《信息检索与数据挖掘》

2019 版

实验要求

实验报告要求

- 格式方面美观简洁即可,不需要严格按照常规实验报告撰写。
- 必须写学号、姓名;电子版。
- 实验报告中要能显示上机内容已做(可附关键代码);对于要求记录结果的问题,请在实验报告上记录结果;对于要求分析的结果,请给出文字分析。
- 实验内容均有编号,在实验报告中请保持编号信息。
- 报告提交方式:报告使用 docx 或 pdf 格式,提交时请将报告与相关实验程序源代码、
 工程文件打包命名为"信息检索与数据挖掘实验报告_姓名_学号"的格式并以此为邮件
 名称发送到 irdm2019@163.com

中国科学技术大学 • 电子工程与信息科学系

目录

目录		I
1. 实验内	容	1
1.1 Rì	语言学习	
1.2 用	R 语言做文本分析	1
1.3 用	R语言做数据挖掘	2
2. 附录 1:	: R 环境安装	3
2.1 基	车 R 环境安装	
2.2 镜·	像选择与包安装	
2.2.1	设定镜像服务器	
2.2.2	安装包(package)	
2.3 Wi	/indows 下的 R 语言编辑器	
附录 2: JAV	/A 有关	9
2.4 JDI	К 安装	9
2.4.1	安装包下载	9
2.4.2	环境变量设置	9
2.5 JDI	x、JRE、JVM 三者间的关系	
2.5.1	JDK (Java Development Kit)	
2.5.2	Java Runtime Environment (JRE)	
2.5.3	JVM (java virtual machine)	
附录: R 语言	言程序的发布	
R 语言做深度	度学习	
2.6 Rì	语言中神经网络的包	
2.6.1	neuralnet	
2.7 Rì	语言中深度学习的包	
2.8 GIT	THUB 十大深度学习项目	
2.9 De	EEP LEARNING MADE EASY WITH R : A GENTLE INTRODUCTION FOR DATA SCIENCE	14
参老文献		14
岁与 人献…		

1. 实验内容

1.1 R 语言学习

- (1) 建议阅读的资料:《R语言实战》(中文版[1]、英文版[2])。
- (2) 矩阵、数据框运算:掌握矩阵运算相关函数的用法。
- (3) 绘图:至少掌握柱状图、散点图、饼图等常规图形的绘制。
- (4) 基本的数据导入/导出:至少掌握导入 EXCEL 文件和 XML 格式数据文件的方法。
- (5) 学习使用编辑器 UltraEdit 或 WinEdt(即 CTex)来写 R 文件。

1.2 用 R 语言做文本分析

- (6) 建议阅读《R语言tm工具包进行文本挖掘实验》[3](与我们给出的示例代码 差别不大);《R语言环境下的文本挖掘---tm包》[4],有关于中文分词的部分;《R 和 Ruby 数据分析之旅》(中文版[5]、英文版 2nd[6])。
- (7) 用文本分析包 tm 创建词项-文档关联矩阵(采用 tm 自带的路透社 20 篇文档的 数据集)并分析。
 - i. 安装 tm 包,在 windows vista 及以后的版本,应该会被安装到 C:\Users\%user_name%\Documents\R\win-library\3.2 目录下(RStudio 安装后包的目录缺省应该是 C:\Program Files\R\R-3.2.3\library)。在该目录下我们可以找到 tm 子目录,在 tm\text\crude 子目录存放的是路透社语料库中选出的 20 篇文档。tm_example1.R、tm_example2.R 示例代码就是针对该子目录下 20 篇文档进行分析。
 - ii. 使用 "help(package="tm")" 查看 tm 包提供的函数; 使用 "vignette("tm")" 查看 tm 自带的说明文档。结合 tm_example1.R 示例代码,给出词项-文档关联(计数) 矩阵。
- iii. 统计数据集中不同词项的出现频度,据此绘制柱状图表示数据集中词项的 PDF, 并记录出现频率最高的 3 个词项。验证在该数据集上词项分布是否满足 Zipf 定律。
- iv. 绘制词项的 Tag 云图(使用包 wordcloud,参考 tm_example3.R 示例代码)
- v. 将上述词项-文档关联(计数)矩阵拓展为词项-文档 TF 矩阵。
- vi. 将上述词项-文档 TF 矩阵拓展为词项-文档 TF-IDF 矩阵。
- vii. 使用 "vignette("tm")" 查看 tm 自带的说明文档。结合 tm_example2.R 示例代码, 了解用 tm 包所支持的格式处理、停用词、词形归并、词干还原等功能。
 比较经过这些处理前后的词项-文档关联(计数)矩阵的变化(矩阵大小、系数程度等);统计出现频率最高的3个停用词在文档集中的出现频率。
- viii. 根据上面得到的词项-文档关联(计数)矩阵,按照布尔检索模型,查询"price AND oil"的返回结果是什么?
- ix. 根据上面得到的词项-文档关联(计数)矩阵,按照向量检索模型,查询"price

AND oil"的返回结果相关度最高的3个文档是什么(结果要排序)?

- x. 根据上面得到的词项-文档关联(计数)矩阵,按照二值独立概率模型,针对查询 "price AND oil",给出 20 个文档的 RSV 值排序结果。
- (8) 基于 tm 包, 取安然公司邮件数据集部分数据(自行决定数据集大小)创建词 项-文档关联矩阵并分析。
 - i. 参考《R 和 Ruby 数据分析之旅》一书第5章"对自己的电子邮件进行挖掘"的内容创建所选定部分数据集对应的词项-文档关联(计数)矩阵。
 - ii. 统计数据集中不同词项的出现频度,据此绘制词项的 Tag 云图(使用包wordcloud,参考tm_example3.R示例代码),并记录出现频率最高的3个词项。
- iii. 将上述词项-文档关联(计数)矩阵拓展为词项-文档 TF 矩阵。
- iv. 将上述词项-文档 TF 矩阵拓展为词项-文档 TF-IDF 矩阵。

1.3 用 R 语言做数据挖掘

- (9) 学习用 R 语言做数据统计与数据挖掘
 - i. 聚类
 - ii. 主成分分析 PCA
- (10) 针对包 tm 自带的路透社 20 篇文档的数据集进行分析。
 - i. 根据词项-文档关联(计数)矩阵进行聚类,记录其结果。
 - ii. 根据词项-文档 TF 矩阵进行聚类,记录其结果。
- iii. 根据词项-文档 TF-IDF 矩阵进行聚类,记录其结果。

2. 附录 1: R 环境安装

2.1 基本 R 环境安装

与起源于贝尔实验室的 S 语言类似, R 也是一种为统计计算和绘图而生的语言和环境, 它是一套开源的数据分析解决方案,由一个庞大且活跃的全球性研究型社区维护。相比于其 他流行的统计和制图软件,如 Microsoft Excel、SAS、IBM SPSS、Stata 以及 Minitab,开源 是最大的特点。。因为它的功能非常丰富,所以文档和帮助文件也相当多。另外,由于许多 功能都是由独立贡献者编写的可选模块提供的,这些文档可能比较零散而且很难找到。

R 可以在 CRAN (Comprehensive R Archive Network) http://cran.r-project.org 上免费下载。Linux、Mac OS X 和 Windows 都有相应编译好的二进制版本,即 RStudio。RStudio 安装,可以从 <u>https://cran.r-project.org/mirrors.html</u> 查找国内的镜像服务器。例如: <u>https://mirrors.tuna.tsinghua.edu.cn/CRAN/bin/windows/base/R-3.2.3-win.exe</u>。

运行 RStudio 后界面会自动打开一个 R Console,这个控制台和 matlab 的命令窗口类似, 输入已定义的函数即可运行。如,输入 help.start() 会打开帮助文档,输入 q() 则退出 RStudio。

🥂 RGui (64-bit)	-		×
文件编辑 查看 其他程序包窗口 帮助			
21 • • • • • • • • • • • • • • • • • • •			
R Console		×	
R version 3.2.3 (2015-12-10) "Wooden Christmas-Tree" Copyright (C) 2015 The R Foundation for Statistical Computing Platform: x86_64-w64-mingw32/x64 (64-bit)		^	
R是自由软件,不带任何担保。 在某些条件下你可以将其自由散布。 用'license()'或'licence()'来看散布的详细条件。			
R是个合作计划,有许多人为之做出了贡献. 用'contributors()'来看合作者的详细情况 用'citation()'会告诉你如何在出版物中正确地引用R或R程序包。			
用'demo()'来看一些示范程序,用'help()'来阅读在线帮助文件,或 用'help.start()'通过HIML浏览器来看帮助文件。 用'q()'退出R.			
>		*	
<	>	- 18	

图 2-1 RStudio 启动界面示意图

2.2 镜像选择与包安装

在 RStudio 安装后,可通过安装称为包(package)的可选模块(同样可从 CRAN 下载) 来增强功能。

2.2.1 设定镜像服务器

CRAN 镜像服务器可以通过 GUI 主菜单"程序包"→"设定 CRAN 镜像"选择。

R	
文件编辑 查看 其他	程序包 窗口 帮助
- 1	加载程序包
P	设定CRAN镜像
K	选择软件库
错误: 找不到对象,	安装程序包
> print	更新程序包
<pre>Iunction (x,) UseMethod("print" <butecode: 0x0d14#<="" pre=""></butecode:></pre>	从本地zip文件安装程序包

图 2-2 在 RStudio 中设定 CRAN 镜像: 主菜单的操作



图 2-3 在 RStudio 中设定 CRAN 镜像:选定镜像类别

HTTP CRAN mirror	
0-Cloud Algeria Argentina (La Plata) Australia (Canberra) Australia (Melbourne) Austria Belgium (Antwerp) Belgium (Ghent) Brazil (BA) Brazil (BA) Brazil (PR) Brazil (PR) Brazil (SP 1) Brazil (SP 1) Brazil (SP 2) Canada (BC) Canada (NS) Canada (ON)	A REAL PROVIDENCE AND
Chile China (Beijing 2) China (Beijing 3) China (Beijing 4) China (Xiamen) Colombia (Cali) Czech Republic Ecuador El Salvador Estonia France (Lyon 1) France (Lyon 2) France (Marseille) France (Montpellier) France (Paris 1) France (Paris 2) Germany (Berlin)	
Germany (G <u+00f6>ttingen) Germany (Münster) Greece Hungary Iceland India Indonesia (Jakarta) Iran Ireland Italy (Milano) Italy (Padua) Italy (Palermo)</u+00f6>	

图 2-4 在 RStudio 中设定 CRAN 镜像:选定一个国内的镜像服务器

如果在某个镜像服务器上找不到某个特定的包(package)可以换一个镜像服务器试试。

2.2.2 安装包 (package)

(1) 通过图像界面从 CRAN 镜像下载并安装包:

安装包可以在 GUI 主菜单"程序包"→"安装程序包"的图形界面中选择;也可以在"R Console" 中用"utils:::menuInstallPkgs()"调出对应的图像界面

R	R Console	
> chooseCRA > utils:::m 试开URL'htt Content typ downloaded	Nmirror() enuInstallPkgs() p://mirror.bjtu.edu.cn/cran/bin/windows/contr e 'application/zip' length 766972 bytes (748 748 KB	1b/3.2/rJava_0.9-8.z\$ KB)
程序包、rJav	a'打开成功,MD5和检查也通过	
F载的二进制	削程序包在	
C:\	Users\cxh\AppData\Local\Temp\RtmpEhhysU\downl	oaded_packages里
		×
		, K
R		
文件 🦸	编辑 查看其他 程序包 窗口 帮助	
	■ ■ ● ● ● ● ● ● ● ● ● ● ● ● ● ● ● ● ● ●	
		_
R	法 なか	
func	tion (x,) 女衣但于已… ethod ("print" 百姓纪克与	
 byt	ecode: 0x0d14	
<env< td=""><td>ironment: nam 从本地zip文件安装</td><td>程序包</td></env<>	ironment: nam 从本地zip文件安装	程序包
□错误	: 无法同'doc\html\rw-ra0.ntml'	JT11 X1ナ大琳
	riade	
	RJafroc	
	rjags	
	rJava	
	RJDBC	
	rje HIDSOCS	
	Rinstatella	
	RISDMY	
	rison	
	RISONIO	
	ristat	
	rivthon	
	rkafka	
	rkafkajars	~
	Trankajaro	
	福元 取消	

(2) 安装已经下载的压缩文件

如果包是通过其他途径已经下载到本地计算机,则可在 Windows 命令行控制台中输入指令进行安装。例如,

cd C:\Program Files\R\R-3.1.0\bin\i386



(3) 直接用 install.packages('package name') 的命令来安装:

这是最方便的方式,但是在使用 install.packages()命令前应该已经正确设置过 CRAN 镜像服 务器地址。一个包仅需安装一次。但和其他软件类似,包经常被其作者更新。使用命令 update.packages() 可以更新已经安装的包。

2.3 Windows 下的 R 语言编辑器

http://www.dataguru.cn/article-2064-1.html

(1) 很自然地就开始用 UltraEdit, 基本上 Windows 下的程序员都用这个, 对各种语法都支持得很好,对R 的支持当然也很不错。到其官网下载关于R 的 wordfile (http://www.ultraedit.com/downloads/extras.html),找到R Scripting (S/S+也行),说是下载r.uew 这个文件,其实只用把页面上的文字全部拷下来,再粘贴到 WORDFILE.UEW 这个文件中去就行,WORDFILE.UEW 的位置可以在"高级-> 配置-> 语法高亮"中查看其路径,然后进入该文件夹,打开这个文件,把r.uew 里的内容全部复制粘贴在最底部,注意最开始的/L20,最好是根据上一个编号改成紧接着的编号。保存 WORDFILE.UEW,重启 UltraEdit 就可以看到后缀为r 的文件中R 的语法都会高亮显示了,括号下面还会自动缩进。【CXH20160303,已验证可行】

(2) Windows 下另一个很常用的编辑器就是 EditPlus, 写脚本的人经常用它,而且韩国人的东西默认的高亮显示都是花花绿绿的很好看。同样也可以提供 R 的高亮显示支持,到其主页(http://www.editplus.com/others.html)下载一个r2.zip,解压后把里面的R.stx 拷到EditPlus的安装文件夹里(可以发现里面有很多.stx 后缀的文件,就是支持各类语言语法的文件),在 "工具 -> 参数选择 -> 文件 -> 语法"中,添加一个新的语法类型,起名为 R,后缀设为 r,在语法文件中选择刚才那个 R.stx 文件,确认保存,重启 EditPlus,就可以发现 R 语言能 高亮显示了。【CXH20160303, editplus 官网目前找不到 r2.zip】

(3) R-WinEdit: R 自带的编辑环境不好用,很不好用。还好,有一个包 R_WinEdt,让你在 WinEdt 环境中写好代码后,可以直接一键发送到 R 中执行。

下载 WinEdt, http://www.winedt.com/

R_WinEdt: http://cran.r-project.org/contrib/extra/winedt/

先安装 WinEdt,然后安装包 RWinEdt 即可。【CXH20160303,未验证】

附录 2: Java 有关

2.4 JDK 安装

2.4.1 安装包下载

有 32bits 的 JDK 和 64bits 的 JDK,不同版本 JDK 下载: http://www.oracle.com/technetwork/java/javase/downloads/jdk8-downloads-2133151.html

32bits 版本

http://download.oracle.com/otn-pub/java/jdk/8u73-b02/jdk-8u73-windows-i586.exe

64bits 版本

http://download.oracle.com/otn-pub/java/jdk/8u73-b02/jdk-8u73-windows-x64.exe

可以安装 32bit 的 jre,并且在环境变量里面把 JAVA_HOME 的值改成 32bit 的 jre 的安装路 径,就可以了。32bit 的 jre 和 64bit 的 jdk 是不会冲突的,他们默认安装在不同的目录下面, JAVA_HOME 指向哪个就会使用哪个。

2.4.2 环境变量设置

安装完 JDK 后需要配置环境变量。可以通过 Windows 图形化操作界面完成,也可以用 命令行程序完成[7] 。

● 图形化操作方式: 计算机→属性→高级系统设置→高级→环境变量



(1)系统变量→新建 JAVA_HOME 变量。
 变量值填写 jdk 的安装目录(本人是 E:\Java\jdk1.7.0)

系统变量→寻找 Path 变量→编辑 在变量值最后输入 %JAVA_HOME%\bin;%JAVA_HOME%\jre\bin; (注意原来 Path 的变量值末尾有没有;号,如果没有,先输入;号再输入上面的代码)

变 里 名(M):	Path
变 量值(V) :	<pre>:/e-Static;%JAVA_HOME%\bin;%JAVA_HOME</pre>
	确定 取消

(2) 系统变量→新建 CLASSPATH 变量

变量值填写 .;%JAVA_HOME%\lib;%JAVA_HOME%\lib\tools.jar(注意最前面有一点) 系统变量配置完毕

变量名(M):	CLASSPATH	
变 里 值(V):	<pre>\$</pre>	-V
	确定 取消	33

(3)检验是否配置成功 运行 cmd 输入 java -version (java 和 -version 之间有空格) 若如图所示 显示版本信息 则说明安装和配置成功。

ava version "1.8.0	ersion 73″			
ava(TM) SE Runtime ava HotSpot(TM) 64-	Environment (build 1.8 Bit Server UM (build 2)	.0_73-b02) 5.73-b02. mixed m	node)	

手工命令行设置环境变量(计算机重启后需要重新设置)
 在命令提示符下输入:

set JAVA_HOME=C:\Program Files\Java\jdk1.8.0_73

set CLASSPATH =%JAVA_HOME%\lib;%JAVA_HOME%\lib\tools.jar

2.5 JDK、JRE、JVM 三者间的关系

JDK (Java Development Kit) 是针对 Java 开发员的产品,是整个 Java 的核心,包括了 Java 运行环境 JRE、Java 工具和 Java 基础类库。Java Runtime Environment (JRE) 是运行 JAVA 程序所必须的环境的集合,包含 JVM 标准实现及 Java 核心类库。JVM 是 Java Virtual Machine (Java 虚拟机)的缩写,是整个 java 实现跨平台的最核心的部分,能够运行以 Java 语言写作的软件程序[8]。

Java各种集成开发工 (JCreator/Ecllpse/JBu	. 具 ullder,	/NetB	eans等)	
JDK(开发工具包)				
JRE(运行环境)运行时势	长库			
JVM(Java虚拟机)				
_{制图/playkid.cn} OS操作	F系	统		

2.5.1 JDK (Java Development Kit)

JDK 是 Java 开发工具包,是 Sun Microsystems 针对 Java 开发员的产品。JDK 中包含 JRE,在 JDK 的安装目录下有一个名为 jre 的目录,里面有两个文件夹 bin 和 lib,在这里可 以认为 bin 里的就是 jvm, lib 中则是 jvm 工作所需要的类库,而 jvm 和 lib 和起来就称为 jre。

JDK 是整个 JAVA 的核心,包括了 Java 运行环境 JRE (Java Runtime Envirnment)、一堆 Java 工具 (javac/java/jdb 等)和 Java 基础的类库 (即 Java API 包括 rt.jar)。

- ① SE(J2SE), standard edition,标准版,是我们通常用的一个版本,从 JDK 5.0 开始, 改名为 Java SE。
- ② EE(J2EE), enterprise edition, 企业版, 使用这种 JDK 开发 J2EE 应用程序, 从 JDK 5.0 开始, 改名为 Java EE。
- ③ ME(J2ME), micro edition, 主要用于移动设备、嵌入式设备上的 java 应用程序, 从 JDK 5.0 开始, 改名为 Java ME。



金字塔结构 JDK=JRE+JVM+其它 运行 Java 程序一般都要求用户的电脑安装 JRE 环境 (Java Runtime Environment);没有 jre, java 程序无法运行;而没有 java 程序, jre 就没有用 武之地。

2.5.2 Java Runtime Environment (JRE)

是运行基于 Java 语言编写的程序所不可缺少的运行环境。也是通过它, Java 的开发者 才得以将自己开发的程序发布到用户手中,让用户使用。

RE 中包含了 Java virtual machine(JVM), runtime class libraries 和 Java application launcher, 这些是运行 Java 程序的必要组件。

与大家熟知的 JDK 不同, JRE 是 Java 运行环境,并不是一个开发环境,所以没有包含 任何开发工具(如编译器和调试器),只是针对于使用 Java 程序的用户。

2.5.3 JVM (java virtual machine)

就是我们常说的 java 虚拟机,它是整个 java 实现跨平台的最核心的部分,所有的 java 程序会首先被编译为.class 的类文件,这种类文件可以在虚拟机上执行。

也就是说 class 并不直接与机器的操作系统相对应,而是经过虚拟机间接与操作系统交互,由虚拟机将程序解释给本地系统执行。

只有 JVM 还不能成 class 的执行,因为在解释 class 的时候 JVM 需要调用解释所需要的 类库 lib,而 jre 包含 lib 类库。

JVM 屏蔽了与具体操作系统平台相关的信息,使得 Java 程序只需生成在 Java 虚拟机上运行的目标代码(字节码),就可以在多种平台上不加修改地运行。

附录: R 语言程序的发布

http://www.zhihu.com/question/24417654

Shiny 很符合你所叙述的要求,它是由 RStudio 公司开发的一个针对 R 的交互式 web 应用。

不需要了解许多 HTML, CSS, JavaScript 方面的知识,一点点极为基础的就好,就可以把自己的 R 脚本变成一个 web 上的应用。比如你写了一个预测感冒的算法,通过 shiny 放到了 web 上,别的用户就可以上传他们自己的数据,来预测感冒或类似的疾病。官网上有很多简单的例子,样子大概是这样的。

yhat 同样是一个很 cool 的工具,你可以把你的 R function 或者某种算法放在网上, yhat 就会提供一个 API 给所有人,别人就可以通过这个接口使用你的函数,载入自己的数据得 到相应的结果。不管你是数据分析师,工程师也好,都会觉得这样的工具很实用,有开源的 帮助对大家都是一件好事。同样,你如果不会建网站,不懂网页前端的知识也没有关系。

另外提一下 OpenCPU, 同样能实现上述功能, 不过更加灵活, 但对 web programming 的要求就要高一些。

coursera 上的一门 data product 的课程详细介绍了 shiny 和 yhat 还有一些其它的东西。

R 语言做深度学习

2.6 R 语言中神经网络的包

http://www.tuicool.com/articles/VNJzYv

R 语言中已经有许多用于神经网络的 package。例如 nnet、AMORE 以及 neuralnet, nnet 提供了最常见的前馈反向传播神经网络算法。AMORE 包则更进一步提供了更为丰富的控制参数,并可以增加多个隐藏层。neuralnet 包的改进在于提供了弹性反向传播算法和更多的激活 函数形式。但以上各包均围绕着 BP 网络,并未涉及到神经网络中的其它拓扑结构和网络模型。而新出炉的 RSNNS 包则在这方面有了极大的扩充。

Stuttgart Neural Network Simulator (SNNS) 是德国斯图加特大学开发的优秀神经网络仿真软件,为国外的神经网络研究者所广泛采用。其手册内容极为丰富,同时支持友好的 Linux 平台。而 RSNNS 则是连接 R 和 SNNS 的工具,在 R 中即可直接调用 SNNS 的函数命令。 require(RSNNS) require(quantmod)

2.6.1 neuralnet

help(package="neuralnet") 帮助文档中有一个 Example

help(package="AMORE")
help(package="nnet")

2.7 R 语言中深度学习的包

darch

http://cran.um.ac.ir/web/packages/darch/index.html

Darch 是建立于 Hinton 和 Salakhutdinov 的 Matlab 代码之上的,其实现方法基于 Hinton 两 篇经典之作"A fast learning algorithm for deep beliefnets" (G. E. Hinton, S. Osindero, Y. W. Teh) 和"Reducing the dimensionality of data with neural networks" (G. E. Hinton, R. R.Salakhutdinov)。 该方法包括了对比散度的预训练和众所周知的训练算法 (如反向传播法或共轭梯度法)的细 调。

deepnet

https://cran.r-project.org/web/packages/deepnet/index.html

Deepnet 实现了一些 Deep Learning 结构和 Neural Network 相关算法, 包括 BP, RBM 训练, Deep Belief Net, Deep Auto-Encoder。作者称后续有时间会继续实现 CNN 和 RNN 算法等。

MXNet

https://mxnet.readthedocs.org/en/latest/R-package/fiveMinutesNeuralNetwork.html

2.8 Github 十大深度学习项目

1 Caffe

Caffe 是利用 Python 和 MATLAB 两者结合创建的深度学习程序库。由伯克利大学的伯 克利视觉与学习中心创办,如果人们仅把 Caffe 当做计算机视觉应用是情有可原的;事实上, 这是一个通用的深度学习程序库,能够开发卷积网络,并创建其他视觉、语音等应用。 2 Data Science IPython Notebooks

这是由 Donne Martin 策划收集的 IPython 笔记本。话题涵盖大数据、Hadoop、scikit-learn 和科学 Python 堆栈以及很多其他方面的内容。至于深度学习,像是 TensorFlow、Theano 和 Caffe 之类的框架也均被涵盖其中,当然还有相关的特定构架和概念等。 3 ConvNetJS

由斯坦福博士生 Andrej Karpathy 写就,他也一直维持着博客更新。ConvNetJS 是一个由 Java 实现的神经网络及其通用模块,并且包括大量的基于浏览器的实例。这些文档和实例数 不胜数,且非常完整。不要让 java 和神经网络结合在一起的想法把你吓跑了,这可是个非常 受欢迎而且非常有用的项目。

4 Keras

Keras 也是一种 Python 深度学习程序库,但它利用了 TensorFlow 和 Theano,这也意味

着它可以在目前已知最受欢迎的 2 个深度学习研发库中的任一个上运行。也是那些越来越 多的被形容为非常高级的库之一,而所有这些高等级库都非常相似:抽象化底层的深度学习 引擎,使使用者可以更快、更便捷以及更灵活地部署神经网络。Keras 支持主流深度学习架 构,自带 30 秒的快速入门指南,并有着完善的文档。

5 MXNet

作为一个深度学习框架,MXNet 旨在灵活和高效,并且通过允许混合使用指令式编程 及符号式编程的技巧来提高代码生产力。这个项目可以和多种语言绑定,如Python,R,以及 Julia。同时 MXNet 还附带了一系列神经网络指南和蓝本。另外值得注意的是,一个相关项 目用 Java 在浏览器环境中实现 MXNet,有兴趣的朋友可以通过该链接 (http://webdocs.cs.ualberta.ca/~bx3/mxnet/classify.html)对一个图形分类模型进行测试。

2.9 Deep Learning Made Easy with R : A Gentle Introduction for Data

Science



- [1] 高涛等译, *R 语言实战*.
- [2] R. I.Kabacoff, *R* in Action data analysis and graphics with *R* (2nd), 2015.
- [3] 炼数成金. *R 语言 tm 工具包进行文本挖掘实验* Available: <u>http://www.dataguru.cn/article-</u> 2641-1.html
- [4] Lemon. *R 语 言 环 境 下 的 文 本 挖 掘 ---tm 包*. Available: <u>http://blog.163.com/zhoulili1987619@126/blog/static/353082012015315104917695/</u>
- [5] 钱昊等译, R 和 Ruby 数据分析之旅.
- [6] S. S. Chang, *Exploring everyday things with R and Ruby*.
- [7] 百度经验. (2016-03-04). *JDK 安装与环境变量配置*. Available: <u>http://jingyan.baidu.com/article/6dad5075d1dc40a123e36ea3.html</u>
- [8] 张德德. (2016-03-04). *JDK、 JRE、 JVM 三者间的关系*. Available: <u>http://playkid.blog.163.com/blog/static/56287260201372113842153/</u>