

首页 (<http://www.open-open.com/>) 代码 (<http://www.open-open.com/code/>) 文档 (<http://www.open-open.com/doc/>) 问答

全部经验分类

Android (/lib/tag/Android) IOS (/lib/tag/IOS) JavaScript (/lib/tag/JavaScript)

(/lib/list/all) 

所有分类 (/lib/list/all) > 软件开发 (/lib/list/1) > 算法 (/lib/list/401)

18种常见数据挖掘算法小结

算法 (/lib/tag/算法) 2015-03-05 16:05:32 发布

您的评价: 3.0

收藏

2收藏

本文所有涉及到的数据挖掘代码的都放在了mygithub上了。

地址链接: <https://github.com/linyiqun/DataMiningAlgorithm>

大概花了将近2个月的时间,自己把18大数据挖掘的经典算法进行了学习并且进行了代码实现,涉及到了决策分类,聚类,链接挖掘,关联挖掘,模式挖掘等等方面。也算是对数据挖掘领域的小小入门了吧。下面就做个小小的总结,后面都是我自己相应算法的博文链接,希望能够帮助大家学习。

1.C4.5算法。C4.5算法与ID3算法一样,都是数学分类算法,C4.5算法是ID3算法的一个改进。ID3算法采用信息增益进行决策判断,而C4.5采用的是增益率。

详细介绍链接: <http://blog.csdn.net/androidlushangderen/article/details/42395865>

2.CART算法。CART算法的全称是分类回归树算法,他是一个二元分类,采用的是类似于熵的基尼指数作为分类决策,形成决策树后之后还要进行剪枝,我自己在实现整个算法的时候采用的是代价复杂度算法,

详细介绍链接: <http://blog.csdn.net/androidlushangderen/article/details/42558235>

3.KNN(K最近邻)算法。给定一些已经训练好的数据,输入一个新的测试数据点,计算包含于此测试数据点的最近的点的分类情况,哪个分类的类型占多数,则此测试点的分类与此相同,所以在这里,有的时候可以复制不同的分类点不同的权重。近的点的权重重点,远的点自然就小点。

详细介绍链接: <http://blog.csdn.net/androidlushangderen/article/details/42613011>

4.Naive Bayes(朴素贝叶斯)算法。朴素贝叶斯算法是贝叶斯算法里面一种比较简单的分类算法,用到了一个比较重要的贝叶斯定理,用一句简单的话概括就是条件概率的相互转换推导。

详细介绍链接: <http://blog.csdn.net/androidlushangderen/article/details/42680161>

5.SVM(支持向量机)算法。支持向量机算法是一种对线性和非线性数据进行分类的方法,非线性数据进行分类的时候可以通过核函数转为线性的情况再处理。其中的一个关键的步骤是搜索最大边缘超平面。

详细介绍链接: <http://blog.csdn.net/androidlushangderen/article/details/42780439>

6.EM(期望最大化)算法。期望最大化算法,可以拆分为2个算法,1个E-Step期望化步骤,和1个M-Step最大化步骤。他是一种算法框架,在每次计算结果之后,逼近统计模型参数的最大似然或最大后验估计。

详细介绍链接: <http://blog.csdn.net/androidlushangderen/article/details/42921789>

7.Apriori算法。Apriori算法是关联规则挖掘算法,通过连接和剪枝运算挖掘出频繁项集,然后根据频繁项集得到关联规则,关联规则的导出需要满足最小置信度的要求。

详细介绍链接: <http://blog.csdn.net/androidlushangderen/article/details/43059211>

8.FP-Tree(频繁模式树)算法。这个算法也有被称为FP-growth算法,这个算法克服了Apriori算法的产生过多候选集的缺点,通过递归的产生频度模式树,然后对树进行挖掘,后面的过程与Apriori算法一致。

详细介绍链接: <http://blog.csdn.net/androidlushangderen/article/details/43234309>

9.PageRank(网页重要性/排名)算法。PageRank算法最早产生于Google,核心思想是通过网页的入链数作为一个网页好快的判定标准,如果1个网页内部包含了多个指向外部的链接,则PR值将会被均分,PageRank算法也会遭到Link Span攻击。

详细介绍链接: <http://blog.csdn.net/androidlushangderen/article/details/43311943>

10.HITS算法。HITS算法是另外一个链接算法,部分原理与PageRank算法是比较相似的,HITS算法引入了权威值和中心值的概念,HITS算法是受用户查询条件影响的,他一般用于小规模的数据链接分析,也更容易遭受到攻击。

详细介绍链接: <http://blog.csdn.net/androidlushangderen/article/details/43311943>

11.K-Means(K均值)算法。K-Means算法是聚类算法，k在这里指的是分类的类型数，所以在开始设定的时候非常关键，算法的原理是首先假定k个分类点，然后根据欧式距离计算分类，然后去同分类的均值作为新的聚簇中心，循环操作直到收敛。

详细介绍链接：<http://blog.csdn.net/androidlushangderen/article/details/43373159>

12.BIRCH算法。BIRCH算法利用构建CF聚类特征树作为算法的核心，通过树的形式，BIRCH算法扫描数据库，在内存中建立一棵初始的CF-树，可以看做数据的多层压缩。

详细介绍链接：<http://blog.csdn.net/androidlushangderen/article/details/43532111>

13.AdaBoost算法。AdaBoost算法是一种提升算法，通过对数据的多次训练得到多个互补的分类器，然后组合多个分类器，构成一个更加准确的分类器。

详细介绍链接：<http://blog.csdn.net/androidlushangderen/article/details/43635115>

14.GSP算法。GSP算法是序列模式挖掘算法。GSP算法也是Apriori类算法，在算法的过程中也会进行连接和剪枝操作，不过在剪枝判断的时候还加上了一些时间上的约束等条件。

详细介绍链接：<http://blog.csdn.net/androidlushangderen/article/details/43699083>

15.PreFixSpan算法。PreFixSpan算法是另一个序列模式挖掘算法，在算法的过程中不会产生候选集，给定初始前缀模式，不断的通过后缀模式中的元素转到前缀模式中，而不断的递归挖掘下去。

详细介绍链接：<http://blog.csdn.net/androidlushangderen/article/details/43766253>

16.CBA(基于关联规则分类)算法。CBA算法是一种集成挖掘算法，因为他是建立在关联规则挖掘算法之上的，在已有的关联规则理论前提下，做分类判断，只是在算法的开始时对数据做处理，变成类似于事务的形式。

详细介绍链接：<http://blog.csdn.net/androidlushangderen/article/details/43818787>

17.RoughSets(粗糙集)算法。粗糙集理论是一个比较新颖的数据挖掘思想。这里使用的是用粗糙集进行属性约简的算法，通过上下近似集的判断删除无效的属性，进行规制的输出。

详细介绍链接：<http://blog.csdn.net/androidlushangderen/article/details/43876001>

18.gSpan算法。gSpan算法属于图挖掘算法领域，主要用于频繁子图的挖掘，相较于其他的图算法，子图挖掘算法是他们的一个前提或基础算法。gSpan算法用到了DFS编码，和Edge五元组，最右路径子图扩展等概念，算法比较的抽象和复杂。

详细介绍链接：<http://blog.csdn.net/androidlushangderen/article/details/43924273>

来源：CSDN博客 作者：Android路上的人

扩展阅读

一文读懂机器学习，大数据/自然语言处理/算法全有了..... (</lib/view/open1420615208000.html>)

数据挖掘18大算法实现以及其他相关经典DM算法 (</lib/view/open1434413922770.html>)

从机器学习谈起 (</lib/view/open1421820307640.html>)

Spark设计理念与基本架构 (</lib/view/open1453899311745.html>)

数据挖掘-Eclat算法 (</lib/view/open1386810819564.html>)

为您推荐

大数据机器学习相关资源 (</lib/view/open1450276344344.html>)

大数据/数据挖掘/推荐系统/机器学习相关资源 (</lib/view/open1450794282980.html>)

大数据/数据挖掘/推荐系统/机器学习相关资源 (</lib/view/open1451549666151.html>)

EM算法 (</lib/view/open1453126027901.html>)

[原]【机器学习基础】随机森林算法 (</lib/view/open1437921406507.html>)

更多

算法 (<https://www.baidu.com/s?wd=site:open-open.com> 算法)

1. 从机器学习谈起, 大数据/自然语言处理/算法全有了..... (<http://www.open-open.com/news/view/1a5abcd>)
2. 一个数据挖掘大牛, 用程序算法做人生选择 (<http://www.open-open.com/news/view/1d79432>)
3. 机器学习常见算法分类汇总 (<http://www.open-open.com/news/view/25821f>)
4. 实际项目中的常见算法 (<http://www.open-open.com/news/view/17b6acb>)
5. 数据挖掘的10种分析方法 (<http://www.open-open.com/news/view/112e77c>)
6. 这些年, 这些挖掘算法, 这些反思 (<http://www.open-open.com/news/view/1dca66>)

同类热门经验

1. 一文读懂机器学习, 大数据/自然语言处理/算法全有了..... (</lib/view/open1420615208000.html>)
2. 各种Java加密算法 (</lib/view/open1397274257325.html>)
3. C语言的学习基础,100个经典的算法 (</lib/view/open1364807445109.html>)
4. MapReduce 模式、算法和用例 (MapReduce Patterns, Algorithms, and Use Cases) (</lib/view/open1330094286171.html>)
5. Java RSA 加密解密算法 入门 (</lib/view/open1346633931975.html>)
6. Mahout算法集 (</lib/view/open1383390452040.html>)

阅读目录

扩展阅读

为您推荐

更多



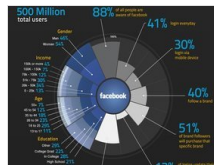
大数据分析工具



数据可视化工具



二手手机



数据可视化



数据分析软件



称重传感器

相关文档 — 更多 (<http://www.open-open.com/doc>)

- 数据挖掘常见算法.ppt (<http://www.open-open.com/doc/view/060174fa1c584d618ac41d625e933f2f>)
- 数据挖掘中分类算法小结.doc (<http://www.open-open.com/doc/view/c51473d961d648e3a384750350407b96>)
- 数据挖掘算法介绍.doc (<http://www.open-open.com/doc/view/ff0add7039a74fef8eef4620a80ffb5b>)
- 数据挖掘十大经典算法.doc (<http://www.open-open.com/doc/view/584103567f5e4ce8a581af798f03e497>)
- 数据挖掘中十大经典算法.pdf (<http://www.open-open.com/doc/view/02609e8cf935432f96e3c0d8f780cf98>)
- 数据挖掘10大算法.pdf (<http://www.open-open.com/doc/view/2327e517845146ba85c160a2cdcf0eaa>)
- 数据挖掘--数据挖掘10大算法.pdf (<http://www.open-open.com/doc/view/0716488a9f91451bbc2fd6c08960490c>)
- 数据挖掘十大算法.pdf (<http://www.open-open.com/doc/view/6a84616b7cfc460f93cec6463bf8ba90>)
- 数据挖掘十大算法.pdf (<http://www.open-open.com/doc/view/b852803c8505421497fe57dd579de2c8>)
- 数据挖掘最常见的十种方法.docx (<http://www.open-open.com/doc/view/66833d9c9b5b41e887571b2dfe2a98d9>)
- 数据挖掘算法.ppt (<http://www.open-open.com/doc/view/27896e5e47aa42c784ced723ed3f237>)
- 非确定图数据的挖掘算法研究.pdf (<http://www.open-open.com/doc/view/16ec9a732cd64f5abe60187d4e89c6a9>)

相关经验 — 更多

(<http://www.open-open.com/lib>)

- 一文读懂机器学习, 大数据/自然语言处理/算法全有了..... (<http://www.open-open.com/lib/view/open1420615208000.html>)
- 数据挖掘18大算法实现以及其他相关经典DM算法 (</lib/view/open1434413922770.html>)
- 从机器学习谈起 (</lib/view/open1421820307640.html>)
- Spark设计理念与基本架构 (</lib/view/open1453899311745.html>)
- 数据挖掘-Eclat算法 (</lib/view/open1386810819564.html>)
- 数据挖掘十大算法 (</lib/view/open1398385932843.html>)
- 十大数据挖掘算法及各自优势 (</lib/view/open1409822530541.html>)
- 数据挖掘领域十大经典算法 (</lib/view/open1454077111683.html>)
- JDBC中5种常见数据库连接 (</lib/view/open1418522870667.html>)
- 数据挖掘入门算法整理 (</lib/view/open1442147884075.html>)
- 数据挖掘 (2): 关联规则FpGrowth

相关讨论 — 更多 (<http://www.open-open.com/solution>)

- 几种查找算法Java实现 (<http://www.open-open.com/solution/view/1415328413558>)
- 算法面试题 (<http://www.open-open.com/solution/view/1320502624734>)
- 开源力量公开课第47期: 数据挖掘入门指引 (免费) (<http://www.open-open.com/solution/view/1394679097848>)
- 搜索算法请教 (<http://www.open-open.com/solution/view/1321583069765>)
- 微信红包的算法实现探讨 (<http://www.open-open.com/solution/view/1423791272482>)
- 排序算法原来是这么排的 (<http://www.open-open.com/solution/view/1337686418683>)
- 阿里面试经历及总结 (数据研发、Java研发方向) (<http://www.open-open.com/solution/view/1426946480648>)

- 非确定图数据的挖掘算法研究.pdf (<http://www.open-open.com/doc/view/16ec9a732cd64f5abe60187d4e89c6a9>) (/lib/view/open1440842048325.html) 算法
- 数据挖掘算法.ppt (<http://www.open-open.com/doc/view/75e29f2bc77c4057b2dcad817069f080>) (/lib/view/open1453126319714.html) 数据挖掘十大算法--Apriori算法
- 数据挖掘原理与算法.pdf (<http://www.open-open.com/doc/view/149c08e510a7468d8385acd6fe582951>) (/lib/view/open1440842334122.html) 数据挖掘（6）：决策树分类算法
- 数据挖掘原理与算法.pdf (<http://www.open-open.com/doc/view/0d763c8d39124ef0bc8215eb09baf22e>) trie的Python包
- 数据挖掘与分析：基本概念与算法（草稿）.pdf (<http://www.open-open.com/doc/view/a5869b2c0d7143e78bf543f81de878da>) 多种非常见数据结构以及NB的各种
- 数据挖掘与分析：基本概念与算法（草稿）.pdf (<http://www.open-open.com/doc/view/4710ddc7c2e740589842171cda9f58db>)
- 数据挖掘算法介绍.ppt (<http://www.open-open.com/doc/view/3162cef574344ef18affdc2709a3820>)
- 数据结构常见算法.ppt (<http://www.open-open.com/doc/view/5076bc75ba9540b6a8472bff72bd601d>)

©2006-2016 深度开源



(<http://www.open-open.com/>)

浙ICP备09019653号-31

(<http://www.miibeian.gov.cn/>) 站长统计

([http://www.cnzz.com/stat/website.php?](http://www.cnzz.com/stat/website.php?web_id=1257892335)

web_id=1257892335)