

# 第五届中国 R 语言会议（上海会场）



自由的统计语言

主办：  
上海财经大学统计与管理学院

协办：  
统计之都

赞助：  
艾瑞咨询  
Mango Solutions  
上海财经大学“数理经济学”教育部重点实验室  
上海市高校一流学科（B类）“统计学”

2012 年 11 月 3 日 - 4 日

## R 语言简介

R 是一个有着统计分析功能及强大作图功能的语言环境和软件系统，由新西兰奥克兰大学统计系的 Ross Ihaka 和 Robert Gentleman 共同创立。R 语言可以看作是由 AT&T 贝尔实验室所创的 S 语言发展出的一种方言。

R 是在 GNU 协议下免费发行的，它的开发及维护现在则由 R 开发核心小组（R Development Core Team）具体负责，这个团队的成员大部分来自大学机构的统计及相关院系。除了这些作者之外，R 还拥有一大批贡献者，他们为 R 编写代码、修正程序缺陷和撰写文档。

R 的功能很大程度上是通过程序包（Package）来实现的，迄今为止，R 语言官网上的程序包数目已经超过 4000 个，广泛地覆盖了数据分析应用到的各类行业和领域。各种统计前沿理论方法的相应计算机程序都会在短时间内以软件包的形式得以实现，这种速度是其它统计软件无法比拟的。

在 KDNuggets 于 2012 年做的“使用何种编程或统计类语言进行分析和数据挖掘”的调查中，R 以 52.5% 的得票率荣登榜首，力压 Python、SQL、JAVA 和 SAS (<http://www.kdnuggets.com/polls/2012/analytics-data-mining-programming-languages.html>)。此外 R 还击败了 Excel 和 Rapidminer（2010 和 2011 年排名第一），在“过去十二个月中你在实际项目中使用的数据挖掘或分析工具”的调查中排名第一。

Rexer Analytics 5th 数据挖掘者调查报告指出：R 语言一直保持上升的势头，牢牢占据工具类的第一名。几乎有一半的调查对象（47%）声称使用 R 语言作为数据挖掘工具。[\(http://www.rexeranalytics.com/Data-Miner-Survey-Results-2011.html\)](http://www.rexeranalytics.com/Data-Miner-Survey-Results-2011.html)。

目前，几乎所有的西方大学与研究机构、以及越来越多的金融机构、制药公司、高科技企业都使用 R。R 的灵活性、开放性以及业界最广泛的支持是其不断完善和发展的根本原因，随着 R 越来越被学术界及业界认可，它也将在数据分析和统计建模中发挥越来越大的作用。

## 上海财经大学统计与管理学院简介

上海财经大学统计与管理学院于 2009 年 7 月创建于原上海财经大学统计学系，主要的学科专业是统计学，包括经济管理统计、金融统计、统计理论与方法、数量金融与风险管理等多个学科方向。上海财经大学统计学科是一个具有悠久历史、成绩斐然的学科。统计学科作为上海财经大学最早设立的系科之一，始建于 1946 年，并一直延续至今。1981 年，本学科点经国务院学位委员会批准为全国第一批统计学硕士学位授予点，1981 年起，本学科点率先招收应用统计研究生，培养的学生中已涌现出了一批有名望的统计专家和高层次的经济管理人员；1986 年，本学科点由国务院学位委员会批准为全国首批统计学博士授予点之一；1996 年经国务院学位委员会批准成为全国首批统计学博士后流动站之一；1997 年，本学科点被评为国家财政部的统计学重点学科，2007 年评为上海市重点学科；2009 年，本学科本科专业被评为国家教育部的特色专业。

上海财经大学统计学科的发展定位于国际化，按国际标准建设和发展本学科。同时，按照研究型大学的建设要求，根据上海财经大学的特点，统计与管理学院重点建设经济管理统计、统计理论与方法、数量金融与风险管理等方向，并为将来发展生物统计、工业统计等方向进行前期建设。统计与管理学院以“夯实数理基础、丰富方法体系、拓宽应用方向”为原则，形成一批具有自身特色的优势学科方向，其中“经济管理统计”和“统计理论与方法”方向在国内名列前茅。新建设方向“数量金融与风险管理”在全国已有一定的知名度。通过本科生、硕士生培养方案的改革，完善统计学科人才培养方案，采用国际通行教材，为全体学生提供优秀的教学服务，培养优质的统计人才。完成和出版了一批具有影响力的高质量科研成果，拥有了一支规模适中、结构合理的学术梯队。

在巩固和发展我院现有的学科基础上，特别是经济管理统计与数理统计，发挥统计与管理学院建院基础学科——经济管理统计学，巩固统计与管理学院的基础学科——数理统计学，创建以管理学（包括经济学）与统计学交叉的新增长点，积极引导、鼓励和支持统计与管理学院的教师从事交叉领域的课题的研究，并凭借上海财经大学经济管理类的学科特点，挖掘学科交叉的新增长点。统计与管理学院以“统计与管理交叉平台”为建设的特色，从数量的角度出发研究金融和管理中的实际问题，并提供所需的统计模型与方法。为社会培养具有数理基础的复合型管理人才和技术人才。

## 统计之都简介\*

“统计之都”（Capital of Statistics，简称 COS）网站成立于 2006 年 5 月，其主旨为传播统计学知识并将其应用于实际领域。纵观现今国内统计学理论和应用的发展，一方面我们不难发现统计学在应用领域的巨大潜力——现代管理、咨询、商业、经济、金融、医药、生物等等，无不需要数据的力量，而另一方面我们也不得不承认，国内统计学的应用很大程度上受理论的制约——无论是应用界的人们对统计学基础理论知识的欠缺，还是学术界所研究的理论对应用领域问题的轻视。

“统计之都”网站便是基于这样的认识而创建的。我们希望，统计理论研究者能充分关注应用问题，而统计应用者也能正确把握统计学基本知识，将统计学这门应用学科真正的潜力开发出来。

“统计之都”为非赢利性质网站，但大力欢迎所有商界和研究领域的朋友与我们在实际应用问题上合作。我们的口号是：

中国统计学门户网站，免费统计学服务平台

我们怀着“十年磨一剑”的决心，要将“统计之都”建成中国的统计学门户网站；我们抱着“己欲立而立人、己欲达而达人”的信条，要将“统计之都”以免费统计学服务平台的形式坚持办下去。我们希望“统计之都”在专业知识体系上有真正的王者风范，在面对用户需求时却又以谦恭的态度为大家服务。

---

\*统计之都网址：<http://cos.name/>

## 艾瑞咨询简介

艾瑞咨询集团（iResearch）是一家专注于网络媒体、电子商务、网络游戏、无线增值等新经济领域，深入研究和了解消费者行为，并为网络行业客户及传统行业客户提供市场调查研究和咨询服务的专业研究机构。公司旨在通过优质的研究咨询服务及可量化的数据产品，帮助客户提高对中国新经济的认知水平、盈利能力和综合竞争力，并推动中国新经济行业的整体发展。

艾瑞咨询集团（iResearch）秉承专业、严谨、客观的工作作风，拥有国内人数最多、最专业的网络行业研究团队，提供权威的行业数据服务，旨在发展成为新经济时代规模最大、客户口碑最好，并持续领先的新经济专业研究机构。

艾瑞咨询集团（iResearch）拥有多个互联网行业的信息资讯网站，为行业人士提供及时的互联网行业信息，主要包括：

艾瑞咨询集团：<http://www.iresearch.com.cn/>

艾瑞网：<http://www.iresearch.cn/>

艾瑞广告先锋：<http://www.iresearchad.com/>

网络媒体精品推荐：<http://www.iwebchoice.com/>

## Mango Solutions简介

Mango Business Solutions Ltd.（简称Mango Solutions或Mango）2002年成立于英国，主要为客户提供定制化的数据分析解决方案、培训及定制的软件产品。公司总部位于英国，业务遍及美国及欧洲。10年来为各行各业的公司、政府部门与学术机构成功地提供了咨询服务，客户中既包含辉瑞制药、诺华、谷歌、索尼这样家喻户晓的跨国企业，也有许多本地的中小型企业与机构。

Mango的团队由建模师，统计分析师与软件工程师组成，既可以为客户解决数学、统计、建模方面的特定问题，如提供数据分析方案、设计与执行数据分析IT平台的搭建，也可以为客户定制开发与数据分析相关的软件工具，也可以提供相关工具的培训，如R, S-PLUS等。

Mango Solutions的中文网站: [www.mango-solutions.com.cn](http://www.mango-solutions.com.cn)

英文网站: [www.mango-solutions.com](http://www.mango-solutions.com)

新浪微博: [weibo.com/mangosolutions](http://weibo.com/mangosolutions)

# 第五届中国 R 语言会议（上海会场）

## 会议指南

### 1. 日程安排

11 月 3 日	注册和报告	上海财经大学，行政楼一楼报告厅
11 月 4 日	报告和讨论	上海财经大学，行政楼一楼报告厅

### 2. 会议议程

演讲	时间
<b>11 月 3 日</b>	
上午（地点：行政楼一楼报告厅）	
周勇（上海财经大学统计与管理学院 常务副院长）：会议致辞	9:00~9:15
冯兴东：R 语言和其他计算机语言的混合编程	9:15~9:50
刘思喆：R 语言环境下的文本挖掘	9:50~10:25
拍照、茶歇	
张翔：短文本分类器与电商品类数据挖掘	10:45~11:15
沈羽、周春英：R 语言在 eBay 搜索引擎反馈与测试中的应用	11:15~11:45
午餐时间	
下午（地点：行政楼一楼报告厅）	
<b>lightning talks：</b> 1. COS（统计之都）社区 2. TopGeek 社区 3. Springer 4. 西安交通大学出版社 5. Mango Solutions 6. eBay 7. 沃尔玛电子商务 8. 传立媒体 9. 美库尔商务信息咨询 10. 思博奥科	13:30~14:30
周扬：基于 R 的地理信息系统	14:30~15:00
茶歇	
肖凯：twitteR 包入门和应用	15:20~15:50
陈钢：系统生物学和转换医学中的 R 语言	15:50~16:20
杭兴宜：R / Bioconductor 在生物多维组学数据整合中的应用	16:20~16:50
第一天会议结束	

演讲	时间
<b>11 月 4 日</b>	
上午（地点：行政楼一楼报告厅）	
陈逸波：基于 RHadoop 的关联规则挖掘	9:00~9:25
李忠：R Case Study from EBAY APD	9:25~9:50
洪健飞：啤酒与尿布的当代版——商品分析在电子商务中的应用	9:50~10:10
茶歇	
廖明：营销分析模型及其在广告界的应用	10:30~10:50
肖嘉敏：网络用户浏览路径分析	10:50~11:15
专题讨论：电子商务与市场营销	11:15~12:00
午餐时间	
下午（地点：行政楼一楼报告厅）	
刘成昊：模型预测的利器——随机森林	13:30~14:00
王雨晨：R 与面向对象统计分析	14:00~14:30
魏太云：R 基础作图与可重复研究	14:30~15:00
专题讨论：R 语言的可视化及可重复研究	15:00~15:30
会议结束	

### 3. 会议机构

**主办单位：**

上海财经大学统计与管理学院

**协办单位：**

统计之都 (<http://cos.name/>)

**赞助单位：**

艾瑞咨询

Mango Solutions

上海财经大学“数理经济学”教育部重点实验室

上海市高校一流学科（B 类）“统计学”

**会议主席：**

王绍立 陈丽云 张翔 李舰

# 上海财经大学 校区分布图

北

R 会议地址



# 演讲摘要



# R 语言和其他计算机语言的混合编程

冯兴东<sup>1,\*</sup>

<sup>1</sup> 上海财经大学统计与管理学院

## 摘要

在我们进行数据模拟的过程中，大量的循环是经常出现的事情。R 语言解读命令的方式经常导致一次循环所需时间越来越长，使得数据模拟变成一个十分耗时的过程。

万幸的是，R 通过提供接口的方式可以实现与其他计算机编程语言（比如C语言、Fortran语言）的混合编程来提高运行效率。同时，在编写C语言或者Fortran语言的过程中，我们也能方便地调用一些R的现成函数，比如产生不同分布的随机数等等。我们将利用这次机会系统地介绍这一混合编程技术，希望能对大家未来的工作中提供有益的帮助。

---

\*电子邮件：[feng.xingdong@mail.shufe.edu.cn](mailto:feng.xingdong@mail.shufe.edu.cn)；上海财经大学统计与管理学院，副教授。

# R 语言环境下的文本挖掘

刘思喆<sup>1,\*</sup>

<sup>1</sup> 统计之都

## 摘要

文本挖掘被描述为“自动化或半自动化处理文本的过程”，包含了文档聚类、文档分类、自然语言处理、文体变化分析及网络挖掘等领域内容，是一种对非结构化数据的处理的有效手段。本报告将系统地介绍以R语言为主要工具的文本挖掘技术以及几个有趣的应用案例。

关键词 Text Mining; Classification; Networks; LDA

---

\*电子邮件：[sunbjt@gmai.com](mailto:sunbjt@gmai.com)；主页：<http://bjt.name>；新浪微博：[@刘思喆](#)

# 短文本分类器与电商品类数据挖掘

张翔<sup>1,\*</sup>

<sup>1</sup> 艾瑞咨询

## 摘要

电子商务和网路视频都是快速发展的行业，大量的商品上架和视频发布带来了极大的信息量。但这些信息的分类程度非常低，条目短，不论人工或是常规文本分类方法都难以分析。应对这种短文本分类的情况本文从实践的角度给出了一种可行的解决方法和初步成果。其中主要运用基于词频的贝叶斯方法，但面对不同情况有很大的优化空间。

**关键词** 文本挖掘；文本分类；电子商务；贝叶斯

---

\*电子邮件：[birdzhangxiang@gmail.com](mailto:birdzhangxiang@gmail.com)

# R 语言在eBay搜索引擎反馈与测试中的应用

沈羽<sup>1,\*</sup> 周春英<sup>2,†</sup>

<sup>1</sup>eBay

<sup>2</sup>eBay

## 摘要

### 1. 用户反馈挖掘—R 文本挖掘应用一文本分类

ebay构建了一个用户意见反馈渠道。每天鼓励用户在站点上发表他们的意见和建议。对于产品部门来说，从大量用户反馈中抽取有价值的信息并更新和改进eBay平台，是一个非常有价值的实践。

在“用户反馈挖掘—R 文本挖掘应用”的讲演中，演讲者将会介绍eBay在实践中抽取用户意见的步骤和实施过程。以及 R 语言、R 工具包及 R 平台在这一过程中的应用。

### 2. R 在试验分析中的应用—Confidence Interval

ebay一直以来都非常重视技术的创新，并坚信试验是产生创新的源泉。为此，ebay搭建了基于真实用户流量的试验平台，不仅支持搜索 / 排序的算法试验、还包括搜索前端UI、广告等不同方面的试验。以搜索算法试验为例，每年都能实现近5%的交易额的增长。交易、用户行为等指标被用来衡量试验的好坏。但在试验评估中，往往遇到一个问题：被测试算法在某个指标上的提高是否具有统计意义，是否仅仅是偶然性的？

演讲将介绍 R 在ebay试验分析中的应用，特别是在评估指标的样本分布模拟、建模和置信区间计算上的应用。具体内容包括在应用过程中 R 语言、R 包、R 平台的使用方法。最后，介绍下 R 与海量数据分析的结合应用，以解决ebay交易量的海量性所带来的问题。

关键词 用户反馈；文本分类；特征抽取

---

\*电子邮件: [yshen1@ebay.com](mailto:yshen1@ebay.com)

†电子邮件: [chunzhou@ebay.com](mailto:chunzhou@ebay.com)

# 基于 R 的地理信息系统

周扬<sup>1,\*</sup>

<sup>1</sup>Mango Solutions

## 摘要

GIS（Geographic Information System，地理信息系统）作为一种将地理信息与数据信息综合展示的平台，在工业界与学术界有着极为广泛的应用和需求。R 作为一种高效的、面向统计和数据、提供和拥有丰富接口的编程语言，在数据多样化处理与展示需求较高的GIS 中有着十分重要的应用。芒果商务咨询在超过十年的数据分析与 R 语言开发过程中，积累了丰富的GIS 和数据可视化经验。本次报告将以GIS作为主线，介绍目前使用较为广泛和成熟的Google Earth（谷歌地球）+ R 技术，以及将来可能成为GIS重要组成部分的 R/ HTML5 技术路线。

关键词 GIS; R; KML; HTML5

---

\*电子邮件：yzhou@mango-solutions.com；主页：<http://yzhou.org/>；新浪微博：[@yummy\\_zhou](#)

# twitteR包入门和应用

肖凯<sup>1,\*</sup>

<sup>1</sup> 上海万达信息

## 摘要

twitter是一种很受全球网络用户欢迎的微型博客服务。由于信息量庞大，不论从学术还是商业的角度，twitter数据都有巨大的分析价值。本报告介绍了如何用R语言的twitteR包来调取twitter数据，并且用四个例子展示了对twitter数据的初步分析结果。

关键词 twitter, 社交网络

---

\*电子邮件：[xccds1977@gmail.com](mailto:xccds1977@gmail.com)；新浪微博：[@xccds](#)

# 系统生物学和转换医学中的R 语言\*

陈钢<sup>1,†</sup>

<sup>1</sup>深圳华大基因研究院

## 摘要

一台3年前的HiSeq 2000测序仪每天大概产生60GB的数据，在华大基因有128台这样的机器在全天候运行，平均每天产生7.6TB的数据。而且正在进行的升级计划完成后，每天产生的数据还要翻一倍，达到15TB。对这些数据进行有效的分析能推动我们对生命的认识，能让我们踏上攻克疾病的征途，有些甚至就能决定一个生命的去留。海量数据的分析是这其中最大的困难，我们将分享R 语言正在如何帮助我们逾越这个障碍，进而推动基础生命科学研究及其在医学健康上的应用。

---

\*R in Systems Biology and Translational Medicine

†电子邮件：[chengangcs@gmail.com](mailto:chengangcs@gmail.com)；主页：<http://gossipcoder.com>

# R /Bioconductor在生物多维组学数据整合中的应用

杭兴宜<sup>1,\*</sup>

<sup>1</sup>思博奥科生物信息科技（北京）有限公司

## 摘要

生命体由无数个细胞组成，而细胞的遗传奥秘遵循着“中心法则”进行传递，由双螺旋的DNA，解链转录RNA，然后再翻译成为蛋白质，由此产生了丰富多样的多层次生物调控机制。随着各种高通量组学技术的发展，目前生物学及生物信息学面临着多层次复杂海量数据的挑战。利用R 和Bioconductor我们将如何开发整合分析工具加以应对？

**关键词** 生物信息学；生物组学；数据整合

---

\*电子邮件：[xingyi.hang@sysbiomics.com](mailto:xingyi.hang@sysbiomics.com)

# 基于RHadoop的关联规则挖掘

陈逸波<sup>1,\*</sup>

<sup>1</sup>杭州数云信息技术有限公司

## 摘要

CRM（客户关系管理）系统是零售企业重要的业务支撑平台，能够帮助企业实现高效而精准的客户服务及市场营销的工作。

先进的CRM系统通常会部署丰富的数据挖掘模块，本报告介绍了其中关联规则挖掘模块的基本模型及主要算法，并描述了我们用以实现分布式关联规则挖掘的R+ Hadoop的策略。

关键词 CRM; 关联规则; RHadoop

---

\*电子邮件: chenyibo.zjumath@gmail.com; 新浪微博: @波波头一

# R Case Study from EBAY APD

李忠<sup>1,\*</sup>

<sup>1</sup>eBay

## 摘要

随着大数据时代的到来, 如何有效的利用大数据来挖掘出对商业有价值的信息变的越来越重要. 针对这个问题, 本次演讲总结了两个用 R 来挖掘大数据的应用案例.

案例1: 由于电商平台的网速变的越来越重要, 直接关系到用户体验和网站的转化率, 本案例用 R 分析了EBAY电商平台在不同时段,不同国家,不同浏览器,不同网页和不同的网页标签网速性能, 对网站优化有着重要的指导意义.

案例2: 由于电商平台的买家来自于世界各地,所以对买家分类以便于个性化服务就变的越来越重要, 本案例用 R, Python 和 kmeans++ 算法对EBAY美国的买家做了一个分类尝试, 并给出了分类结果和潜在的个性化服务商机.

---

\*电子邮件: zholi@ebay.com

# 啤酒与尿布的当代版：商品分析在电子商务中的应用

洪健飞<sup>1,\*</sup>

<sup>1</sup> 沃尔玛电子商务产品管理部

## 摘要

数据分析之前以及指标体系的搭建。

具体的案例，包括商品生命周期、商品价值分析、礼券使用率分析、品牌地图。

---

\*沃尔玛电子商务产品管理部，数据分析与挖掘副总监。毕业于复旦大学和中国社会科学院。有5年B2C行业数据分析经验。工作经历包括：易观国际，从事电信行业分析工作。当当网，从事用户行为分析和挖掘。24券，从事BI服务系统建设和数据分析挖掘。沃尔玛电子商务，从事BI服务系统建设和数据分析挖掘。

# 营销分析模型及其在广告界的应用

廖明<sup>1,\*</sup>

<sup>1</sup> 传立媒体 (Mindshare)

## 摘要

1895年，John Wanamaker 感慨到“我一半的广告被浪费了，问题是我不知道是哪一半”。100多年过去了，大多数广告主依然为之困惑。幸运的是，随着各种营销数据的丰富和统计模型的发展，商业决策者可以借助营销分析模型去找出市场的驱动力，衡量投资回报率，并做出最优的科学决策。

本报告主要介绍统计学模型在营销和广告界的主流应用，包括跨媒体优化，产品组合优化，定价，促销策略等。报告同时也讨论这些问题的复杂性和挑战，以及R可以帮助研究人员面对这些挑战。

**关键词** 市场营销；广告；投资回报率；跨媒体研究；非线性模型；动态模型；贝叶斯模型

---

\*电子邮件：[mliao1@yahoo.com](mailto:mliao1@yahoo.com)；廖明毕业于Duke大学，获统计与决策科学博士。曾在尼尔森，联合利华等公司从事营销分析和消费者洞察工作。现在传立媒体任董事总经理，领导营销策划部门，为各行业客户提供分析及战略咨询服务。

# 网络用户浏览路径分析

肖嘉敏<sup>1,\*</sup>

<sup>1</sup> 艾瑞咨询集团

## 摘要

基于连续性网络用户行为数据，深度分析和挖掘目标用户的网络行为特征。涉及网站分析，用户行为分析。

通过日志文件还原用户浏览路径，转化为时间序列，利用序列分析方法探索用户上网模式，并对用户上网模式进行分类。其中包括序列可视化，相似性度量，以及共同模式的识别。在分析过程中提炼分析方法，进行 R 与 SQL 方法对比，并应用于其它案例，如 QQ 群分析。

---

\*电子邮件：jiaminshaw@gmail.com

# 模型预测的利器：随机森林

刘成昊<sup>1,\*</sup>

<sup>1</sup>浙江大学计算机学院

## 摘要

随机森林在回归和分类问题上有很好的预测性能并且十分通用，被广泛的应用在数据挖掘的各个领域。此次演讲将介绍随机森林的核心思想，结合不同领域数据的特点，以及真实数据存在的一些问题（数据缺失、数据失衡），探讨如何提高随机森林的性能。

---

\*电子邮件：[twinsken@gmail.com](mailto:twinsken@gmail.com)

# R 与面向对象统计分析

王雨晨<sup>1,\*</sup>

<sup>1</sup>华东师范大学金融与统计学院

## 摘要

报告介绍了R中的三个面向对象系统S3、S4和R5。S3并不遵循严格的面向对象方式，而是依赖一系列命名规则来实现基于泛型函数的面向对象功能。由于R中大部分的统计分析功能用S3实现，所以理解S3系统的基本规则有助于对R中统计功能的理解。报告举例说明了如何利用S3对象来实现更简单和准确的统计分析。S4和R5是更严格的面向对象系统，提供了类似JAVA的更“主流”的面向对象的功能。报告主要介绍其特色和基本语法，并举例说明这三个系统之间的不同。

关键词 面向对象；S3；S4；类；方法；泛型函数

---

\*电子邮件：ycwang0712@gmail.com；新浪微博：@Teary\_技术宅

# R基础作图与可重复研究

魏太云<sup>1,\*</sup>

<sup>1</sup> 中国人民大学

## 摘要

### 1. Base图形：

直方图、条形图、茎叶图、散点图、平滑散点图、散点图矩阵、饼图、玫瑰图、箱线图、二维箱线图、小提琴图、条件密度图、等高图、颜色等高图、三维透视图、四瓣图、颜色图、矩阵图、热图、社会网络图、马赛克图、星状图、蛛网图、雷达图、脸谱图、平行坐标图、调和曲线图、地图。

### 2. knitr介绍：

knitr介绍，包括编辑器配置、基本语法、图片、代码、动画输出参数等。

案例展示，包括 `LATEX+ knitr -> pdf, markdown + knitr -> Html`

---

\*电子邮件：[weitaiyun@gmail.com](mailto:weitaiyun@gmail.com)

宣传页



## COS（统计之都）活动回顾

统计之都（下文简称 COS）虽以网站和论坛起家繁荣，但是随着越来越多喜爱统计的朋友们加入，大家对于线下活动和书稿撰写翻译等等的需求也越来越旺。目前，COS 的线下活动从一年两次的 R 会议（春季北京、秋季上海），逐渐发展到沙龙、交流会、竞赛、讲座、培训等等。我们希望更多的新鲜血液可以就近加入 COS 的线下活动中。

COS 线下活动总结：

1. 中国 R 语言会议：目前已开展到第五届，分别在中国人民大学（北京）和华东师范大学、上海财经大学（上海）举行，一般为期两天。历届会议纪要和幻灯片共享都可以在 COS 主站上找到：
  1. 第五届中国 R 语言会议：<http://cos.name/chinar/chinar-2012/>
  2. 第四届中国 R 语言会议：<http://cos.name/chinar/chinar-2011/>
  3. 第三届中国 R 语言会议：<http://cos.name/chinar/chinar-2010/>
  4. 第二届中国 R 语言会议：<http://cos.name/chinar/chinar-2009/>
  5. 第一届中国 R 语言会议：<http://cos.name/chinar/chinar-2008/>
2. 北上广（深）三地沙龙：目前我们定期在北京、上海和广州深圳开展线下沙龙活动。不同于规模庞大的 R 语言会议，沙龙形式更为轻巧，注重讨论交流。北京的沙龙已经进行了三期，上海的沙龙进行了五期，而广深沙龙则刚刚完成第一期。  
参与方式：每期沙龙开始前将在微博和论坛上进行通知，同时我们将建立参与过沙龙的朋友们的邮件列表。  
新浪微博：@统计之都  
COS 论坛：[cos.name/cn/](http://cos.name/cn/)  
沙龙邮箱：[salon@cos.name](mailto:salon@cos.name)
3. 比赛、交流会、讲座：除了定期的活动之外，我们还将不定期举行或者和其他组织共同举办一些交流会和讲座。北京的交流会主要以中国人民大学为中心，而上海的则更偏重于沟通合作。目前，我们主办或协办过的活动包括：
  1. 数据挖掘竞赛
  2. 科普讲座：与科学松鼠会合作，分别在北京和上海联合举办过统计科普讲座——别让数字吓到你
  3. 校内交流会：不定期在学校内举办统计建模、统计实际应用、统计学出国等经验交流会，包括中国人民大学和南开大学等。
4. 培训：我们亦致力于 R 语言的普及，已有培训包括
  1. 七周七语言：与 Top Geek 合作开展七周七语言系列的 R 语言培训，由资深 useR 李舰主讲。
  2. Supstat 夏令营：2012 年 9 月，我们第一次尝试开展了为时两天的 Supstat 夏令营活动，由二十多名 useR 相互交流讨论，系统而深入的讨论了 R 语言的方方面面。

COS 撰写和翻译书籍：

1. 撰写  
现代统计图形，R 忍者秘籍，Play Econometrics with R
2. 翻译  
[The Art of R programming, RExcel, ggplot 2, R in a Nutshell, R in Action](http://www.cadenza.info/)

**Mango Solutions** 精于数据分析及数据相关的服务，为客户提供精心定制的培训、咨询、与软件，协助客户挖掘复杂统计分析的能量，来支持关键性的商业决策。我们的总部位于英国，在美国、瑞士、中国设有分支机构。

Mango 掌握数据分析行业的前沿技术，客户涵盖制药、金融、互联网、农业、食品等多种行业，积累了宝贵的技术经验与资源，竭力提供高质量的客户体验。

## 公司业务

- 数据分析方案咨询
- 统计编程
- R 语言技术支持
- 数据分析外包
- R 语言培训
- 软件开发

## 产品

### Navigator

基于 NONMEM 的数据处理及报告的系统。用户可以上传他们的 NONMEM 模型，在分布式的网格平台上运行这些模型，并对运行的结果生成自动化的图表报告。更方便的是，用户可以使用 Navigator 把这些图表报告自动插入到他们的 Word 文档中，或把它们单独下载为 Word 或 PDF 格式。

### ModSpace

Mango 开发的 ModSpace 系统为企业与组织的知识共享与管理提供了新的方案。ModSpace 是一个集知识共享、搜索、与版本控制于一体的新型知识管理软件。ModSpace 最初被制药公司用于他们的 M&S 团队，用于药物研发中的统计模型的共享与管理，但经过产品化的开发，它也可适用于多种通用类型的环境，比如公司内部文档的管理与共享。

### iNCAS

可以为初次临床研究确定安全并有效的药物剂量提供帮助。通过与 GSK 全球安全委员会的紧密合作，Mango 开发出一个能够集中储存实验数据，并能够以统一的各式展示的应用程序。通过这个应用程序，研究人员可以使用一个简单、统一的图形界面来对研究数据进行分析，以此来进行预测建模并出安全有效的药物剂量。

## 客户



### 英国（总部）

地址：2 Methuen Park, Chippenham, Wiltshire, UK  
电话：+44 - (0)1249 - 705 450  
电子邮件：info@mango-solutions.com

### 中国（分公司）

地址：上海徐汇区肇嘉浜路 1065 号飞雕国际大厦 1607C  
电话：+86 - (0)21 - 5178 1325  
电子邮件：fshao@mango-solutions.com

## 服务

### 金融与保险

金融、保险公司与机构极度依赖基于高质量数据做出的快速反应。Mango 帮助金融与保险客户完成各种统计算法及应用程序。我们的服务包括小至为 Excel 的使用提供便利的咨询服务，到大型企业级软件的开发。

### 制药

新药研发是一项需要大量时间与劳动的工作。一种新药从研发到投放市场可能要经历二十年的时间、耗资几十亿美元。Mango 为制药与生物科技公司提供相关的服务与技术，可以帮助他们减少新药研发所需的时间。我们所提供的帮助包括软件系统、咨询、培训等。

### 能源

能源公司使用多种不同的统计技术来流程化他们的工作活动，以提高效率，增加投资回报。从管道分析，到能源交易方案，Mango 使统计学的力量为能源公司所用。

### 医学图像分析

Mango Imaging 为 MRI (Magnetic Resonance Imaging) 与 PET (Positron Emission Tomography) 研究提供全面的量化图像分析与统计推理服务。我们为不同的形态及各类的实验建立了不同的数据分析管道，并应用紧密跟踪的质量监控手段。

### 食品与感官分析

感官分析由应用统计分析方法与实验设计组成。消费品测试的结果会被逐条记录，随后据此找出有意义的数据，并做出推论。Mango 开发了 R 与 S PLUS 的应用程序，使研究人员能够通过简单的菜单点击就可以实现复杂、可定制的统计分析。

### 客户关系管理

Mango 可以利用统计分析技术帮助市场推广组织与部门使他们的市场活动更加有效与经济。我们使用复杂的统计算法来识别最有利润的目标客户，以此来节省大量的时间与人力。

### 其他行业

统计与数据分析可以应用在众多行业。我们在互联网、快速消费品、农业研究等众多领域都有诸多案例。如果您觉得我们可能对您的业务有所帮助，请尽快与我们联系。我们非常乐意与您探讨可能的解决方案。

### R 语言培训与咨询

我们是世界上最大、最专业的 R 语言服务提供商。我们为各种公司与机构提供 R 语言培训与技术咨询、支持。为企业定制的培训课程受到众多客户的好评。





Start from 2005

**Series Editors:**

Robert Gentleman  
Kurt Hornik  
Giovanni Parmigiani

**Printed book**

Never out of stock with  
Print-on-Demand

**eBook**

Available from libraries offering  
Springer's eBook Collection, or for  
individual purchase via online  
bookstores.

A free preview is available on  
SpringerLink.

► [springer.com/ebooks](http://springer.com/ebooks)

**Kindle edition**

**Publishing Editor:**

Michelle Feng,  
Springer Beijing Office  
[michelle.feng@springer.com](mailto:michelle.feng@springer.com)

中国•北京

海淀区北四环西路 58 号  
理想国际大厦 408-412 室  
邮编 100080  
电话 +86 10 8267 0211 分机 875  
[www.springer.com.cn](http://www.springer.com.cn)

## The 5th China-R Conference Shanghai, Nov. 2012

TITLES DISPLAY	YEAR
Interactive and Dynamic Graphics for Data Analysis	2007
Bayesian Core: A Practical Approach to Computational Bayesian Statistics	2007
Applied Spatial Data Analysis with R	2008
Applied Econometrics with R	2008
Wavelet Methods in Statistics with R	2008
Statistical Methods for Environmental Epidemiology with R	2008
Lattice	2008
Bayesian Computation with R	2009
Applied Statistical Genetics with R	2009
R Through Excel	2009
ggplot2	2009
Introduction to Probability Simulation and Gibbs Sampling with R	2010
An Introduction to Applied Multivariate Analysis with R	2011
Data Mining with Rattle and R	2011
Six Sigma with R	2012
Graphical Models with R	2012
Modeling Psychophysical Data in R	2012
Biostatistics with R	2012
Behavioral Research Data Analysis with R	2012
Competing Risks and Multistate Model with R	2012

Find More Books in Use R! Series and Book Intro  
[on springer.com/series/6991](http://springer.com/series/6991)