



北京邮电大学

BEIJING UNIVERSITY OF POSTS AND TELECOMMUNICATIONS



Data Analytics with R

语言数据分析



相随相伴、谓之关联

艾新波 / 2018 • 北京



课程体系



R语言数据分析



上部：论道



- 第1章 气象万千、数以等观
- 第2章 所谓学习、归类而已
- 第3章 格言联璧话学习
- 第4章 源于数学、归于工程



中部：执具



- 第5章 工欲善其事必先利其器
- 第6章 基础编程
- 第7章 数据对象



- 第8章 人人都爱tidyverse
- 第9章 最美不过数据框



下部 博术



- 第10章 观数以形
- 第11章 相随相伴、谓之关联
- 第12章 既是世间法、自当有分别
- 第13章 方以类聚、物以群分
- 第14章 庐山烟雨浙江潮

数据分析简史

1805

Linear regression

1936

Linear Discriminant Analysis

1951

k-Nearest Neighbors algorithm

1957

k-Means

1958

Logistic Regression

1984

Classification And Regression Tree

1986

Back Propagation

1993

Apriori

1995

Support Vector Machine

2001

Random Forest

2006

Deep Learning

2016

AlphaGo

啤酒与尿布的故事



啤酒与尿布，或许是有史以来推销数据挖掘最好的广告之一~

啤酒与尿布的故事：故事版本一

沃尔玛在美国的一家分店发生过这样一件趣事：在一个夏季，管理者发现在那段时间里婴儿尿布和啤酒的销量次第拔高。这如果在一般的商店也许就会被忽略，但沃尔玛超市的管理者没有轻易放过。他们立即对这个现象进行了分析和讨论，并且派出了专门的队伍在卖场内进行全天候的观察。最后，谜底终于水落石出：原来，购买这两种产品的顾客一般都是年龄在25至35周岁的青年男子，由于孩子尚在哺乳期，所以每天下班后他们都会遵太太的命令到超市里为孩子购买婴儿纸尿裤，每当这个时候，他们大都会为自己顺带买回几瓶啤酒。

沃尔玛的管理者立即针对此现象采取了行动：将卖场内原来相隔很远的妇婴用品区与酒类饮料区的空间距离拉近，减少顾客的行走时间。

啤酒与尿布的故事：故事版本二

故事情节大致相同。

故事的主角依然是**沃尔玛**。

但是地点变成了德国，并且，心理学家给出了煞有介事的解释：

因为德国是大男子主义盛行的国家，男人要面子，觉着拿着尿布实在难受，所以他一定要拎着一箱啤酒，出门的时候让别人以为是顺带买了尿布的。

啤酒与尿布的故事：故事版本三

经过考证，这个故事最可信的版本发生在一家名为**奥斯卡（Osco）**的连锁杂货店里。这家企业的数据分析人员确实模糊地记得他们曾经看到了商店里纸尿裤和啤酒的销售之间有一种联系。但是，这些分析人员提出如何寻找这种联系，事实上并不是一个努力工作的年轻程序员偶然间发现的。最重要的是，这个结论被认为有些反常，因为在奥斯卡的连锁店里，纸尿裤和啤酒从来没有放在一起陈列过（甚至并不是它所有的连锁店都能销售啤酒）。

——《数据分析竞争法——企业赢之道》P157-158

啤酒与尿布的真实距离

某位网友收集的数据：

| 调查时间 | 店址 | 啤酒与尿布的距离 |
|-------------------|-----------------|------------|
| 2009年7月21日 | 易初莲花石佛营店 | 45米 |
| <u>2009年7月28日</u> | <u>沃尔玛万达广场店</u> | <u>91米</u> |
| 2009年8月21日 | 家乐福双井店 | 78米 |
| 2009年9月28日 | 京客隆甜水园店 | 120米 |

购物篮分析

Robustness assessment of urban rail transit based on complex network theory: A case study of the Beijing Subway - Google Chrome

www.sciencedirect.com/science/article/pii/S0925753515001484

ScienceDirect

Download PDF

ELSEVIER

Robustness as...
network theory

Yuhao Yang^a, Yong...

Show more

http://dx.doi.org/10.1016

Highlights

- The Beijing Subv
- A new weighted composite index was developed to assess node importance in a

Other users also viewed these articles

Download 3 PDFs

- ☒ Use of a human factors classification framework to identify causal factors for m...
Original Research Article
Safety Science, Volume 79, November 2015, Pages 163–174
Rebecca J. Mitchell, , Ann Williamson, , Brett Molesworth,
Download PDF
- ☒ Passenger-demands-oriented train scheduling for an urban rail transit network
Original Research Article
Transportation Research Part C: Emerging Technologies, Volume 60, November 2015, Pages 1–23
Yihui Wang, , Tao Tang, , Bin Ning, , Ton J.J. van den Boom, , Bart De Schutter,
Download PDF
- ☒ Using network theory to explore the complexity of subway construction acciden...
Original Research Article
Safety Science, Volume 64, April 2014, Pages 127–136
Zhipeng Zhou, , Javier Irizarry, , Qiming Li,
Download PDF

Do not show again

Sign in

Feedback

购物篮分析

✓ 商品已成功加入购物车！



爱是唯一的真相：找寻前世今生的灵魂伴侣

数量：1

假如你数据工程师/科学家
如何实现？

购买了该商品的用户还购买了



4册你若盛开清风自来+不生气
的智慧+淡定的女人最幸福

¥23.80

加入购物车



前世今生 轮回的生命启示

¥22.00

加入购物车



前世今生 生命轮回的启示 布
莱恩魏斯 励志与成功哲学 书

¥22.40

加入购物车



正版 遇见未知的自己 我心安
处是幸福 愿你与这世界温暖相

¥36.80

加入购物车



轮回：前世今生来生缘

¥27.20

加入购物车



正版包邮 你若盛开清风自来
+不生气的智慧+淡定的女人

¥24.80 赠京豆

加入购物车

站在算法设计者的角度：面临的问题

蛮力搜索带来的组合爆炸（可能的 $A \Rightarrow B$ ）： $3^k - 2^{k+1} + 1$

| 商品个数 | 可能的规则数 |
|------|---|
| 2 | 2 |
| 4 | 50 |
| 10 | 57002 |
| 20 | 3484687250 (3.5e9) |
| 100 | 515377520732011324202202224420402268886864624842 (5.2e47) |

站在算法设计者的角度：解决的办法



Apriori

A decorative blue border with rounded corners and a dashed line inside. Two thin blue lines intersect to form a crosshair, one positioned above and to the right of the Chinese text, and the other below and to the left of the English text.

谢谢聆听
Thank you

教师个人联系方式

艾新波

手机: 13641159546

QQ: 23127789

微信: 13641159546

E-mail: 13641159546@126.com

axb@bupt.edu.cn

地址: 北京邮电大学科研楼917室

课程网址: <https://github.com/byaxb/RDataAnalytics>

