

机器学习(MachineLearning)&深度学习(DeepLearning)资料(链接版 2Chapters503+358 共 861 份, 单击小标题可跳转其网站)

Github.com 官网 ty4z2008 廖君 2015-04-04 14:45:23 发布至今仍在更新 2016.10.9-12 秦陇纪 QinDragon 整理 您的评价: 9.0

注:本文转自 Github 官网: <https://github.com/ty4z2008/Qix/blob/master/dl.md> 之一 503 份,

<https://github.com/ty4z2008/Qix/blob/master/dl2.md> 之二 358 份资料, 机器学习资料篇目一共 500 条, 篇目二正在更新, 《机器学习(Machine Learning)&深度学习(Deep Learning)资料(Chapter 1&2)》链接版 2Chapters503+358 共 861 份, 单击小标题可跳转其网站。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。

1. 《Brief History of 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.

2. 《Deep Learning in Neural Networks: An Overview》

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.

3. 《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.

4. 《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.

5. 《Machine Learning is Fun!》

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/>

6. 《R 语言参考卡片》

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

7. 《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比较了 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>

8. 《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>

9. 《The LION Way: Machine Learning plus Intelligent Optimization》

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/>

10. 《深度学习与统计学习理论》

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.

11. 《计算机科学中的数学》

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

12. 《信息时代的计算机科学理论(Foundations of Data Science)》

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

13. 《Data Science with R》

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。

14. 《Twenty Questions for Donald Knuth》

介绍:这并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, AI

请关注“数据精简 DataSimp”公众号! 14 时 7 分 3 秒 2016 年 10 月 12 日星期三

秦陇纪 2010 数据简化 DataSimp 个人新浪博客 blog.sina.com.cn/inxcn, 微博 weibo.com/INXCN, 网盘 <http://pan.baidu.com/s/1i4D2GO9>, 分享人工智能大数据文章。欢迎大家参加线下聚会, 线上点赞互粉、跟顶转帖、下载学习、转发 QQ 空间、微信朋友圈、公众号。

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。

15. 《Automatic Construction and Natural-Language Description of Nonparametric Regression Models》

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了麻省理工 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...

16. 《ICLR 2014 论文集》

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

17. 《Introduction to Information Retrieval》

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>

18. 《Machine learning in 10 pictures》

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

19. 《雅虎研究院的数据集汇总》

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。

20. 《An Introduction to Statistical Learning with Applications in R》

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>

21. 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。

22. My deep learning reading list

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。

23. Cross-Language Information Retrieval

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

24. 探索推荐引擎内部的秘密, 第 1 部分: 推荐引擎初探

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

25. 《Advice for students of machine learning》

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."

26. 分布式并行处理的数据

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

27. 《“机器学习”是什么?》

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

28. 《2014 年国际机器学习大会 ICML 2014 论文》

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办。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

29. 《Machine Learning for Industry: A Case Study》

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

请关注“数据精简 DataSimp”公众号! 14 时 7 分 3 秒 2016 年 10 月 12 日星期三

秦陇纪 2010 数据简化 DataSimp 个人新浪博客 blog.sina.com.cn/inxcn, 微博 weibo.com/INXCN, 网盘 <http://pan.baidu.com/s/1i4D2GO9>, 分享人工智能大数据文章。欢迎大家参加线下聚会, 线上点赞互粉、跟顶转帖、下载学习、转发 QQ 空间、微信朋友圈、公众号。

Machines for Pattern Recognition, Some Notes on Applied Mathematics for Machine Learning

30. 100 Best GitHub: Deep Learning

介绍:100 Best GitHub: Deep Learning

31. 《UFLDL-斯坦福大学 Andrew Ng 教授“Deep Learning”教程》

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

32. *《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,毛塞顿开。

33. 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

34. 《Machine Learning Summer School》

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)

35. 《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

36. 斯坦福《自然语言处理》课程视频

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。

37. 《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。

38. 《Recommending music on Spotify with deep learning》

介绍:利用卷积神经网络做音乐推荐。

39. 《Neural Networks and Deep Learning》

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。

40. 《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。

41. 《Machine Learning Theory: An Introductory Primer》

介绍:机器学习最基本的入门文章,适合零基础者

42. 《机器学习常见算法分类汇总》

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。

43. 《机器学习经典论文/survey 合集》

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。

44. 《机器学习视频库》

介绍:视频由加州理工学院(Caltech)出品。需要英语底子。

45. 《机器学习经典书籍》

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。

46. 《16 Free eBooks On Machine Learning》

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。

47. 《A Large set of Machine Learning Resources for Beginners to 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

48. 《机器学习最佳入门学习资料汇总》

介绍:入门的书真的很多,而且我已经帮你找齐了。

49. 《Sibyl》

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。

50. 《Deep Learning》

介绍:Yoshua Bengio, Ian Goodfellow, Aaron Courville 著

51. 《Neural Network & Text Mining》

请关注“数据精简 DataSimp”公众号! 14 时 7 分 3 秒 2016 年 10 月 12 日星期三

秦陇纪 2010 数据简化 DataSimp 个人新浪博客 blog.sina.com.cn/inxcn, 微博 weibo.com/INXCN, 网盘 <http://pan.baidu.com/s/1i4D2GO9>, 分享人工智能大数据文章。欢迎大家参加线下聚会,线上点赞互粉、跟顶转帖、下载学习、转发 QQ 空间、微信朋友圈、公众号。

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

52. 《前景目标检测 1 (总结)》

介绍:计算机视觉入门之前景目标检测 1 (总结)

53. 《行人检测》

介绍:计算机视觉入门之行人检测

54. 《Deep Learning – important resources for learning and understanding》

介绍:Important resources for learning and understanding . Is awesome

55. 《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

56. 《Neural Networks and Deep Learning》

介绍:在线 Neural Networks and Deep Learning 电子书

57. 《Python 网页爬虫 & 文本处理 & 科学计算 & 机器学习 & 数据挖掘兵器谱》

介绍:python 的 17 个关于机器学习的工具

58. 《神奇的伽玛函数(上)》

介绍:下集在这里《神奇的伽玛函数(下)》

59. 《分布式机器学习的故事》

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

60. 《机器学习提升之道 (Level-Up Your Machine Learning)》

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。

61. 《Machine Learning Surveys》

介绍:机器学习各个方向综述的网站

62. 《Deep Learning Reading list》

介绍:深度学习阅资源列表

63. 《Deep Learning: Methods and Applications》

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

64. 《Machine Learning Summer School 2014》

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)

65. 《Sibyl: 来自 Google 的大规模机器学习系统》

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

66. 《Building a deeper understanding of images》

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。

67. 《Bayesian network 与 python 概率编程实战入门》

介绍:贝叶斯学习。如果不是很清楚看看概率编程语言与贝叶斯方法实践

68. 《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。"

69. 《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)》

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。

70. 《文本与数据挖掘视频汇总》

介绍: Videolectures 上最受欢迎的 25 个文本与数据挖掘视频汇总

71. 《怎么选择深度学习的 GPU》

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>

72. 《对话机器学习大神 Michael Jordan: 深度模型》

介绍:对话机器学习大神 Michael Jordan

73. 《Deep Learning 和 Knowledge Graph 引爆大数据革命》

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

74. 《Deep Learning 教程翻译》

请关注“数据精简 DataSimp”公众号! 14 时 7 分 3 秒 2016 年 10 月 12 日星期三

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

75. 《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!

76. 《UFLDL Tutorial》

介绍:这是斯坦福大学做的一免费课程(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

77. 《Toronto Deep Learning Demos》

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

78. 《Deep learning from the bottom up》

介绍:机器学习模型,阅读这个内容需要有一定的基础。

79. 《R 工具包的分类汇总》

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

80. 《机器学习常见算法分类汇总》

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。

81. 《Deep Learning (深度学习) 学习笔记整理系列》

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)

Deep Learning (深度学习) 学习笔记整理系列之(八)

82. 《Tutorials Session A - Deep Learning for Computer Vision》

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

83. 《FudanNLP》

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。

84. 《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

85. 《机器学习周刊》

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

86. 《线性代数》

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

87. 《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。

88. 《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。

89. 《Entanglement-Based Quantum Machine Learning》

介绍:应对大数据时代,量子机器学习的第一个实验 paper 下载

90. 《How a Math Genius Hacked OkCupid to Find True Love》

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

请关注“数据精简 DataSimp”公众号! 14 时 7 分 3 秒 2016 年 10 月 12 日星期三

秦陇纪 2010 数据简化 DataSimp 个人新浪博客 blog.sina.com.cn/inxcn, 微博 weibo.com/INXCN, 网盘 <http://pan.baidu.com/s/1i4D2GO9>, 分享人工智能大数据文章。欢迎大家参加线下聚会,线上点赞互粉、跟顶转帖、下载学习、转发 QQ 空间、微信朋友圈、公众号。

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！

91. 《Underactuated Robotics》

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！

92. 《mllib 实践经验(1)》

介绍:mllib 实践经验分享

93. 《Google Turns To Deep Learning Classification To Fight Web Spam》

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

94. 《NLP 常用信息资源》

介绍:NLP 常用信息资源* 《NLP 常用信息资源》

95. 《机器学习速查表》

介绍:机器学习速查表

96. 《Best Papers vs. Top Cited Papers in Computer Science》

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

97. 《InfiniTAM: 基于深度图像的体数据集成框架》

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。

98. 《Hacker's guide to Neural Networks》

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

99. 《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

100. 《Deep Learning Sentiment Analysis for Movie Reviews using Neo4j》

介绍：使用 Neo4j 做电影评论的情感分析。

101. 《DeepLearning.University – An Annotated Deep Learning Bibliography》

介绍：不仅是资料，而且还对有些资料做了注释。

102. 《A primer on deeping learning》

介绍：深度学习入门的初级读本

103. 《Machine learning is teaching us the secret to teaching 》

介绍：机器学习教会了我们什么？

104. 《scikit-learn: 用于机器学习的 Python 模块》

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

105. 《对话机器学习大神 Michael Jordan: 解析领域中各类模型》

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，乔丹教授对此一一做了解释和展望。

106. 《A*搜索算法的可视化短教程》

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

107. 《基于云的自然语言处理开源项目 FudanNLP》

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

108. 《吴立德《概率主题模型&数据科学基础》》

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

109. 《机器学习入门资源不完全汇总》》

介绍：好东西的干货真的很多

110. 《收集从 2014 年开始深度学习文献》

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

111. 《EMNLP 上两篇关于股票趋势的应用论文 》

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。

112. 《Bengio 组（蒙特利尔大学 LISA 组）深度学习教程 》

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。

113. 《学习算法的 Neural Turing Machine 》

请关注“数据精简 DataSimp”公众号！14 时 7 分 3 秒 2016 年 10 月 12 日星期三

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

114. 《Learning to Rank for Information Retrieval and Natural Language Processing》

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

115. 《Rumor has it: Identifying Misinformation in Microblogs》

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

116. 《R 机器学习实践》

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。

117. 《大数据分析: 机器学习算法实现的演化》

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

118. 《图像处理, 分析与机器视觉》

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》

119. 《LinkedIn 最新的推荐系统文章 Browsemaps》

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!

120. 《初学者如何查阅自然语言处理(NLP)领域学术资料》

介绍: 初学者如何查阅自然语言处理(NLP)领域学术资料

121. 《树莓派的人脸识别教程》

介绍: 用树莓派和相机模块进行人脸识别

122. 《利用深度学习与大数据构建对话系统》

介绍: 如何利用深度学习与大数据构建对话系统

123. 《经典论文 Leo Breiman: Statistical Modeling: The Two Cultures》

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。

124. 《Reproducing Kernel Hilbert Space》

介绍: RKHS 是机器学习中重要的概念, 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。

125. 《Hacker's guide to Neural Networks》

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。

126. 《【语料库】语料库资源汇总》

介绍: 【语料库】语料库资源汇总

127. 《机器学习算法之旅》

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。

128. 《Reproducible Research in Computational Science》

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

129. 《NYU 2014 年的深度学习课程资料》

介绍: NYU 2014 年的深度学习课程资料, 有视频

130. 《计算机视觉数据集不完全汇总》

介绍: 计算机视觉数据集不完全汇总

131. 《Machine Learning Open Source Software》

介绍: 机器学习开源软件

132. 《LIBSVM》

介绍: A Library for Support Vector Machines

133. 《Support Vector Machines》

介绍: 数据挖掘十大经典算法之一

134. 《100 Best GitHub: Deep Learning》

请关注“数据精简 DataSimp”公众号! 14 时 7 分 3 秒 2016 年 10 月 12 日星期三

介绍: github 上面 100 个非常棒的项目

135. 《加州大学欧文分校(UCI)机器学习数据集仓库》

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

136. 《Andrej Karpathy 个人主页》

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.

137. 《Andrej Karpathy 的深度强化学习演示》

介绍: Andrej Karpathy 的深度强化学习演示, 论文在这里

138. 《CIKM 数据挖掘竞赛夺冠算法-陈运文》

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。

139. 《Geoffrey E. 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。

140. 《自然语言处理的深度学习理论与实际》

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

141. 《用大数据和机器学习做股票价格预测》

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.

142. 《关于机器学习的若干理论问题》

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。

143. 《深度学习在自然语言处理的应用》

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

144. 《Undergraduate machine learning at UBC》

介绍: 机器学习课程

145. 《人脸识别必读的 N 篇文章》

介绍: 人脸识别必读文章推荐

146. 《推荐系统经典论文文献及业界应用》

介绍: 推荐系统经典论文文献

147. 《人脸识别必读的 N 篇文章》

介绍: 人脸识别必读文章推荐

148. 《第十二届中国"机器学习及其应用"研讨会 PPT》

介绍: 第十二届中国"机器学习及其应用"研讨会 PPT

149. 《统计机器学习》

介绍: 统计学习是关于计算机基于数据构建的概率统计模型并运用模型对数据进行预测和分析的一门科学, 统计学习也成为统计机器学习。课程来自上海交通大学

150. 《机器学习导论》

介绍: 机器学习的目标是对计算机编程, 以便使用样本数据或以往的经验来解决给定的问题。

151. 《CIKM 2014 主题报告的幻灯片》

介绍: CIKM 2014 Jeff Dean、Qi Lu、Gerhard Weikum 的主题报告的幻灯片, Alex Smola、Limsoon Wong、Tong Zhang、Chih-Jen Lin 的 Industry Track 报告的幻灯片

152. 《人工智能和机器学习领域有趣的开源项目》

介绍: 部分中文列表

153. 《机器学习经典算法详解及 Python 实现--基于 SMO 的 SVM 分类器》

介绍: 此外作者还有一篇元算法、AdaBoost python 实现文章

154. 《Numerical Optimization: Understanding L-BFGS》

介绍: 加州伯克利大学博士 Aria Haghighi 写了一篇超赞的数值优化博文, 从牛顿法讲到拟牛顿法, 再讲到 BFGS 以及 L-BFGS, 图文并茂, 还有伪代码。强烈推荐。

155. 《简明深度学习方法概述 (一)》

介绍: 还有续集简明深度学习方法概述 (二)

156. 《R language for programmers》

介绍: R 语言程序员私人定制版

157. 《谷歌地图解密: 大数据与机器学习的结合》

介绍: 谷歌地图解密

158. 《空间数据挖掘常用方法》

请关注“数据精简 DataSimp”公众号! 14 时 7 分 3 秒 2016 年 10 月 12 日星期三

秦陇纪 2010 数据简化 DataSimp 个人新浪博客 blog.sina.com.cn/inxcn, 微博 weibo.com/INXCN, 网盘 <http://pan.baidu.com/s/1i4D2GO9>, 分享人工智能大数据文章。欢迎大家参加线下聚会, 线上点赞互粉、跟顶转帖、下载学习、转发 QQ 空间、微信朋友圈、公众号。

介绍:空间数据挖掘常用方法

159. 《Use Google's Word2Vec for movie reviews》

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

160. 《PyNLPIR》

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)

161. 《深度卷积神经网络下围棋》

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

162. 《NIPS 审稿实验》

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。

163. 《2014 年最佳的大数据,数据科学文章》

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

164. 《机器学习经典算法详解及 Python 实现--线性回归(Linear Regression)算法》

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

165. 《2014 中国大数据技术大会 33 位核心专家演讲 PDF》

介绍:2014 中国大数据技术大会 33 位核心专家演讲 PDF 下载

166. 《使用 RNN 和 Paragraph Vector 做情感分析》

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。

167. 《NLPIR/ICTCLAS2015 分词系统大会上的技术演讲》

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

168. 《Machine Learning is Fun!》

介绍:Convex Neural Networks 解决维数灾难

169. 《CNN 的反向求导及练习》

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

170. 《正则表达式优化成 Trie 树》

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

171. 《Deep learning Reading List》

介绍:深度学习阅读清单

172. 《Caffe》

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)

173. 《GoogLeNet 深度学习模型的 Caffe 复现》

介绍:2014 ImageNet 冠军 GoogLeNet 深度学习模型的 Caffe 复现模型,GoogleNet 论文.

174. 《LambdaNet, Haskell 实现的开源人工神经网络库》

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。

175. 《百度余凯&张潼机器学习视频》

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。

176. 《杨强在 TEDxNanjing 谈智能的起源》

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

请关注“数据精简 DataSimp”公众号! 14 时 7 分 3 秒 2016 年 10 月 12 日星期三

秦陇纪 2010 数据简化 DataSimp 个人新浪博客 blog.sina.com.cn/inxcn, 微博 weibo.com/INXCN, 网盘 http://pan.baidu.com/s/1i4D2GO9, 分享人工智能大数据文章。欢迎大家参加线下聚会,线上点赞互粉、跟顶转帖、下载学习、转发 QQ 空间、微信朋友圈、公众号。

177. 《深度 RNN/LSTM 用于结构化学习 0)序列标注 Connectionist Temporal Classification ICML06》

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

178. 《Deep Learning 实战之 word2vec》

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

179. 《Machine learning open source 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

180. 《机器学习入门者学习指南》

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

181. 《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

182. 《2014 年的《机器学习日报》大合集》

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.

183. 《Image classification with deep learning 常用模型》

介绍:这是一篇关于图像分类在深度学习中的文章

184. 《自动语音识别:深度学习方法》

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

185. 《NLP 中的中文分词技术》

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

186. 《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

187. 《书籍推荐:Advanced Structured Prediction》

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

188. 《An Introduction to Matrix Concentration Inequalities》

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

189. 《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

190. 《A Brief Overview of Deep Learning》

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

191. 《A Deep Dive into Recurrent Neural Nets》

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

192. 《机器学习:学习资源》

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

193. 《Statistical foundations of machine learning》

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。

194. 《A Deep Learning Tutorial: 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

195. 《Research priorities for robust and beneficial artificial intelligence》

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

请关注“数据精简 DataSimp”公众号! 14 时 7 分 3 秒 2016 年 10 月 12 日星期三

秦陇纪 2010 数据简化 DataSimp 个人新浪博客 blog.sina.com.cn/inxcn, 微博 weibo.com/INXCN, 网盘 <http://pan.baidu.com/s/1i4D2GO9>, 分享人工智能大数据文章。欢迎大家参加线下聚会,线上点赞互粉、跟顶转帖、下载学习、转发 QQ 空间、微信朋友圈、公众号。

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。

196. 《metacademy》

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

197. 《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。

198. 《浅析人脸检测之 Haar 分类器方法》

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。

199. 《如何成为一位数据科学家》

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

200. 《Deep learning from the bottom up》

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

201. 《Hands-On Data Science with R Text Mining》

介绍:主要是讲述了利用 R 语言进行数据挖掘

202. 《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

203. 《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

204. 《Learning Deep Architectures for AI》

介绍:一本学习人工智能的书籍, 作者是 Yoshua Bengio, 相关国内报道

205. 《Geoffrey E. Hinton 个人主页》

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

206. 《PROBABILITY THEORY: THE LOGIC OF SCIENCE》

介绍:概率论: 数理逻辑书籍

207. 《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

208. 《ICLR 2015 会议的 arXiv 稿件合集》

介绍:在这里你可以看到最近深度学习有什么新动向。

209. 《Introduction to Information Retrieval》

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

210. 《Information Geometry and its Applications to Machine Learning》

介绍:信息几何学及其在机器学习中的应用

211. 《Legal Analytics – Introduction to the Course》

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。

212. 《文本上的算法》

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

213. 《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

214. 《Deep Learning on Hadoop 2.0》

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

215. 《Practical recommendations for gradient-based training of deep architectures》

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

216. 《Machine Learning With Statistical And Causal Methods》

介绍: 用统计和因果方法做机器学习 (视频报告)

请关注“数据精简 DataSimp”公众号! 14 时 7 分 3 秒 2016 年 10 月 12 日星期三

217. 《Machine Learning Course 180'》

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。

218. 《回归(regression)、梯度下降(gradient descent)》

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

219. 《美团推荐算法实践》

介绍: 美团推荐算法实践, 从框架, 应用, 策略, 查询等分析

220. 《Deep Learning for Answer Sentence Selection》

介绍: 深度学习用于问答系统答案句的选取

221. 《Learning Semantic Representations Using Convolutional Neural Networks for Web Search 》

介绍: CNN 用于 WEB 搜索, 深度学习在文本计算中的应用

222. 《Awesome Public Datasets》

介绍: Awesome 系列中的公开数据集

223. 《Search Engine & Community》

介绍: 一个学术搜索引擎

224. 《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

225. 《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

226. 《Topic modeling 的经典论文》

介绍: Topic modeling 的经典论文, 标注了关键点

227. 《Move Evaluation in Go Using Deep Convolutional Neural Networks》

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

228. 《机器学习周刊第二期》

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

229. 《Learning more like a human: 18 free eBooks on Machine Learning》

介绍: 18 free eBooks on Machine Learning

230. 《Recommend :Hang Li Home》

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

231. 《DEEPLARNING.UNIVERSITY – AN ANNOTATED 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

232. 《MLMU.cz - Radim Řehůřek - Word2vec & friends (7.1.2015)》

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

233. 《Introducing streaming k-means in Spark 1.2》

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。

234. 《LDA 入门与 Java 实现》

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。

235. 《AMiner - Open Science Platform》

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。

236. 《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

237. 《机器学习公开课汇总》

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

请关注“数据精简 DataSimp”公众号! 14 时 7 分 3 秒 2016 年 10 月 12 日星期三

秦陇纪 2010 数据简化 DataSimp 个人新浪博客 blog.sina.com.cn/inxcn, 微博 weibo.com/INXCN, 网盘 <http://pan.baidu.com/s/1i4D2GO9>, 分享人工智能大数据文章。欢迎大家参加线下聚会, 线上点赞互粉、跟顶转帖、下载学习、转发 QQ 空间、微信朋友圈、公众号。

238. 《A First Course in Linear Algebra》

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

239. 《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。

240. 《Inverting a Steady-State》

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

241. 《机器学习入门书单》

介绍:机器学习入门书籍, 具体介绍

242. 《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

243. 《Rise of the Machines》

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

244. 《实例详解机器学习如何解决问题》

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实际问题。这篇文章是美团的实际环境中的实战篇

245. 《Gaussian Processes for Machine Learning》

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

246. 《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

247. 《Blocks》

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。

248. 《Introduction to Machine Learning》

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。

249. 《Collaborative Feature Learning from Social Media》

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。

250. 《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。

251. 《Empower Your Team to Deal with Data-Quality Issues》

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)

252. 《中文分词入门之资源》

介绍:中文分词入门之资源。

253. 《Deep Learning Summit, San Francisco, 2015》

介绍:15 年旧金山深度学习峰会视频集萃,国内云盘

254. 《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

255. 《A Fast and Accurate Dependency Parser using Neural Networks》

介绍:来自 Stanford, 用神经网络实现快速准确的依存关系解析器

256. 《Which GPU(s) to Get for Deep Learning: My Experience and Advice for Using GPUs in Deep Learning》

介绍:做深度学习如何选择 GPU 的建议

257. 《Sparse Linear Models》

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

请关注“数据精简 DataSimp”公众号! 14 时 7 分 3 秒 2016 年 10 月 12 日星期三

超过样本数)的线性模型,13 年同主题报告、讲义.

258. 《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

259. 《Adam Szeidl》

介绍: social networks course

260. 《Building and deploying large-scale machine learning pipelines》

介绍: 大规模机器学习流程的构建与部署.

261. 《人脸识别开发包》

介绍: 人脸识别二次开发包, 免费, 可商用, 有演示、范例、说明书.

262. 《Understanding Natural Language with 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.

263. 《The NLP Engine: A Universal Turing Machine for NLP》

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.

264. 《The Probabilistic Relevance Framework: BM25 and Beyond》

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.

265. 《Introduction to ARMA Time Series Models – simplified》

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.

266. 《Encoding Source Language with Convolutional Neural Network for Machine Translation》

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

267. 《Spices form the basis of food pairing in Indian cuisine》

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

268. 《HMM 相关文章索引》

介绍: HMM 相关文章

269. 《Zipf's and Heap's law》

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

270. 《I am Jürgen 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.

271. 《学术种子网站: AcademicTorrents》

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

272. 《机器学习交互速查表》

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

273. 《A Full Hardware Guide to Deep Learning》

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

274. 《行人检测(Pedestrian Detection)资源》

介绍: Pedestrian Detection paper & data

275. 《A specialized face-processing network consistent with the representational geometry of monkey face patches》

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.

276. 《Neural Net in C++ Tutorial》

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.

277. 《How to Choose a Neural Network》

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

278. 《Deep Learning (Python, C/C++, Java, Scala, Go)》

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

请关注“数据精简 DataSimp”公众号! 14 时 7 分 3 秒 2016 年 10 月 12 日星期三

秦陇纪 2010 数据简化 DataSimp 个人新浪博客 blog.sina.com.cn/inxcn, 微博 weibo.com/INXCN, 网盘 <http://pan.baidu.com/s/1i4D2GO9>, 分享人工智能大数据文章. 欢迎大家参加线下聚会, 线上点赞互粉、跟顶转帖、下载学习、转发 QQ 空间、微信朋友圈、公众号。

279. 《Deep Learning Tutorials》
介绍:深度学习教程
280. 《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》
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.
281. 《FaceNet: A Unified Embedding for Face Recognition and Clustering》
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.
282. 《MLlib 中的 Random Forests 和 Boosting》
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.
283. 《Sum-Product Networks(SPN) 》
介绍:华盛顿大学 Pedro Domingos 团队的 DNN, 提供论文和实现代码.
284. 《Neural Network Dependency Parser》
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.
285. 《神经网络语言模型》
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], SENNA 等重要变形, 总结的特别好.
286. 《Classifying Spam Emails using Text and Readability Features》
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.
287. 《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
288. 《IPOL Journal · Image Processing On Line》
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.
289. 《Machine learning classification over encrypted data》
介绍:出自 MIT, 研究加密数据高效分类问题.
290. 《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.
291. 《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.
292. 《Hands-on with 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.
293. 《The Natural Language Processing Dictionary》
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.
294. 《PageRank Approach to Ranking National Football Teams》
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.
295. 《R Tutorial》
介绍:R 语言教程, 此外还推荐一个 R 语言教程 An Introduction to R.
296. 《Fast unfolding of communities in large networks》
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.
297. 《NUML》
介绍: 一个面向 .net 的开源机器学习库, github 地址
298. 《synaptic.Js》
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
299. 《Machine learning for package users with R (1): Decision Tree》
介绍: 决策树
300. 《Deep Learning, The Curse of Dimensionality, and Autoencoders》
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

请关注“数据精简 DataSimp”公众号! 14 时 7 分 3 秒 2016 年 10 月 12 日星期三

秦陇纪 2010 数据简化 DataSimp 个人新浪博客 blog.sina.com.cn/inxcn, 微博 weibo.com/INXCN, 网盘 <http://pan.baidu.com/s/1i4D2GO9>, 分享人工智能大数据文章. 欢迎大家参加线下聚会, 线上点赞互粉、跟帖转帖、下载学习、转发 QQ 空间、微信朋友圈、公众号。

301. 《Advanced Optimization and Randomized Methods》
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)
302. 《CS231n: Convolutional Neural Networks for Visual Recognition》
介绍: "面向视觉识别的 CNN"课程报告集锦. 近百篇, 内容涉及图像识别应用的各个方面
303. 《Topic modeling with LDA: MLlib meets GraphX》
介绍: 用 Spark 的 MLlib+GraphX 做大规模 LDA 主题抽取.
304. 《Deep Learning for Multi-label Classification》
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
305. 《Google DeepMind publications》
介绍: DeepMind 论文集锦
306. 《kaldi》
介绍: 一个开源语音识别工具包, 它目前托管在 sourceforge 上面
307. 《Data Journalism Handbook》
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
308. 《Data Mining Problems in Retail》
介绍: 零售领域的数据挖掘文章.
309. 《Understanding Convolution in Deep Learning》
介绍: 深度学习卷积概念详解, 深入浅出.
310. 《pandas: powerful Python data analysis toolkit》
介绍: 非常强大的 Python 的数据分析工具包.
311. 《Text Analytics 2015》
介绍: 2015 文本分析(商业)应用综述.
312. 《Deep Learning libraries and first 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.
313. 《DEEP learning》
介绍: MIT 的 Yoshua Bengio 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.
314. 《simplebayes》
介绍: Python 下开源可持久化朴素贝叶斯分类库.
315. 《Paracel》
介绍: Paracel is a distributed computational framework designed for machine learning problems, graph algorithms and scientific computing in C++.
316. 《HanLP: Han Language processing》
介绍: 开源汉语言处理包.
317. 《Simple Neural Network implementation in Ruby》
介绍: 使用 Ruby 实现简单的神经网络例子.
318. 《Hacker's guide to Neural Networks》
介绍: 神经网络黑客入门.
319. 《The Open-Source Data Science Masters》
介绍: 好多数据科学家名人推荐, 还有资料.
320. 《Text Understanding from Scratch》
介绍: 实现项目已经开源在 github 上面 Crepe
321. 《Improving Distributional Similarity with Lessons Learned from Word Embeddings》
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.
322. 《CS224d: Deep Learning for Natural Language Processing》
介绍: Stanford 深度学习与自然语言处理课程, Richard Socher 主讲.
323. 《Math Essentials in Machine Learning》
介绍: 机器学习中的重要数学概念.
324. 《Improved Semantic Representations From Tree-Structured Long Short-Term Memory Networks》
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.
325. 《Statistical Machine Learning》
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.
326. 《AM207: Monte Carlo Methods, Stochastic Optimization》

请关注“数据精简 DataSimp”公众号! 14 时 7 分 3 秒 2016 年 10 月 12 日星期三

秦陇纪 2010 数据简化 DataSimp 个人新浪博客 blog.sina.com.cn/inxcn, 微博 weibo.com/INXCN, 网盘 <http://pan.baidu.com/s/1i4D2GO9>, 分享人工智能大数据文章. 欢迎大家参加线下聚会, 线上点赞互粉、跟帖转帖、下载学习、转发 QQ 空间、微信朋友圈、公众号。

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 Kaynig-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.

327. 《生物医学的 SPARK 大数据应用》

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.

328. 《ACL Anthology》

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.

329. 《Twitter Sentiment Detection via Ensemble Classification Using Averaged Confidence Scores》

介绍:论文+代码:基于集成方法的 Twitter 情感分类, 实现代码.

330. 《NIPS 2014 CIML workshop》

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.

331. 《CS231n: Convolutional Neural Networks for Visual Recognition》

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.

332. 《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.

333. 《Back-to-Basics Weekend Reading - Machine Learning》

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

334. 《A Probabilistic Theory of Deep Learning》

介绍:莱斯大学(Rice University)的深度学习概率理论.

335. 《Nonsensical beer reviews via 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.

336. 《Deep Learning for Natural Language Processing (without Magic)》

介绍:视频+讲义:深度学习用于自然语言处理教程(NAACL13).

337. 《Introduction to Data Analysis using Machine Learning》

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

338. 《Beyond Short Snippets: Deep Networks for Video Classification》

介绍:基于 CNN+LSTM 的视频分类, google 演示.

339. 《How does Quora use machine learning in 2015?》

介绍:Quora 怎么用机器学习.

340. 《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.

341. 《Parallel Machine Learning with scikit-learn and IPython》

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

342. 《Intro to machine learning with scikit-learn》

介绍:DataSchool 的机器学习基本概念教学.

343. 《DeepCLn》

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

344. 《An Inside Look at the Components of a Recommendation Engine》

介绍:基于 Mahout 和 Elasticsearch 的推荐系统.

345. 《Forecasting in Economics, Business, Finance and Beyond》

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

346. 《Time Series Econometrics - A Concise Course》

介绍:Francis X. Diebold 的《时序计量经济学》.

347. 《A comparison of open source tools for sentiment analysis》

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.

348. 《Pattern Recognition And Machine Learning》

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

349. 《Probabilistic Data Structures for Web Analytics and Data Mining》

介绍:用于 Web 分析和数据挖掘的概率数据结构.

350. 《Machine learning in navigation devices: detect maneuvers using accelerometer and gyroscope》

请关注“数据精简 DataSimp”公众号! 14 时 7 分 3 秒 2016 年 10 月 12 日星期三

介绍:机器学习在导航上面的应用.

351. 《Neural Networks Demystified》

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.

352. 《swirl + DataCamp》

介绍:{swirl}数据训练营:R&数据科学在线交互教程.

353. 《Learning to Read with Recurrent Neural Networks》

介绍:关于深度学习和 RNN 的讨论 Sequence to Sequence Learning with Neural Networks.

354. 《深度强化学习 (Deep Reinforcement Learning) 的资源》

介绍:Deep Reinforcement Learning.

355. 《Machine Learning with Scikit-Learn》

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

356. 《PDNN》

介绍:PDNN: A Python Toolkit for Deep Learning.

357. 《Introduction to Machine Learning》

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.

358. 《Big Data Processing》

介绍:大数据处理课.内容覆盖流处理、MapReduce、图算法等.

359. 《Spark MLlib: 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.

360. 《Picture: A Probabilistic Programming Language for Scene Perception》

介绍:以往上千行代码概率编程(语言)实现只需 50 行.

361. 《Beautiful plotting in R: A ggplot2 cheatsheet》

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》.

362. 《Using Structured Events to Predict Stock Price Movement: An Empirical Investigation》

介绍:用结构化模型来预测实时股票行情.

363. 《International Joint Conference on Artificial Intelligence Accepted paper》

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.

364. 《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.

365. 《Distributed (Deep) Machine Learning Common》

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

366. 《Reinforcement Learning: An Introduction》

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.

367. 《Free ebook: Microsoft Azure Essentials: Azure Machine Learning》

介绍:免费书:Azure ML 使用精要.

368. 《A Deep Learning Tutorial: From Perceptrons to Deep Networks》

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

369. 《Machine Learning is Fun! - The world's easiest introduction to Machine Learning》

介绍:有趣的机器学习: 最简明入门指南,中文版.

370. 《A Brief Overview of Deep Learning》

介绍:深度学习简明介绍,中文版.

371. 《Wormhole》

介绍:Portable, scalable and reliable distributed machine learning.

372. 《convnet-benchmarks》

介绍:CNN 开源实现横向评测,参评框架包括 Caffe 、 Torch-7、 CuDNN 、 cudaconvnet2 、 fbfft、 Nervana Systems 等, NervanaSys 表现突出.

373. 《This catalogue lists resources developed by faculty and students of the Language Technologies Institute.》

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.

374. 《Sentiment Analysis on Twitter》

介绍:Twitter 情感分析工具 SentiTweet,视频+讲义.

375. 《Machine Learning Repository @ Wash U》

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

376. 《Machine learning cheat sheet》

介绍:机器学习速查表.

377. 《Spark summit east 2015 agenda》

请关注“数据精简 DataSimp”公众号! 14 时 7 分 3 秒 2016 年 10 月 12 日星期三

介绍:最新的 Spark summit 会议资料.

378. 《Spark summit east 2015 agenda》

介绍:最新的 Spark summit 会议资料.

379. 《Learning Spark》

介绍:Ebook Learning Spark.

380. 《Advanced Analytics with Spark, Early Release Edition》

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

381. 《国内机器学习算法及应用领域人物篇:唐杰》

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.

382. 《国内机器学习算法及应用领域人物篇:杨强》

介绍:迁移学习的国际领军人物.

383. 《国内机器学习算法及应用领域人物篇:周志华》

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.

384. 《国内机器学习算法及应用领域人物篇:王海峰》

介绍:信息检索,自然语言处理,机器翻译方面的专家.

385. 《国内机器学习算法及应用领域人物篇:吴军》

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.

386. 《Cat Paper Collection》

介绍:喵星人相关论文集.

387. 《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.

388. 《Building a new trends experience》

介绍:Twitter 新 trends 的基本实现框架.

389. 《Storm Blueprints: Patterns for Distributed Real-time Computation》

介绍:Storm 手册,国内有中文翻译版本,谢谢作者.

390. 《SmileMiner》

介绍:Java 机器学习算法库 SmileMiner.

391. 《机器翻译学术论文写作方法和技巧》

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.

392. 《神经网络训练中的 Tricks 之高效 BP (反向传播算法)》

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.

393. 《我和 NLP 的故事》

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.

394. 《The h Index for Computer Science 》

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.

395. 《Structured Learning for Taxonomy Induction with Belief Propagation》

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) .

396. 《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.

397. 《deep net highlights from 2014》

介绍:deep net highlights from 2014.

398. 《Fast R-CNN》

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

399. 《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.

400. 《The Computer Vision Industry 》

请关注“数据精简 DataSimp”公众号! 14 时 7 分 3 秒 2016 年 10 月 12 日星期三

秦陇纪 2010 数据简化 DataSimp 个人新浪博客 blog.sina.com.cn/inxcn, 微博 weibo.com/INXCN, 网盘 <http://pan.baidu.com/s/1i4D2GO9>, 分享人工智能大数据文章。欢迎大家参加线下聚会,线上点赞互粉、跟顶转帖、下载学习、转发 QQ 空间、微信朋友圈、公众号。

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.

401. 《Seaborn: statistical data visualization》

介绍:Python 版可视化数据统计开源库.

402. 《IPython lecture notes for OCW MIT 18.06》

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.

403. 《Canova: A Vectorization Lib for ML》

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.

404. 《DZone Refcardz: Distributed Machine Learning with Apache Mahout》

介绍:快速入门: 基于 Apache Mahout 的分布式机器学习.

405. 《Learning scikit-learn: Machine Learning in Python》

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.

406. 《Lightning fast Machine Learning with Spark》

介绍:基于 Spark 的高效机器学习,视频地址.

407. 《How we're using machine learning to fight shell selling》

介绍:WePay 用机器学习对抗信用卡"shell selling"诈骗.

408. 《Data Scientists Thoughts that Inspired Me》

介绍:16 位数据科学家语录精选.

409. 《Deep learning applications and challenges in big data analytics》

介绍:深度学习在大数据分析领域的应用和挑战.

410. 《Free book:Machine Learning,Mathematics》

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.

411. 《Object detection via a multi-region & semantic segmentation-aware CNN model》

介绍:一篇关于 CNN 模型对象识别 Paper.

412. 《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.

413. 《Highway Networks》

介绍:用 SGD 能高效完成训练的大规模(多层)深度网络 HN.

414. 《What I Read For Deep-Learning》

介绍:深度学习解读文章.

415. 《An Introduction to Recommendation Engines》

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

416. 《Stanford Machine Learning》

介绍:Andrew Ng 经典机器学习课程笔记.

417. 《ICLR 2015》

介绍:ICLR 2015 见闻录,博客的其他机器学习文章也不错.

418. 《Stanford Machine Learning》

介绍:推荐系统"个性化语义排序"模型.

419. 《The More Excited We Are, The Shorter We Tweet》

介绍:激情时分更惜字——MIT 的最新 Twitter 研究结果.

420. 《苏州大学人类语言技术研究论文主页》

介绍:苏州大学人类语言技术研究相关论文.

421. 《Neural Turing Machines implementation》

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.

422. 《Computer Vision - CSE 559A, Spring 2015》

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

423. 《Mining of Massive Datasets》

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.

424. 《Learning Deep Learning》

介绍:一个深度学习资源页,资料很丰富.

425. 《Learning Deep Learning》

介绍:免费电子书"Learning Deep Learning".

请关注“数据精简 DataSimp”公众号! 14 时 7 分 3 秒 2016 年 10 月 12 日星期三

秦陇纪 2010 数据简化 DataSimp 个人新浪博客 blog.sina.com.cn/inxcn, 微博 weibo.com/INXCN, 网盘 <http://pan.baidu.com/s/1i4D2GO9>, 分享人工智能大数据文章. 欢迎大家参加线下聚会, 线上点赞互粉、跟顶转帖、下载学习、转发 QQ 空间、微信朋友圈、公众号。

426. 《Tutorial: Machine Learning for Astronomy with Scikit-learn》
介绍:Machine Learning for Astronomy with scikit-learn.
427. 《An Introduction to Random Forests for Beginners》
介绍:免费电子书"随机森林入门指南".
428. 《Top 10 data mining algorithms in plain English》
介绍:白话数据挖掘十大算法.
429. 《An Inside Look at the Components of a Recommendation Engine》
介绍:基于 Mahout 和 Elasticsearch 的推荐系统,国内译版.
430. 《Advances in Extreme Learning Machines》
介绍:博士学位论文:ELM 研究进展.
431. 《10-minute tour of pandas》
介绍:Pandas 十分钟速览,ipn.
432. 《Data doesn't grow in tables: harvesting journalistic insight from documents》
介绍:面向数据新闻的文本挖掘.
433. 《Time-lapse Mining from Internet Photos》
介绍:用网络图片合成延时视频(SIGGRAPH 2015).
434. 《The Curse of Dimensionality in classification》
介绍:分类系统的维数灾难.
435. 《Deep Learning vs Big Data: Who owns what?》
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>)
436. 《A Primer on Predictive Models》
介绍:预测模型入门.
437. 《Demistifying LSTM Neural Networks》
介绍:深入浅出 LSTM.
438. 《ICLR 2015》
介绍:2015 年 ICLR 会议视频与讲义.
439. 《On Visualizing Data Well》
介绍:Ben Jones 的数据可视化建议.
440. 《Decoding Dimensionality Reduction, PCA and SVD》
介绍:解读数据降维/PCA/SVD.
441. 《Supervised learning superstitions cheat sheet》
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.
442. 《DopeLearning: A Computational Approach to Rap Lyrics Generation》
介绍:基于 RankSVM 和 DNN 自动(重组)生成 Rap 歌词.
443. 《An Introduction to Random Indexing》
介绍:随机索引 RI 词空间模型专题.
444. 《VDiscover》
介绍:基于机器学习的漏洞检测工具 VDiscover.
445. 《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.
446. 《CVPR 2015 paper》
介绍:2015 年国际计算机视觉与模式识别会议 paper.
447. 《What are the advantages of different classification algorithms?》
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.
448. 《Results for Microsoft COCO Image Captioning Challenge》
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.
449. 《Caffe con Troll: Shallow Ideas to Speed Up Deep Learning》
介绍:基于 Caffe 的加速深度学习系统 CcT.
450. 《Low precision storage for deep learning》
介绍:深度学习(模型)低精度(训练与)存储.
451. 《Model-Based Machine Learning (Early Access)》
介绍:新书预览:模型机器学习.
452. 《Regret Analysis of Stochastic and Nonstochastic Multi-armed Bandit Problems》
介绍:免费电子书多臂老虎机,此外推荐 Introduction to Bandits: Algorithms and Theory.

请关注“数据精简 DataSimp”公众号! 14 时 7 分 3 秒 2016 年 10 月 12 日星期三

453. 《Kaggle R Tutorial on Machine Learning》

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》.

454. 《Deep Learning (深度学习) 学习笔记整理系列》

介绍:Deep Learning (深度学习) 学习笔记整理系列.

455. 《Introduction to Neural Machine Translation with GPUs》

介绍:神经(感知)机器翻译介绍.

456. 《Andrew Ng: Deep Learning, Self-Taught Learning and Unsupervised Feature Learning》

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.

457. 《Recurrent Neural Network Training with Dark Knowledge Transfer》

介绍:论文:通过潜在知识迁移训练 RNN.

458. 《Show Me The Money》

介绍:面向金融数据的情感分析工具.

459. 《pyLDAvis》

介绍:(Python)主题模型交互可视化库 pyLDAvis.

460. 《Logistic Regression and Gradient Descent》

介绍:Logistic 回归与优化实例教程.

461. 《贾扬清微信讲座记录》

介绍:贾扬清(谷歌大脑科学家、caffe 缔造者)微信讲座记录.

462. 《sketch》

介绍:Theano/Blocks 实现 RNN 手写字符串生成 sketch.

463. 《Web Scale Document Clustering: Clustering 733 Million Web Pages》

介绍:基于 TopSig 的海量(7 亿+)网页聚类.

464. 《NAACL 2015 Proceedings on ACL Anthology》

介绍:NAACL 2015 论文 papers.

465. 《Stock Forecasting With Machine Learning - Seven Possible Errors》

介绍:机器学习预测股市的七个问题.

466. 《Are there any good resources for learning about neural networks?》

介绍:神经网络学习资料推荐.

467. 《A Critical Review of Recurrent Neural Networks for Sequence Learning》

介绍:面向序列学习的 RNN 综述.

468. 《Handling and Processing Strings in R》

介绍:R 文本处理手册.

469. 《Must-watch videos about Python》

介绍:“必看”的 Python 视频集锦.

470. 《The Google Stack》

介绍:Google(基础结构)栈.

471. 《Randomized Algorithms for Matrices and Data》

介绍:矩阵和数据的随机算法(UC Berkeley 2013).

472. 《Intermediate R》

介绍:DataCamp 中级 R 语言教程.

473. 《Topology Without Tears》

介绍:免费电子书:轻松掌握拓扑学,中文版.

474. 《Information Theory, Pattern Recognition, and Neural Networks》

介绍:Book,video.

475. 《Scikit-learn》

介绍:Scikit-learn 是基于 Scipy 为机器学习建造的一个 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

476. 《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.

477. 《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.

478. 《Nilearn》

请关注“数据精简 DataSimp”公众号! 14 时 7 分 3 秒 2016 年 10 月 12 日星期三

秦陇纪 2010 数据简化 DataSimp 个人新浪博客 blog.sina.com.cn/inxcn, 微博 weibo.com/INXCN, 网盘 <http://pan.baidu.com/s/1i4D2GO9>, 分享人工智能大数据文章. 欢迎大家参加线下聚会, 线上点赞互粉、跟顶转帖、下载学习、转发 QQ 空间、微信朋友圈、公众号。

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。

479. 《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。

480. 《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。

481. 《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。你使用他来通过很多种的方式来替代自己的数据。

482. 《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。

483. 《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。

484. 《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。

485. 《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。

486. 《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。

487. 《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类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。

488. 《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。

489. 《nolearn》

介绍:这个程序包容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。

490. 《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。

491. 《Feature Forge》

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。)

492. 《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。

493. 《Python 学习机器样品》

介绍:用亚马逊的机器学习建造的简单软件收集。

494. 《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。

495. 《Dimension Reduction》

请关注“数据精简 DataSimp”公众号! 14 时 7 分 3 秒 2016 年 10 月 12 日星期三

秦陇纪 2010 数据简化 DataSimp 个人新浪博客 blog.sina.com.cn/inxcn, 微博 weibo.com/INXCN, 网盘 <http://pan.baidu.com/s/1i4D2GO9>, 分享人工智能大数据文章。欢迎大家参加线下聚会, 线上点赞互粉、跟顶转帖、下载学习、转发 QQ 空间、微信朋友圈、公众号。

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

496. 《Datasets Used For Benchmarking Deep Learning Algorithms》

介绍:deeplearning.net 整理的深度学习数据集列表.

497. 《Golang Natural Language Processing》

介绍:Go 语言编写的自然语言处理工具.

498. 《Rehabilitation of Count-based Models for Word Vector Representations》

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

499. 《Three Aspects of Predictive Modeling》

介绍:预测模型的三个方面.

500. 《CS224d: Deep Learning for Natural Language Processing》

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

501. 《Google Computer Vision research at CVPR 2015》

介绍:CVPR2015 上 Google 的 CV 研究列表.

502. 《Using Deep Learning to Find Basketball Highlights》

介绍:利用(Metamind)深度学习自动发现篮球赛精彩片段.

503. 《Learning Deep Features for Discriminative Localization》

介绍:对本土化特征学习的分析

504. 《Image Scaling using Deep Convolutional Neural Networks》

介绍:使用卷积神经网络的图像缩放.

505. 《Proceedings of The 32nd International Conference on Machine Learning》

介绍:ICML2015 论文集,优化 4 个+稀疏优化 1 个;强化学习 4 个,深度学习 3 个+深度学习计算 1 个;贝叶斯非参、高斯过程和学习理论 3 个;还有计算广告和社会选择.ICML2015 Sessions.

506. 《Image Scaling using Deep Convolutional Neural Networks》

介绍:使用卷积神经网络的图像缩放.

507. 《Microsoft researchers accelerate computer vision accuracy and improve 3D scanning models》

介绍:,第 28 届 IEEE 计算机视觉与模式识别(CVPR)大会在美国波士顿举行.微软研究员们在大会上展示了比以往更快更准的计算机视觉图像分类新模型,并介绍了如何使用 Kinect 等传感器实现在动态或低光环境的快速大规模 3D 扫描技术.

508. 《Machine Learning for Humans》

介绍:(文本)机器学习可视化分析工具.

509. 《A Plethora of Tools for Machine Learning》

介绍:机器学习工具包/库的综述/比较.

510. 《The art of visualizing visualizations: a best practice guide》

介绍:数据可视化最佳实践指南.

511. 《MIT Machine Learning for Big Data and Text Processing Class Notes - Day 1》

介绍:Day 1、Day 2、Day 3、Day 4、Day 5.

512. 《Getting “deep” about “deep learning”》

介绍:深度学习之“深”——DNN 的隐喻分析.

513. 《Mixture Density Networks》

介绍:混合密度网络.

514. 《Interview Questions for Data Scientist Positions》

介绍:数据科学家职位面试题.

515. 《Accurately Measuring Model Prediction Error》

介绍:准确评估模型预测误差.

516. 《Continually updated Data Science Python Notebooks》

介绍:Continually updated Data Science Python Notebooks.

517. 《How to share data with a statistician》

介绍:How to share data with a statistician.

518. 《The Eyescream Project NeuralNets dreaming natural images》

介绍:来自 Facebook 的图像自动生成.

519. 《How to share data with a statistician》

介绍:How to share data with a statistician.

520. 《A Neural Conversational Model》

介绍:(Google)神经(感知)会话模型.

521. 《The 50 Best Masters in Data Science》

请关注“数据精简 DataSimp”公众号! 14 时 7 分 3 秒 2016 年 10 月 12 日星期三

介绍:The 50 Best Masters in Data Science.

522. 《NLP 常用信息资源》

介绍:NLP 常用信息资源.

523. 《Conditional Random Fields as Recurrent Neural Networks》

介绍:语义图像分割的实况演示,通过深度学习技术和概率图模型的语义图像分割.

524. 《Fully Convolutional Networks for Semantic Segmentation》

介绍:Caffe 模型/代码:面向图像语义分割的全卷积网络,模型代码.

525. 《Growing Pains for Deep Learning》

介绍:深度学习——成长的烦恼.

526. 《Clustering Text Data Streams – A Tree based Approach with Ternary Function and Ternary Feature Vector 》

介绍:基于三元树方法的文本流聚类.

527. 《Foundations and Advances in Data Mining》

介绍:Free Ebook:数据挖掘基础及最新进展.

528. 《The Deep Learning Revolution: Rethinking Machine Learning Pipelines》

介绍:深度学习革命.

529. 《The Definitive Guide to Do Data Science for Good》

介绍:数据科学(实践)权威指南.

530. 《Microsoft Academic Graph》

介绍:37G 的微软学术图谱数据集.

531. 《Challenges and Opportunities Of Machine Learning In Production》

介绍:生产环境(产品级)机器学习的机遇与挑战.

532. 《Neural Nets for Newbies》

介绍:神经网络入门.

533. 《A Nearly-Linear Time Framework for Graph-Structured Sparsity》

介绍:来自麻省理工的结构化稀疏论文.

534. 《Optimal and Adaptive Algorithms for Online Boosting》

介绍:来自雅虎的机器学习小组关于在线 Boosting 的论文 .

535. 《Top 20 Python Machine Learning Open Source Projects》

介绍:20 个最热门的开源(Python)机器学习项目.

536. 《The Parallel C++ Statistical Library for Bayesian Inference: QUESO》

介绍:C++并行贝叶斯推理统计库 QUESO,github code.

537. 《《Deep learning》 Yann LeCun, Yoshua Bengio, Geoffrey Hinton (2015) 》

介绍:Nature:LeCun/Bengio/Hinton 的最新文章《深度学习》,Jürgen Schmidhuber 的最新评论文章《Critique of Paper by "Deep Learning Conspiracy" (Nature 521 p 436)》 .

538. 《Palladium》

介绍:基于 Scikit-Learn 的预测分析服务框架 Palladium.

539. 《Advances in Structured Prediction》

介绍:John Langford 和 Hal Daume III 在 ICML2015 上关于 Learning to Search 的教学讲座幻灯片.

540. 《100 open source Big Data architecture papers for data professionals》

介绍:读完这 100 篇论文 就能成大数据高手,国内翻译.

541. 《Social Media & Text Analytics》

介绍:NLP 课程《社交媒体与文本分析》精选阅读列表.

542. 《Machine Learning for Developers》

介绍:写给开发者的机器学习指南.

543. 《Hot news detection using Wikipedia》

介绍:基于维基百科的热点新闻发现.

544. 《Harvard Intelligent Probabilistic Systems Group》

介绍:(Harvard)HIPS 将发布可扩展/自动调参贝叶斯推理神经网络.

545. 《An Empirical Exploration of Recurrent Network Architectures》

介绍:面向上下文感知查询建议的层次递归编解码器.

546. 《Efficient Training of LDA on a GPU by Mean-for-Mode Estimation》

介绍:GPU 上基于 Mean-for-Mode 估计的高效 LDA 训练.

547. 《From the Lab to the Factory: Building a Production Machine Learning Infrastructure》

介绍:从实验室到工厂——构建机器学习生产架构.

548. 《6 Useful Databases to Dig for Data (and 100 more)》

介绍:适合做数据挖掘的 6 个经典数据集(及另外 100 个列表).

请关注“数据精简 DataSimp”公众号! 14 时 7 分 3 秒 2016 年 10 月 12 日星期三

549. 《Deep Networks for Computer Vision at Google – ILSVRC2014》
介绍:Google 面向机器视觉的深度学习.
550. 《How to choose a machine learning API to build predictive apps》
介绍:构建预测类应用时如何选择机器学习 API.
551. 《Exploring the shapes of stories using Python and sentiment APIs》
介绍:Python+情感分析 API 实现故事情节(曲线)分析.
552. 《Movie selection using R》
介绍:(R)基于 Twitter/情感分析的口碑电影推荐,此外推荐分类算法的实证比较分析.
553. 《A Tutorial on Graph-based Semi-Supervised Learning Algorithms for NLP》
介绍:CMU(ACL 2012)(500+页)面向 NLP 基于图的半监督学习算法.
554. 《Arbitrariness of peer review: A Bayesian analysis of the NIPS experiment》
介绍:从贝叶斯分析 NIPS, 看同行评审的意义.
555. 《Basics of Computational Reinforcement Learning》
介绍:(RLDM 2015)计算强化学习入门.
556. 《Deep Reinforcement Learning》
介绍:David Silver 的深度强化学习教程.
557. 《On Explainability of Deep Neural Networks》
介绍:深度神经网络的可解释性.
558. 《The Essential Spark Cheat Sheet》
介绍:Spark 快速入门.
559. 《Machine Learning for Sports and Real Time Predictions》
介绍:TalkingMachines:面向体育/政治和实时预测的机器学习.
560. 《CS224W: Social and Information Network Analysis Autumn 2014》
介绍:Stanford 社交网络与信息网络分析课程资料+课设+数据.
561. 《RL Course by David Silver》
介绍:David Silver(DeeMind)的强化学习课程,slide.
562. 《Faster deep learning with GPUs and Theano》
介绍:基于 Theano/GPU 的高效深度学习.
563. 《Introduction to R Programming》
介绍:来自微软的.
564. 《Golang:Web Server For Performing Sentiment Analysis》
介绍:(Go)情感分析 API 服务 Sentiment Server.
565. 《A Beginner's Guide to Restricted Boltzmann Machines》
介绍:受限波尔兹曼机初学者指南.
566. 《KDD2015 十年最佳论文》
介绍:Mining and Summarizing Customer Reviews ,Mining High-Speed Data Streams,Optimizing Search Engines using Clickthrough Data.
567. 《Nvidia Deep Learning Courses》
介绍:Nvidia 深度学习课程.
568. 《Deep Learning Summer School 2015》
介绍:2015 年深度学习暑期课程, 推荐讲师主页.
569. 《百度深度学习的图像识别进展》
介绍:这是一篇关于百度文章《基于深度学习的图像识别进展: 百度的若干实践》的摘要,建议两篇文章结合起来阅读.
570. 《Machine Learning Methods in Video Annotation》
介绍:视频标注中的机器学习技术.
571. 《Training Recurrent Neural Networks》
介绍:博士论文:(Ilya Sutskever)RNN 训练.
572. 《On Explainability of Deep Neural Networks》
介绍:深度神经网络的灰色区域: 可解释性问题, 中文版.
573. 《Machine Learning Libraries in GoLang by Category》
介绍:Golang 实现的机器学习库资源汇总.
574. 《A Statistical View of Deep Learning》
介绍:深度学习的统计分析.
575. 《Deep Learning For NLP - Tips And Techniques》
介绍:面向 NLP 的深度学习技术与技巧.

请关注“数据精简 DataSimp”公众号! 14 时 7 分 3 秒 2016 年 10 月 12 日星期三

576. 《CrowdFlower Competition Scripts: Approaching NLP》
介绍:Kaggle's CrowdFlower 竞赛 NLP 代码集锦.
577. 《CS224U: Natural Language Understanding》
介绍:斯坦福的自然语言理解课程.
578. 《Deep Learning and Shallow Learning》
介绍:Deep Learning 与 Shallow Learning 介绍
579. 《A First Encounter with Machine Learning》
介绍:这是一本机器学习的电子书,作者 Max Welling 先生在机器学习教学上面有着丰富的经验,这本书小但精致.
580. 《Click Models for Web Search》
介绍:由荷兰阿姆斯特丹大学 & 谷歌瑞士著.
581. 《Hinton CSC321 课程/Deep Learning/Notes on CNN/Python/Theano/CUDA/OpenCV/...》
介绍:介绍个乐于总结和翻译机器学习和计算机视觉类资料的博客,包含的内容: Hinton 的 CSC321 课程的总结; Deep Learning 综述; Notes on CNN 的总结; python 的原理总结; Theano 基础知识和练习总结; CUDA 原理和编程; OpenCV 一些总结.
582. 《Which Algorithm Family Can Answer My Question?》
介绍:针对具体问题(应用场景)如何选择机器学习算法(系列).
583. 《Free Data Science Books》
介绍:数据科学免费书分类集合
584. 《Tutorial 4: Deep Learning for Speech Generation and Synthesis》
介绍:深度学习在语音合成最新进展有哪些? 推荐 MSRA 的 Frank Soong 老师关于语音合成的深度学习方法的录像和幻灯片与以及谷歌的 LSTM-RNN 合成介绍,论文
585. 《The Art of Data Science》
介绍:新书(可免费下载):数据科学的艺术
586. 《Pattern Recognition and Machine Learning》
介绍:模式识别与机器学习书籍推荐,本书是微软剑桥研究院大神 Bishop 所写,算是最为广为认知的机器学习教材之一,内容覆盖全面,难度中上,适合研究生中文版 or 备份
587. 《an introduction to visualizing DATA》
介绍:数据可视化介绍(23 页袖珍小册子)
588. 《That's So Annoying!!!: A Lexical and Frame-Semantic Embedding Based Data Augmentation Approach to Automatic Categorization of Annoying Behaviors using #petpeeve Tweets * 》
介绍:这篇论文荣获 EMNLP2015 的最佳数据/资源奖优秀奖,标注的推特数据集
589. 《26 Things I Learned in the Deep Learning Summer School》
介绍:作者在深度学习的思考.
590. 《Data-Visualization Tools & Books》
介绍:数据可视化常用工具软件资源汇总
591. 《Machine Learning and Probabilistic Graphical Models Course》
介绍:Buffalo 大学教授 Sargur Srihari 的“机器学习和概率图模型”的视频课程
592. 《Understanding Machine Learning: From Theory to Algorithms》
介绍:耶路撒冷希伯来大学教授 Shai Shalev-Shwartz 和滑铁卢大学教授 Shai Ben-David 的新书 Understanding Machine Learning: From Theory to Algorithms,此书写的比较偏理论,适合对机器学习理论有兴趣的同学选读
593. 《Machine Learning Checklist》
介绍:机器学习学习清单
594. 《NLP 界有哪些神级人物? 》
介绍:知乎上面的一篇关于 NLP 界有哪些神级人物? 提问。首推 Michael Collins
595. 《机器学习温和指南》
介绍:机器学习与 NLP 专家、MonkeyLearn 联合创始人&CEO Raúl Garreta 面向初学者大体概括使用机器学习过程中的重要概念,应用程序和挑战,旨在让读者能够继续探寻机器学习知识。
596. 《Gradient Boosted Regression Trees》
介绍:(IPN)基于 Scikit-Learn 的 GBRT(Gradient Boost Regression Tree)教程, slide
597. 《Apache SINGA : Distributed Deep Learning System》
介绍: 无需做深度学习就能用的分布式深度学习软件.
598. 《E-commerce Recommendation with Personalized Promotion》
介绍: 在亚马逊数据和众包 Mechanical Turk 上,实现了来自彩票和拍卖的机制,以收集用户对产品的乐意购买价格(WTP, willingness-to-pay)训练集。 E-commerce Recommendation with Personalized Promotion [Zhao,RecSys15] 回归模型预测未知 WTP,提升卖家利润和消费者满意度
599. 《Scalable Machine Learning》

请关注“数据精简 DataSimp”公众号! 14 时 7 分 3 秒 2016 年 10 月 12 日星期三

介绍:来自伯克利分校的大规模机器学习.

600. 《机器学习资料大汇总》

介绍:来自 52ml 的机器学习资料大汇总.

601. 《Automatic Summarization》

介绍:这本书的作者 McKeown 是 2013 年世界首个数据科学院(位于哥伦比亚大学)主任,她亦是 ACL、AAAI 和 ACM Fellow .

602. 《Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing》

介绍:EMNLP-15 文本摘要若干.

603. 《Recommender Systems (Machine Learning Summer School 2014 @ CMU)》

介绍:来自 Netflix 的 Xavier Amatriain 在 Summer School 2014 @ CMU 上长达 4 小时的报告,共 248 页,是对推荐系统发展的一次全面综述,其中还包括 Netflix 在个性化推荐方面的一些经验介绍.

604. 《BigData Stream Mining》

介绍:(ECML PKDD 2015)大数据流挖掘教程,此外推荐 ECML PKDD 2015 Tutorial 列表.

605. 《Deep learning on Spark with Keras》

介绍:Spark 上的 Keras 深度学习框架 Elephas.

606. 《Prof. Surya Ganguli - The statistical physics of deep learning》

介绍:Surya Ganguli 深度学习统计物理学.

607. 《(系统/算法/机器学习/深度学习/图模型/优化/...)在线视频课程列表》

介绍:(系统/算法/机器学习/深度学习/图模型/优化/...)在线视频课程列表.

608. 《Introduction to Topic Modeling in Python》

介绍:(PyTexas 2015)Python 主题建模.

609. 《Large Scale Distributed Deep Learning on Hadoop Clusters》

介绍:Hadoop 集群上的大规模分布式机器学习.

610. 《Top Deep Learning Employers Based On LinkedIn Data》

介绍:基于 LinkedIn 数据得出的深度学习热门"东家"排行.

611. 《Neural Net in C++ Tutorial》

介绍:(c++)神经网络手把手实现教程.

612. 《Large-scale CelebFaces Attributes (CelebA) Dataset》

介绍:香港中文大学汤晓鸥教授实验室公布的大型人脸识别数据集: Large-scale CelebFaces Attributes (CelebA) Dataset 10K 名人, 202K 脸部图像, 每个图像 40 余标注属性.

613. 《Unsupervised Feature Learning in Computer Vision》

介绍:面向机器视觉的无监督特征学习,Ross Goroshin's webpage.

614. 《Scheduled Sampling for Sequence Prediction with Recurrent Neural Networks》

介绍:谷歌研究院 Samy Bengio 等人最近写的 RNN 的 Scheduled Sampling 训练方法论文.

615. 《Essential Machine Learning Algorithms in a nutshell》

介绍:机器学习基本算法简要入门.

616. 《A Huge List of Machine Learning And Statistics Repositories》

介绍:Github 机器学习/数学/统计/可视化/深度学习相关项目大列表.

617. 《Information Processing and Learning》

介绍:CMU 的信息论课程.

618. 《Scheduled sampling for sequence prediction with recurrent neural networks》

介绍:谷歌研究院 Samy Bengio 等人最近写的 RNN 的 Scheduled Sampling 训练方法论文.

619. 《基于 Hadoop 集群的大规模分布式深度学习》

介绍:基于 Hadoop 集群的大规模分布式深度学习.

620. 《Learning both Weights and Connections for Efficient Neural Networks 习》

介绍:来自斯坦福大学及 NVIDIA 的工作,很实在很实用.采用裁剪网络连接及重训练方法,可大幅度减少 CNN 模型参数.针对 AlexNet、VGG 等模型及 ImageNet 数据,不损失识别精度情况下,模型参数可大幅度减少 9-13 倍.

621. 《Apache Singa --A General Distributed Deep Learning Platform》

介绍:无需做深度学习就能用的分布式深度学习软件,github.

622. 《24 Ultimate Data Scientists To Follow in the World Today》

介绍:当今世界最 NB 的 25 位大数据科学家,通过他们的名字然后放在 google 中搜索肯定能找到很多很棒的资源译文.

623. 《Deep Learning for NLP - Lecture October 2015》

介绍:Nils Reimers 面向 NLP 的深度学习(Theano/Lasagne)系列教程.

624. 《Connection between probability theory and real analysis》

介绍:主讲人是陶哲轩,资料 Probability: Theory and Examples,笔记.

请关注“数据精简 DataSimp”公众号! 14 时 7 分 3 秒 2016 年 10 月 12 日星期三

秦陇纪 2010 数据简化 DataSimp 个人新浪博客 blog.sina.com.cn/inxcn, 微博 weibo.com/INXCN, 网盘 <http://pan.baidu.com/s/1i4D2GO9>, 分享人工智能大数据文章. 欢迎大家参加线下聚会, 线上点赞互粉、跟顶转帖、下载学习、转发 QQ 空间、微信朋友圈、公众号。

625. 《Data Science Learning Resources》

介绍:数据科学(学习)资源列表.

626. 《8 Tactics to Combat Imbalanced Classes in Your Machine Learning Dataset》

介绍:应对非均衡数据集分类问题的八大策略.

627. 《Top 20 Data Science MOOCs》

介绍:重点推荐的 20 个数据科学相关课程.

628. 《Recurrent Neural Networks》

介绍:递归神经网络.

629. 《Histograms of Oriented Gradients》

介绍:(HOG)学习笔记.

630. 《Computational modelling courses》

介绍:计算建模/计算神经学课程汇总.

631. 《How We Use Deep Learning to Classify Business Photos at Yelp》

介绍:(Yelp)基于深度学习的商业图片分类.

632. 《Neural Networks and Deep Learning》

介绍:免费在线书《Neural Networks and Deep Learning》神经网络与深度学习。目前提供了前四章的草稿，第一章通过手写数字识别的例子介绍 NN，第二章讲反向传播算法，第三章讲反向传播算法的优化，第四章讲 NN 为什么能拟合任意函数。大量 python 代码例子和交互动画，生动有趣.中文版

633. 《Books to Read if You Might Be Interested in Data Science》

介绍:数据科学大咖荐书(入门).

634. 《Deep Learning for NLP resources》

介绍:NLP 深度学习资源列表.

635. 《GitXiv》

介绍:很多 arXiv 上面知名论文可以在这个网站找到 github 的项目链接.

636. 《Learning Multi-Domain Convolutional Neural Networks for Visual Tracking》

介绍:深度学习在视觉跟踪的探索.

637. 《Beginners Guide: Apache Spark Machine Learning Scenario With A Large Input Dataset》

介绍:Spark 机器学习入门实例——大数据集(30+g)二分类.

638. 《Semantic Scholar》

介绍:保罗艾伦人工智能实验室表示, Google Scholar 是十年前的产物, 他们现在想要做进一步的提高. 于是推出了全新的, 专门针对科学家设计的学术搜索引擎 Semantic Scholar.

639. 《Semi-Supervised Learning》

介绍:半监督学习,Chapelle.篇篇都是经典,作者包括 Vapnik,Bengio,Lafferty,Jordan.此外推荐 Xiaojin (Jerry) Zhu 编写的 Introduction to Semi-Supervised Learning.

介绍:Spark 机器学习入门实例——大数据集(30+g)二分类.

640. 《Free Resources for Beginners on Deep Learning and Neural Network》

介绍:为入门者准备的深度学习与神经网络免费资源.

641. 《TensorFlow is an Open Source Software Library for Machine Intelligence》

介绍:Google 开源最新机器学习系统 TensorFlow,此外提供 TensorFlow 白皮书 white paper of tensorflow

2015.hacker news,Google 大牛解读 TensorFlow

642. 《Ves: Distributed machine learning platform》

介绍:三星开源的快速深度学习应用程序开发分布式平台.

643. 《DMTK: Microsoft Distributed Machine Learning Toolkit 》

介绍:分布式机器学习工具包.

644. 《Semantics Approach to Big Data and Event Processing》

介绍:语义大数据——大数据/事件处理的语义方法.

645. 《LSTM(Long Short Term Memory)和 RNN(Recurrent)学习教程》

介绍:LSTM(Long Short Term Memory)和 RNN(Recurrent)学习教程.

646. 《Marvin: A minimalist GPU-only N-dimensional ConvNet framework》

介绍:Princeton Vision Group 的深度学习库开源.

647. 《Ufora is a compiled, automatically parallel subset of python for data science and numerical computing》

介绍:基于 AWS 的自动分布式科学计算库 Ufora, Why I Open Sourced Five Years of Work.

648. 《Deep Learning and Deep Data Science - PyCon SE 2015》

介绍:(PyCon SE 2015)深度学习与深度数据科学.

649. 《Zhi-Hua Zhou Papers》

介绍:推荐南京大学机器学习与数据挖掘研究所所长——周志华教授的 Google 学术主页.

请关注“数据精简 DataSimp”公众号! 14 时 7 分 3 秒 2016 年 10 月 12 日星期三

秦陇纪 2010 数据简化 DataSimp 个人新浪博客 blog.sina.com.cn/inxcn, 微博 weibo.com/INXCN, 网盘 <http://pan.baidu.com/s/1i4D2GO9>, 分享人工智能大数据文章. 欢迎大家参加线下聚会, 线上点赞互粉、跟帖转帖、下载学习、转发 QQ 空间、微信朋友圈、公众号。

650. 《Advanced Linear Models for Data Science》
介绍:免费书:面向数据科学的高级线性模型.
651. 《Net2Net: Accelerating Learning via Knowledge Transfer》
介绍:基于知识迁移的神经网络高效训练 Net2Net.
652. 《徐亦达机器学习课程 Variational Inference》
介绍:徐亦达机器学习课程 Variational Inference.
653. 《Learning the Architecture of Deep Neural Networks》
介绍:深度神经网络结构学习.
654. 《Multimodal Deep Learning》
介绍:来自斯坦福大学的 Multimodal Deep Learning papers.
655. 《深度学习简析,TensorFlow,Torch,Theano,Mxnet》
介绍:深度学习简析,TensorFlow,Torch,Theano,Mxnet.
656. 《"Notes Essays —CS183C: Technology-enabled Blitzscaling—Stanford University》
介绍:这个专栏是一个 stanford 学生做的 CS183c 课程的一个 note, 该课程是由 Reid Hoffman 等互联网 boss 级人物开设的, 每节课请一位巨头公司的相关负责人来做访谈, 讲述该公司是怎么 scale 的. 最新两期分别请到了雅虎的梅姐和 airbnb 创始人 Brian Chesky. .
657. 《Natural Language Understanding with Distributed Representation》
介绍:基于分布式表示的自然语言理解(100+页),论文.
658. 《Recommender Systems Handbook》
介绍:推荐系统手册.
659. 《Understanding LSTM Networks》
介绍:理解 LSTM 网络翻译.
660. 《Machine Learning at Quora》
介绍:机器学习在 quora 中的应用.
661. 《On Learning to Think: Algorithmic Information Theory for Novel Combinations of Reinforcement Learning Controllers and Recurrent Neural World Models》
介绍:思维学习——RL+RNN 算法信息论.
662. 《The 5 Ways Data Scientists Keep Learning After College》
介绍:数据科学家毕业后继续学习的 5 种方式.
663. 《Deep Learning in Neural Networks: An Overview》
介绍:深度学习在神经网络的应用.
664. 《Contextual Learning》
介绍:上下文学习,代码.
665. 《Machine Learning For Complete Beginners》
介绍:机器学习零基础入门,代码.
666. 《2015 年中国计算机学会 (CCF) 优秀博士学位论文》
介绍:2015 年度 CCF 优秀博士学位论文奖论文列表.
667. 《Learning to Hash Paper, Code and Dataset》
介绍:Learning to Hash Paper, Code and Dataset.
668. 《Neural networks with Theano and Lasagne》
介绍:(PyData2015)基于 Theano/Lasagne 的 CNN/RNN 教程,github.
669. 《神经网络与深度学习讲义》
介绍:复旦大学邱锡鹏老师编写的神经网络与深度学习讲义,ppt.
670. 《Microsoft Open Sources Distributed Machine Learning Toolkit》
介绍:微软亚洲研究院开源分布式机器学习工具包.
671. 《语音识别的技术原理是什么?》
介绍:语音识别的技术原理浅析
672. 《Michael I. Jordan》
介绍:迈克尔·I.乔丹的主页.根据主页可以找到很多资源.迈克尔·I.乔丹是知名的计算机科学和统计学学者, 主要研究机器学习和人工智能.他的重要贡献包括指出了机器学习与统计学之间的联系, 并推动机器学习界广泛认识到贝叶斯网络的重要性.
673. 《Geoff 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.通过他的主页可以发掘到很多 Paper 以及优秀学生的 paper,此外推荐他的学生 Yann Lecun 主页
674. 《Yoshua Bengio》

请关注“数据精简 DataSimp”公众号! 14 时 7 分 3 秒 2016 年 10 月 12 日星期三

介绍:Yoshua Bengio 是机器学习方向的牛人,如果你不知道可以阅读对话机器学习大神 Yoshua Bengio (上),对话机器学习大神 Yoshua Bengio (下)

675. 《Large Scale Deep Learning within google》

介绍:google 大规模深度学习应用演进

676. 《Deep Learning: An MIT Press Book in Preparation》

介绍:MIT 出版的深度学习电子书,公开电子书

677. 《A Mathematical Theory of Deep Convolutional Neural Networks for Feature Extraction》

介绍:深度卷积神经网络(CNN)提取特征的数学理论

678. 《Microsoft Research Asia: Kaiming He》

介绍:推荐微软亚洲研究院何恺明主页

679. 《Speech and Language Processing (3rd ed. draft)》

介绍:《语音与语言处理》第三版(草稿)

680. 《LSA 311: Computational Lexical Semantics - Summer 2015》

介绍:Stanford 新课"计算词汇语义学"

681. 《上海交大张志华老师的统计机器学习与机器学习导论视频》

介绍:上海交大张志华老师的统计机器学习与机器学习导论视频链接:密码: r9ak .概率基础

682. 《Computational Linguistics and Deep Learning》

介绍:computational linguistics and deep learning 视频,推荐 Deep Learning: An Introduction from the NLP Perspective

683. 《Black Hat USA 2015 - Deep Learning On Disassembly》

介绍:(BlackHat2015)深度学习应用之流量鉴别(协议鉴别/异常检测),[slide]](<https://www.blackhat.com/docs/us-15/materials/us-15-Wang-The-Applications-Of-Deep-Learning-On-Traffic-Identification.pdf>),[material]](<https://www.blackhat.com/docs/us-15/materials/us-15-Wang-The-Applications-Of-Deep-Learning-On-Traffic-Identification-wp.pdf>)

684. 《LibRec: A Java Library for Recommender Systems》

介绍:一个推荐系统的 Java 库

685. 《Multi-centrality Graph Spectral Decompositions and their Application to Cyber Intrusion Detection》

介绍:多中心图的谱分解及其在网络入侵检测中的应用(MC-GPCA&MC-GDL)

686. 《Computational Statistics in Python》

介绍:用 Python 学计算统计学

687. 《New open-source Machine Learning Framework written in Java》

介绍:datumbbox-framework——Java 的开源机器学习框架,该框架重点是提供大量的机器学习算法和统计检验,并能够处理中小规模的数据集

688. 《Awesome Recurrent Neural Networks》

介绍:递归神经网络 awesome 系列,涵盖了书籍,项目,paper 等

689. 《Pedro Domingos》

介绍:Pedro Domingos 是华盛顿大学的教授,主要研究方向是机器学习与数据挖掘.在 2015 年的 ACM webinar 会议,曾发表了关于盘点机器学习领域的五大流派主题演讲.他的个人主页拥有很多相关研究的 paper 以及他的教授课程.

690. 《Video resources for machine learning》

介绍:机器学习视频集锦

691. 《Deep Machine Learning libraries and frameworks》

介绍:深度机器学习库与框架

692. 《大数据/数据挖掘/推荐系统/机器学习相关资源》

介绍:这篇文章内的推荐系统资源很丰富,作者很有心,摘录了《推荐系统实战》内引用的论文.

693. 《Bayesian Methods in Astronomy: Hands-on Statistics》

介绍:(天文学)贝叶斯方法/MCMC 教程——统计实战

694. 《Statistical Learning with Sparsity: The Lasso and Generalizations》

介绍:免费书:统计稀疏学习,作者 Trevor Hastie 与 Rob Tibshirani 都是斯坦福大学的教授,Trevor Hastie 更是在统计学学习上建树很多

695. 《The Evolution of Distributed Programming in R》

介绍:R 分布式计算的进化,此外推荐(R)气候变化可视化,(R)马尔可夫链入门

696. 《neon workshop at Startup.ML: Sentiment Analysis and Deep Reinforcement Learning》

介绍:Nervana Systems 在 Startup.ML 的主题研讨会——情感分析与深度强化学习

697. 《Understanding Convolution in Deep Learning》

介绍:深度学习卷积概念详解.

698. 《Python libraries for building recommender systems》

介绍:Python 推荐系统开发库汇总.

请关注“数据精简 DataSimp”公众号! 14 时 7 分 3 秒 2016 年 10 月 12 日星期三

秦陇纪 2010 数据简化 DataSimp 个人新浪博客 blog.sina.com.cn/inxcn, 微博 weibo.com/INXCN, 网盘 <http://pan.baidu.com/s/1i4D2GO9>, 分享人工智能大数据文章. 欢迎大家参加线下聚会, 线上点赞互粉、跟顶转帖、下载学习、转发 QQ 空间、微信朋友圈、公众号。

699. 《Neural networks class - Université de Sherbrooke》

介绍:超棒的神经网络课程,深入浅出介绍深度学习,由 Hugo Larochelle(Yoshua Bengio 的博士生,Geoffrey Hinton 之前的博士后)主讲,强烈推荐.

700. 《CS231n: Convolutional Neural Networks for Visual Recognition》

介绍:斯坦福新课程,面向视觉识别的卷积神经网络(Fei-Fei Li & Andrej Karpathy),slides+video,homework.

701. 《NIPS 2015 Deep Learning Symposium Part I》

介绍:NIPS 2015 会议总结第一部分,第二部分.

702. 《python 机器学习入门资料梳理》

介绍:python 机器学习入门资料梳理.

703. 《Reading Text in the Wild with Convolutional Neural Networks》

介绍:牛津大学著名视觉几何组 VGG 在 IJCV16 年首卷首期: Reading Text in the Wild with Convolutional Neural Networks,Jaderberg。这篇期刊文章融合了之前两篇会议(ECCV14,NIPS14ws),定位和识别图片中的文本(叫 text spotting)。端到端系统:检测 Region + 识别 CNN。论文、数据和代码.

704. 《Yet Another Computer Vision Index To Datasets (YACVID)》

介绍:计算机视觉的一个较大的数据集索引,包含 387 个标签,共收录了 314 个数据集合,点击标签云就可以找到自己需要的库了.

705. 《Why SLAM Matters, The Future of Real-Time SLAM, and Deep Learning vs SLAM》

介绍:Tombone 对 ICCV SLAM workshop 的总结: the future of SLAM, SLAM vs deep learning 重点介绍了 monoSLAM 和 LSD-SLAM,而且讨论了 feature-based 和 feature-free method 的长短。在全民 deep learning 做 visual perception 的时候,再来读读 CV 中的 geometry.

706. 《Python based Deep Learning Framework by Nervana™》

介绍:Nervana Systems 的开源深度学习框架 neon 发布.

707. 《ImageNet and MS COCO Visual Recognition Challenges video and slider》

介绍:ICCV 2015 的 ImageNet 比赛以及 MS COCO 竞赛联合研讨会的幻灯片和视频.

708. 《An Introduction to Machine Learning with Python》

介绍:Python 机器学习入门.

709. 《Neural Enquirer: Learning to Query Tables with Natural Language》

介绍:Neural Enquirer 第二版.

710. 《Deep Learning - Taking machine learning to the next level》

介绍:[Google]基于 TensorFlow 的深度学习/机器学习课程.

711. 《100 "must read" R-bloggers' posts for 2015》

介绍:R-bloggers 网站 2015"必读"的 100 篇文章,R 语言学习的福音.

712. 《Machine Learning: a Probabilistic Perspective》

介绍:推荐书籍:<机器学习: 概率视角>,样章 Undirected graphical models Markov random fields.

713. 《Deep learning Book》

介绍:这是一本在线的深度学习书籍,合著者有 Ian Goodfellow, Yoshua Bengio 和 Aaron Courville.如果你是一位新入门的学员可以先看这本书籍 Yoshua Bengio: How can one get started with machine learning?

714. 《UFLDL Recommended Readings》

介绍:UFLDL 推荐的深度学习阅读列表.

715. 《CSE 705: Deep Learning (Spring 2015)》

介绍:纽约州立大学布法罗分校 2015 年春季机器学习课程主页.

716. 《Theano is a Deep learning Python library 》

介绍: Theano 是主流的深度学习 Python 库之一,亦支持 GPU,入门比较难.推荐 Theano tutorial,Document

717. 《Statistical Language Models Based On Neural Networks》

介绍:博士论文:神经网络统计语言模型.

718. 《文本数据的机器学习自动分类方法(上)》

介绍:文本数据的机器学习自动分类方法(下).

719. 《Pixel Recurrent Neural Networks》

介绍:用 RNN 预测像素,可以把被遮挡的图片补充完整.

720. 《Computational Network Toolkit (CNTK)》

介绍:微软研究院把其深度学习工具包 CNTK,想进一步了解和学 CNTK 的同学可以看前几天公布的《CNTK 白皮书》An Introduction to Computational Networks and the Computational Network Toolkit.

721. 《Kalman and Bayesian Filters in Python》

介绍:卡尔曼滤波器教材,用尽量少的数学和推导,传授直觉和经验,全部 Python 示例,内容覆盖卡尔曼滤波器、扩展卡尔曼滤波,无迹卡尔曼滤波等,包括练习和参考答案

722. 《Statistical inference for data science》

请关注“数据精简 DataSimp”公众号! 14 时 7 分 3 秒 2016 年 10 月 12 日星期三

秦陇纪 2010 数据简化 DataSimp 个人新浪博客 blog.sina.com.cn/inxcn, 微博 weibo.com/INXCN, 网盘 <http://pan.baidu.com/s/1i4D2GO9>, 分享人工智能大数据文章. 欢迎大家参加线下聚会, 线上点赞互粉、跟顶转帖、下载学习、转发 QQ 空间、微信朋友圈、公众号。

介绍:在线免费书:面向数据科学的统计推断, R 示例代码, 很不错 GitHub.

723. 《Learning Deep Architectures for AI》

介绍:这本书是由 Yoshua Bengio 撰写的教程,其内容包含了学习人工智能所使用的深度学习架构的学习资源,书中的项目已停止更新 DeepLearnToolbox.

724. 《Machine Learning Tutorials》

介绍:这是一份机器学习和深度学习教程, 文章和资源的清单。这张清单根据各个主题进行撰写, 包括了许多与深度学习有关的类别、计算机视觉、加强学习以及各种架构.

725. 《Data science ipython notebooks》

介绍:这是由 Donne Martin 策划收集的 IPython 笔记本。话题涵盖大数据、Hadoop、scikit-learn 和科学 Python 堆栈以及很多其他方面的内容。至于深度学习, 像是 TensorFlow、Theano 和 Caffe 之类的框架也均被涵盖其中, 当然还有相关的特定构架和概念等.

726. 《Open Source Deep Learning Server》

介绍:开源的深度学习服务,DeepDetect 是 C++实现的基于外部机器学习/深度学习库(目前是 Caffe)的 API。给出了图片训练(ILSVRC)和文本训练(基于字的情感分析, NIPS15)的样例, 以及根据图片标签索引到 ElasticSearch 中 github.

727. 《Data Mining, Analytics, Big Data, and Data Science》

介绍:这是国外的一个科技频道,涵盖了数据挖掘,分析以及数据科学类的文章.偶尔还有机器学习精选.

728. 《Data Mining And Statistics: What's The Connection?》

介绍:经典论文:数据挖掘与统计学.

729. 《(NIPS'2015 Tutorial)Yoshua Bengio 深度学习》

介绍:NIPS'2015 Tutorial by Yoshua Bengio.

730. 《(NENO:Python based Deep Learning Framework》

介绍:Nervana Systems 的开源深度学习框架 neon 发布.

731. 《(Matt Might:Reading for graduate students》

介绍:犹他州大学 Matt Might 教授推荐的研究生阅读清单.

732. 《Awesome Public Datasets》

介绍:开放数据集.

733. 《Introduction to Probability - The Science of Uncertainty》

介绍:(edX)不确定性的科学——概率论导论(MITx).

734. 《R software and tools for everyday use》

介绍:R 语言开发常用软件/工具推荐.

735. 《Implementing Dynamic memory networks》

介绍:动态记忆网络实现.

736. 《Deeplearning4j 中文主页》

介绍:英文主页

737. 《Big Data Analysis Learning Resources: 50 Courses, Blogs, Tutorials, And More For Mastering Big Data Analytics》

介绍:50 个大数据分析最佳学习资源(课程、博客、教程等)

738. 《A Full Hardware Guide to Deep Learning》

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

739. 《Deep Residual Networks》

介绍:kaiming 开源作品

740. 《The Definitive Guide to Natural Language Processing》

介绍:自然语言处理(NLP)权威指南

741. 《Evaluating language identification performance》

介绍:如何在社交媒体上做语言检测? 没有数据怎么办? 推特官方公布了一个十分难得的数据集: 12 万标注过的 Tweets, 有 70 种语言

742. 《ICLR 2016 Accepted Papers》

介绍:深度学习和机器学习重要会议 ICLR 2016 录取文章

743. 《Machine Learning: An In-Depth, Non-Technical Guide - Part 1》

介绍:机器学习——深度非技术指南

744. 《Data Storytelling 101: Helpful Tools for Gathering Ideas, Designing Content & More》

介绍:数据叙事入门指南——创意生成/数据采集/内容设计相关资源推荐

745. 《WikiTableQuestions: a Complex Real-World Question Understanding Dataset》

介绍:WikiTableQuestions——复杂真实问答数据集

746. 《Big Data: 35 Brilliant And Free Data Sources For 2016》

请关注“数据精简 DataSimp”公众号! 14 时 7 分 3 秒 2016 年 10 月 12 日星期三

介绍:(2016 版)35 个超棒的免费大数据源

747. 《SPARKNET: training deep networks in spark》

介绍:Ion Stoica 和 Michael I. Jordan 两位大家首次联手发文, CAFFE 和 SPARK 完美结合, 分布式深度学习混搭模式! github

748. 《DeepLearning.University – An Annotated Deep Learning Bibliography | Memkite》

介绍:深度学习(分类)文献集

749. 《Learning Deep Learning》

介绍:深度学习阅读列表

750. 《Awesome42 The easiest way to find R packages》

介绍:探索 R 包的好网站 Awesome 42

751. 《MLbase:Distributed Machine Learning Made Easy》

介绍:MLbase 是 Prof. Dr. Tim Kraska 的一个研究项目, MLbase 是一个分布式机器学习管理系统

752. 《Deep Learning At Scale and At Ease》

介绍:分布式深度学习平台 SINGA 介绍

753. 《Learn All About Apache Spark (100x Faster than Hadoop MapReduce)》

介绍:Spark 视频集锦

754. 《R For Deep Learning (I): Build Fully Connected Neural Network From Scratch》

介绍:R 语言深度学习第一节:从零开始

755. 《A Visual Introduction to Machine Learning》

介绍:图解机器学习

756. 《Citation Network Dataset》

介绍:AMiner 论文引用数据集(v7:2,244,021 papers and 4,354,534 citation relationships)

757. 《Best Free Machine Learning Ebooks》

介绍:10 本最佳机器学习免费书

758. 《International Conference on Computer Vision (ICCV) 2015, Santiago》

介绍:ICCV15 视频集

759. 《CaffeOnSpark Open Sourced for Distributed Deep Learning on Big Data Clusters》

介绍:.(Yahoo)基于 Hadoop/Spark 的分布式 Caffe 实现 CaffeOnSpark

760. 《A Short Introduction to Learning to Rank》

介绍:Learning to Rank 简介

761. 《Global Deep learning researcher》

介绍:全球深度学习专家列表,涵盖研究者主页

762. 《Top Spark Ecosystem Projects》

介绍:Spark 生态顶级项目汇总

763. 《Proceedings of the 21st International Conference on Intelligent User Interfaces》

介绍:ACM IUI'16 论文集 Conference Navigator - Proceedings

764. 《Machine Learning: An In-Depth, Non-Technical Guide - Part 1》

介绍:深入机器学习,2,3,4

765. 《Oxford Deep Learning》

介绍:Nando de Freitas 在 Oxford 开设的深度学习课程,课程 youtube 地址,Google DeepMind 的研究科学家,此外首页:computervisiontalks 的内容也很丰富,如果你是做机器视觉方面的研究,推荐也看看其他内容.肯定收获也不小.还有,这位 youtube 主页顶过的视频也很有份量

766. 《Neural Networks for Machine Learning》

介绍:Geoffrey Hinton 在 Coursera 开设的 MOOC

767. 《Deep Learning News》

介绍:深度学习领域的 Hacker news.紧跟深度学习的新闻、研究进展和相关的创业项目。从事机器学习,深度学习领域的建议每天看一看

768. 《Maxout Networks》

介绍:Maxout 网络剖析

769. 《Advances in Neural Information Processing Systems》

介绍:NIPS 领域的会议 paper 集锦

770. 《Machine learning applications in genetics and genomics》

介绍:机器学习在生物工程领域的应用,如果你从事生物工程领域,可以先阅读一篇文章详细介绍

771. 《Deep Learning in Bioinformatics》

介绍:深度学习在生物信息学领域的应用

772. 《A Few Useful Things to Know about Machine Learning》

请关注“数据精简 DataSimp”公众号! 14 时 7 分 3 秒 2016 年 10 月 12 日星期三

秦陇纪 2010 数据简化 DataSimp 个人新浪博客 blog.sina.com.cn/inxcn, 微博 weibo.com/INXCN, 网盘 <http://pan.baidu.com/s/1i4D2GO9>, 分享人工智能大数据文章。欢迎大家参加线下聚会, 线上点赞互粉、跟顶转帖、下载学习、转发 QQ 空间、微信朋友圈、公众号。

介绍:一些关于机器学习需要知道知识,对于刚刚入门机器学习的同学应该读一读

773. 《Cambridge Machine Learning Group》

介绍:剑桥大学机器学习用户组主页,网罗了剑桥大学一些机器学习领域专家与新闻

774. 《Randy Olson's data analysis and machine learning projects》

介绍:Randy Olson's 的一些数据分析与机器学习项目库,是学习实践的好材料

775. 《GoLearn:Golang machine learning library》

介绍:Golang 机器学习库,简单,易扩展

776. 《Swift Ai》

介绍:用 Swift 开发苹果应用的倒是很多,而用来做机器学习的就比较少了.Swift Ai 在这方面做了很多聚集.可以看看

777. 《Please explain Support Vector Machines (SVM) like I am a 5 year old》

介绍:如何向一位 5 岁的小朋友解释支持向量机(SVM)

778. 《reddit Machine learning》

介绍:reddit 的机器学习栏目

779. 《ComputerVision resource》

介绍:计算机视觉领域的一些牛人博客,超有实力的研究机构等的网站链接.做计算机视觉方向的朋友建议多关注里面的资源

780. 《Multimedia Laboratory Homepage》

介绍:香港中文大学深度学习研究主页,此外研究小组对 2013 年 deep learning 的最新进展和相关论文做了整理,其中 useful links 的内容很受益

781. 《Search Engines that Learn from Their Users》

介绍:这是一篇关于搜索引擎的博士论文,对现在普遍使用的搜索引擎 google, bing 等做了分析.对于做搜索类产品的很有技术参考价值

782. 《Deep Learning Books》

介绍:深度学习书籍推荐(毕竟这类书比较少).

783. 《Deep Learning Books》

介绍:深度学习书籍推荐(毕竟这类书比较少).

784. 《Towards Bayesian Deep Learning: A Survey》

介绍:贝叶斯定理在深度学习方面的研究论文.

785. 《Revisiting Distributed Synchronous SGD》

介绍:来自谷歌大脑的重温分布式梯度下降.同时推荐大规模分布式深度网络

786. 《Research Issues in Social Computing》

介绍:社交计算研究相关问题综述.

787. 《What are some important areas of research in social computing right now?》

介绍:社交计算应用领域概览,里面有些经典论文推荐

788. 《Collaborative Filtering Recommender Systems》

介绍:协同过滤在推荐系统应用.

789. 《Content-Boosted Collaborative Filtering for Improved Recommendations》

介绍:协同过滤在内容推荐的研究.

790. 《Unifying User-based and Item-based Collaborative Filtering Approaches by Similarity Fusion》

介绍:协同过滤经典论文.

791. 《Item-based Collaborative Filtering Recommendation Algorithms》

介绍:协同过滤算法.

792. 《Amazon.com Recommendations Item-to-Item Collaborative Filtering》

介绍:亚马逊对于协同过滤算法应用.

793. 《Collaborative Filtering for Implicit Feedback Datasets》

介绍:协同过滤的隐式反馈数据集处理.

794. 《Tutorials, papers and code for computer graphics, fractals and demoscene》

介绍:计算机图形,几何等论文,教程,代码.做计算机图形的推荐收藏.

795. 《ELEN 6886 Sparse Representation and High-Dimensional Geometry》

介绍:推荐哥伦比亚大学课程,稀疏表示和高维几何.12 年由 Elsevier、13 年至今由 PAMI(仍由 Elsevier 赞助)设立的青年研究者奖(Young Researcher Award)授予完成博士学位后七年内取得杰出贡献的;由 CV 社区提名,在 CVPR 会议上宣布.2015 年得主是哥大助理教授 John Wright, 09 年《健壮人脸识别的稀疏表示法》引用已超 5K.

796. 《Software engineer how to learning Machine learning》

介绍:CMU 机器学习系著名教授 Alex Smola 在 Quora 对于《程序员如何学习 Machine Learning》的建议: Alex 推荐了不少关于线性代数、优化、系统、和统计领域的经典教材和资料.

请关注“数据精简 DataSimp”公众号! 14 时 7 分 3 秒 2016 年 10 月 12 日星期三

秦陇纪 2010 数据简化 DataSimp 个人新浪博客 blog.sina.com.cn/inxcn, 微博 weibo.com/INXCN, 网盘 <http://pan.baidu.com/s/1i4D2GO9>, 分享人工智能大数据文章.欢迎大家参加线下聚会,线上点赞互粉、跟顶转帖、下载学习、转发 QQ 空间、微信朋友圈、公众号。

797. 《Book review: Fundamentals of Deep Learning》

介绍: 书籍推荐,深度学习基础.源码

798. 《Learning from Big Code》

介绍: 软件工程领域现在也对机器学习和自然语言处理很感兴趣,有人推出了“大代码”的概念,分享了不少代码集合,并且觉得 ML 可以用在预测代码 Bug, 预测软件行为, 自动写新代码等任务上。大代码数据集下载

799. 《Object Detection》

介绍: 深度学习进行目标识别的资源列表: 包括 RNN、MultiBox、SPP-Net、DeepID-Net、Fast R-CNN、DeepBox、MR-CNN、Faster R-CNN、YOLO、DenseBox、SSD、Inside-Outside Net、G-CNN

800. 《Deep Learning: Course by Yann LeCun at Collège de France 2016(Slides in English)》

介绍: Yann LeCun 2016 深度学习课程的幻灯片 (Deep Learning Course by Yann LeCun at Collège de France 2016) 百度云密码: cwsml 原地址

801. 《Stanford HCI Group》

介绍: 斯坦福人机交互组五篇 CHI16 文章。1.众包激励机制的行为经济学研究: 批量结算比单任务的完成率高。2.在众包专家和新手间建立联系: 微实习。3.词嵌入结合众包验证的词汇主题分类 (如猫、狗属于宠物)。4.词嵌入结合目标识别的活动预测。5.鼓励出错以加快众包速度。

802. 《Learn Data Science》

介绍: 自学数据科学

803. 《CS224D Lecture 7 - Introduction to TensorFlow》

介绍: 本课是 CS224D 一节介绍 TensorFlow 课程, ppt,DeepDreaming with TensorFlow

804. 《Leaf - Machine Learning for Hackers》

介绍: Leaf 是一款机器学习的开源框架, 专为黑客打造, 而非为科学家而作。它用 Rust 开发, 传统的机器学习, 现今的深度学习通吃。Leaf

805. 《MXnet:Flexible and Efficient library for deep learning》

介绍: GTC 2016 视频, MXnet 的手把手深度学习 tutorial,相关参考资料 MXNet Tutorial for NVidia GTC 2016.

806. 《OpenAI Gym: Toolkit for developing, comparing reinforcement learning algorithms》

介绍: OpenAI Gym: 开发、比较强化学习算法工具箱

807. 《conference-iclr-2016 Papers and Code》

介绍: 机器学习会议 ICLR 2016 论文的代码集合

808. 《probabilistic graphical models principles and techniques》

介绍: 此书是斯坦福大学概率图模型大牛 Daphne Koller 所写, 主要涉及的是贝叶斯网络和马尔科夫逻辑网络的 learning 和 inference 问题, 同时又对 PGM 有深刻的理论解释, 是学习概率图模型必看的书籍。难度中上, 适合有一些 ML 基础的研究生.备份地址

809. 《Information Theory, Inference, and Learning Algorithms》

介绍: 此书是剑桥大学著名信息论专家 David MacKay 所写, 出发角度与很多机器学习的书籍都不一样, inference 和 MCMC 那章写的最好, 难度中。适合研究生和本科生。

810. 《Convex Optimization – Boyd and Vandenberghe》

介绍: 非常好的 Convex Optimization 教材, 覆盖了各种 constrained 和 unconstrained optimization 方法, 介绍了 convex 优化的基本概念和理论。难度中, 适合对优化和机器学习有一定基础的人群

811. 《Machine - Learning - Tom Mitchell》

介绍: 本书是 CMU 机器学习系主任 Mitchell 早年写的机器学习教科书, 年代有点久远.难度不高。适合初学者, 本科生, 研究生

812. 《Learning with Kernels Support Vector Machines, Regularization, Optimization, and Beyond》

介绍: 本书设计学习 Kernel 和 SVM 的各种理论基础, 需要较强的数学功底, 适合对 kernel 和 SVM 感兴趣的同同学选读 Learning with Kernels PPT, 参考 Learning with Kernels

813. 《The Elements of Statistical Learning: Data Mining, Inference, and Prediction》

介绍: 斯坦福统计系三位大神的统计学习教科书, 偏统计和学习理论, 需要对线性代数、统计和概率论有一定基础、难度高、适合研究生

814. 《Data Mining: Practical Machine Learning Tools and Techniques》

介绍: 本书是著名机器学习工具 Weka 作者撰写的应用机器学习指导书、非常实用、难度低、适合文科和各种应用科学做参考

815. 《Foundations of Statistical Natural Language Processing》

介绍: 本书也是一本比较受欢迎的 NLP 教科书, 难度一般, 主要覆盖统计 NLP 方法, 是斯坦福的另一位大牛 Chirs Manning 所写

816. 《Speech and Language Processing》

介绍: 在北美 NLP 最常用的教材, 斯坦福 Jurafsky 所写的自然语言处理入门教程, 覆盖面较为全面, 难度中低。适合本科生和研究生

请关注“数据精简 DataSimp”公众号! 14 时 7 分 3 秒 2016 年 10 月 12 日星期三

秦陇纪 2010 数据简化 DataSimp 个人新浪博客 blog.sina.com.cn/inxcn, 微博 weibo.com/INXCN, 网盘 <http://pan.baidu.com/s/1i4D2GO9>, 分享人工智能大数据文章。欢迎大家参加线下聚会, 线上点赞互粉、跟顶转帖、下载学习、转发 QQ 空间、微信朋友圈、公众号。

817. 《Natural Language Processing with Python - NLTK》

介绍: 实战型教程, 著名工具 NLTK 作者的著作, 适合本科生和入门者边动手边学

818. 《nlp compromise》

介绍: 自然语言处理学习工具包, 帮助你理解什么是自然语言处理

819. 《Multi-Way, Multilingual Neural Machine Translation with a Shared Attention Mechanism》

介绍: 基于 Attention 共享机制的多通道/多语言感知机器翻译模型

820. 《CS229 - Machine Learning, Spring 2016, Stanford》

介绍: 斯坦福 2016 年机器学习春季课程

821. 《What is deep learning?》

介绍: 我们深度学习, 做数据处理。究竟深度学习可以用来做什么? 什么场景下引入深度学习要必传统的数据处理方式要好? 初学者可以先看看这篇文章

822. 《KDD16 accept papers》

介绍: KDD16 论文是双轨制: 分为研究和应用数据科学(之前叫工业&政府), 均是 10 页。前者录取 142 篇(口头 70+展报 72), 后者录取 66 篇(口头 40+展报 26), 共计 208 篇。去年录取率: 160/819=19.5%。高产作者: 九篇 Hui Xiong; 七篇 Jieping Ye; 五篇 Hengshu Zhu。高产机构: 十篇 Tsinghua University; 九篇 Carnegie Mellon University、Rutgers; 八篇 Stanford University。高频词: learning online networks [评论来自: 学生古]

823. 《Medical Data for Machine Learning》

介绍: 医疗数据

824. 《Matlab Hacks》

介绍: Matlab 小技巧

825. 《Fundamentals of TensorFlow》

介绍: TensorFlow 基本原理

826. 《A Course in Machine Learning》

介绍: 电子书, 机器学习课程

827. 《ICML 2016 Tutorials》

介绍: 国际机器学习会议 PPT 提供下载

828. 《TensorFlow: A system for large-scale machine learning》

介绍: 大规模机器学习系统: TensorFlow, 本论文已经投稿到 OSDI 2016

829. 《Globally Normalized Transition-Based Neural Networks》

介绍: 这篇论文提出的模型改进, 可用于任何包含 transition system 的 feed-forward NN 上。他们将这个模型用于了儿大 NLP sequence tagging 的任务上: POS speech tagging, dependency parsing 和 sentence compression, 结果发现有时候甚至比 RNN 还要好。

830. 《The best explanation of Convolutional Neural Networks on the Internet!》

介绍: CNNs 详解, 同时推荐 A Beginner's Guide To Understanding Convolutional Neural Networks

831. 《Python Data Visualization》

介绍: Python 数据可视化教程

832. 《EMNLP 2016 accepted papers》

介绍: 2016 年 EMNLP 会议接收的论文下载

833. 《Summaries and notes on Deep Learning research papers》

介绍: 深度学习领域论文笔记与总结

834. 《Deep Language Modeling for Question Answering using Keras》

介绍: Keras 是一个极简和高度模块化的神经网络库, 这是使用 Keras 构建深度语言模型使用问答系统 Github, 此外推荐 Keras resource

835. 《An Actor-Critic Algorithm for Sequence Prediction》

介绍: Yoshua Bengio 最新论文: 用于序列预测的 actor-critic 算法

836. 《Ranking Relevance in Yahoo Search》

介绍: KDD2016 工业界最佳论文, 另外学术界: 最佳论文 FRAUDAR: Bounding Graph Fraud in the Face of Camouflage; 最佳学生论文: TRIEST: Counting Local and Global Triangles in Fully-Dynamic Streams with Fixed Memory Size;

837. 《Top Deep Learning Projects》

介绍: Github top deep learning project

838. 《深度卷积网络 CNN 与图像语义分割》

介绍: 深度卷积网络 CNN 与图像语义分割

839. 《How Convolutional Neural Networks Work》

介绍: 卷积神经网络内部分析.

840. 《Deep Reinforcement Learning survey》

请关注“数据精简 DataSimp”公众号! 14 时 7 分 3 秒 2016 年 10 月 12 日星期三

介绍: 深度强化学习.

841. 《Python Data Science Tutorials》

介绍: Python 数据科学教程,推荐 Data Science Dictionary: P-Values.Interesting Python Tutorials

842. 《Full Resolution Image Compression with Recurrent Neural Networks》

介绍: google 新论文,使用循环神经网络进行高分辨率图像压缩处理

843. 《PArallel Distributed Deep LEarning》

介绍: 百度开源了 PaddlePaddle (PArallel Distributed Deep LEarning), 是一个易于使用的, 高效的, 灵活的, 可扩展的深度学习平台.

844. 《Neural Network Architectures》

介绍: 神经网络架构演进史: 全面回顾从 LeNet5 到 ENet 十余种架构, 深度神经网络与深度学习是非常强大的流行算法. 许多运用它们的成功典范依赖于对神经网络架构的精心设计. 作者在这篇文章中重新回顾一下最近几年的深度学习背景下神经网络设计的历史. 中文翻译

845. 《A Tutorial for Reinforcement Learning》

介绍: 强化学习教程

846. 《CSCC11 Introduction to Machine Learning and Data Mining》

介绍: 多伦多大学机器学习与数据挖掘 CSCC11 学习笔记。

847. 《Awesome 3D reconstruction list》

介绍: 图像三维重建相关资源大列表

848. 《The Mathematics of Machine Learning》

介绍: 机器学习需要哪些数学?

849. 《Mathematics of Machine Learning》

介绍: MIT 2015 年秋季课程——机器学习数学 resource

850. 《The Mathematics of Machine Learning》

介绍: 机器学习需要哪些数学?

851. 《A Complete Tutorial on Tree Based Modeling from Scratch (in R & Python)》

介绍: 一篇非常全面的树模型介绍文章, 节点分裂、树模型的关键参数及如何避免过拟合、和线性模型的比较、随机森林、Bagging、Boosting、GBM 和 XGBoost 实战都是干货满满, 真要说缺点什么, 那可能就是没有公式推导。

852. 《Phd-Level Machine Learning Courses》

介绍: 机器学习高级课程

853. 《Introduction to Recommender Systems: A 4-hour lecture》

介绍: Xavier Amatriain 在 CMU 的 Machine Learning 暑期学校上的讲授的内容

854. 《Paper Collection of Real-Time Bidding》

介绍: 实时竞价相关论文汇总

855. 《GoogleNews-vectors-negative300-SLIM》

介绍: GoogleNews 词向量模型精简版(235M)

856. 《Super-Resolution.Benchmark》

介绍: 超分辨率及其评价相关资源汇总

857. 《Awesome Network Analysis》

介绍: 网络分析相关资源大列表

858. 《Data Resources: Datasets ,Center for Data on the Mind》

介绍: 认知科学数据集大列表

859. 《Evaluation of Deep Learning Toolkits》

介绍: 评估了 Caffe, CNTK, TensorFlow, Theano, 和 Torch 等一些流行的深度学习工具箱各方面的特性。

860. 《Paper Weekly》

介绍: PaperWeekly 每周会分享 N 篇当下最流行、最有趣的 NLP paper, 旨在用最精炼的话说明白 paper 的贡献和创新

861. 《Stat212b: Topics Course on Deep Learning》

介绍: 这个主题目的是提出建立高维数据稳定表达, 比如图像, 文本和音频。我们将深入到深度学习这个主题, 从监督和无监督的学习讨论最新进展。特别重点将放在卷积结构, 不变性学习, 无监督学习和非凸优化。