

## R语言的前世今生(转)

5月前 阅读(17) 原文 [违法举报](#)

最近因病休养在家,另外也算是正式的离开Snack Studio.终于有了大把可以自由支配的时间.可以自主的安排.最近闲暇的时间总算是恶补了不少前段时间行业没有时间关注的新事物.看着行业里引领潮流的东西,除非为数不多的天才能够创造得出,估计大多数人跟我一样,看着这些目不暇给的潮流总是觉得多少有些不知所措.但这并不妨碍从一个学习者的角度来了解这些新东西.类似本篇即将要谈到的R语言.

当一旦涉及一个新的领域.或一门新的语言.总是很容易陷入一些具体细节中而无法从更高的角度看到一门语言形成的背后的行业的背景、以及解决现实中存在的问题和适用的场景 以及未来的发展趋势.这点的缺乏也会导致我们看不清设计一门新的语言的作者的初衷是什么?类似我最近实践R语言时.当我很快实现一个简单的Demo原型后.虽然有些成就感但依然还是一知半解.所以本篇不谈R的技术细节.就来聊聊R语言背后的一些故事.

R语言是什么?

R是一种适用于统计分析计算和图像处理的语言.受S语言和Scheme语言影响发展而来.早期R是基于S语言的一个GNU项目,所以也可以当作S语言的一种实现,通常用S语言编写的代码都可以不作任何修改的在R环境下运行。R的语法是来自Scheme.

R语言如何而来的?

R本来是由来自新西兰奥克兰大学的Ross Ihaka和Robert Gentleman开发.[因两人名字都是以R 开头 所以也因此形象称为R]



PS: [左为Ross Ihaka 右为Robert Gentleman]

其实在最早的时候,R语言还没有发明之前,Ross Ihaka从一本书中了解到了Scheme语言.Ross 对Scheme 非常感兴趣.大约与此同时,他正好获得一版新S语言[后面说明]的源代码.此时他注意到Scheme和S语言二者之间的异同点,后来Ross开始准备用Scheme向别人演示词法作用域.但由于手边没有Scheme的书,就采用S语言来演示结果却失败了.由此让他萌生了改进S语言的想法.

在相当长的一段时间之后,Ross Ihaka[[链接为个人站点](#)]和Robert Gentleman[[链接为个人站点](#)]在奥克大学成了同事.他们都对统计计算十分感兴趣.而且试图为他们的实验室寻找一个更好的软件.因为在商业软件

中一直找不到他们想要的这是Ross 联想到当初S语言后自己一直一直想改进S语言的想法.于是他们决定基于此自己开发一种语言.这就是两人合作后才产生R语言.

可以说S语言就是R的father.S语言时一门用来数据分析和图形化的高级语言.早在1998年美国计算机协会ACM就授予了S语言的主要设计者John M.chambers“软件系统奖”用来表彰S取得的成就.这是迄今为止众多统计软件中“唯一”被ACM授奖的统计系统。当时ACM是这样评价S语言的： 永久地改变了人们分析、可视化、处理数据的方式； 是一个优雅的、被广泛接受的、不朽的软件系统。

**历年ACM授予软件系统奖的列表，这些优秀的软件系统与我们的生活息息相关：**

**1983年， Unix**

**1986年， TeX**

**1989年， PostScript**

**1991年， TCP/IP**

**1995年， World-Wid-Web**

**1997年， Tcl/Tk**

**1998年， S**

**1999年， The Apache Group**

**2002年， Java**

而Scheme语言时LISP语言一个方言或者说成是一个变种.与其他LISP不同的是Scheme是可以编译成机器码的.Scheme诞生于1975年的MIT，对于这个有着近30年历史的编程语言来说.在国外得到广泛应用.就像国内的C++/Java那样收到商业领域的青睐.Scheme的主要特征是可以像操作数据一样操作函数调用.其主要目的是训练人的机器化思维.以其简洁的语言环境和大量的脑力思考而著称.而后期演化的R正式参考了Scheme的语法.

在1993年.Ross和Robert将R的部分二进制文件放到卡耐基.梅隆大学统计系的Statlib中.并在S语言的新闻列表上发布了一个公告.随后有人开始下载使用并提出一些反馈.其中以苏黎世理工学院的MartinM最为突出,Martin在邮件中极力劝说两位原作者公布源代码.让R成为自由软件.终于在Martin劝说下于1995年6月R预言的源代码正式发布到自由软件协会的FTP上.

同年S语言的许可证被MathSoft公司买断，S-PLUS成为其公司的主打数据分析产品。由于S-PLUS继承了S语言的优秀血统，所以被世界各国的统计学家广泛使用.

随着R语言的进一步开发.程序版本的归档又成了一个问题.维也纳工业大学的Kurt Hornik承担了这个任务.在维也纳建立了R程序的归档.这使得程序版本的发布变得更加规范.同时世界各地也出现了R程序的镜像.类似Statlib等.

随着时间推移.于1997年中期R核心团队正式成立. 包含11位早起成员.包括现在R语言版本依然还是由“R开发核心团队”负责开发.截止到目前我在写这篇文章时也就是2013年的现在.R核心团队已经达到20人.成员主要来自世界各地的大学如牛津大学.加拿大西安大略大学等.也有来自企业的成员.比如AT&T实验室的Simon Urbanek等.由于R语言自身扩展性非常强.随着发展和使用使用人数增多.因此也吸引了大量用户编写的自定义的函数包供更多人使用.这些附件包可以冲世界各地的CRAN镜像上下载.

截止到目前.R语言源程序已经更迭了超过了70以上的版本.目前最新版本是3.0.0. 而源程序大小也有1997年R核心团队成立时的959K增加到今天的51.5M[Windows版本].从版本更新和文件大小来看.R的发展速度的确非常之快.而且整个软件体系一直都保持者非常小的优势.这也几乎是任何一门商业软件都无法比拟的.

而随着R语言的发展, 其实就在R语言开放源码的1993年, S语言的许可证被MathSoft公司买断, S-PLUS成为其公司的主打数据分析产品。由于S-PLUS继承了S语言的优秀血统, 所以被世界各国的统计学家广泛使用。但好景不长, 1997年R语言正式成为GNU项目, 大量优秀统计学家加入到R语言开发的行列。随着R语言的功能逐渐强大, S-PLUS的用户渐渐地转到了同承一脉的R语言。S语言的发明人之一——John M. Chambers最后也成为R语言的核心团队成员。S-PLUS这款优秀的软件也几经易手, 最后花落TIBCO公司.

现金R的邮件列表依然还是有苏黎世理工学院的Martin Macher提供支持.这些邮件列表主要包含四大类:R-announce 消息发布、R-help附加包、R-help帮助、R-develR程序的开发.

R除了官方文档外.还有创办于2001年的一个刊物R News. 该刊物主要用来介绍R的最新特征.CRAN的动态.附加包的说明短文、编程技巧、手册和FAQ中没有介绍的小提示以及R在数据分析中应用实例.自2009年R News更名为The R Journal.

虽然说R语言诞生于新西兰.但后来的服务器架设到奥地利.而现在Windows 主程序的维护者在加拿大.而windows 附加包维护者在德国.Mac os版本维护者在美国.邮件列表维护者在瑞士..就是这样的一个形式松散却有着共同目标的群体.数十年间以志愿者的身份坚持不懈的推送R语言的发展.众多统计学或相关领域的程序员也贡献自己的力量.讲大量统计方法以附加包的形式发布出来.使其他不擅长变成的用户也可以以最快的速度运用上最新的统计方法.而封闭源代码的商业源代码则很难有此这样推送力量.R语言发展到今天正式开源社区不断推送下 组建完善并得到进一步发展. 并成就了R语言的今天.

其实R语言早已是又来已久.如果算上第一个正式版本.可见它的出现俨然比Java还要早.R确切的说是一门用于统计计算和作图的语言, 它不单是一门语言, 更是一个数据计算与分析的环境。统计计算领域有三大工具: SAS、SPSS、S, R正是受S语言和Scheme语言影响发展而来。其最主要的特点是免费、开源、各种各样的模块十分齐全, 在R的综合档案网络CRAN中, 提供了大量的第三方功能包, 其内容涵盖了从统计计算到机器学习, 从金融分析到生物信息, 从社会网络分析到自然语言处理, 从各种数据库各种语言接口到高性能计算模型, 可以说无所不包, 无所不容, 这也是为什么R正在获得越来越多各行各业的从业人员喜爱的一个重要原因.

而现在随着数据呈现几何数量级的增加.数据挖掘需求的增长而使R日益得到普及.它虽源于S但其发展却远远地超过了S.google首席经济学家Hal Varian说: "R的最让人惊艳之处在于你可以通过修改它来做所有的事情, 而你已经拥有大量可用的工具包, 这无疑让你站在巨人的肩膀上工作。"R语言也在目前机器学习、统计计算、高性能计算得到广泛应用.大数据产品销售商, 比如数据仓库与 Hadoop 数据过滤器有

其喜欢R语言.R 进过调整, 以便 Hadoop 集群的每个节点都可以对 Hadoop 集群上存储在 Hadoop 分布式文件系统的数据进行本地 R 分析, 并对这些计算的结果进行整合, 类似 MapReduce 对非结构化数据的操作.

在2012年初Oracle 也加入了 R 语言行列.推出Advanced Analytics 工具, 作为 Oracle 数据库与 R 分析引擎之间的桥接.Advanced Analytics 是 Oracle 在其 11g R2 数据库中部署的 Data Mining 附件。当 R 程序员需要运行统计例程时, 他们可以在数据挖掘工具箱中调用等同的 SQL 函数, 并在该数据库中运行。如果没有这样的 SQL 函数, 遍历数据库节点 (如果为集群) 的嵌入式 R 引擎将运行 R 例程, 收集汇总数据并作为结果将其返回 R 控制台。

另外, Oracle为其Big Data Appliance提供了一个名为 R Connector for Hadoop 工具, 这是一个在 Oracle Exa x86 集群上运行的 Cloudera CDH3 Hadoop 境。该连接器可让 R 控制台与在Big Data Appliance 运行的Hadoop 分布式文件系统和NoSQL 数据库进行通信。

随着big data 大数据时代来临.R语言起源统计分析.现在服务于数据.未来也会逐渐随着数据分布渗透到各个行业...

**版权声明:** 本文内容来源于网络, 版权归原作者, 此博客不拥有其著作权, 亦不承担相应法律责任。文本页已经标记具体来源原文地址, 请点击[原文](#)查看来源网址, 站内文章以及资源内容站长不承诺其正确性, 如侵犯了您的权益, 请联系站长如有侵权请联系站长, 将立刻删除

## R语言的前世今生(转)-相关文章

### 1. Windows下利用虚拟机运行FSL 安装和配置

FSL是牛津大学FMRIB开发的用于分析功能磁共振影像的科研软件包。运行要求: windows7或vista操作系统, 10G硬盘空间, 4G内存。从FSL官网上下载Centos6的虚拟盘 (FSLVm6\_64bit.zip) 解压。用VMPlayer打开其中的FSLVm6\_64.vmx, 运行虚拟机。下载适用于CentOS6的FSL文件包 (fsl-5.0.4-centos6\_64.tar.gz) ,

### 2. 第十六届“二十一世纪的计算”学术研讨会 牛津大学肿瘤成像学教授Michael Brady主题演讲

Computing andHealthcare牛津大学肿瘤成像学教授Michael Brady主题演讲\\ title=\\第十六届“二十一世纪的计算”学术研讨会 牛津大学肿瘤成像学教授Michael Brady主题演讲\\大家好。我想首先告诉大家我的结论, 这样大家可以看手机了。计算科学已经在改变医疗行业, 但是它的潜力是无限

### 3. 【PgSQL安装 (含配置)】PostgreSQL简称PgSQL, 是1980以加利福尼亚大学开发的DBMS, 严格遵守标准SQL。

【下载地址】<http://www.enterprisedb.com/products-services-training/pgdownload#window>  
S..... 【记】端口号5432 【配置路径】PostgreSQL9.5datapostgresql.conf 【可选 学习环境配置】使用“listen\_addresses”作为

### 4. 牛津大学xDeepMind自然语言处理

<http://study.163.com/course/introduction/1004336028.htm>

## 5. 2020厦门大学综述翻译：3D点云深度学习（Remote Sensing期刊）

目录摘要1、引言：2、点云深度学习的挑战3、基于结构化网格的学习3.1 基于体素3.2 基于多视图3.3 高维晶格4、直接在点云上进行的深度学习4.1 PointNet4.2 局部结构计算方法4.2.1 不探索局部相关性的方法4.2.2 探索局部相关性的方法4.3 基于图5、基准数据集5.1 3D模型数据集5.2 3D室内

## 6. 计算机专业-世界大学学术排名,QS排名,U.S.NEWS排名

2015年美国大学计算机专业排名计算机专业介绍：计算机涉及的领域非常广泛，其分支学科也是非常多。所以在美国将主要的专业方向分为人工智能，程序应用，计算机系统(Systems)以及计算机理论(theory)这四个部分。每个部分又有很多的研究方向，诸如人工智能与知识工程，模

## 7. 世界名校网络课程大盘点,美国大学CS专业十三大研究方向,世界50所知名大学提供开放课程

世界名校网络课程大盘点 加州大学伯克利分校<http://webcast.berkeley.edu/> 加州大学伯克利分校与斯坦福大学、麻省理工学院等一同被誉为美国工程科技界的学术领袖，其常年位居泰晤士高等教育杂志全球大学排行前十名。作为美国第一的公立大学，伯克利分校提供了学校许多优

## 8. 语义Web和本体开发相关技术

在技术实现方面，语义Web和本体理论的研究日趋成熟，已经有许多成熟的工具或程序接口，诸如Jena、OWL API等API是系统实现的关键技术。这里介绍系统的可行性分析以及系统开发设计的关键技术。1 JenaJena[69]是来自于惠普实验室语义Web研究项目的开放资源，是用于创建语义Web应