

硕 士 学 位 论 文

企业销售数据挖掘与预测模型研究

**Research on Data Mining and Forecasting Model of
Pharmaceutical Enterprise Sales**

作 者 姓 名: 许 铭

学 科、专 业: 计 算 机 科 学 与 技 术

学 号: 21624016

指 导 教 师: 金 博 教授

完 成 日 期: 2019 年 3 月 26 日

大连理工大学

Dalian University of Technology

大连理工大学学位论文独创性声明

作者郑重声明：所呈交的学位论文，是本人在导师的指导下进行研究工作所取得的成果。尽我所知，除文中已经注明引用内容和致谢的地方外，本论文不包含其他个人或集体已经发表的研究成果，也不包含其他已申请学位或其他用途使用过的成果。与我一同工作的同志对本研究所做的贡献均已在论文中做了明确的说明并表示了谢意。

若有不实之处，本人愿意承担相关法律责任。

学位论文题目：企业销售数据挖掘与预测模型研究

作者签名： 日期： 年 月 日

摘 要

本模版是根据大连理工大学硕士学位论文格式规范制作的 \LaTeX 硕士学位论文模板。

本模板是基于北京大学、清华大学、哈尔滨工业大学等高校的硕博士论文模板，并按照大连理工大学硕士学位论文格式规范开发的 \LaTeX 论文模板，经过完善和修改，目前已经基本满足了论文规范的要求，而且易用性良好，功能强大。不过，可能还存在着一些问题，欢迎大家积极使用本模版，反馈遇到的问题，以便不断对其进行改进。

当然这个模板仅仅是一个开始，希望有更多的人能够参与进来，不断改进准确性、易用性和较好的可维护性，造福需要的兄弟姐妹们。总体上来说，当前这个模板还是很值得推荐使用的。

本模板的目的旨在推广 \LaTeX 这一优秀的排版软件在大工（尤其是数学相关专业）的应用，为广大同学提供一个方便、美观的论文模板，减少论文撰写格式方面的麻烦。

以下顺便补充一些研究生院所提供的 Word 模版中的注意事项（略去已经嵌入到此模版中的内容）：

1. 论文摘要是学位论文的缩影，文字要简练、明确。内容要包括目的、方法、结果和结论。单位制一律换算成国际标准计量单位制，除特殊情况外，数字一律用阿拉伯数码。文中不允许出现插图。重要的表格可以写入；

2. 篇幅以一页为限，字数为 600-800 字（工程硕士、MBA、EMBA、MPA 等专业学位论文字数为 400-500 字）；

3. 摘要正文后，列出 3-5 个关键词。关键词请尽量用《汉语主题词表》等词表提供的规范词。关键词词间用分号间隔，末尾不加标点，3-5 个。

关键词：写作规范；排版格式；硕士学位论文； \LaTeX 模版

Research on Data Mining and Forecasting Model of Pharmaceutical Enterprise Sales

Abstract

This is a L^AT_EX template of master degree thesis of Dalian University of Technology, which is built according to the required format.

内容应与“中文摘要”对应。使用第三人称，最好采用现在时态编写。

Key Words: Write Criterion; Typeset Format; Master's Degree Paper; X₃L^AT_EX Template

目 录

摘 要	I
Abstract	III
引 言	1
0.1 Windows 操作系统	1
0.1.1 编译运行	1
0.2 Linux 操作系统 (以 Ubuntu 为例)	1
0.2.1 编译运行	3
0.3 XeTeX 简介	3
0.3.1 历史	3
0.4 引言内容要求	4
1 数据介绍与数据分析	5
1.1 数据集一特征提取与特征分析	5
1.1.1 数据集一的描述	5
1.2 数据集二特征提取与特征分析	6
1.3 正文	7
2 列表, 三线表, 插图, 参考文献	9
2.1 列表	9
2.2 表	10
2.2.1 三线表	10
2.2.2 表格的列按小数点对齐	12
2.2.3 多列三线表	13
2.2.4 多行三线表	14
2.2.5 宽度控制	15
2.2.6 斜线表头	16
2.3 图	18
2.3.1 双图并列	19
2.3.2 两子图并列	21
2.4 数学	22
2.4.1 math	22

2.4.2 傅立叶变换.....	22
2.5 参考文献.....	24
3 注意事项	26
3.1 存在的问题	26
3.2 将来的工作	26
结 论	27
参考文献.....	29
附录 A 附录内容名称.....	31
攻读硕士学位期间发表学术论文情况	33
致 谢	35
大连理工大学学位论文版权使用授权书.....	37

引言

0.1 Windows 操作系统

强烈建议下载最新版本的 $\text{CT}_{\text{E}}\text{X}$ 。最新版本号为 v2.9.2.164。建议选择完整版本下载，因为本硕士论文模板使用的某些宏包比较新。不然可能会造成编译错误。

0.1.1 编译运行

需要注意的是，在 WinEdit 中必须在每个 tex 文件的开始添加如下的两行：

```
% !TEX TS-program = XeLaTeX
% !TEX encoding = UTF-8 Unicode
```

否则文件会变成乱码。

以本模版为例，在 Windows 下的编译过程是这样的：

- (1) 打开 main.tex 文件；
- (2) 先点击 WinEdt 工具栏上的 $\text{Xe}_{\text{L}}\text{A}_{\text{T}}\text{E}_{\text{X}}$ 按钮 (可能在 Acrobat Reader 按钮的下拉菜单中)；
- (3) 再点击 WinEdt 工具栏上的 $\text{Bib}_{\text{E}}\text{X}$ 按钮；
- (4) 再点击 WinEdt 工具栏上的 $\text{Xe}_{\text{L}}\text{A}_{\text{T}}\text{E}_{\text{X}}$ 按钮两到三遍；
- (5) 最后点击 WinEdt 工具栏上的 Acrobat Reader 按钮就可以看到输出的 PDF 文档了。

0.2 Linux 操作系统 (以 Ubuntu 为例)

First things first, 首先的工作是安装一个合适的 $\text{Xe}_{\text{L}}\text{A}_{\text{T}}\text{E}_{\text{X}}$ 编译系统。这个问题并不难解决，现在主流的 $\text{L}_{\text{A}}\text{T}_{\text{E}}\text{X}$ 编译系统均已经包含了对 $\text{Xe}_{\text{L}}\text{A}_{\text{T}}\text{E}_{\text{X}}$ 的支持 (包括 xeCJK 中文宏包)，并不需要自己额外再进行安装。在 Linux 下推荐使用 $\text{T}_{\text{E}}\text{XLive}$ ，目前最新版本为 $\text{T}_{\text{E}}\text{XLive}$ 2011。下面以在 Ubuntu 下的本地安装为例，简要的说明 $\text{T}_{\text{E}}\text{XLive}$ 的安装及配置过程，高玩们请主动绕行：

- (1) 下载 $\text{T}_{\text{E}}\text{Xlive}$ 2011 镜像，点击[这里](#)进入下载列表。如果你有检查文件完整性的习惯的话，这个列表还提供了 md5 和 sha256 校验值；
- (2) 安装 perl-tk 包，以便使用图形界面进行安装。在终端中输入命令 `sudo apt-get install perl-tk`；
- (3) 挂载下载好的 iso 镜像，`sudo mkdir /mnt/texlive` (在 /mnt 下创建 texlive 文件夹)，`sudo mount -o loop texlive2011.iso /mnt/texlive` (挂载 texlive2011.iso)。进入 /mnt/texlive 目

录，输入命令 `sudo ./install-tl -gui` 之后出现图形界面。之后的操作就比较简单了，可以去掉不用的语言包以节省磁盘空间，注意选择最后一项 **Create symlinks in system directories**，让安装程序自动创建语法链接。确定安装，耐心等待进度条到头；

- (4) 配置环境变量。在终端中输入 `sudo gedit /etc/bash.bashrc`，在此文件末尾添加

```
PATH=/usr/local/texlive/2011/bin/i386-linux:$PATH; export PATH
MANPATH=/usr/local/texlive/2011/texmf/doc/man:$MANPATH; export
MANPATH
INFOPATH=/usr/local/texlive/2011/texmf/doc/info:$INFOPATH;
export INFOPATH
```

在 `/etc/manpath.config` 文件的 `# set up PATH to MANPATH mapping` 这行下面的列表后增加一条：

```
MANPATH_MAP /usr/local/texlive/2011/bin/i386-linux
/usr/local/texlive/2011/texmf/doc/man
```

在 `/etc/manpath.config` 文件的 `# set up PATH to MANPATH mapping` 这行下面的列表后增加一条：

```
MANPATH_MAP /usr/local/texlive/2011/bin/i386-linux
/usr/local/texlive/2011/texmf/doc/man
```

至此安装过程结束。

以上 **T_EXLive** 安装过程摘自某位筒子的博客文摘，原始链接位于 [wordpress](#) 空间，访问有问题，不过 [百度空间](#) 有转载，虽然百度搜不着什么玩意。

接下来我们需要安装一套中文字体，你可以使用 **Windows** 下的方正、华文或者中易字体，但要注意选择的字体最好是包含宋体、黑体、楷体和仿宋的完整套装。不过由于这些字体在 **PDF** 浏览器中的显示效果并不好，所以建议选用 **Adobe** 的中文字体。安装及配置过程如下：

- (1) 下载 **Adobe** 中文字体，点击 [这里](#) 进入下载页面；
- (2) 将下载的字体拷至 `/usr/share/fonts/truetype/adobe` 目录，如果没有请以管理员身份新建；
- (3) 刷新字体缓存，在终端中输入 `sudo fc-cache -fv`。这时，你可以通过 `fc-list :lang=zh-cn |sort` 命令来查看字体是否安装成功，注意 `fc-list` 后有个空格；
- (4) 你可能还需要在终端中运行 `sudo apt-get install poppler-data cmap-adobe-cns1 cmap-adobe-gb1` 命令来解决 **Adobe** 中文字体在 **PDF** 文件中不显示的情况。

这样，我们就配置好了中文字体，当然这没什么特别的，网上教程一大把。

之后我们需要一个类似于 WinEdt 的集成编译环境。在 Ubuntu 软件中心中，我们能很容易的安装 T_EXmaker 和 T_EXworks，两者功能差不多，T_EXmaker 更强大一些。当然，你也可以自己配置 VIM 下的 L^AT_EX 编译环境，在此就不赘述了。

0.2.1 编译运行

在安装并配置好编译环境之后，接下来的工作就是如何编译 X_YL^AT_EX 文件，生成所需的 PDF 文档了。以本模版为例，在 T_EXworks 编译过程是这样的：

- (1) 打开 main.tex 文件；
- (2) 将工具栏上的编译命令切换至 X_YL^AT_EX 后，点击运行；
- (3) 再将工具栏上的编译命令切换至 BibT_EX 后，点击运行；
- (4) 再将工具栏上的编译命令切换至 X_YL^AT_EX 后，点击运行，这里需要运行两到三遍；
- (5) 如果编译没有错误，就可以看到输出的 PDF 文件了。

对于 T_EXmaker，首先需要在【选项】【配置 T_EXmaker】【命令】中将第一行的 latex 改成 xelatex，之后用 L^AT_EX 作为 X_YL^AT_EX 命令执行即可，其他的和上面类似。

0.3 XeTeX 简介

X_YT_EX（英文发音为“zee-T_EX”）是一种使用 Unicode 的 T_EX 排版引擎，并支持一些现代字体技术，例如 OpenType。其作者和维护者是 Jonathan Kew，并以 X11 自由软件许可证发布。

虽然 X_YT_EX 最初只是为 Mac OS X 所开发，但它现在在各主要平台上都可以运作。它原生的支持 Unicode，并默认其输入文件为 UTF-8 编码。X_YT_EX 可以在不进行额外配置的情况下直接使用操作系统中安装的字体，因此可以直接利用 OpenType，Graphite 中的高级特性，例如额外的字形，花体，合字，可变的文本粗细等等。X_YT_EX 提供了对 OpenType 中本地排版约定（locl 标签）的支持，也允许向字体传递 OpenType 的元标签。它亦支持使用包含特殊数学字符的 Unicode 字体排版数学公式，例如使用 Cambria Math 或 Asana Math 字体代替传统的 T_EX 字体。

0.3.1 历史

2004 年 4 月，发布了 X_YT_EX 的第一个版本，这个版本只支持 Mac OS X，并包括了内建的 AT_T 和 Unicode 支持。2005 年，加入了对 OpenType 的支持。在 2006 年 BachoT_EX 期间，发布了第一个支持 Linux 的版本，并在数月后由 Akira Kakuto 移植到了 Microsoft Windows 上，其跨平台版本最终包含在 T_EXL^Aive 2007 中。另外，从 2.7 版开始，MiK_T_EX

也包含了 X_YTeX。

作为 T_EXLive 的一部分, X_YTeX 支持大多数为 L^AT_EX, OpenType, TrueType 和 PostScript 字体开发的宏包, 而无需特别的安装和设定。

0.4 引言内容要求

以下给出研究生院对引言内容的要求, 格式的要求已经嵌入到本模版中:

- (1) 引言包含的内容有说明论文的主题和选题的范围、对本论文研究主要范围内已有文献的评述以及说明本论文所要解决的问题;
- (2) 注意不要与摘要内容雷同;
- (3) 建议与相关历史回顾、前人工作的文献评论、理论分析等相结合, 如果引言部分省略, 该部分内容在正文中单独成章, 标题改为绪论, 用足够的文字叙述。

特别注意: 是否如实引用前人结果反映的是学术道德问题, 应明确写出同行相近的和已取得的成果, 避免抄袭之嫌。

1 数据介绍与数据分析

1.1 数据集一特征提取与特征分析

1.1.1 数据集一的描述

数据一为 kaggle 竞赛中一个连锁超市的数据，该比赛是通过各种特征来预测接下来的销量，从而进行销售指导，平衡供求关系，减少额外的配送费用，其提供的原始数据集有如下几个文件：

表 1.1 数据一原始信息表

Tab. 1.1 data1

原始文件	文件描述
train.csv	训练数据，包含日期，店铺，售出的商品
stores.csv	各个店铺的详细信息，比如商店的位置和商店的类型
items.csv	商品的特征，比如是否易腐坏，商品类别等
transactions.csv	训练数据中每个店铺的销售数额
oil.csv	每日的原油价格
holidays_events.csv	假期信息

其中 train.csv 文件中包括如下几列：id,date,store_nbr,item_nbr,unit_sales,onpromotion; 分别表示日期，商店 id，商品 id，销量和是否正在促销。对该文件的数据进行可视化，如图 1.1 所示，可以发现如下几点信息：

1. 随着时间的增加，销量整体呈现上升趋势。
2. onpromotion 是否在促销列有着大约五分之一的缺失值，且促销产品占比不大。
3. 数据中一共有 54 个店铺和大量的商品。

其中 stores.csv 文件中包括如下几列：store_nbr,city,state,type,cluster 分别代表着商店 id，商店所在的城市，商店所在的州，商店的类型，其中，商店的类型包括（A，B，C，D）四个类型，以及相似商店的聚类结果。对其进行可视化，可以发现如下几点信息：

1. 各个商店分布在不同的州，市
2. C 类型和 D 类型的店铺较多
3. 一共有 17 个不同的 cluster

其中 items.csv 文件中包括如下几列：item_nbr,family,class,perishable 分别代表着商品 id，商品家族，商品的类别，易腐程度。如下几点信息：

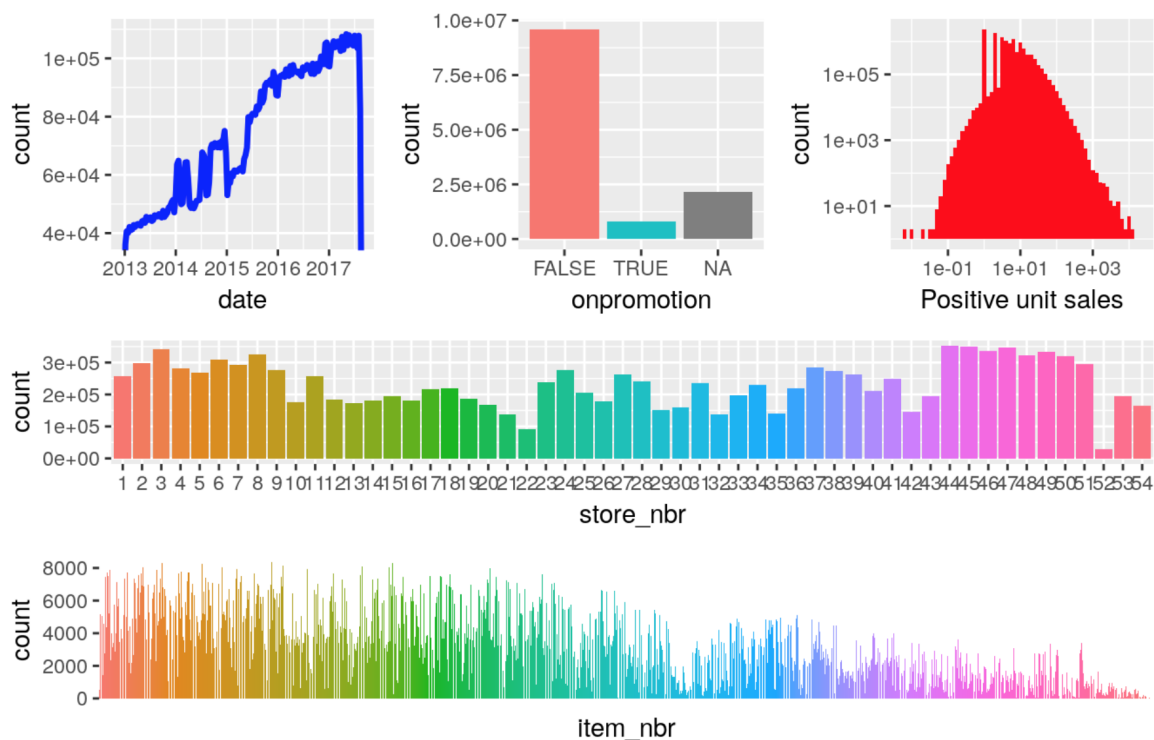


图 1.1 train.csv 数据可视化图

Tab. 1.1 description of train.csv

1. 各个商店分布在不同的州，市
2. C 类型和 D 类型的店铺较多
3. 一共有 17 个不同的 cluster

1.2 数据集二特征提取与特征分析

中英文摘要和关键字也位于 /preface/cover.tex 文件中，分别定义在 cabstract, eabstract, ckeywords, ekeywords 中，替换成自己的即可。

这里附上研究生院对摘要和关键字的要求：

1. “摘要”是摘要部分的标题，不可省略。论文摘要是学位论文的缩影，文字要简练、明确。内容要包括目的、方法、结果和结论。单位制一律换算成国际标准计量单位制，除特殊情况外，数字一律用阿拉伯数码。文中不允许出现插图，重要的表格可以写入；
2. 关键词请尽量用《汉语主题词表》等词表提供的规范词。关键词之间用全角分号间隔，末尾不加标点；
3. 英文摘要和中文摘要对应，但不要逐字翻译。英文关键字使用半角分号间隔，末

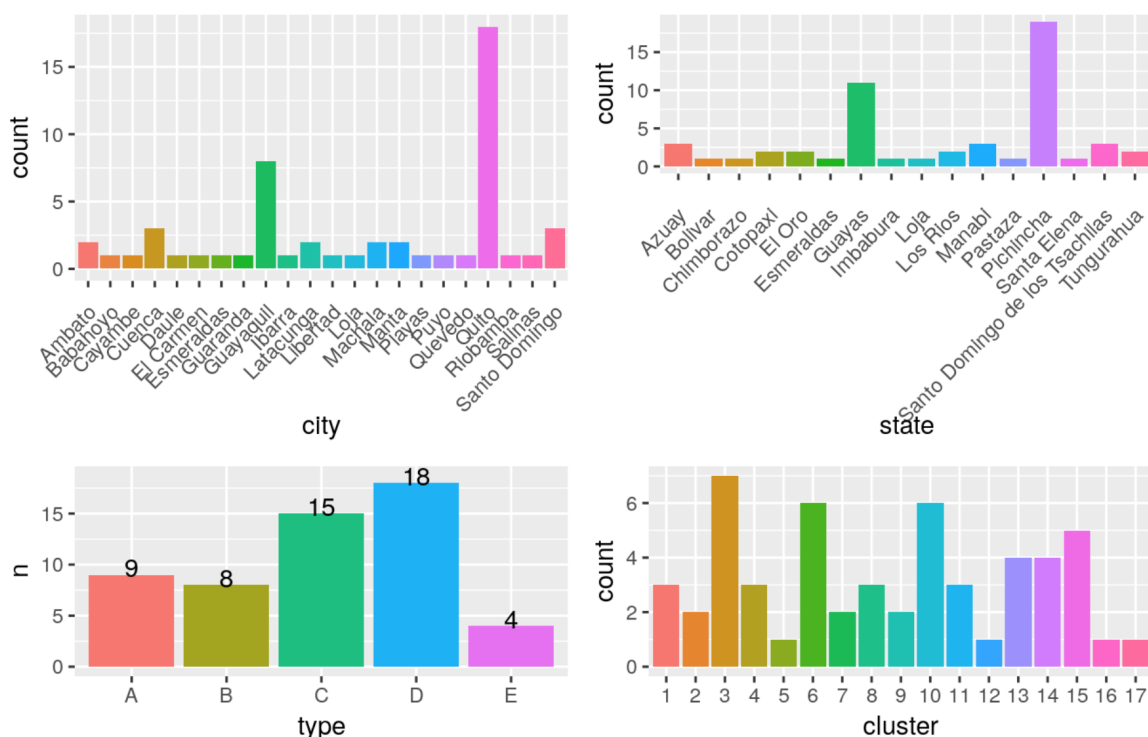


图 1.2 stores.csv 数据可视化图

Tab. 1.2 description of stores.csv

尾同样不加标点。

1.3 正文

正文部分包括了引言 (chap00.tex)、正文内容章节 (chap01.tex、chap02.tex、……)、结论 (conclusion.tex) 三个部分，均位于 body 文件夹中。同时位于 body 文件夹下的还有 BibTeX 参考文献文件 (reference.bib)。

正文内容章节以 chapXX.tex 形式为文件名，从 01 开始计数，使得文件名序号即为章节序号。这些正文内容章节需要依次写入 main.tex 文件中，格式为 `\include{body / chapXX}`。

所有的图片放在 figure 文件夹中。

同样，附上研究生院对正文的要求：

“正文”不可省略。

正文是硕士学位论文的主体，要着重反映研究生自己的工作，要突出新的见解，例如新思想、新观点、新规律、新研究方法、新结果等。正文一般可包括：理论分析；试验装置和测试方法；对试验结果的分析讨论及理论计算结果的比较等。

正文要求论点正确，推理严谨，数据可靠，文字精练，条理分明，文字图表清晰整

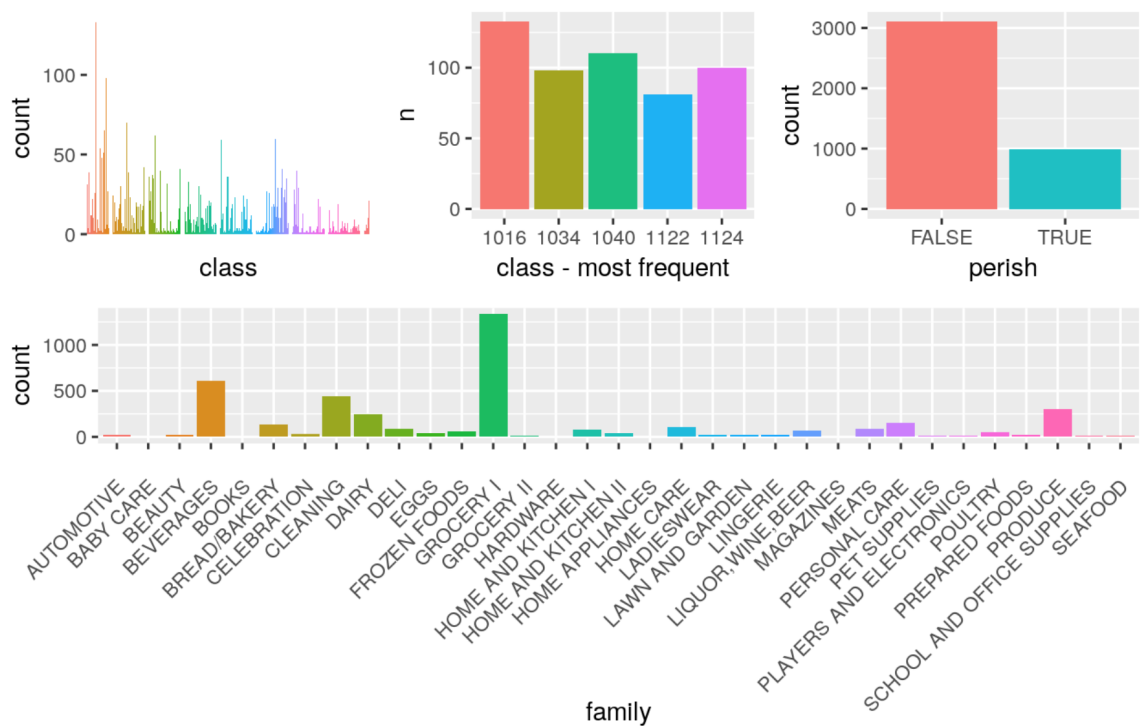


图 1.3 items.csv 数据可视化图

Tab. 1.3 description of items.csv

齐，计算单位采用国务院颁布的《统一公制计量单位中文名称方案》中规定和名称。各类单位、符号必须在论文中统一使用，外文字母必须注意大小写，正斜体。简化字采用正式公布过的，不能自造和误写。利用别人研究成果必须附加说明。引用前人材料必须引证原著文字。在论文的行文上，要注意语句通顺，达到科技论文所必须具备的“正确、准确、明确”的要求。

2 列表，三线表，插图，参考文献

2.1 列表

如果不是搞昆虫学,那就是连着几个月记着写那本大书《文化史上德谟克利特和柏拉图两个流派》,或者是《形态学的发展》,或者是《应用生物学中的统计方法》,再不然是一九五一年至一九五二年编写的教程。几十、几百页都是这种枯燥无味、事务性的记载,每天五至七行¹:

- (1) 鉴定袋蛾——二十分
- (2) 给斯拉瓦写信——二小时四十五分
- (3) 植物保护小组开会——二小时二十五分

上述是默认的列表样式。源代码如下:

```
\begin{enumerate}
\item 鉴定袋蛾——二十分
\item 给斯拉瓦写信——二小时四十五分
\item 植物保护小组开会——二小时二十五分
\end{enumerate}
```

enumerate 环境就是列表环境。每条`\item`后面跟一个空格,然后就是具体的条目。

默认的样式是按照 (1),(2),(3) 来排序的,如果想按照英文字母 (a),(b),(c) 或者罗马数字 (i),(ii),(iii) 这样的顺序呢,只需要在`\begin{enumerate}`后面加一个参数,参数放在方括号内。比如:

- (a) 鉴定袋蛾——二十分
- (b) 给斯拉瓦写信——二小时四十五分
- (c) 植物保护小组开会——二小时二十五分

源代码是:

```
\begin{enumerate}[(a)]
\item 鉴定袋蛾——二十分
\item 给斯拉瓦写信——二小时四十五分
\item 植物保护小组开会——二小时二十五分
\end{enumerate}
```

如上,方括号的中参数是可以更改的。**a** 代表小写字母, **A** 代表大写字母, **1** 代表数字, **i** 代表小写罗马数字, **I** 代表大写罗马数字。这些参数可以加上圆括号,也可以加上

¹格拉宁:《奇特的一生》页 20

一个点 (英文句号)。**[a)]**: 列表的标签就会变成 a)、b)、c)。**[1.]**: 列表的标签就会变成 1、2、3。

罗马数字的例子:

- i. 鉴定袋蛾——二十分
- ii. 给斯拉瓦写信——二小时四十五分
- iii. 植物保护小组开会——二小时二十五分

源代码:

```
\begin{enumerate}[i.]  
\item 鉴定袋蛾——二十分  
\item 给斯拉瓦写信——二小时四十五分  
\item 植物保护小组开会——二小时二十五分  
\end{enumerate}
```

2.2 表

2.2.1 三线表

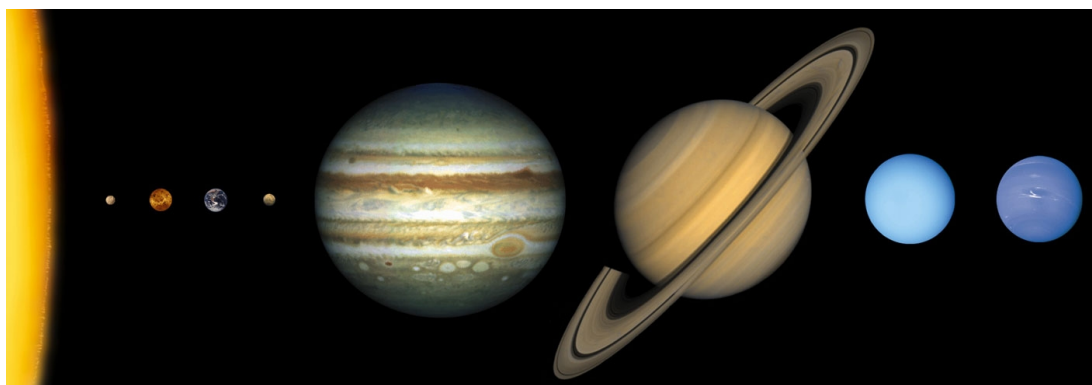


图 2.1 最左侧是太阳，向右依序为水星、金星、地球、火星、木星、土星、天王星与海王星

Fig. 2.1 Outward from the Sun, the planets are Mercury, Venus, Earth, Mars, Jupiter, Saturn, Uranus and Neptune.

天文学家在太阳系内以天文单位 (AU) 来测量距离。1AU 是地球到太阳的平均距离，大约是149 598 000公里 (93 000 000英里)。冥王星与太阳的距离大约是 39AU，木星则约是 5.2AU。最常用在测量恒星距离的长度单位是光年，1 光年大约相当于63 240天文单位¹。

图 2.1展示了太阳系的各行星的位置。

¹维基百科：太阳系

表 2.1 行星数据表

Tab. 2.1 Planet

Planet	Size(Earth=1)	Weight(Earth=1)	Radius
Mercury	0.056	0.055	0.3871
Venus	0.857	0.815	0.7233
Earth	1.00	1.000	1.0000
Mars	0.151	0.107	1.5237
Jupiter	1321	317.832	5.2026
Saturn	755	95.16	9.5549
Uranus	63	14.54	19.2184
Neptune	58	17.15	30.1104

表 2.1 就是最简单的三线表。源代码如下：

```

\begin{table}[htbp]
  \bicaption[tab:xingxing]{ 表 }{ 行星数据表 }{Tab.}{Planet}
  \centering
  \vspace{0.2cm}
  \zhongwu
  \begin{tabular}{cccc}
    \toprule
    Planet & Size(Earth=1) & Weight(Earth=1) & Radius \\
    \midrule
    Mercury & 0.056 & 0.055 & 0.3871 \\
    Venus & 0.857 & 0.815 & 0.7233 \\
    Earth & 1.00 & 1.000 & 1.0000 \\
    Mars & 0.151 & 0.107 & 1.5237 \\
    Jupiter & 1321 & 317.832 & 5.2026 \\
    Saturn & 755 & 95.16 & 9.5549 \\
    Uranus & 63 & 14.54 & 19.2184 \\
    Neptune & 58 & 17.15 & 30.1104 \\
    \bottomrule
  \end{tabular}
\end{table}

```

表格和插图通常需要占据大块空间,所以在文字处理软件中用户经常需要调整它们的位置。**table** 环境可以自动完成这样的任务;这种自动调整位置的环境称作浮动环境

(float), 下一节里还会介绍插图浮动环境¹。

htbp 选项用来指定表格的理想位置, 这几个字母分别代表 **here, top, bottom, float page**, 也就是就这里、页顶、页尾、浮动页 (专门放浮动环境的单独页面)。我们可以使用这几个字母的任意组合, 四个字母都写上表示放哪里都无所谓; 一般不推荐单独使用 **h**, 因为 **L^AT_EX** 自以为它的排版算法是最完美的, 不愿意被束缚手脚。

\centering 用来使表格居中; **\bicaption** 命令设置表格标题, **L^AT_EX** 会自动给浮动环境的标题加上编号。

它的官方使用说明为:

```
\bicaption[label]{ 中文短标题 }{ 中文标题 }{Tab.}{ 英文标题 }
```

可选参数 **label** 用来作为交叉引用链接。例如表 2.1 中的 **lable** 为 **tab:xingxing**。这里的标签一般为英文。中文短标题一般没什么用, 可以随意填。最简单就是“表”。

在表格环境中, 标题必须位于表格的上方。而在图片环境中, 标题的位置必须位于图片的下方。

tabular 环境提供了最简单的表格功能。它用 **&** 来分列, 用 **** 来换行; 每列可以采用居中、居左、居右等横向对齐方式, 分别用 **l**、**c**、**r** 来表示。

三线表的三条横线就分别用 **\toprule**、**\midrule**、**\bottomrule** 等命令表示。

\vspace{0.2cm} 是用来控制表格标题与表格正文的垂直间距的, 请在插入表格时务必添加。**\zhongwu** 是用来调整表格内容的行距的。

2.2.2 表格的列按小数点对齐

以表 2.1 为例, 想把其中的第三列按小数点对齐²。先看一下效果:

在表 2.2 中, 我们调整了原来四列数的对齐方式。原来是 **cccc**, 现在是 **lcSr**。第一列左对齐, 第二列不变, 还是居中对齐, 第四列右对齐。值得注意的是第三列, 这里新引入了一个参数 **S**, 含义就是这一列的数字按照小数点对齐。一定是大写的 **S**。另外, 第三列的列头 **Weight(Earth=1)** 两边也加上了大括号, 因为这不是数字。在使用参数 **S** 的时候, 不是数字的行需要用大括号括起来, 不然会造成编译错误。

```
\begin{table}[htbp]
  \bicaption[tab:xiaoshu]{ 表 }{ 行星数据表 }{Tab.}{Planet}
  \centering
  \vspace{0.2cm}
  \zhongwu
```

¹包太雷: **L^AT_EX NOTES**——雷太赫排版系统简介

²参见宏包 **siunitx**

表 2.2 行星数据表

Tab. 2.2 Planet

Planet	Size(Earth=1)	Weight(Earth=1)	Radius
Mercury	0.056	0.055	0.3871
Venus	0.857	0.815	0.7233
Earth	1.00	1.000	1.0000
Mars	0.151	0.107	1.5237
Jupiter	1321	317.832	5.2026
Saturn	755	95.16	9.5549
Uranus	63	14.54	19.2184
Neptune	58	17.15	30.1104

```

\begin{tabular}{lcr}
\toprule
Planet & Size(Earth=1) & {Weight(Earth=1)} & Radius \\
\midrule
Mercury & 0.056 & 0.055 & 0.3871 \\
Venus & 0.857 & 0.815 & 0.7233 \\
Earth & 1.00 & 1.000 & 1.0000 \\
Mars & 0.151 & 0.107 & 1.5237 \\
Jupiter & 1321 & 317.832 & 5.2026 \\
Saturn & 755 & 95.16 & 9.5549 \\
Uranus & 63 & 14.54 & 19.2184 \\
Neptune & 58 & 17.15 & 30.1104 \\
\bottomrule
\end{tabular}
\end{table}

```

2.2.3 多列三线表

在三线表中，有些列的列头会横跨好几列的数据。一般使用 **multicolumn** 命令。它的用法是：

```
\multicolumn{ 列数 }{ 对齐方式 }{ 表格内容 }
```

“列数”是指这一列横跨的列数，在表 2.3 是 2 列，就填“2”；“对齐方式”从 **lcr** 三者中选其一即可，在表 2.3 中是 **c**。“表格内容”填入自己的内容。一般还会在这一列的下面画一小横线，以示辨识。使用 **cmidrule** 命令。在表 2.3 中，由于横跨的是第 2 列和

第 3 列，因此 `cmidrule` 的参数是 2-3。

表 2.3 不同操作系统下的 \LaTeX

Tab. 2.3 OS with \LaTeX

Test	Common Tools	
	Distribution	Editor
Windows	MikTeX	TexMakerX
Mac OS	MacTeX	TeXShop
Linux/Unix	TeX Live	TeXworks

```
\begin{table}[htbp]
\bicaption[tab: linux ]{\ 不同操作系统下的\LaTeX}\{Tab.}\{OS with \LaTeX}\}
\centering
\vspace{0.2cm}
\zhongwu
\begin{tabular}{ccc}
\toprule
Test & \multicolumn{2}{c}{Common Tools} \\
\cmidrule{2-3}
OS & Distribution & Editor \\
\midrule
Windows & MikTeX & TexMakerX \\
Mac OS & MacTeX & TeXShop \\
Linux/Unix & TeX Live & TeXworks \\
\bottomrule
\end{tabular}
\end{table}
```

2.2.4 多行三线表

既然有多列三线表，多行三线表也是用类似的方法解决。我们把表 2.3 来改造一下，相对应的，一般使用 `multirow` 命令。它的用法是：

```
\multirow{行数}{*}{表格内容}
```

“行数”是指竖向跨的行数，在表 2.4 中是 2 行，中间有个星号，表示自然宽度。

```
\begin{table}[htbp]
```


表 2.4 不同操作系统下的 \LaTeX Tab. 2.4 OS with \LaTeX

OS	Common Tools	
	Distribution	Editor
Windows	MikTeX	TexMakerX
Mac OS	MacTeX	TeXShop
Linux/Unix	TeX Live	TeXworks

```

\caption[tab:unix]{不同操作系统下的\LaTeX}{Tab.}{OS with \LaTeX}
\centering
\vspace{0.2cm}
\zhongwu
\begin{tabular}{ccc}
\toprule
\multirow{2}{*}{OS} & \multicolumn{2}{c}{Common Tools} \\
\cmidrule{2-3}
& Distribution & Editor \\
\midrule
Windows & MikTeX & TexMakerX \\
Mac OS & MacTeX & TeXShop \\
Linux/Unix & TeX Live & TeXworks \\
\bottomrule
\end{tabular}
\end{table}

```

2.2.5 宽度控制

有时候表格中的某行太长了，需要折行。可以使用 `tabularx` 宏包的同名环境，其语法如下：

```

\begin{tabularx}{表格总宽度}{对齐方式}
...
\end{tabularx}

```

“表格总宽度”最好用 `textwidth` 乘以某个系数表示。例如 `0.8\textwidth` 表示表格宽度是版芯宽度的 0.8 倍。这样出来的效果比较好看。对齐方式除了原有的 `l,c,r` 之外，多了一个 `X`，表示某列可以折行。

表 2.5 墙上的 44 句话
Tab. 2.5 Mikko Kuorinki

People	Says
Elias Canetti	If you were alone, you would cut yourself in two, so that one part would shape the other.
Franz Kafka	In the struggle between yourself and the world, second the world.

```
\begin{table}[htbp]
\centering
\bicaption[tab:figure]{ 表 }{ 墙上的44句话 }{Tab.}{Mikko Kuorinki}
\vspace{0.2cm}
\zhongwu
\begin{tabularx}{0.8\textwidth}{X}
\toprule
People & Says \\
\midrule
Elias Canetti & If you were alone, you would cut yourself
in two, so that one part would shape the other.\\
Franz Kafka & In the struggle between yourself and the world,
second the world.\\
\bottomrule
\end{tabularx}
\end{table}
```

2.2.6 斜线表头

还是有些童鞋的表示三线表不实用啊，非要回归到原来的斜线表头去。我们可以使用宏包 **diagbox** 提供的命令轻松完成。不过呢，出来的表格很 **ugly** 罢了。

diagbox 是宏包提供的主要命令。它可以带有两个必选参数，表示要生成斜线表头的两部分内容。默认斜线是从西北到东南方向的。

需要注意的是，使用斜线表格后就不能使用三线表的三条横线，不然请看表 2.6 的下场。正确的做法是使用最原始的 **hline**，见表 2.7。

```
\begin{table}[htbp]
\bicaption[tab:diagbox]{ 表 }{ 斜线表头 }{Tab.}{Diagbox}
\centering
\vspace{0.2cm}
\zhongwu
```

表 2.6 斜线表头

Tab. 2.6 Diagbox

Times \ Day	Mon	Tue	Wed
Morning	used	used	
Afternoon		used	used

```

\begin{tabular}{|l|ccc|}
\toprule
\diagbox{Times}{Day} & Mon & Tue & Wed \\
\midrule
Morning & used & used & \\
Afternoon & & used & used \\
\bottomrule
\end{tabular}
\end{table}

```

表 2.7 斜线表头

Tab. 2.7 Diagbox

Times \ Day	Mon	Tue	Wed
Morning	used	used	
Afternoon		used	used

```

\begin{table}[htbp]
\caption{表 2.6 斜线表头 Tab. 2.6 Diagbox}
\centering
\vspace{0.2cm}
\zhongwu
\begin{tabular}{|l|ccc|}
\hline
\diagbox{Times}{Day} & Mon & Tue & Wed \\
\hline
Morning & used & used & \\
Afternoon & & used & used \\
\hline
\end{tabular}

```

```
\end{tabular}
\end{table}
```

2.3 图

我来北京十一年，上学，上班，也积累了不少同学同事，但我一次他们的婚礼都没有参加过。刚毕业的时候这种邀请很多，好象是种翻天覆地日新月异的见证，不怕那些同学们伤心，我也知道人家是真心邀请的，但我真觉得我没跟他们谁好到真有必要参加那些婚礼，所以我就不要去。这次黄总的婚礼通知的突然，但我却很想去见证一下，我的罪恶太多，正好也让天主顺便宽恕宽恕我，沾沾喜气¹。

我挺佩服黄总他们俩的，他们就真的仅仅在婚礼当天才拍所谓的“婚纱照”。我觉得这样挺好。身为摄影部的美女，周围全是靠摄影吃饭的人，的确没必要出去花冤枉钱就为了拍几张照片。



图 2.2 婚礼

Fig. 2.2 Wedding

```
\begin{figure}[htbp]
\centering
\includegraphics[scale=0.6]{wedding.jpg}
\bicaption[fig:wedding]{婚礼}{婚礼}{Fig.}{Wedding}
\end{figure}
```

¹蚌病生珠：阳光下的婚礼

论文使用的图片都放在 **figure** 文件夹中，插图浮动环境是 **figure**，基本命令是 **includegraphics**，而在图片环境中，标题的位置必须位于图片的下方。

includegraphics 的基本参数见表 2.8。

表 2.8 插图命令参数

Tab. 2.8 Parameter

参数	说明
width=x,height=y	宽度和高度, 绝对尺寸, 可用任意长度单位。
scale=s	缩放比。绝对尺寸和缩放比用一种即可, 同时使用两者, 绝对尺寸起作用。
keepaspectratio	保持图形比例。宽度和高度通常设置一个即可, 否则图形比例会失调, 除非再加上此选项, 这样图形宽度和高度都不超过指定参数。
angle=a	逆时针旋转角度, 单位是度。

对于图 2.2, 只使用了 **scale** 这一个参数, 缩放因子是 0.6。当然, 也可以直接指定图形的宽度和高度。图 2.1 的源代码如下:

```
\begin{figure}[htbp]
\centering
\includegraphics[width=\textwidth {}, keepaspectratio]{sun.jpg}
\bicaption[fig:sun]{图}{最左侧是太阳, 向右依序为水星、金星}
{Fig.}{Outward from the Sun, the planets are Mercury, Venus, Earth, Mars, Jupiter, Saturn, Uranus and Neptune.}
\end{figure}
```

可以看到, 图 2.1 的宽度指定为版芯的宽度, 然后使用了保持宽高比这个选项。

2.3.1 双图并列

温文敦厚的新郎, 美丽可爱的新娘。

Alan 和 Cher 这一对从校园时代就相偎相依一直到步入婚礼的殿堂。

我一直觉得, 这样的情侣是最最难得的, 两个人之间最珍贵的东西得以一直保存、延续, 直至在未来某个时候升华成为生命中不可名状的一种记忆和体验。这个世界上有太多因为坚持或者不坚持, 执着或者不执着导致的有始无终。能够沿路陪伴, 最终成为眷属, 也算是不大不小的奇迹。

两个人的婚礼誓词很肉麻, 很感人, 写在小纸片上, 认真的读出来, 直到读到对方

流下感动的泪水，直到在场的嘉宾都用掌声回应这份真情¹。



图 2.3 新郎

Fig. 2.3 Bridegroom



图 2.4 新娘

Fig. 2.4 Brige

```
\begin{figure}[\htbp]
  \centering
  \begin{minipage}{0.4\textwidth}
    \centering
    \includegraphics[keepaspectratio]{lang.jpg}
    \bcaption[fig:lang]{图}{新郎}{Fig.}{Bridegroom}
  \end{minipage}
  \begin{minipage}{0.4\textwidth}
    \centering
    \includegraphics[keepaspectratio]{liang.jpg}
    \bcaption[fig:niang]{图}{新娘}{Fig.}{Brige}
  \end{minipage}
\end{figure}
```

如果想要两幅并排的插图各有自己的标题,可以在 figure 环境中使用两个 minipage 环境,每个里面插入一幅图(见图 2.3和图 2.4)。不用 minipage 的话,因为插图标题的缺省宽度是整个行宽;两幅插图就会上下排列。

这里指定了每个 minipage 的宽度为 0.4 倍的版芯宽度。当然,也可以自己指定,只是两个宽度加起来不超过版芯宽度就可以了。

¹蚌病生珠：罗兰湖畔

2.3.2 两子图并列

有你在，开水瓶里永远都有水喝；冰箱里永远有一袋应急的速冻饺子；阳台的晾衣架上我昨天换下的衣服已经有了今天太阳的味道；小猫也不用害怕得病和不舒服，它们有最负责的家庭医生；还有别人根本见都没见过的那些点心和饼干；还有，你是我见过的少有的照片和本人都好看的女孩¹。



(a) 超人 A

(b) 超人 A

图 2.5 小超人老师

Fig. 2.5 Judy

```
\begin{figure}[htbp]
  \centering
  \subfigure[超人A]{
    \label{fig:1a}
    \includegraphics[keepaspectratio]{chao.jpg}
  }
  \hspace{20pt}
  \subfigure[超人A]{
    \label{fig:1b}
    \includegraphics[keepaspectratio]{ren.jpg}
  }
  \bicaption[fig:judy]{图}{小超人老师}{Fig.}{Judy}
\end{figure}
```

¹蚌病生珠:Two-year anniversary

如果想要两幅并排的图片共享一个标题, 并且各有自己的子标题, 可以使用 `subcaption` 宏包。如图 2.5, 子图的标题用命令 `subcaption` 即可。

2.4 数学

2.4.1 math

\mathbf{A} , \mathcal{A} , \mathfrak{A} , \mathbb{A}

`$\$ \backslash \mathrm{mathbf}{A}, \backslash \mathrm{mathcal}{A}, \backslash \mathrm{mathfrak}{A}, \backslash \mathrm{mathbb}{A} \$$`

2.4.2 傅立叶变换

在现代数学中有一个很容易被外行误解的词汇：信号 (signal)。当数学家们说起「一个信号」的时候，他们脑海中想到的并不是交通指示灯所发出的闪烁光芒或者手机屏幕顶部的天线图案，而是一段可以具体数字化的信息，可以是声音，可以是图像，也可是遥感测量数据。简单地说，它是一个函数，定义在通常的一维或者多维空间之上。譬如一段声音就是一个定义在一维空间上的函数，自变量是时间，因变量是声音的强度，一幅图像是定义在二维空间上的函数，自变量是横轴和纵轴坐标，因变量是图像像素的色彩和明暗，如此等等¹。

在数学上，关于一个信号最基本的问题在于如何将它表示和描述出来。按照上面所说的办法，把一个信号理解成一个定义在时间或空间上的函数是一种自然而然的表示方式，但是它对理解这一信号的内容来说常常不够。例如一段声音，如果单纯按照定义在时间上的函数来表示，它画出来是这个样子的：

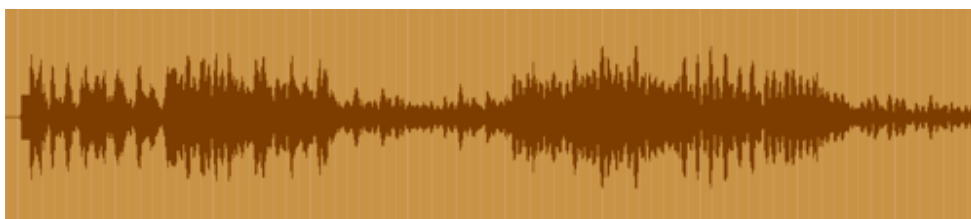


图 2.6 波形图

Fig. 2.6 Wave

图 2.6 通常被称为波形图。毫无疑问，它包含了关于这段声音的全部信息。但是同样毫无疑问的是，这些信息几乎没法从上面这个「函数」中直接看出来，事实上，它只不过是巴赫的小提琴无伴奏 *Partita No.3* 的序曲开头几个小节。图 2.7 是巴赫的手稿，从某种意义上说来，它也构成了对上面那段声音的一个「描述」：

¹木遥：不确定性原理的前世今生·数学篇(一)

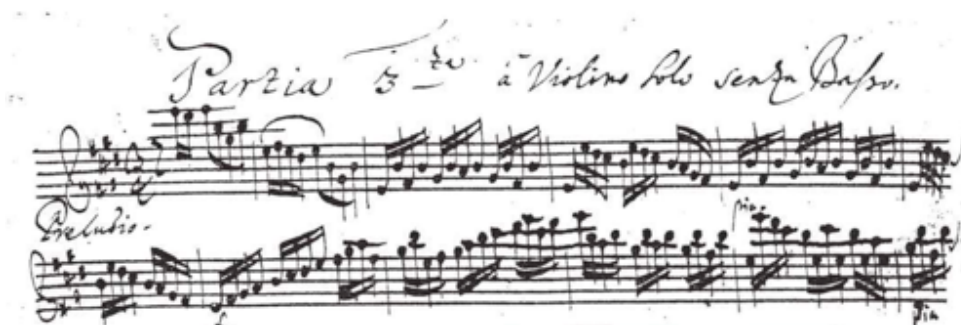


图 2.7 巴赫手稿

Fig. 2.7 Partita No.3

这两种描述之间的关系是怎样的呢？第一种描述刻划的是具体的信号数值，第二种描述刻划的是声音的高低（即声音震动的频率）。人们直到十九世纪才渐渐意识到，在这两种描述之间，事实上存在着一种对偶的关系，而这一点并不显然。

1807 年，法国数学家傅立叶 (J. Fourier) 提出了一个崭新的观念：任何一个函数都可以表达为一系列不同频率的简谐振动（即简单的三角函数）的叠加。

用今天的语言来描述，傅立叶的发现实际上是在说：任何一个信号都可以用两种方式来表达，一种就是通常意义上的表达，自变量是时间或者空间的坐标，因变量是信号在该处的强度，另一种则是把一个信号「展开」成不同频率的简单三角函数（简谐振动）的叠加，于是这就相当于把它看作是定义在所有频率所组成的空间（称为频域空间）上的另一个函数，自变量是不同的频率，因变量是该频率所对应的简谐振动的幅度。

这两个函数一个定义在时域（或空域）上，一个定义在频域上，看起来的样子通常截然不同，但是它们是在以完全不同的方式殊途同归地描述着同一个信号。它们就象是两种不同的语言，乍一听完全不相干，但是其实可以精确地互相翻译。在数学上，这种翻译的过程被称为「傅立叶变换」。

傅立叶变换是一个数学上极为精美的对象：

- (1) 它是完全可逆的，任何能量有限的时域或空域信号都存在唯一的频域表达，反之亦然。
- (2) 它完全不损伤信号的内在结构：任何两个信号之间有多少相关程度（即内积），它们的频域表达之间也一定有同样多的相关程度。
- (3) 它不改变信号之间的关联性：一组信号收敛到一个特定的极限，它们的频域表达也一定收敛到那个极限函数的频域表达。

在傅立叶变换的所有这些数学性质中，最不寻常的是这样一种特性：一个在时域或空域上看起来很复杂的信号（譬如一段声音或者一幅图像）通常在频域上的表达会很简

单。这里「简单」的意思是说作为频域上的函数，它只集中在很小一块区域内，而很大一部分数值都接近于零。

一个在空域中看起来占满全空间的信号，从频域中看起来很可能只不过占用了极小一块区域，而大部分频率是被浪费了的。这就导出了一个极为有用的结论：一个看起来信息量很大的信号，其实可以只用少得多的数据来加以描述。只要对它先做傅立叶变换，然后只记录那些不接近零的频域信息就可以了，这样数据量就可以大大减少。

基本上，这正是今天大多数数据压缩方法的基础思想。在互联网时代，大量的多媒体信息需要在尽量节省带宽和时间的前提下被传输，所以数据压缩从来都是最核心的问题之一。而今天几乎所有流行的数据压缩格式，无论是声音的 mp3 格式还是图像的 jpg 格式，都是利用傅立叶变换才得以发明的。从这个意义上说来，几乎全部现代信息社会都建立在傅立叶的理论的基础之上。

2.5 参考文献

硕士论文写了 3 周。90 多页英文，昏天黑地没日没夜写到想吐。好在有几个欧洲博士后帮忙改语法错。改的他们也很想哭。后来已经功成名就论文无数 ACM Fellow 英国 Fellow of Royal Society 的老板来给我们讲，写论文最重要的是写 Introduction。写 Introduction 就和写童话一样¹。

1. 有一条巨龙抓走了公主 (介绍你的问题为什么值得研究)
2. 巨龙是多么多么多么难打 (强调你的研究的重要性)
3. 王子提着一把金光闪闪的剑而不是破斧子烂长矛登场 (你的方法好在哪里，别人 sui 在哪里)
4. 王子是如何打败巨龙 (你的方法简介)
5. 从此王子和公主幸福的生活在一起。(解决了问题)

老板说写论文就是写童话嘛。其余的也不过就是把这些东西细节讲一讲。做研究很简单的。听完我就不想再做研究了。合着我写到吐血掉头发的时候大牛都把写论文当给小盆友写童话。

我们的一切知识都是从经验开始^[1]，这是没有任何怀疑的^{[2][3]}；因为，如果不是对象激动我们的感官，一则由它们自己引起表象，一则使我们的知性活动运作起来，对这些表象加以比较，把它们粘结或分开，^[4,5] 这样把感性印象的原始素材加工成称之为经验的对象知识，那么知识能力又该由什么来唤起活动呢？^[1,6-8] 所以按照时间，我们没有任何知识是先行于经验的，一切知识都是从经验开始的。

只要是中文文献，图书，期刊，会议，专利等等需要为每个条目增加一个域：

¹理 cici: 硕士论文你有哪些经验与收获？耗时多久？

```
language={c},
```

对于参考文献 [1], 原先的 bib 文件是这样的:

```
@article{ 李秋零1999 ,  
  title ={ 康德何以步安瑟尔谟的后尘? },  
  author={ 李秋零 },  
  journal={ 中国人民大学学报 },  
  volume={2},  
  year={1999}  
}
```

但是由于是中文文献, 需要增加一个语言域, 就变成下列样式:

```
@article{ 李秋零1999 ,  
  title ={ 康德何以步安瑟尔谟的后尘? },  
  author={ 李秋零 },  
  language={c},  
  journal={ 中国人民大学学报 },  
  volume={2},  
  year={1999}  
}
```

3 注意事项

请直接双面打印 PDF 文件，空白页已经按要求留出。打印时，缩放页面的选项设为“无”，否则页面会缩小。

参考文献的 bib 文件的条目的名称是不允许出现空格的。

3.1 存在的问题

本模板还有如下一些问题：

- (1) 目录页中的“目录”与第一级标题的垂直间距过大；
- (2) 英文摘要的页码是罗马数字三，其实应该为二；
- (3) 原创性声明页没有页眉的那一横线。

3.2 将来的工作

有关子图处理的宏包仍然是 **subfigure**。该宏包已经好久没有更新了，被 **subcaption** 宏包所代替。但是目前而言，后者与 **ccaption** 宏包发生冲突。因此等到不远的某一天，这个 **bug** 被修订之后，还是希望使用后一宏包的。

结 论

结论是理论分析和实验结果的逻辑发展，是整篇论文的归宿。结论是在理论分析、试验结果的基础上，经过分析、推理、判断、归纳的过程而形成的总观点。结论必须完整、准确、鲜明、并突出与前人不同的新见解。

参考文献

- [1] 李秋零. 康德何以步安瑟尔谟的后尘?[J]. 中国人民大学学报, 1999, 2.
- [2] 邓晓芒, 赵林. 西方哲学史 [M]. 高等教育出版社, 2005.
- [3] 邓晓芒. 什么是艺术作品的本源?——海德格尔与马克思美学思想的一个比较 [J]. 哲学研究, 2000, 10(8):58–64.
- [4] 欧进萍, 何政. 钢筋混凝土结构基于地震损伤性能的设计 [J]. 地震工程与工程振动, 1999, 19(1): 21–30.
- [5] 欧进萍, 牛荻涛, 杜修力. 设计用随机地震动的模型及其参数确定 [J]. 地震工程与工程振动, 1991, 11(3):45–53.
- [6] BRAUN J, MURDOCH D. A first course in statistical programming with R[M]. Cambridge University Press Cambridge, 2007.
- [7] KELTON W, SADOWSKI R, SADOWSKI D. Simulation with ARENA[M]. McGraw–Hill Boston, MA, 2002.
- [8] STRAWDERMAN R. Monte carlo methods in statistical physics[J]. Journal of the American Statistical Association, 2001, 96(454):778–778.

附录 A 附录内容名称

以下内容可放在附录之内：

1. 正文内过于冗长的公式推导；
2. 方便他人阅读所需的辅助性数学工具或表格；
3. 重复性数据和图表；
4. 论文使用的主要符号的意义和单位；
5. 程序说明和程序全文。

这部分内容可省略。

攻读硕士学位期间发表学术论文情况

- [1] 作者 1, 作者 2. 论文题目. 大连理工大学网络学刊, 2010 年. 主办单位: 大连理工大学研究生院. (本硕士学位论文第 X 章)

致 谢

学位论文中不得书写与论文工作无关的人和事，对导师的致谢要实事求是。一同工作的同志对本研究所做的贡献应在论文中做明确的说明并表示谢意。这部分内容不可省略。

在这里，向所有协助测试的同学、朋友表示感谢。

大连理工大学学位论文版权使用授权书

本人完全了解学校有关学位论文知识产权的规定，在校攻读学位期间论文工作的知识产权属于大连理工大学，允许论文被查阅和借阅。学校有权保留论文并向国家有关部门或机构送交论文的复印件和电子版，可以将本学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印、或扫描等复制手段保存和汇编本学位论文。

学位论文题目： 企业销售数据挖掘与预测模型研究

作者签名： _____ 日期： _____ 年 ____ 月 ____ 日

导师签名： _____ 日期： _____ 年 ____ 月 ____ 日

