN近期热议话题,主题涉及机器学习、NLP、SNA等。下载最简单的方法, 通过BT软件, RSS订阅各集合即可

- [《机器学习交互速查表》](#)
- http://scikit-learn.org/stable/tutorial/machine_learning_map/index.html

介绍: Scikit-Learn官网提供, 在原有的Cheat Sheet基础上加上了Scikit-Learn相关文档的链接, 方便浏览

- [《A Full Hardware Guide to Deep Learning》](#)
- <https://timdettmers.wordpress.com/2015/03/09/deep-learning-hardware-guide/>

介绍: 深度学习的全面硬件指南, 从GPU到RAM、CPU、SSD、PCIe

- [《行人检测\(Pedestrian Detection\)资源》](#)
- <http://hi.baidu.com/susongzhi/item/085983081b006311eafe38e7>

介绍: Pedestrian Detection paper & data

- [《A specialized face-processing network consistent with the representational geometry of monkey face patches》](#)
- <http://arxiv.org/abs/1502.01241>

介绍: 【神经科学碰撞人工智能】在脸部识别上你我都是专家, 即使细微的差别也能辨认。研究已证明人类和灵长类动物在面部加工上不同于其他物种, 人类使用梭状回面孔区(FFA)。Khaligh-Razavi等通过计算机模拟出人脸识别的FFA活动, 堪称神经科学与人工智能的完美结合。

- [《Neural Net in C++ Tutorial》](#)
- <https://vimeo.com/19569529>

介绍: 神经网络C++教程, 本文介绍了用可调节梯度下降和可调节动量法设计和编码经典BP神经网络, 网络经过训练可以做出惊人和美妙的东西出来。此外作者博客的其他文章也很不错。

- [《How to Choose a Neural Network》](#)
- <http://deeplearning4j.org/neuralnetworktable.html>

介绍: deeplearning4j官网提供的实际应用场景NN选择参考表, 列举了一些典型问题建议使用的神经网络

- [《Deep Learning \(Python, C/C++, Java, Scala, Go\)》](#)
- <https://github.com/yusugomori/DeepLearning>

介绍: 一个深度学习项目, 提供了Python, C/C++, Java, Scala, Go多个版本的代码

- [《Deep Learning Tutorials》](#)
- <http://deeplearning.net/tutorial/>

介绍: 深度学习教程, github

- [《自然语言处理的发展趋势——访卡内基梅隆大学爱德华·霍威教授》](#)
- <http://www.ccf.org.cn/resources/1190201776262/2015/03/12/15.pdf>

介绍:自然语言处理的发展趋势——访卡内基梅隆大学爱德华·霍威教授.

- [《FaceNet: A Unified Embedding for Face Recognition and Clustering》](#)
- <http://arxiv.org/abs/1503.03832>

介绍:Google对Facebook DeepFace的有力回击—— FaceNet, 在LFW(Labeled Faces in the Wild)上达到99.63%准确率(新纪录), FaceNet embeddings可用于人脸识别、鉴别和聚类.

- [《MLlib中的Random Forests和Boosting》](#)
- <http://databricks.com/blog/2015/01/21/random-forests-and-boosting-in-mllib.html>

介绍:本文来自Databricks公司网站的一篇博客文章, 由Joseph Bradley和Manish Amde撰写, 文章主要介绍了Random Forests和Gradient-Boosted Trees (GBTs) 算法和他们在MLlib中的分布式实现, 以及展示一些简单的例子并建议该从何处上手. [中文版](#).

- [《Sum-Product Networks\(SPN\) 》](#)
- <http://spn.cs.washington.edu/index.shtml>

介绍:华盛顿大学Pedro Domingos团队的DNN, 提供论文和实现代码.

- [《Neural Network Dependency Parser》](#)
- <http://nlp.stanford.edu/software/nndep.shtml>

介绍:基于神经网络的自然语言依存关系解析器(已集成至Stanford CoreNLP), 特点是超快、准确, 目前可处理中英文语料, 基于 [《A Fast and Accurate Dependency Parser Using Neural Networks》](#) 思路实现.

- [《神经网络语言模型》](#)
- <http://www.flickering.cn/nlp/2015/03/%E6%88%91%E4%BB%AC%E6%98%AF%E8%BF%99%E6%A0%B7%E7%90%86%E8%A7%A3%E8%AF%AD%E8%3%E7%A5%9E%E7%BB%8F%E7%BD%91%E7%BB%9C%E8%AF%AD%E8%A8%80%E6%A8%A1%E5%9E%8B/>

介绍:本文根据神经网络的发展历程, 详细讲解神经网络语言模型在各个阶段的形式, 其中的模型包含NNLM[Bengio, 2003]、Hierarchical NNLM[Bengio, 2005]、Log-Bilinear[Hinton, 2007]、SENN等变形, 总结的特别好.

- [《Classifying Spam Emails using Text and Readability Features》](#)
- http://www.elg.uottawa.ca/~nat/Courses/csi5387_Winter2014/paper13.pdf

介绍:经典问题的新研究: 利用文本和可读性特征分类垃圾邮件.

- [《BCI Challenge @ NER 2015》](#)
- <https://github.com/alexandrebarachant/bci-challenge-ner-2015>

介绍:Kaggle脑控计算机交互(BCI)竞赛优胜方案源码及文档, 包括完整的数据处理流程, 是学习Python数据处理和Kaggle经典参赛框架的绝佳实例

- [《IPOL Journal · Image Processing On Line》](#)
- <http://www.ipol.im/>

介绍:IPOL (在线图像处理) 是图像处理和图像分析的研究期刊, 每篇文章都包含一个算法及相应的代码、Demo和实验文档. 文本和源码是经过同行评审的. IPOL是开放的科学和可重复的研究期刊. 我一直想做点类似的工作, 拉近产品和技术之间的距离.

- 《Machine learning classification over encrypted data》
- <http://eprint.iacr.org/2014/331>

介绍:出自MIT, 研究加密数据高效分类问题.

- 《purine2》
- <https://github.com/purine/purine2>

介绍:新加坡LV实验室的神经网络并行框架Purine: A bi-graph based deep learning framework,支持构建各种并行的架构, 在多机多卡, 同步更新参数的情况下基本达到线性加速. 12块Titan 20小时可以完成Googlenet的训练.

- 《Machine Learning Resources》
- <http://michal.io/machine-learning-resources/>

介绍:这是一个机器学习资源库,虽然比较少,但蚊子再小也是肉,有突出部分.此外还有一个由zheng Rui整理的机器学习资源.

- 《Hands-on with machine learning》
- <https://github.com/cjdd3b/nicar2015/tree/master/machine-learning>

介绍:Chase Davis在NICAR15上的主题报告材料, 用Scikit-Learn做监督学习的入门例子.

- 《The Natural Language Processing Dictionary》
- <http://www.cse.unsw.edu.au/~billw/nlpdict.html>

介绍:这是一本自然语言处理的词典,从1998年开始到目前积累了成千上万的专业词语解释,如果你是一位刚入门的朋友,可以借这本词典让自己成长更快.

- 《PageRank Approach to Ranking National Football Teams》
- <http://arxiv.org/abs/1503.01331>

介绍:通过分析1930年至今的比赛数据, 用PageRank计算世界杯参赛球队排行榜.

- 《R Tutorial》
- <http://cyclismo.org/tutorial/R/>

介绍:R语言教程,此外还推荐一个R语言教程An Introduction to R.

- 《Fast unfolding of communities in large networks》
- <http://arxiv.org/abs/0803.0476>

介绍:经典老文, 复杂网络社区发现的高效算法, Gephi中的Community detection即基于此.

- 《NUML》
- <http://numl.net/>

介绍: 一个面向 .net 的开源机器学习库,github地址

- 《synaptic.js》
- <http://synaptic.juancazala.com/>

介绍: 支持node.js的JS神经网络库, 可在客户端浏览器中运行, 支持LSTM等 [github地址](#)

- 《Machine learning for package users with R (1): Decision Tree》
- <http://tjo-en.hatenablog.com/entry/2015/03/20/191614>

介绍: 决策树

- 《Deep Learning, The Curse of Dimensionality, and Autoencoders》
- <http://www.kdnuggets.com/2015/03/deep-learning-curse-dimensionality-autoencoders.html>

介绍: 讨论深度学习自动编码器如何有效应对维数灾难, [国内翻译](#)

- 《Advanced Optimization and Randomized Methods》
- <http://www.cs.cmu.edu/~suvrit/teach/>

介绍: CMU的优化与随机方法课程, 由A. Smola和S. Sra主讲, 优化理论是机器学习的基石, 值得深入学习 [国内云\(视频\)](#)

- 《CS231n: Convolutional Neural Networks for Visual Recognition》
- <http://cs231n.stanford.edu/reports.html>

介绍: “面向视觉识别的CNN”课程设计报告集锦.近百篇, 内容涉及图像识别应用的各个方面

- 《Topic modeling with LDA: MLlib meets GraphX》
- <http://databricks.com/blog/2015/03/25/topic-modeling-with-lda-mllib-meets-graphx.html>

介绍: 用Spark的MLlib+GraphX做大规模LDA主题抽取.

- 《Deep Learning for Multi-label Classification》
- <http://arxiv.org/abs/1502.05988>

介绍: 基于深度学习的多标签分类, 用基于RBM的DBN解决多标签分类(特征)问题

- 《Google DeepMind publications》
- <http://deepmind.com/publications.html>

介绍: DeepMind论文集锦

- 《kaldi》
- <http://kaldi-asr.org/>

介绍: 一个开源语音识别工具包, 它目前托管在[sourceforge](#)上面

- 《Data Journalism Handbook》
- <http://datajournalismhandbook.org/>

介绍: 免费电子书《数据新闻手册》, 国内有热心的朋友翻译了[中文版](#), 大家也可以[在线阅读](#)

- 《Data Mining Problems in Retail》
- <https://highlyscalable.wordpress.com/2015/03/10/data-mining-problems-in-retail/>

介绍: 零售领域的数据挖掘文章.

- [《Understanding Convolution in Deep Learning》](#)
- <https://timdettmers.com/2015/03/26/convolution-deep-learning/>

介绍: 深度学习卷积概念详解,深入浅出.

- [《pandas: powerful Python data analysis toolkit》](#)
- <http://pandas.pydata.org/>

介绍: 非常强大的Python的数据分析工具包.

- [《Text Analytics 2015》](#)
- <http://breakthroughanalysis.com/2015/03/23/text-analytics-2015/>

介绍: 2015文本分析(商业)应用综述.

- [《Deep Learning libraries and first experiments with Theano》](#)
- <http://www.slideshare.net/VincenzoLomonaco/deep-learning-libraries-and-rst-experiments-with-theano>

介绍: 深度学习框架、库调研及Theano的初步测试体会报告.

- [《DEEP learning》](#)
- <http://www.deeplearningbook.org/>

介绍: MIT的Yoshua Bengio, Ian Goodfellow, Aaron Courville著等人讲深度学习的新书, 还未定稿, 线上提供Draft chapters收集反馈, 超赞! 强烈推荐.

- [《simplebayes》](#)
- <https://github.com/hickeroar/simplebayes>

介绍: Python下开源可持久化朴素贝叶斯分类库.

- [《Paracel》](#)
- <http://paracel.io/>

介绍: Paracel is a distributed computational framework designed for machine learning problems, graph algorithms and scientific computing in C++.

- [《HanLP: Han Language processing》](#)
- <http://hanlp.linrunsoft.com/>

介绍: 开源汉语言处理包.

- [《Simple Neural Network implementation in Ruby》](#)
- <http://www.rubylab.io/2015/03/18/simple-neural-network-impenentation-in-ruby/>

介绍: 使用Ruby实现简单的神经网络例子.

- [《Hacker's guide to Neural Networks》](#)
- <https://karpathy.github.io/neuralnets/>

介绍:神经网络黑客入门.

- [《The Open-Source Data Science Masters》](#)
- <http://datasciencemasters.org/>

介绍:好多数据科学家名人推荐,还有资料.

- [《Text Understanding from Scratch》](#)
- <http://arxiv.org/abs/1502.01710>

介绍:实现项目已经开源在github上面[Crepe](#)

- [《Improving Distributional Similarity with Lessons Learned from Word Embeddings》](#)
- <https://levyomer.files.wordpress.com/2015/03/improving-distributional-similarity-tacl-2015.pdf>

介绍:作者发现,经过调参,传统的方法也能和word2vec取得差不多的效果.另外,无论作者怎么试,GloVe都比不过word2vec.

- [《CS224d: Deep Learning for Natural Language Processing》](#)
- <http://cs224d.stanford.edu/index.html>

介绍:Stanford深度学习与自然语言处理课程,Richard Socher主讲.

- [《Math Essentials in Machine Learning》](#)
- http://courses.washington.edu/css490/2012.Winter/lecture_slides/02_math_essentials.pdf

介绍:机器学习中的重要数学概念.

- [《Improved Semantic Representations From Tree-Structured Long Short-Term Memory Networks》](#)
- <http://arxiv.org/abs/1503.00007>

介绍:用于改进语义表示的树型LSTM递归神经网络,句子级相关性判断和情感分类效果很好.[实现代码](#).

- [《Statistical Machine Learning》](#)
- <http://www.stat.cmu.edu/~larry/=sml/>

介绍:卡耐基梅隆Ryan Tibshirani和Larry Wasserman开设的机器学习课程,先修课程为机器学习(10-715)和中级统计学(36-705),聚焦统计理论和方法在机器学习领域应用.

- [《AM207: Monte Carlo Methods, Stochastic Optimization》](#)
- <http://am207.org/>

介绍:《哈佛大学蒙特卡洛方法与随机优化课程》是哈佛应用数学研究生课程,由V. Karygin-Fittkau、P. Protopapas主讲,Python程序示例,对贝叶斯推理感兴趣的朋友一定要看看,提供授课视频及[课上IPN讲义](#).

- [《生物医学的SPARK大数据应用》](#)
- <http://spark-summit.org/wp-content/uploads/2015/03/SSE15-40-Danford.pdf>

介绍:生物医学的SPARK大数据应用.并且伯克利开源了他们的big data genomics系统[ADAM](#),其他的内容可以关注一下[官方主页](#).

- [《ACL Anthology》](#)
- <http://aclanthology.info/>

介绍:对自然语言处理技术或者机器翻译技术感兴趣的亲们,请在提出自己牛逼到无以伦比的idea(自动归纳翻译规律、自动理解语境、自动识别语义等等)之前,请通过谷歌学术简单搜一下,如果谷歌不可用,这个网址有这个领域几大顶会的论文列表,切不可断章取义,胡乱假设.

- [《Twitter Sentiment Detection via Ensemble Classification Using Averaged Confidence Scores》](#)
- http://www.uni-weimar.de/medien/webis/publications/papers/stein_2015b.pdf

介绍:论文+代码:基于集成方法的Twitter情感分类,实现代码.

- [《NIPS 2014 CIML workshop》](#)
- <http://ciml.chalearn.org/schedule>

介绍:NIPS CiML 2014的PPT,NIPS是神经信息处理系统进展大会的英文简称.

- [《CS231n: Convolutional Neural Networks for Visual Recognition》](#)
- <http://cs231n.stanford.edu/reports.html>

介绍:斯坦福的深度学习课程的Projects 每个人都要写一个论文级别的报告 里面有一些很有意思的应用 大家可以看看 .

- [《A Speed Comparison Between Flexible Linear Regression Alternatives in R》](#)
- <http://www.sumsar.net/blog/2015/03/a-speed-comparison-between-flexible-linear-regression-alternatives-in-r/>

介绍:R语言线性回归多方案速度比较具体方案包括lm()、nls()、glm()、bayesglm()、nls()、mle2()、optim()和Stan's optimizing()等.

- [《Back-to-Basics Weekend Reading - Machine Learning》](#)
- <http://www.allthingsdistributed.com/2015/04/machine-learning.html>

介绍:文中提到的三篇论文(机器学习那些事、无监督聚类综述、监督分类综述)都很经典,Domnigos的机器学习课也很精彩

- [《A Probabilistic Theory of Deep Learning》](#)
- <http://arxiv.org/abs/1504.00641>

介绍:莱斯大学(Rice University)的深度学习概率理论.

- [《Nonsensical beer reviews via Markov chains》](#)
- <http://www.gregreda.com/2015/03/30/beer-review-markov-chains/>

介绍:基于马尔可夫链自动生成啤酒评论的开源Twitter机器人,github地址.

- [《Deep Learning for Natural Language Processing \(without Magic\)》](#)
- <http://nlp.stanford.edu/courses/NAACL2013/>

介绍:视频+讲义:深度学习用于自然语言处理教程(NAAACL13).

- [《Introduction to Data Analysis using Machine Learning》](#)
- <https://www.youtube.com/watch?v=U4lYsLgNgoY&hd=1>

介绍:用机器学习做数据分析,David Taylor最近在McGill University研讨会上的报告,还提供了一系列讲机器学习方法的ipn,很有价值 [GitHub](#).国内

- [《Beyond Short Snippets: Deep Networks for Video Classification》](#)
- <http://arxiv.org/abs/1503.08909>

介绍:基于CNN+LSTM的视频分类,google演示.

- [《How does Quora use machine learning in 2015?》](#)
- <http://www.quora.com/How-does-Quora-use-machine-learning-in-2015/answer/Xavier-Amatriain>

介绍:Quora怎么用机器学习.

- [《Amazon Machine Learning – Make Data-Driven Decisions at Scale》](#)
- <https://aws.amazon.com/cn/blogs/aws/amazon-machine-learning-make-data-driven-decisions-at-scale/>

介绍:亚马逊在机器学习上面的一些应用,代码示例.

- [《Parallel Machine Learning with scikit-learn and IPython》](#)
- https://github.com/ogrisel/parallel_ml_tutorial

介绍:并行机器学习指南(基于scikit-learn和IPython).notebook

- [《Intro to machine learning with scikit-learn》](#)
- <http://blog.kaggle.com/2015/04/08/new-video-series-introduction-to-machine-learning-with-scikit-learn/>

介绍:DataSchool的机器学习基本概念教学.

- [《DeepCLn》](#)
- <https://github.com/hughperkins/DeepCL>

介绍:一个基于OpenGL实现的卷积神经网络,支持Linux及Windows系.

- [《An Inside Look at the Components of a Recommendation Engine》](#)
- <https://www.mapr.com/blog/inside-look-at-components-of-recommendation-engine>

介绍:基于Mahout和Elasticsearch的推荐系统.

- [《Forecasting in Economics, Business, Finance and Beyond》](#)
- <http://www.ssc.upenn.edu/~fdiebold/Teaching221/econ221.html>

介绍:Francis X. Diebold的《(经济|商业|金融等领域)预测方法》.

- [《Time Series Econometrics - A Concise Course》](#)
- <http://www.ssc.upenn.edu/~fdiebold/Teaching706/econ706Penn.html>

介绍:Francis X. Diebold的《时序计量经济学》.

- [《A comparison of open source tools for sentiment analysis》](#)
- <http://fotiad.is/blog/sentiment-analysis-comparison/>

介绍:基于Yelp数据集的开源情感分析工具比较,评测覆盖Naive Bayes、SentiWordNet、CoreNLP等.

- [《Pattern Recognition And Machine Learning》](#)
- http://vdisk.weibo.com/s/ayG13we2u_sAZ

介绍:国内Pattern Recognition And Machine Learning读书会资源汇总,各章pdf讲稿,博客.

- [《Probabilistic Data Structures for Web Analytics and Data Mining 》](#)
- <https://highlyscalable.wordpress.com/2012/05/01/probabilistic-structures-web-analytics-data-mining/>

介绍:用于Web分析和数据挖掘的概率数据结构.

- [《Machine learning in navigation devices: detect maneuvers using accelerometer and gyroscope》](#)
- <https://blindmotion.github.io/2015/04/11/ml-in-navigation/>

介绍:机器学习在导航上面的应用.

- [《Neural Networks Demystified 》](#)
- <https://www.youtube.com/user/Taylor34/videos>

介绍:Neural Networks Demystified系列视频, Stephen Welch制作, 纯手绘风格, 浅显易懂,国内云.

- [《swirl + DataCamp 》](#)
- <https://www.datacamp.com/swirl-r-tutorial>

介绍:{swirl}数据训练营:R&数据科学在线交互教程.

- [《Learning to Read with Recurrent Neural Networks 》](#)
- <http://blog.terminal.com/recurrent-neural-networks-deep-net-optimization-1stm/>

介绍:关于深度学习和RNN的讨论 [Sequence to Sequence Learning with Neural Networks](#).

- [《深度强化学习 \(Deep Reinforcement Learning \) 的资源》](#)
- <http://wanghaitao8118.blog.163.com/blog/static/13986977220153811210319/>

介绍:Deep Reinforcement Learning.

- [《Machine Learning with Scikit-Learn》](#)
- https://github.com/jakevdp/sklearn_pycon2015

介绍:(PyCon2015)Scikit-Learn机器学习教程,Parallel Machine Learning with scikit-learn and IPython.

- [《PDNN》](#)
- <http://www.cs.cmu.edu/~ymiao/pdnnk.html>

介绍:PDNN: A Python Toolkit for Deep Learning.

- [《Introduction to Machine Learning》](#)
- <http://alex.smola.org/teaching/10-701-15/index.html>

介绍:15年春季学期CMU的机器学习课程, 由Alex Smola主讲, 提供讲义及授课视频, 很不错.国内镜像.

- [《Big Data Processing》](#)
- <http://www.st.ewi.tudelft.nl/~hauff/TI2736-B.html>

介绍:大数据处理课.内容覆盖流处理、MapReduce、图算法等.

- [《Spark MLlib: Making Practical Machine Learning Easy and Scalable》](#)
- <https://www.hakkalabs.co/articles/spark-mlib-making-practical-machine-learning-easy-and-scalable>

介绍:用Spark MLlib实现易用可扩展的机器学习,国内镜像.

- [《Picture: A Probabilistic Programming Language for Scene Perception》](#)
- http://mrkulk.github.io/www_cvpr15/

介绍:以往上千行代码概率编程(语言)实现只需50行.

- [《Beautiful plotting in R: A ggplot2 cheatsheet》](#)
- <http://zevross.com/blog/2014/08/04/beautiful-plotting-in-r-a-ggplot2-cheatsheet-3/>

介绍:ggplot2速查小册子,另外一个,此外还推荐 [《A new data processing workflow for R: dplyr, magrittr, tidyr, ggplot2》](#) .

- [《Using Structured Events to Predict Stock Price Movement: An Empirical Investigation》](#)
- <http://emnlp2014.org/papers/pdf/EMNLP2014148.pdf>

介绍:用结构化模型来预测实时股票行情.

- [《International Joint Conference on Artificial Intelligence Accepted paper》](#)
- <http://ijcai-15.org/index.php/accepted-papers>

介绍:国际人工智能联合会议录取论文列表,大部分论文可使用Google找到.

- [《Why GEMM is at the heart of deep learning》](#)
- <http://petewarden.com/2015/04/20/why-gemm-is-at-the-heart-of-deep-learning/>

介绍:一般矩阵乘法(GEMM)对深度学习的重要性.

- [《Distributed \(Deep\) Machine Learning Common》](#)
- <https://github.com/dmlc>

介绍:A Community of awesome Distributed Machine Learning C++ projects.

- [《Reinforcement Learning: An Introduction》](#)
- <http://webdocs.cs.ualberta.ca/~sutton/book/the-book.html>

介绍:免费电子书<强化学习介绍>,第一版(1998),第二版(2015草稿),相关课程资料,Reinforcement Learning.

- [《Free ebook: Microsoft Azure Essentials: Azure Machine Learning》](#)
- http://blogs.msdn.com/b/microsoft_press/archive/2015/04/15/free-ebook-microsoft-azure-essentials-azure-machine-learning.aspx

介绍:免费书:Azure ML使用精要.

- [《A Deep Learning Tutorial: From Perceptrons to Deep Networks》](#)
- <http://www.toptal.com/machine-learning/an-introduction-to-deep-learning-from-perceptrons-to-deep-networks>

介绍:A Deep Learning Tutorial: From Perceptrons to Deep Networks.

- [《Machine Learning is Fun! - The world's easiest introduction to Machine Learning》](#)
- <https://medium.com/@ageitgey/machine-learning-is-fun-80ea3ec3c471>

介绍:有趣的机器学习：最简明入门指南,中文版.

- [《A Brief Overview of Deep Learning》](#)
- yue.blogspot.com/2015/01/a-brief-overview-of-deep-learning.html

介绍:深度学习简明介绍,中文版.

- [《Wormhole》](#)
- <https://github.com/dmlc/wormhole>

介绍:Portable, scalable and reliable distributed machine learning.

- [《convnet-benchmarks》](#)
- <https://github.com/soumith/convnet-benchmarks>

介绍:CNN开源实现横向评测,参评框架包括Caffe、Torch-7、CuDNN、cudacnn2、fbfft、Nervana Systems等,NervanaSys表现突出.

- [《This catalogue lists resources developed by faculty and students of the Language Technologies Institute.》](#)
- http://islp21.is.cs.cmu.edu:3000/lti_catalogue

介绍:卡耐基梅隆大学计算机学院语言技术系的资源大全,包括大量的NLP开源软件工具包,基础数据集,论文集,数据挖掘教程,机器学习资源.

- [《Sentiment Analysis on Twitter》](#)
- <https://github.com/mayank93/Twitter-Sentiment-Analysis>

介绍:Twitter情感分析工具SentiTweet,视频+讲义.

- [《Machine Learning Repository @ Wash U》](#)
- <http://machinelearning.wustl.edu/mlpapers/venues>

介绍:华盛顿大学的Machine Learning Paper Repository.

- [《Machine learning cheat sheet》](#)
- <https://github.com/soulmachine/machine-learning-cheat-sheet>

介绍:机器学习速查表.

- [《Spark summit east 2015 agenda》](#)
- <http://spark-summit.org/east>

介绍:最新的Spark summit会议资料.

- 《Spark summit east 2015 agenda》
- <http://spark-summit.org/east>

介绍:最新的Spark summit会议资料.

- 《Learning Spark》
- <http://pan.baidu.com/s/1eQkybJG>

介绍:Ebook Learning Spark.

- 《Advanced Analytics with Spark, Early Release Edition》
- <http://pan.baidu.com/s/1jGot9qe>

介绍:Ebook Advanced Analytics with Spark, Early Release Edition.

- 《国内机器学习算法及应用领域人物篇:唐杰》
- <http://keg.cs.tsinghua.edu.cn/jietang/>

介绍:清华大学副教授,是图挖掘方面的专家。他主持设计和实现的Arnetminer是国内领先的图挖掘系统,该系统也是多个会议的支持商.

- 《国内机器学习算法及应用领域人物篇:杨强》
- <http://www.cse.ust.hk/~qyang/>

介绍:迁移学习的国际领军人物.

- 《国内机器学习算法及应用领域人物篇:周志华》
- <http://cs.nju.edu.cn/zhouzh/>

介绍:在半监督学习, multi-label学习和集成学习方面在国际上有一定的影响力.

- 《国内机器学习算法及应用领域人物篇:王海峰》
- http://ir.hit.edu.cn/~wanghaifeng/whf_pub.htm

介绍:信息检索,自然语言处理,机器翻译方面的专家.

- 《国内机器学习算法及应用领域人物篇:吴军》
- <http://www.cs.jhu.edu/~junwu/>

介绍:吴军博士是当前Google中日韩文搜索算法的主要设计者。在Google其间,他领导了许多研发项目,包括许多与中文相关的产品和自然语言处理的项目,他的[新个人主页](#).

- 《Cat Paper Collection》
- http://www.eecs.berkeley.edu/~junyanz/cat/cat_papers.html

介绍:喵星人相关论文集.

- 《How to Evaluate Machine Learning Models, Part 1: Orientation》
- <http://blog.dato.com/how-to-evaluate-machine-learning-models-part-1-orientation>

介绍:如何评价机器学习模型系列文章,How to Evaluate Machine Learning Models, Part 2a: Classification Metrics,How to Evaluate Machine Learning Models, Part 2b: Ranking and Regression Metrics.

- 《Building a new trends experience》
- <https://blog.twitter.com/2015/building-a-new-trends-experience>

介绍:Twitter新trends的基本实现框架.

- 《Storm Blueprints: Patterns for Distributed Real-time Computation》
- <https://www.packtpub.com/big-data-and-business-intelligence/storm-blueprints-patterns-distributed-real-time-computation>

介绍:Storm手册, 国内有[中文翻译版本](#),谢谢作者.

- 《SmileMiner》
- <https://github.com/haifengl/smile>

介绍:Java机器学习算法库SmileMiner.

- 《机器翻译学术论文写作方法和技巧》
- http://nlp.csai.tsinghua.edu.cn/~ly/talks/cwmt14_tut.pdf

介绍:机器翻译学术论文写作方法和技巧, Simon Peyton Jones的[How to write a good research paper](#)同类视频[How to Write a Great Research Paper](#),how to paper talk.

- 《神经网络训练中的Tricks之高效BP (反向传播算法)》
- <http://blog.csdn.net/zouxy09/article/details/45288129>

介绍:神经网络训练中的Tricks之高效BP,博主的其他博客也挺精彩的.

- 《我和NLP的故事》
- <http://www.52cs.org/?p=499>

介绍:作者是NLP方向的硕士,短短几年内研究成果颇丰,推荐新入门的朋友阅读.

- 《The h Index for Computer Science》
- <http://www.cs.ucla.edu/~palsberg/h-number.html>

介绍:UCLA的Jens Palsberg根据Google Scholar建立了一个计算机领域的H-index牛人列表,我们熟悉的各个领域的大牛绝大多数都在榜上,包括1位诺贝尔奖得主,35位图灵奖得主,近百位美国工程院/科学院院士,300多位ACM Fellow,在这里推荐的原因是大家可以在google通过搜索牛人的名字来获取更多的资源,这份资料很宝贵.

- 《Structured Learning for Taxonomy Induction with Belief Propagation》
- http://ttic.uchicago.edu/~mbansal/papers/acl14_structuredTaxonomy.pdf

介绍:用大型语料库学习概念的层次关系,如鸟是鹦鹉的上级,鹦鹉是虎皮鹦鹉的上级.创新性在于模型构造,用因子图刻画概念之间依存关系,因引入兄弟关系,图有环,所以用有环扩散 (loopy propagation) 迭代计算边际概率 (marginal probability) .

- 《Bayesian analysis》
- <http://www.stata.com/stata14/bayesian-analysis/>

介绍: 这是一款贝叶斯分析的商业软件,官方写的[贝叶斯分析的手册](#)有250多页,虽然R语言 已经有类似的[项目](#),但毕竟可以增加一个可选项.

- [《deep net highlights from 2014》](#)
- <http://www.quora.com/Boris-Babenko/Posts/deep-net-highlights-from-2014>

介绍:deep net highlights from 2014.

- [《Fast R-CNN》](#)
- <http://arxiv.org/pdf/1504.08083v1.pdf>

介绍:This paper proposes Fast R-CNN, a clean and fast framework for object detection.

- [《Fingerprinting Images for Near-Duplicate Detection》](#)
- <https://realpython.com/blog/python/fingerprinting-images-for-near-duplicate-detection/>

介绍:图像指纹的重复识别,作者[源码](#),国内[翻译版本](#).

- [《The Computer Vision Industry 》](#)
- <http://www.cs.ubc.ca/~lowe/vision.html>

介绍:提供计算机视觉、机器视觉应用的公司信息汇总.应用领域包括:自动辅助驾驶和交通管理、眼球和头部跟踪、影视运动分析、影视业、手势识别、通用视觉系统、各种工业自动化和检验、医药和生物、移动设备目标识别和AR、人群跟踪、摄像、安全监控、生物监控、三维建模、web和云应用.

- [《Seaborn: statistical data visualization》](#)
- <https://github.com/mwaskom/seaborn>

介绍:Python版可视化数据统计开源库.

- [《IPython lecture notes for OCW MIT 18.06》](#)
- <http://www.juanklopper.com/opencourseware/mathematics-2/ipython-lecture-notes/>

介绍:麻省理工Gilbert Strang线性代数课程笔记,Gilbert Strang《Linear Algebra》课程主页[视频+讲义](#).

- [《Canova: A Vectorization Lib for ML》](#)
- <http://deeplearning4j.org/canova.html>

介绍:面向机器学习/深度学习的数据向量化工具Canova,[github](#), 支持CSV文件、MNIST数据、TF-IDF/Bag of Words/word2vec文本向量化.

- [《DZone Refcardz: Distributed Machine Learning with Apache Mahout》](#)
- <http://java.dzone.com/articles/dzone-refcardz-distributed>

介绍:快速入门:基于Apache Mahout的分布式机器学习.

- [《Learning scikit-learn: Machine Learning in Python》](#)
- <http://nbviewer.ipython.org/github/gmnonce/scikit-learn-book/tree/master/>

介绍:基于scikit-learn讲解了一些机器学习技术,如SVM, NB, PCA, DT, 以及特征工程、特征选择和模型选择问题.

- [《Lightning fast Machine Learning with Spark》](#)
- <https://speakerdeck.com/nivdul/lightning-fast-machine-learning-with-spark>

介绍:基于Spark的高效机器学习,[视频地址](#).

- [《How we're using machine learning to fight shell selling》](#)
- <http://blog.wepay.com/how-were-using-machine-learning-to-fight-shell-selling/>

介绍:WePay用机器学习对抗信用卡"shell selling"诈骗.

- [《Data Scientists Thoughts that Inspired Me》](#)
- <http://www.datasciencecentral.com/profiles/blog/show?id=6448529:BlogPost:273276>

介绍:16位数据科学家语录精选.

- [《Deep learning applications and challenges in big data analytics》](#)
- <http://www.journalofbigdata.com/content/2/1/1>

介绍:深度学习在大数据分析领域的应用和挑战.

- [《Free book:Machine Learning,Mathematics》](#)
- <http://resrc.io/list/10/list-of-free-programming-books/#machine-learning>

介绍:免费的机器学习与数学书籍,除此之外还有其他的[免费编程书籍](#),编程语言,设计,操作系统等.

- [《Object detection via a multi-region & semantic segmentation-aware CNN model》](#)
- <http://arxiv.org/pdf/1505.01749.pdf>

介绍:一篇关于CNN模型对象识别Paper.

- [《A Statistical View of Deep Learning \(V\): Generalisation and Regularisation》](#)
- <http://blog.shakirm.com/2015/05/a-statistical-view-of-deep-learning-v-generalisation-and-regularisation/>

介绍:深度学习的统计分析V:泛化和正则化.

- [《Highway Networks》](#)
- <http://arxiv.org/abs/1505.00387>

介绍:用SGD能高效完成训练的大规模(多层)深度网络HN.

- [《What I Read For Deep-Learning》](#)
- <http://www.erogol.com/what-i-read-for-deep-learning/>

介绍:深度学习解读文章.

- [《An Introduction to Recommendation Engines》](#)
- <http://dataconomy.com/an-introduction-to-recommendation-engines>

介绍:Coursera上的推荐系统导论 (Introduction to Recommender Systems) 公开课.

- [《Stanford Machine Learning》](#)
- <http://www.holehouse.org/mlclass/index.html>

介绍:Andrew Ng经典机器学习课程笔记.

- [《ICLR 2015》](#)
- http://yaroslavvb.blogspot.de/2015/05/iclr-2015_12.html

介绍:ICLR 2015见闻录,博客的其他机器学习文章也不错.

- [《Stanford Machine Learning》](#)
- <http://www.cripac.ia.ac.cn/People/sw/Xu2015PSR.pdf>

介绍:推荐系统"个性化语义排序"模型.

- [《The More Excited We Are, The Shorter We Tweet》](#)
- <http://senseable.mit.edu/tweetbursts/>

介绍:激情时分更惜字——MIT的最新Twitter研究结果.

- [《苏州大学人类语言技术研究论文主页》](#)
- <http://hlt.suda.edu.cn/paper.html>

介绍:苏州大学人类语言技术研究相关论文.

- [《Neural Turing Machines implementation》](#)
- <http://arxiv.org/abs/1505.00387>

介绍:实现神经图灵机(NTM),项目地址,此外推荐相关神经图灵机算法.

- [《Computer Vision - CSE 559A, Spring 2015》](#)
- http://www.cse.wustl.edu/~furukawa/cse559a/2015_spring/

介绍:华盛顿大学的机器视觉(2015),参考资料Computer Vision: Algorithms and Applications.

- [《Mining of Massive Datasets》](#)
- <http://www.mmds.org/>

介绍:"Mining of Massive Datasets"发布第二版,Jure Leskovec, Anand Rajaraman, Jeff Ullman 新版增加Jure Leskovec作为合作作者,新增社交网络图数据挖掘、降维和大规模机器学习三章,电子版依旧免费.

- [《Learning Deep Learning》](#)
- <http://rt.dgyblog.com/ref/ref-learning-deep-learning.html>

介绍:一个深度学习资源页,资料很丰富.

- [《Learning Deep Learning》](#)
- <http://vdisk.weibo.com/s/ayG13we2ler9b>

介绍:免费电子书"Learning Deep Learning".

- [《Tutorial: Machine Learning for Astronomy with Scikit-learn》](#)

- http://www.astroml.org/sklearn_tutorial/index.html

介绍:Machine Learning for Astronomy with scikit-learn.

- 《An Introduction to Random Forests for Beginners》
- <http://info.salford-systems.com/an-introduction-to-random-forests-for-beginners>

介绍:免费电子书"随机森林入门指南".

- 《Top 10 data mining algorithms in plain English》
- <http://rayli.net/blog/data/top-10-data-mining-algorithms-in-plain-english/>

介绍:白话数据挖掘十大算法.

- 《An Inside Look at the Components of a Recommendation Engine》
- <https://www.mapr.com/blog/inside-look-at-components-of-recommendation-engine#.VVmZ5vmqqko>

介绍:基于Mahout和Elasticsearch的推荐系统,国内译版.

- 《Advances in Extreme Learning Machines》
- <https://aaltodoc.aalto.fi/bitstream/handle/123456789/15585/isbn9789526061498.pdf>

介绍:博士学位论文:ELM研究进展.

- 《10-minute tour of pandas》
- <https://vimeo.com/59324550>

介绍:Pandas十分钟速览,ipn.

- 《Data doesn't grow in tables: harvesting journalistic insight from documents》
- <http://pudo.org/blog/2015/05/15/document-mining.html>

介绍:面向数据新闻的文本挖掘.

- 《Time-lapse Mining from Internet Photos》
- <http://grail.cs.washington.edu/projects/timelapse/>

介绍:用网络图片合成延时视频(SIGGRAPH 2015).

- 《The Curse of Dimensionality in classification》
- <http://www.visiondummys.com/2014/04/curse-dimensionality-affect-classification/>

介绍:分类系统的维数灾难.

- 《Deep Learning vs Big Data: Who owns what?》
- <http://www.computervisionblog.com/2015/05/deep-learning-vs-big-data-who-owns-what.html>

介绍:深度学习vs.大数据——从数据到知识：版权的思考,[翻译版](<http://www.csdn.net/article/2015-05-19/2824707>)

- [《A Primer on Predictive Models》](#)
- <http://www.nature.com/ctg/journal/v5/n1/abs/ctg201319a.html>

介绍:预测模型入门.

- [《Demistifying LSTM Neural Networks》](#)
- <http://blog.terminal.com/demistifying-long-short-term-memory-lstm-recurrent-neural-networks/>

介绍:深入浅出LSTM.

- [《ICLR 2015》](#)
- <https://www.youtube.com/playlist?list=PLhiWXaTdsWB8PnrVZquVyqIRFWXM4ijYz>

介绍:2015年ICLR会议视频与讲义.

- [《On Visualizing Data Well》](#)
- <http://dataremixed.com/2015/05/on-visualizing-data-well/>

介绍:Ben Jones的数据可视化建议.

- [《Decoding Dimensionality Reduction, PCA and SVD》](#)
- <http://bigdata-madesimple.com/decoding-dimensionality-reduction-pca-and-svd/>

介绍:解读数据降维/PCA/SVD.

- [《Supervised learning superstitions cheat sheet》](#)
- http://ryancompton.net/assets/ml_cheat_sheet/supervised_learning.html

介绍:IPN:监督学习方法示例/对比参考表,覆盖logistic回归, 决策树, SVM, KNN, Naive Bayes等方法.

- [《DopeLearning: A Computational Approach to Rap Lyrics Generation》](#)
- <http://arxiv.org/abs/1505.04771>

介绍:基于RankSVM和DNN自动(重组)生成Rap歌词.

- [《An Introduction to Random Indexing》](#)
- https://www.sics.se/~mange/papers/RI_intro.pdf

介绍:随机索引RI词空间模型专题.

- [《VDiscover》](#)
- <http://www.vdiscover.org/>

介绍:基于机器学习的漏洞检测工具VDiscover.

- [《Minerva》](#)
- <https://github.com/dmlc/minerva>

介绍:深度学习系统minerva. 拥有python编程接口. 多GPU几乎达到线性加速. 在4块GPU上能在4天内将GoogLeNet训练到68.7%的top-1以及89.0%的top-5准确率. 和同为dmlc项目的cxxnet相比, 采用动态数据流引擎, 提供更多灵活性. 未来将和cxxnet一起整合为mxnet项目, 互取优势.

- 《CVPR 2015 paper》
- <http://www.cv-foundation.org/openaccess/CVPR2015.py>

介绍:2015年国际计算机视觉与模式识别会议paper.

- 《What are the advantages of different classification algorithms?》
- <http://www.quora.com/What-are-the-advantages-of-different-classification-algorithms/answer/Xavier-Amatriain>

介绍:Netflix工程总监眼中的分类算法 : 深度学习优先级最低,中文版.

- 《Results for Microsoft COCO Image Captioning Challenge》
- <https://www.codalab.org/competitions/3221#results>

介绍:Codalab图像标注竞赛排行+各家论文,Reddit上flukeskywalker整理了各家技术[相关论文](#).

- 《Caffe con Troll: Shallow Ideas to Speed Up Deep Learning》
- <http://arxiv.org/abs/1504.04343>

介绍:基于Caffe的加速深度学习系统CcT.

- 《Low precision storage for deep learning》
- <http://arxiv.org/abs/1412.7024>

介绍:深度学习(模型)低精度(训练与)存储.

- 《Model-Based Machine Learning (Early Access)》
- <http://www.mbmlbook.com/index.html>

介绍:新书预览:模型机器学习.

- 《Regret Analysis of Stochastic and Nonstochastic Multi-armed Bandit Problems》
- <http://www.princeton.edu/~sbubeck/SurveyBCB12.pdf>

介绍:免费电子书多臂老虎机,此外推荐[Introduction to Bandits: Algorithms and Theory](#).

- 《Kaggle R Tutorial on Machine Learning》
- <https://www.datacamp.com/courses/kaggle-tutorial-on-machine-learning-the-sinking-of-the-titanic>

介绍:基于Kaggle's Titanic Competition的交互式R机器学习教程,介绍 [《Interactive R Tutorial: Machine Learning for the Titanic Competition》](#) .

- 《Deep Learning (深度学习) 学习笔记整理系列》
- <http://suanfazu.com/t/deep-learning/9401>

介绍:Deep Learning (深度学习) 学习笔记整理系列.

- 《Introduction to Neural Machine Translation with GPUs 》
- <http://devblogs.nvidia.com/parallelforall/introduction-neural-machine-translation-with-gpus/>

介绍:神经(感知)机器翻译介绍.

- [《Andrew Ng: Deep Learning, Self-Taught Learning and Unsupervised Feature Learning》](#)
- <https://www.youtube.com/watch?v=n1ViNeWhC24&hd=1>

介绍:Andrew Ng关于深度学习/自学习/无监督特征学习的报告,国内云.

- [《Recurrent Neural Network Training with Dark Knowledge Transfer》](#)
- <http://arxiv.org/abs/1505.04630>

介绍:论文:通过潜在知识迁移训练RNN.

- [《Show Me The Money》](#)
- <https://github.com/chrischris292/ShowMeTheMoney>

介绍:面向金融数据的情感分析工具.

- [《pyLDAvis》](#)
- <https://github.com/bmabey/pyLDAvis>

介绍:(Python)主题模型交互可视化库pyLDAvis.

- [《Logistic Regression and Gradient Descent》](#)
- <http://nbviewer.ipynb.org/github/folkman/learningwithdata/blob/master/Logistic%20Gradient%20Descent.ipynb>

介绍:Logistic回归与优化实例教程.

- [《贾扬清微信讲座记录》](#)
- <http://pan.baidu.com/s/1dDGVl53>

介绍:贾扬清(谷歌大脑科学家、caffe缔造者)微信讲座记录.

- [《sketch》](#)
- <https://github.com/udibr/sketch>

介绍:Theano/Blocks实现RNN手写字符串生成sketch.

- [《Web Scale Document Clustering: Clustering 733 Million Web Pages》](#)
- <http://chris.de-vries.id.au/2015/05/web-scale-document-clustering.html>

介绍:基于TopSig的海量(7亿+)网页聚类.

- [《NAACL 2015 Proceedings on ACL Anthology》](#)
- <http://aclweb.org/anthology/N/N15/>

介绍:NAACL 2015 论文papers.

- [《Stock Forecasting With Machine Learning - Seven Possible Errors》](#)
- <http://www.anlytcs.com/2015/05/stock-forecasting-with-machine-learning.html>

介绍:机器学习预测股市的七个问题.

- [《Are there any good resources for learning about neural networks?》](#)
- http://www.reddit.com/r/MachineLearning/comments/378but/are_there_any_good_resources_for_learning_about/

介绍:神经网络学习资料推荐.

- [《A Critical Review of Recurrent Neural Networks for Sequence Learning》](#)
- <http://arxiv.org/abs/1506.00019v1>

介绍:面向序列学习的RNN综述.

- [《Handling and Processing Strings in R》](#)
- http://gastonsanchez.com/Handling_and_Processing_Strings_in_R.pdf

介绍:R文本处理手册.

- [《Must-watch videos about Python》](#)
- <https://github.com/s16h/py-must-watch>

介绍:"必看"的Python视频集锦.

- [《The Google Stack》](#)
- <http://malteschwarzkopf.de/research/assets/google-stack.pdf>

介绍:Google(基础结构)栈.

- [《Randomized Algorithms for Matrices and Data》](#)
- <http://cs.stanford.edu/people/mmahoney/f13-stat260-cs294/>

介绍:矩阵和数据的随机算法(UC Berkeley 2013).

- [《Intermediate R》](#)
- <https://www.datacamp.com/courses/intermediate-r>

介绍:DataCamp中级R语言教程.

- [《Topology Without Tears》](#)
- <http://www.topologywithouttears.net/>

介绍:免费电子书:轻松掌握拓扑学,中文版.

- [《Information Theory, Pattern Recognition, and Neural Networks》](#)
- http://www.inference.phy.cam.ac.uk/itprnn_lectures/

介绍:Book,video.

- [《Scikit-learn》](#)
- www.github.com/scikit-learn/scikit-learn

介绍:Scikit-learn 是基于Scipy为机器学习建造的的一个Python模块，他的特色就是多样化的分类，回归和聚类的算法包括支持向量机，逻辑回归，朴素贝叶斯分类器，随机森林，Gradient Boosting，聚类算法和DBSCAN。而且也设计出了Python numerical和scientific libraries Numpy and Scipy

- 《Pylearn2》
- www.github.com/lisa-lab/pylearn2

介绍:Pylearn是一个让机器学习研究简单化的基于Theano的库程序。

- 《NuPIC》
- www.github.com/numenta/nupic

介绍:NuPIC是一个以HTM学习算法为工具的机器智能平台。HTM是皮层的精确计算方法。HTM的核心是基于时间的持续学习算法和储存和撤销的时空模式。NuPIC适合于各种各样的问题,尤其是检测异常和预测的流数据来源。

- 《Nilearn》
- www.github.com/nilearn/nilearn

介绍:Nilearn 是一个能够快速统计学习神经影像数据的Python模块。它利用Python语言中的scikit-learn 工具箱和一些进行预测建模，分类，解码，连通性分析的应用程序来进行多元的统计。

- 《PyBrain》
- www.github.com/pybrain/pybrain

介绍:Pybrain是基于Python语言强化学习，人工智能，神经网络库的简称。 它的目标是提供灵活、容易使用并且强大的机器学习算法和进行各种各样的预定义的环境中测试来比较你的算法。

- 《Pattern》
- www.github.com/clips/pattern

介绍:Pattern 是Python语言下的一个网络挖掘模块。它为数据挖掘，自然语言处理，网络分析和机器学习提供工具。它支持向量空间模型、聚类、支持向量机和感知机并且用KNN分类法进行分类。

- 《Fuel》
- www.github.com/mila-udem/fuel

介绍:Fuel为你的机器学习模型提供数据。他有一个共享如MNIST, CIFAR-10 (图片数据集), Google's One Billion Words (文字)这类数据集的接口。你使用他通过很多种的方式来替代自己的数据。

- 《Bob》
- www.github.com/idiap/bob

介绍:Bob是一个免费的信号处理和机器学习的工具。它的工具箱是用Python和C++语言共同编写的，它的设计目的是变得更加高效并且减少开发时间，它是由处理图像工具,音频和视频处理、机器学习和模式识别的大量软件包构成的。

- 《Skdata》
- www.github.com/jaberg/skdata

介绍:Skdata是机器学习和统计的数据集的库程序。这个模块对于玩具问题，流行的计算机视觉和自然语言的数据集提供标准的Python语言的使用。

- 《MILK》
- www.github.com/luispedro/milk

介绍:MILK是Python语言下的机器学习工具包。它主要是在很多可得到的分类比如SVMS,K-NN,随机森林,决策树中使用监督分类法。它还执行特征选择。这些分类器在许多方面相结合,可以形成不同的例如无监督学习、密切关系金传播和由MILK支持的K-means聚类等分类系统。

- 《IEPY》
- www.github.com/machinalis/iepy

介绍:IEPY是一个专注于关系抽取的开源性信息抽取工具。它主要针对的是需要对大型数据集进行信息提取的用户和想要尝试新的算法的科学家。

- 《Quepy》
- www.github.com/machinalis/quepy

介绍:Quepy是通过改变自然语言问题从而在数据库查询语言中进行查询的一个Python框架。他可以简单的被定义为在自然语言和数据库查询中不同类型的问题。所以,你不用编码就可以建立你自己的一个用自然语言进入你的数据库的系统。现在Quepy提供对于Sparql和MQL查询语言的支持。并且计划将它延伸到其他的数据库查询语言。

- 《Hebel》
- www.github.com/hannes-brt/hebel

介绍:Hebel是在Python语言中对于神经网络的深度学习的一个库程序,它使用的是通过PyCUDA来进行GPU和CUDA的加速。它是最重要的神经网络模型的工具而且能提供一些不同的活动函数的激活功能,例如动力,涅斯捷罗夫动力,信号丢失和停止法。

- 《mlxtend》
- www.github.com/rasbt/mlxtend

介绍:它是一个由有用的工具和日常数据科学任务的扩展组成的一个库程序。

- 《nolearn》
- www.github.com/dnouri/nolearn

介绍:这个程序包容纳了大量能对你完成机器学习任务有帮助的实用程序模块。其中大量的模块和scikit-learn一起工作,其它的通常更有用。

- 《Ramp》
- www.github.com/kvh/ramp

介绍:Ramp是一个在Python语言下制定机器学习中加快原型设计的解决方案的库程序。他是一个轻型的pandas-based机器学习中可插入的框架,它现存的Python语言下的机器学习和统计工具(比如scikit-learn, rpy2等) Ramp提供了一个简单的声明性语法探索功能从而能够快速有效地实施算法和转换。

- 《Feature Forge》
- www.github.com/machinalis/featureforge

介绍:这一系列工具通过与scikit-learn兼容的API,来创建和测试机器学习功能。这个库程序提供了一组工具,它会让你在许多机器学习程序使用中很受用。当你使用scikit-learn这个工具时,你会感觉到受到了很大的帮助。(虽然这只能在你有不同的算法时起作用。)

- 《REP》
- www.github.com/yandex/rep

介绍:REP是以一种和谐、可再生的方式为指挥数据移动驱动所提供的一种环境。它有一个统一的分类器包装来提供各种各样的操作,例如TMVA, Sklearn, XGBoost, uBoost等等。并且它可以在一个群体以平行的方式训练分类器。同时它也提供了一个交互式的情节。

- 《Python 学习机器样品》
- www.github.com/awsllabs/machine-learning-samples

介绍:用亚马逊的机器学习建造的简单软件收集。

- [《Python-ELM》](#)
- www.github.com/dclambert/Python-ELM

介绍:这是一个在Python语言下基于scikit-learn的极端学习机器的实现。

- [《Dimension Reduction》](#)
- <http://forum.memect.com/thread/dimension-reduction/>

介绍:电子书降维方法,此外还推荐Dimensionality Reduction A Short Tutorial、Matlab Toolbox for Dimensionality Reduction、Unsupervised Kernel Dimension Reduction

- [《Datasets Used For Benchmarking Deep Learning Algorithms》](#)
- <http://deeplearning.net/datasets/>

介绍:deeplearning.net整理的深度学习数据集列表.

- [《Golang Natural Language Processing》](#)
- <https://github.com/advancedlogic/go-freeling>

介绍:Go语言编写的自然语言处理工具.

- [《Rehabilitation of Count-based Models for Word Vector Representations》](#)
- <http://arxiv.org/abs/1412.4930>

介绍:词频模型对词向量的反击,参考Improving Distributional Similarity with Lessons Learned from Word Embeddings 。

- [《Three Aspects of Predictive Modeling》](#)
- http://static1.squarespace.com/static/51156277e4b0b8b2ffe11c00/t/55344152e4b0ff30bb9ec163/1429487954122/ASA_Kuhn.pdf

介绍:预测模型的三个方面.

- [《CS224d: Deep Learning for Natural Language Processing》](#)
- <http://cs224d.stanford.edu/>

介绍:斯坦福大学深度学习与自然语言处理课程,部分课程笔记[词向量](#)、[引言](#)

- [《Google Computer Vision research at CVPR 2015》](#)
- <http://googleresearch.blogspot.jp/2015/06/google-computer-vision-research-at-cvpr.html>

介绍:CVPR2015上Google的CV研究列表.

- [《Using Deep Learning to Find Basketball Highlights》](#)
- <http://public.hudl.com/bits/archives/2015/06/05/highlights/>

介绍:利用(Metamind)深度学习自动发现篮球赛精彩片段.

- [《Learning Deep Features for Discriminative Localization》](#)

- <http://arxiv.org/abs/1512.04150>

介绍:对本土化特征学习的分析