

ChaoSimple

博客园 首页 新随笔 联系 管理

随笔- 124 文章- 1 评论- 47

昵称: ChaoSimple
园龄: 5年7个月
粉丝: 309
关注: 3
+加关注

随笔分类

.NET(6)
C# 基础(25)
C++/MFC(1)
DevExpress(1)
Java(8)
LINQ(3)
Linux系统管理(2)
Python(11)
SQL Server(4)
方法论(1)
机器学习和数据挖掘(24)
计算机文化(2)
科研相关(1)
快速开发(1)
其他（无关技术）(3)
设计模式和UML建模(10)
数据结构(1)
数学(6)
算法设计(4)
统计学习(1)
图形图像(1)

随笔档案

2015年8月 (1)
2015年5月 (1)
2014年12月 (6)
2014年11月 (3)
2014年10月 (8)
2014年9月 (1)
2014年1月 (2)
2013年11月 (5)
2013年10月 (4)
2013年9月 (1)
2013年8月 (2)
2013年7月 (9)
2013年6月 (13)
2013年5月 (6)
2013年4月 (9)
2013年3月 (11)
2013年2月 (3)
2013年1月 (1)
2012年12月 (6)
2012年11月 (7)
2012年9月 (4)
2012年8月 (7)
2012年6月 (2)
2012年5月 (12)

文章分类

[Markdown]
Socket(1)

数据归一化和两种常用的归一化方法

数据标准化（归一化）处理是数据挖掘的一项基础工作，不同评价指标往往具有不同的量纲和量纲单位，这样的情况会影响到数据分析的结果，为了消除指标之间的量纲影响，需要进行数据标准化处理，以解决数据指标之间的可比性。原始数据经过数据标准化处理后，各指标处于同一数量级，适合进行综合对比评价。以下是两种常用的归一化方法：

一、min-max标准化（Min-Max Normalization）

也称为离差标准化，是对原始数据的线性变换，使结果值映射到[0 - 1]之间。转换函数如下：

$$x^* = \frac{x - \min}{\max - \min}$$

其中max为样本数据的最大值，min为样本数据的最小值。这种方法有个缺陷就是当有新数据加入时，可能导致max和min的变化，需要重新定义。

二、Z-score标准化方法

这种方法给予原始数据的均值（mean）和标准差（standard deviation）进行数据的标准化。经过处理的数据符合标准正态分布，即均值为0，标准差为1，转化函数为：

$$x^* = \frac{x - \mu}{\sigma}$$

其中 μ 为所有样本数据的均值， σ 为所有样本数据的标准差。

参考文献：

http://webdataanalysis.net/data-analysis-method/data-normalization/

分类: [机器学习和数据挖掘](#)

好文要顶

关注我

收藏该文







ChaoSimple

关注 - 3

粉丝 - 309

+加关注

6

0

« 上一篇: [KNN算法简介](#)

» 下一篇: [常用检索系统简单说明](#)

posted @ 2013-07-31 10:52 ChaoSimple 阅读(196971) 评论(0) 编辑 收藏

刷新评论 刷新页面 返回顶部

注册用户登录后才能发表评论，请 [登录](#) 或 [注册](#)，[访问](#)网站首页。

【推荐】50万行VC++源码: 大型组态工控、电力仿真CAD与GIS源码库
【推荐】腾讯云免费实验室，1小时搭建人工智能应用
【推荐】高性能云服务器2折起，0.73元/日节省80%运维成本
【新闻】H3 BPM体验平台全面上线
【推荐】具有实战经验的前端全栈开发课程

http://www.cnblogs.com/chaosimple/p/3227271.html

1/2

**最新IT新闻:**

- 王川出任迅雷董事长 小米系要让迅雷告别“旧时代”
 - 360回怼“92年女生”: 黑公关所为 疑背后有人搞事
 - 果小美完成5000万美元C1轮融资 祥峰领投
 - 腾讯告《最萌英雄》抄袭LOL胜诉 获赔100万
 - 苏宁超市双十二战报出炉: 隐形眼镜成最大黑马
- » 更多新闻...

**最新知识库文章:**

- 以操作系统的角度述说线程与进程
 - 软件测试转型之路
 - 门内门外看招聘
 - 大道至简，职场上做人做事做管理
 - 关于编程，你的练习是不是有效的？
- » 更多知识库文章...

Copyright ©2017 ChaoSimple