



南京大學

研究生畢業論文
(申請碩士學位)

論文題目 一種基於多维時間序列分析的音乐推荐系統研究與實現

作者姓名 王守濤

學科、專業方向 計算機技術

指導教師 徐鋒 教授

研究方向 可信計算

2014 年 5 月

学 号 : MF1133033

论文答辩日期 : 2014 年 6 月 1 日

指 导 教 师 : (签字)

Research and Implementation of A Music Recommender System Based on Multidimensional Time Series Analysis

by

Shoutao Wang

Directed by

Professor Xu Feng

Computer Technology

May 2014

*Submitted in partial fulfilment of the requirements
for the degree of Master in Computer Technology*

南京大学研究生毕业论文中文摘要首页用纸

毕业论文题目：一种基于多维时间序列分析的音乐推荐系统研究与实现

计算机技术 专业 2014 级硕士生姓名：王守涛

指导教师（姓名、职称）：徐锋 教授

摘 要

近些年来，互联网相关技术得到了快速发展，而信息过载 [19]和选择悖论 [16]的问题也愈发显著。为了提升信息的利用率，推荐系统应运而生并已广泛应用在电子商务、音乐电台、个性化阅读等领域。音乐作为一种强力的沟通和自我表达工具已经成为人们日常生活中一种非常重要的消费品,而相关研究也表明人们平时的听歌行为要远远多于读书、看电视、看电影等行为 [15]。为了解决人们在音乐消费过程中遇到的信息过载和选择悖论问题，一些个性化音乐推荐系统被开发了出来，比如Last.fm [10]、Pandora [13]、豆瓣音乐 [7] 等。然而，大多数音乐推荐系统没有充分考虑到用户听歌行为与其所处上下文环境之间的密切关系，导致做出的推荐不符合用户的即时需求，而一些考虑用户上下文的工作对用户行为历史的分析也不够细致和全面。此外，用户的未来行为受到其长期行为、即时行为 and 中期行为的共同影响，但目前未见工作将这几方面的影响综合起来加以考察。因此，本文给出了一种基于多维时间序列分析的个性化音乐推荐系统以全面细致地对用户行为进行分析。本文的主要贡献和工作如下所示：

1. 提出了一种基于多维时间序列分析的音乐推荐方法。该方法首先使用主题模型将歌曲表示成由若干隐含主题构成的概率分布，进而将用户行为建模为多维时间序列并对该序列进行分析和预测，从而实现了用户对用户行为的定量分析。

2. 从更细的粒度对用户听歌行为进行分析和预测，认为用户的听歌行为在每个隐含主题上都对应着一个时间序列，进而通过对这些时间序列的分析全面考察每一个隐含主题的贡献和作用。

3. 给出了一种基于用户三期行为的综合音乐推荐方法，综合考虑了用户长期行为、中期行为以及即时行为对用户未来行为的影响和作用。

4. 实现了一个原型系统，验证了本文所述方法的可行性。其中，本文利用分布式实时计算框架Storm实现了所述方法的并行化，提升了系统的执行效率。

关键词： 音乐推荐系统; 主题模型; 多维时间序列; 综合推荐

南京大学研究生毕业论文英文摘要首页用纸

THESIS: Research and Implementation of A Music Recommender System Based on Multidimensional Time Series Analysis

SPECIALIZATION: Computer Technology

POSTGRADUATE: Shoutao Wang

MENTOR: Professor Xu Feng

Abstract

In recent years, internet-related technologies have grown rapidly. However, information overload problem and paradox of choice become much more obviously. In order to enhance the utilization of information, recommender systems appear and have been applied to many fields, such as e-commerce, music, and and reading. Being a powerful communication and self-express tool, music has been an important consumer goods in people's daily life and previous research has also indicated that participants listened to music more often than other activities, such as watching television, reading books and watching movies. To alleviating the information overload problem and paradox of choice in music field, some music recommender systems have been developed, such as Lastfm, Douban, Pandora. However, most of recent music recommender systems do not take the close relationship between user's behavior and his/her current context into consideration, leading to bad recommendations that do not match user's current requirement. Although there are some researchers have done some related works, they are not meticulous and comprehensive enough. Besides, user's future behaviors are affected by his/her long-term behaviors, immediately behaviors and middle-term behaviors at the same time and there are few work giving a composite idea on them. In order to get better recommendations, we propose a music recommender system based on multidimensional time series analysis that can analyze the user's behavior more meticulous and comprehensive. Our main contribution and work are as follows:

1. A method for music recommendation based on multidimensional time series analysis was proposed. The method firstly using topic models to model a song as a probability distribution that is constituted by some latent topics and then it model the

user's behavior as a multidimensional time series. By analyzing the time series, the method can forecast the user's future behavior in a quantitative way.

2. User's behavior was analyzed in finer granularity. With the basic idea of thinking every latent topic corresponds to a time series, we analyze every time series and forecast the future values leading to comprehensive analyzing the influence of all latent topics.

3. A composite music recommender method was proposed. The method takes into consideration user's long-term behavior, immediate behavior and middle-term behavior at the same time.

4. A prototype that is used to verify the feasibility of our recommender methods was implemented. Specially, we paralleled the method with an distributed realtime computation framework-Storm, making the system running faster.

Keywords: Music Recommender System, Topic Model, Multidimensional Time Series, Composite Recommendation

目 录

目录	iv
第一章 背景介绍	1
1.1 系统要求	3
1.2 下载与安装	3
1.3 本模板打包的文件	4
1.4 编译方法	4
1.4.1 Windows 平台下	5
1.4.2 Linux 平台下	7
1.5 问题反馈	7
第二章 论文撰写	9
2.1 文档类选项	9
2.2 封面和摘要页的字符串	9
2.3 内容撰写	9
2.4 插图	9
2.5 数学公式	10
2.6 参考文献	10
2.7 其它问题	11
第三章 常见问题	12
3.1 表格	12
3.2 脚注	12
第四章 总结与展望	13
附录 A 博士(硕士)学位论文编写格式规定(试行)	14
A.1 适用范围	14
A.2 引用标准	14

A.3	印制要求	14
A.4	编写格式	14
A.5	前置部分	14
A.5.1	封面（博士学位论文国图版用）	14
A.5.2	题名	15
A.5.3	前言	15
A.5.4	摘要	16
A.5.5	关键词	16
A.5.6	目次页	16
A.5.7	插图和附表清单	16
A.5.8	符号、标志、缩略词、首字母缩写、计量单位、名词、术语等的注释表 符号、标志、缩略词、首字母缩写、计量单位、名词、术语等的注释说明汇集表，应置于图表清单之后。	16
A.6	主体部分	16
A.6.1	格式	16
A.6.2	序号	17
A.6.3	绪论	17
A.6.4	正文	17
A.6.5	结论	19
A.6.6	致谢	20
A.6.7	参考文献表	20
A.7	附录	20
A.8	结尾部分(必要时)	21
	参考文献	22
	简历与科研成果	24
	致谢	25

表 格

1.1 信息过载问题主流解决方案	1
------------------------	---

插图

第一章 背景介绍

近些年来，互联网相关技术得到快速发展，信息呈指数级增长，这造成了严重的信息过载(Information Overload)问题 [19]。特别是随着大数据时代的来临，信息的增长速度已经远远超过人们可以理解的速度，而海量信息中更是包含着很多对人们来说无用的冗余信息以及会干扰人们抉择的错误信息。面对海量的信息和众多的选择，人们往往无所适从而陷入到选择悖论(Paradox of Choice) [16]之中，或无法做出合理的选择，或需要消耗很大的精力才能做出正确选择。为了解决信息过载问题，减轻人们在做出抉择时所承受的沉重负担，信息分类、搜索引擎和推荐系统等技术应运而生，如表1.1所示。信息分类技

表 1.1: 信息过载问题主流解决方案

名称	特点	案例
信息分类	分门别类地组织信息	Yahoo!、58同城
搜索引擎	根据关键字进行查询	Google、百度
推荐系统	分析用户行为历史，主动推荐	Amazon、淘宝

术是通过将互联网上的各种信息分门别类的组织起来以提高人们查询的效率，比如早期的Yahoo!和国内的58同城的信息分类网站。然而，信息分类技术往往依赖于人工，可扩展性差，随着互联网上信息的增加，这显然不合时宜。搜索引擎技术通过建立索引的方式将互联网上的网页组织起来，然后根据用户输入的关键字进行查询，比如Google和百度。显然，搜索引擎相对于信息分类技术来说已经有了很大的进步，能够适应互联网的快速发展。但是，人们很多时候要么不愿意费时费力去输入关键字，要么无法准确地用关键字去描述自己的想法，比如“今天看什么电影好呢?”、“去哪儿吃饭呢?”等，此时搜索引擎就无能为力了。为了解放人们的双手以及深度挖掘人们的内在偏好和需求，推荐系统技术应运而生。推荐系统本质上是一种信息过滤系统，其通过对用户行为历史的分析挖掘出用户的行为偏好，进而帮助用户将海量信息中的无用信息过滤掉，进而将符合用户偏好的信息推荐给用户 [5]。目前，推荐系统已经在各个领域得到了广泛的应用，比如电子商务领域的亚马逊和淘宝，在线视频领域的Netflix和优酷，在线音乐领域的Lastfm [10] 和豆瓣电台 [7]以及个性化阅读领域Flipboard和无觅阅读等。图1.1展示了电子商务领域Amazon的商品推荐系统界面和视频领域PPTV的在线视频推荐系统界面。音乐推荐系统是推荐系统在音乐

领域的应用，其旨在将人们从海量的音乐中解脱出来，通过分析用户的收听习惯以及歌曲本身的特征来为用户推荐符合口味的歌曲 [17]。音乐推荐与传统物品推荐的区别在于歌曲时长比较短、消费代价低、可重复消费以及具有强烈的情感性等特点，而人们的听歌行为往往对上下文环境往往更为敏感。比如，在不同的天气、不同的情绪以及不同的场合，人们的音乐喜好都会有所不同。因此，在做音乐推荐的时候需要充分的考虑歌曲和人们收听行为的特殊性。为了满足人们对音乐的需求，一些个性化的音乐推荐系统被设计和实现出来，如国外的Lastfm [10]、Pandora [13]以及国内的豆瓣电台 [7]、虾米音乐 [18]，图1.2展示了豆瓣电台和虾米音乐的推荐界面。这些音乐推荐系统首先建立自己的曲库，然后分析歌曲的特征和用户的听歌习惯继而为用户做出推荐。其中，Pandora [13] 是当今最流行的音乐推荐系统之一，其通过Music Genome Project将400种属性分配给每一首歌曲，进而按照歌曲的相似程度为用户做出推荐。从可见的资料看，Last.fm [10]、豆瓣 [7] 主要是基于“相似的人往往具有相似的行为”的假设进行推荐。其中，虾米音乐 [18]还给出了推荐的同时还给出了做出推荐的原因，这在一定程度上提升了用户对推荐结果的接收程度。文献 [15]表明，人们在日常生活中的听歌行为要远远多于读书、看电影等行为，这说明音乐已经成为人们生活中不可缺少的一部分。为了提升人们在大数据时代消费音乐产品的体验，越来越多的工作开始关注推荐系统在音乐领域的应用，一些有效的音乐推荐算法也被提了出来。按照参考用户行为的长短，可以将这些算法分为基于用户长期行为的音乐推荐、基于用户即时行为的音乐推荐以及基于用户中期行为的音乐推荐三类。其中，基于用户长期行为和即时行为的音乐推荐已经得到了比较广泛地研究，而基于用户中期行为的音乐推荐由于考虑了上下文环境对用户行为的影响，目前也开始得到越来越多的关注。但是，目前关于用户中期行为的音乐推荐算法还存在对用户行为刻画不够细致、全面的问题，加之在一定程度上依赖于其他用户的行为，导致推荐的效果不够理想。为了提高基于中期行为的音乐推荐算法的效果，本文给出了一致基于多维时间序列分析的个性化音乐推荐方法，该方法将用户行为建模为一个多维时间序列，从而实现对用户行为细致、全面地分析。此外，用户的未来行为除了受中期行为的影响之外，还受其长期行为和即时行为的影响。但是，目前未见将这三种行为综合起来考虑来预测用户未来行为的工作。因此，本文在给出一种基于多维时间序列分析的音乐推荐方法的基础上，进一步给出一种基于用户三期行为的综合音乐推荐方法，从而全面地考察用户长期行为、即时行为 and 中期行为的贡献和作用。在给出基于多维时间序列分析的音乐推荐方法和基于用户三期行为的综合推荐方法之后，本文通过实验验证了所给方法能够取得比参考方法更高的推荐效果

和更低的推荐误差。进一步地，本文基于上述方法实现了一个原型系统且在该系统中实现了所给方法的并行化，这验证了所给方法的可行性。本文结构按照如下方式进行组织：第2章介绍与本文工作相关的一些工作，包括常用的推荐算法、常用的物品相似度计算方法、推荐系统的评测指标以及文本处理、时间序列分析的相关理论等；第3章描述了本文所提的一种基于时间序列分析的个性化音乐推荐方法，包括问题的定义、方法的执行流程等；第4章描述了用户听歌行为的局部性特征和全局性特征并以此为基础给出一个基于时间序列分析的个性化音乐推荐系统框架；第5章给出了原型系统的相关实现技术；第6章对全文进行总结并给出对未来工作的展望。

下面是 TODO 列表：

- （高优先级）增加本科生论文格式，当 `bachelor` 作为模板参数时生成本科生毕业论文格式。
- （低优先级）提供 UTF-8 格式，在此格式下提供对 Linux 系统的支持。

1.1 系统要求

NJUthesis 宏包可以在目前大多数的 $\text{T}_{\text{E}}\text{X}$ 系统中使用，例如 $\text{C}_{\text{T}}\text{E}_{\text{X}}$ 、 $\text{M}_{\text{i}}\text{K}_{\text{T}}\text{E}_{\text{X}}$ 、 $\text{t}_{\text{e}}\text{T}_{\text{E}}\text{X}$ 、 $\text{f}_{\text{p}}\text{T}_{\text{E}}\text{X}$ 。

NJUthesis 宏包通过 `ctex` 宏包来获得中文支持。`ctex` 宏包提供了一个统一的中文 $\text{L}_{\text{A}}\text{T}_{\text{E}}\text{X}$ 文档框架，底层支持 CCT 和 CJK 两种中文 $\text{L}_{\text{A}}\text{T}_{\text{E}}\text{X}$ 系统。最新的 `ctex` 宏包可以从 <http://www.ctex.org> 网站下载。

此外，NJUthesis 宏包还使用了宏包 `amsmath`、`amsthm`、`amsfonts`、`amssymb`、`bm`、`geometry`、`enumitem` 和 `hyperref`。目前大多数的 $\text{T}_{\text{E}}\text{X}$ 系统中都包含有这些宏包。即使您系统中的宏包不全，您也可以到 CTAN 下载到这些宏包。

最新的 $\text{C}_{\text{T}}\text{E}_{\text{X}}$ 套装（2.4.1 以上版本）中包含了以上列出的各种宏包，用户无需额外的设置即可使用。

1.2 下载与安装

NJUthesis 宏包的最新版本可以从 <http://njuthesis.googlecode.com/> 网站或者 <http://share.solrex.org/njuthesis/> 网站下载。

对此宏包最简单的使用方法是直接修改 `template.tex` 和 `template.bib` 文件。如果您仅仅想安装宏包，可以将宏包文件放置到 $\text{T}_{\text{E}}\text{X}$ 系统的 `localtexmf/tex/latex/njuthesis` 目录下，然后刷新 $\text{T}_{\text{E}}\text{X}$ 系统的文件名数据库。

关于安装过程的问题可以参考 $\text{C}_{\text{T}}\text{E}_{\text{X}}\text{-FAQ}$ 以及其他 $\text{L}_{\text{A}}\text{T}_{\text{E}}\text{X}$ 教材。

1.3 本模板打包的文件

关键文件：

NJUthesis.cls 文档类文件，是 NJUthesis 宏包的主要文件，用户在文件头 `\documentclass` 声明文档类为 NJUthesis 即可使用此文档类格式化文档。一般不建议用户手工修改该文件。

NJUthesis.cfg 配合文档类的控制文件，里面主要是一些字符串的定义。用户可以编辑修改该文件。

template.tex 随宏包包含的一个模板文件，该模板文件指示了如何使用宏包提供的文档类。用户可以使用该文件作为毕业论文源文件的很好的开端。

template.bib 随宏包包含的模板文件相关的 BibTeX 文献数据库，当用户使用 BibTeX 时可以以其为一个很好的开端；当用户使用自定义的参考文献时，可以在模板文件中注释掉对其的引用并忽略该文件的存在。

辅助文件：

NJUthesis.bst 默认的 BibTeX 样式文件，如果想修改样式，例如使用 `IEEEtran.bst` 作为 BibTeX 样式，只需要将 `IEEEtran.bst` 更名为 `NJUthesis.bst` 即可。

NJUthesis.pdf 宏包帮助文档，其中有对宏包的说明以及对宏包源文件的完全注释，适合高级用户修改文档类文件时参考。

1.4 编译方法

下面以编译模板文件 `template.tex` 生成 `template.pdf` 为例，给出在不同情况下应该使用的编译命令流程。

1.4.1 Windows 平台下

1.4.1.1 标准编译命令

如果您使用 BibTeX 管理参考文献，并且将自己的参考文件条目均添加到 tempalte.bib 中，编译命令应该为：

```
set ARTICLE=template
latex %ARTICLE%.tex
bibtex %ARTICLE%
latex %ARTICLE%.tex
gbk2uni %ARTICLE%
latex %ARTICLE%.tex
dvipdfmx %ARTICLE%.dvi
pause
```

其中，bibtex 命令完成对参考文献的处理，gbk2uni 命令完成对 pdf 书签编码的处理，latex 命令生成 dvi 文件，dvipdfmx 命令生成 pdf 文件。如果您修改了模板文件名，例如将 template.tex 修改成 thesis-yang.tex，那么您只需将第一句中的变量定义修改成 thesis-yang 即可。

您可以将上面的命令保存成一个文件（并且建议这样做），例如 make.bat，这样每次编译时双击该文件，或者在该目录的 Windows 命令行下输入 make 即可完成编译。

1.4.1.2 不使用 BibTeX

尽管我们强烈建议您使用 BibTeX 管理参考文献，如果您不使用 BibTeX 而是自己手写参考文献条目，那么您可以使用以下方式编译模板。

首先在模板文件 template.tex 中注释掉引用 BibTeX 数据库的语句，即：

```
\bibliography{template}
```

并加入手写的参考文献条目。

其次使用下面命令编译模板文件：

```
set ARTICLE=template
latex %ARTICLE%.tex
latex %ARTICLE%.tex
gbk2uni %ARTICLE%
latex %ARTICLE%.tex
dvi2pdf %ARTICLE%.dvi
pause
```

1.4.1.3 使用 pdflatex 命令

我们强烈不建议这样做，主要原因是本模板中文封面中的南京大学校徽和毛笔校名均使用的是 eps 图片，使用 pdflatex 无法对其进行正常处理。如果您希望使用 pdflatex，那么您至少需要将这两张 eps 图片转成 jpg, png 或者 pdf 图片。

出于对矢量图的偏好，我们目前不愿意提供与 pdflatex 兼容的图片格式，也不提供 pdflatex 的编译流程，因此您需要自己处理与 pdflatex 有关的兼容问题。

1.4.1.4 清理垃圾文件

latex 编译时会生成很多垃圾文件，下面的命令可以清理除 pdf 之外的垃圾文件，您可以将其保存成一个 clean.bat 的批处理文件，即可双击执行。

```
del /Q *.aux *.toc *.lon *.lor *.lof *.ilg *.idx *.ind
del /Q *.out *.log *.exa *.nav *.snm *.bbl *.blg *.spl
del /Q *.lot *.bak *~ *.dvi *.ps
pause
```

如果想同时清除 pdf 文件，您可以在后面加上 *.pdf，并保存成 dist-clean.bat 的批处理文件。

1.4.1.5 使用 cygwin 时的 Makefile

在 Windows 下使用 cygwin 时，用 Makefile 来对项目进行编译是再自然不过的想法。下面给出一个编译本模板可以使用的 Makefile。值得注意的是 Makefile 的缩进需要强制使用制表符。


```
ARTICLE=template
ARTICLE_SRC=$(ARTICLE).tex $(ARTICLE).bib CASThesis.cfg
IMAGES=

all: article

article: $(ARTICLE).pdf

$(ARTICLE).pdf: $(ARTICLE_SRC) $(IMAGES)
    latex $(ARTICLE).tex
    bibtex $(ARTICLE)
    latex $(ARTICLE).tex
    gbkt2uni $(ARTICLE)
    latex $(ARTICLE).tex
    dvipdfmx $(ARTICLE).dvi

clean:
    rm -f *.aux *.toc *.lon *.lor *.lof *.ilg *.idx *.ind
    rm -f *.out *.log *.exa *.nav *.snm *.bbl *.blg *.spl
    rm -f *.lot *.bak *~ *.dvi *.ps

distclean: clean
    rm -f *.pdf
```

1.4.2 Linux 平台下

Void.

1.5 问题反馈

用户在使用中遇到问题、需要增加某种功能或者提供一个 patch，都可以到本项目的 Issue 列表中提出，Issue 列表的地址在：

<http://code.google.com/p/njuthesis/issues/list>

欢迎大家反馈自己的使用情况，使我们可以不断改进宏包，但是在提问前请最好检查自己使用的是否为最新版本，以及列表中类似问题是否已经被提出。总之，感谢您的支持!

第二章 论文撰写

2.1 文档类选项

NJUthesis 文档类提供有以下一些选项:

phd/master 当选择 **phd** 时, 生成南大博士论文, 包含国家图书馆格式的封面, 但不包括书脊, 书脊需要单独制作; 选择 **master** 时, 生成南大硕士论文。

oneside/twoside 当选择 **oneside** 时, 生成单面打印格式论文; 当选择 **twoside** 时, 生成双面打印格式论文。双面打印格式会将封面、摘要起始页、章起始页全部放在奇数页上。

dvipdfm/dvips/pdflatex 兼容的编译命令选项, 最好使用默认的 **dvipdfm**。如果您对默认的命令有意见, 那么我们认为您充分了解这几个命令的不同, 那么您足以自己修改图片格式和决定编译流程, 所以使用其它选项时请不要完全使用本文中的帮助。

2.2 封面和摘要页的字符串

NJUthesis 宏包的设置都保存在 `NJUthesis.cfg` 文件中。用户可以在 `.tex` 中通过宏包提供的命令修改设置。对于常用的设置修改, 如培养单位名称、专业名称等, 可以直接在 `NJUthesis.cfg` 文件中进行。各培养单位可以修改后提供本单位统一的 `NJUthesis.cfg` 文件供本单位用户使用。

2.3 内容撰写

请阅读文献《[The Not So Short Introduction to L^AT_EX 2_ε](#)》[12], 或者其有点儿老的中文翻译版《[一份不太简短的 L^AT_EX 2_ε 介绍](#)》[4]。

2.4 插图

请阅读文献《[Using Imported Graphics in L^AT_EX and pdfL^AT_EX](#)》[14], 或者其有点儿老的中文翻译版《[L^AT_EX 2_ε 插图指南](#)》[3]。

2.5 数学公式

请阅读文献《The LaTeX Companion》第 8 章“[Higher Mathematics](#)”[11]。

2.6 参考文献

本节主要介绍参考文献的格式，在 NJUthesis 中控制参考文献的主要有两个地方：

```
\RequirePackage[sort&compress, numbers]{natbib}
```

和

```
\bibliographystyle{NJUthesis}
```

第一个条目是 natbib 宏包，其规定了参考文献的排序和展示等格式，第二个条目是 NJUthesis.bst 文件，规定了参考文献条目的展示格式。期望使用下面的参考文献格式时，必须使用 BibTeX 来处理参考文献。

在 numeric mode 下（正如本文所用，将 numbers 作为 natbib 参数），参考文献的标准引用方式是使用方括号加文献序号的方式，例如 [6]，这时候使用默认的

```
\cite{key}
```

命令即可。但也可以使用 author year 的方式引用，例如 Al-Osh 和 Alzaid [6]，这时候使用

```
\citete{key}
```

命令。

但是，如果您希望在 numeric mode 下使用 authoryear mode（将 authoryear 作为 natbib 参数）的标签样式，例如 Al-Osh 和 Alzaid (1987)，可以使用我们自定义的

```
\citenju{key}
```

命令，下面是一个例子：

```
\Citenju{Al-Osh1987}~引入了~XXX,  
\citenju{Duffie2001}~介绍了~XXX。  
XXX~可见~\citenjup{Freeland1998},  
XX~可见~(\citenjupr{Xu2007}; \citenjupr{Wang1999})。
```

Al-Osh 和 Alzaid (1987) 引入了 XXX, Duffie (2001) 介绍了 XXX。XXX 可见 (Freeland, 1998), XX 可见 (徐亮 等, 2007; 王立洪 和 王金德, 1999)。

2.7 其它问题

请到南京大学小百合论坛科技文献排版讨论区（网址：<http://bbs.nju.edu.cn/bbsdpc?board=Tex>），或者水木清华论坛 L^AT_EX 讨论区（网址：<http://www.newsmth.net/bbsdpc.php?board=Tex>）提问。

第三章 常见问题

3.1 表格

问题 1. 页眉里论文题目和各章标题中的字母均为大写，不能实现大小写的区别，而我写的论文需要在页眉中出现的标题中区分英文字母的大小写比如：YBaCuO而不是YBACUO。

回答 1. 在NJUthesis.cfg 文件中加上

```
\renewcommand\title[2][\NJUT@value@title]{%
  \def\njut@value@title{#2}
  \def\njut@value@titlemark{#1}}
\def\chaptermark#1{\markboth {{\ifnum \c@secnumdepth>\m@ne
  \if@mainmatter\CTEXthechapter \quad\fi
  \fi #1}}{}}%
\def\sectionmark#1{\markright{{\ifnum \c@secnumdepth >\z@
  \CTEXthesection \quad \fi #1}}}
```

3.2 脚注

问题 2. 如果在章节标题中加入注脚，则不仅会出现在本章首页的页脚，也会出现在目录的页脚，不知是否能够让其不要出现在目录的页脚中。

回答 2. 可以使用如下的命令来定义章节的标题

```
\chapter[出现在目录和页眉的标题]{出现在正文的标题\footnote{这个
不会出现在目录中。}}
```

section、subsection 等命令也有类似的用法。

第四章 总结与展望

附录 A 博士(硕士)学位论文编写格式规定(试行)

A.1 适用范围

本规定适用于博士学位论文编写，硕士学位论文编写应参照执行。

A.2 引用标准

GB7713科学技术报告、学位论文和学术论文的编写格式。

GB7714文后参考文献著录规则。

A.3 印制要求

论文必须用白色纸印刷，并用A4(210mm×297mm)标准大小的白纸。纸的四周应留足空白边缘，上方和左侧应空边25mm以上，下方和右侧应空边20mm以上。除前置部分外，其它部分双面印刷。

论文装订不要用铁钉，以便长期存档和收藏。

论文封面与封底之间的中缝（书脊）必须有论文题目、作者和学校名。

A.4 编写格式

论文由前置部分、主体部分、附录部分(必要时)、结尾部分(必要时)组成。

前置部分包括封面，题名页，声明及说明，前言，摘要(中、英文)，关键词，目次页，插图和附表清单(必要时)，符号、标志、缩略词、首字母缩写、单位、术语、名词解释表(必要时)。

主体部分包括绪论(作为正文第一章)、正文、结论、致谢、参考文献表。

附录部分包括必要的各种附录。

结尾部分包括索引和封底。

A.5 前置部分

A.5.1 封面（博士论文国图版用）

1. 封面是论文的外表面，提供应有的信息，并起保护作用。

2. 封面上应包括下列内容:

- (a) 分类号在左上角注明分类号, 便于信息交换和处理。一般应注明《中国图书资料分类法》的类号, 同时应注明《国际十进分类法UDC》的类号;
- (b) 密级在右上角注明密级;
- (c) “博士学位论文”用大号字标明;
- (d) 题名和副题名用大号字标明;
- (e) 作者姓名;
- (f) 学科专业名称;
- (g) 研究方向;
- (h) 导师姓名, 职称;
- (i) 日期包括论文提交日期和答辩日期;
- (j) 学位授予单位。

A.5.2 题名

1. 题名是以最恰当、最简明的词语反映论文中最重要的特定内容的逻辑组合。

题名所用每一词语必须考虑到有助于选定关键词和编写题录、索引等二次文献可以提供检索的特定实用信息。

题名应避免使用不常见的缩略词、首字母缩写字、字符、代号和公式等。

题名一般不宜超过20字。

论文应有外文题名, 外文题名一般不宜超过10个实词。

2. 可以有副题名。
3. 题名在整本论文中不同地方出现时, 应完全相同。

A.5.3 前言

前言是作者对本论文基本特征的简介, 如论文背景、主旨、目的、意义等并简述本论文的创新性成果。

A.5.4 摘要

1. 摘要是论文内容不加注释和评论的简单陈述。
2. 论文应有中、英文摘要，中、英文摘要内容应相同。
3. 摘要应具有独立性和自含性，即不阅读论文的全文，便能获得必要的信息，摘要中有数据、有结论，是一篇完整的短文，可以独立使用，可以引用，可以用于推广。摘要的内容应包括与论文同等量的主要信息，供读者确定有无必要阅读全文，也供文摘等二次文献引用。摘要的重点是成果和结论。
4. 中文摘要一般在1500字，英文摘要不宜超过1500实词。
5. 摘要中不用图、表、化学结构式、非公知公用的符号和术语。

A.5.5 关键词

关键词是为了文献标引工作从论文中选取出来用于表示全文主题内容信息款目的单词或术语。

每篇论文选取3—8个词作为关键词，以显著的字符另起一行，排在摘要的左下方。在英文摘要的左下方应标注与中文对应的英文关键词。

A.5.6 目次页

目次页由论文的章、节、附录等的序号、名称和页码组成，另页排在摘要的后面。

A.5.7 插图和附表清单

论文中如图表较多，可以分别列出清单并置于目次页之后。

图的清单应有序号、图题和页码。表的清单应有序号、表题和页码。

A.5.8 符号、标志、缩略词、首字母缩写、计量单位、名词、术语等的注释表符号、标志、缩略词、首字母缩写、计量单位、名词、术语等的注释说明汇集表，应置于图表清单之后。

A.6 主体部分

A.6.1 格式

主体部分由绪论开始，以结论结束。主体部分必须由另页右页开始。每一

章必须另页开始。全部论文章、节、目的格式和版面安排要划一，层次清楚。

A.6.2 序号

1. 论文的章可以写成：第一章。节及节以下均用阿拉伯数字编排序号，如1.1，1.1.1等。
2. 论文中的图、表、附注、参考文献、公式、算式等一律用阿拉伯数字分别分章依序连续编排序号。其标注形式应便于互相区别，一般用下例：图1.2；表2.3；附注1)；文献[4]；式(6.3)等。
3. 论文一律用阿拉伯数字连续编页码。页码由首页开始，作为第1页，并为右页另页。封页、封二、封三和封底不编入页码，应为题名页、前言、目次页等前置部分单独编排页码。页码必须标注在每页的相同位置，便于识别。
4. 附录依序用大写正体A、B、C ---编序号，如：附录A。附录中的图、表、式、参考文献等另行编序号，与正文分开，也一律用阿拉伯数字编码，但在数码前题以附条序号，如图A.1；表B.2；式(B.3)；文献[A.5]等。

A.6.3 绪论

绪论（综述）：简要说明研究工作的目的、范围、相关领域的前人工作和知识空白、理论基础和分析，研究设想、研究方法和实验设计、预期结果和意义等。一般在教科书中有的知识，在绪论中不必赘述。

绪论的内容应包括论文研究方向相关领域的最新进展、对有关进展和问题的评价、本论文研究的命题和技术路线等；绪论应表明博士生对研究方向相关的学科领域有系统深入的了解，论文具有先进性和前沿性；

为了反映出作者确已掌握了坚实的基础理论和系统的专门知识，具有开阔的科学视野，对研究方案作了充分论证，绪论应单独成章，列为第一章，绪论的篇幅应达1~2万字，不得少于1万字；绪论引用的文献应在100篇以上，其中外文文献不少于60应按正文中引用的先后排列。

A.6.4 正文

论文的正文是核心部分，占主要篇幅。正文必须实事求是，客观真切，准确完备，合乎逻辑，层次分明，简便可读。

正文的每一章(除绪论外)应有小结, 在小结中应明确阐明作者在本章中所做的工作, 特别是创新性成果。凡本论文要用的基础性内容或他人的成果不应单独成章, 也不应作过多的阐述, 一般只引结论、使用条件等, 不作推导;

A.6.4.1 图

图包括曲线图、构造图、示意图、图解、框图、流程图、记录图、布置图、地图、照片、图版等。

图应具有“自明性”, 即只看图、图题和图例, 不阅读正文, 就可以理解图意。

图应编排序号。每一图应有简短确切的图题, 连同图号置于图下。必要时, 应将图上的符号、标记、代码, 以及实验条件等, 用最简练的文字, 横排于图题下方, 作为图例说明。

曲线图的纵、横坐标必须标注“量、标准规定符号、单位”。此三者只有在不必要标明(如无量纲等)的情况下方可省略。坐标上标注的量的符号和缩略词必须与正文一致。

照片图要求主题和主要显示部分的轮廓鲜明, 便于制版。如用放大缩小的复制品, 必须清晰, 反差适中。照片上应该有表示目的物尺寸的标度。

A.6.4.2 表

表的编排, 一般是内容和测试项目由左至右横读, 数据依序竖排。表应有自明性。

表应编排序号。

每一表应有简短确切的表题, 连同标号置于表上。必要时, 应将表中的符号、标记、代码, 以及需要说明事项, 以最简练的文字, 横排于表题下, 作为表注, 也可以附注于表下。表内附注的序号宜用小号阿拉伯数字并加圆括号置于被标注对象的右上角, 如: $x \times x1$), 不宜用“*”, 以免与数学上共轭和物质转移的符号相混。

表的各栏均应标明“量或测试项目、标准规定符号、单位”。只有在无必要标注的情况下方可省略。表中的缩略词和符号, 必须与正文中一致。

表内同一栏的数字必须上下对齐。表内不宜用“同上”, “同左”和类似词, 一律填入具体数字或文字。表内“空白”代表未测或无此项, “—”或“---”(因“—”可能与代表阴性反应相混)代表未发现, “0”代表实测结果确为零。

如数据已绘成曲线图, 可不再列表。

A.6.4.3 数学、物理和化学式

正文中的公式、算式或方程式等应编排序号，序号标注于该式所在行(当有续行时，应标注于最后一行)的最右边。

较长的式，另行居中横排。如式必须转行时，只能在 $+$ 、 $-$ 、 \times 、 \div 、 $\frac{1}{2}$ 、 $\frac{3}{4}$ 处转行。上下式尽可能在等号“=”处对齐。

小数点用“.”表示。大于999的整数和多于三位数的小数，一律用半个阿拉伯数字字符的小间隔分开，不用千位撇。对于纯小数应将0列于小数点之前。

示例：应该写成94 652.023 567；0.314 325，不应写成94, 652.023, 567；.314, 325

应注意区别各种字符，如：拉丁文、希腊文、俄文、德文花体、草体；罗马数字和阿拉伯数字；字符的正斜体、黑白体、大小写、上下脚标(特别是多层次，如“三踏步”)、上下偏差等。

A.6.4.4 计量单位

报告、论文必须采用国务院发布的《中华人民共和国法定计量单位》，并遵照《中华人民共和国法定计量单位使用方法》执行。使用各种量、单位和符号，必须遵循附录B所列国家标准的规定执行。单位名称和符号的书写方式一律采用国际通用符号。

A.6.4.5 符号和缩略词

符号和缩略词应遵照国家标准的有关规定执行。如无标准可循，可采纳本学科或本专业的权威性机构或学术团体所公布的规定；也可以采用全国自然科学名词审定委员会编印的各学科词汇的用词。如不得不引用某些不是公知公用的、且又不易为同行读者所理解的、或系作者自定的符号、记号、缩略词、首字母缩写字等时，均应在第一次出现时一一加以说明，给以明确的定义。

A.6.5 结论

报告、论文的结论是最终的、总体的结论，不是正文中各段的小结的简单重复。结论应该准确、完整、明确、精炼。在结论中要清楚地阐明论文中有那些自己完成的成果，特别是创新性成果；

如果不可能导出应有的结论，也可以没有结论而进行必要的讨论。可以在结论或讨论中提出建议、研究设想、仪器设备改进意见、尚待解决的问题等。

A.6.6 致谢

可以在正文后对下列方面致谢：

国家科学基金、资助研究工作的奖学金基金、合作单位、资助或支持的企业、组织或个人；

协助完成研究工作和提供便利条件的组织或个人；

在研究工作中提出建议和提供帮助的人；

给予转载和引用权的资料、图片、文献、研究思想和设想的所有者；

其他应感谢的组织或个人。

A.6.7 参考文献表

A.6.7.1 专著著录格式

主要责任者，其他责任者，书名，版本，出版地：出版者，出版年

例：1. 刘少奇，论共产党员的修养，修订2版，北京：人民出版社，1962

A.6.7.2 连续出版物中析出的文献著录格式

析出文献责任者，析出文献其他责任者，析出题名，原文献题名，版本：文献中的位置。

例：2. 李四光，地壳构造与地壳运动，中国科学，1973 (4)：400—429

A.6.7.3 参考文献采用顺序编码制，按论文正文所引用文献出现的先后顺序连续编码。

A.7 附录

附录是作为报告、论文主体的补充项目，并不是必需的。

1. 下列内容可以作为附录编于报告、论文后，也可以另编成册；
 - (a) 为了整篇论文材料的完整，但编入正文又有损于编排的条理和逻辑性，这一材料包括比正文更为详尽的信息、研究方法和技术更深入的叙述，建议可以阅读的参考文献题录，对了解正文内容有用的补充信息等；
 - (b) 由于篇幅过大或取材于复制品而不便于编入正文的材料；
 - (c) 不便于编入正文的罕见珍贵资料；

- (d) 对一般读者并非必要阅读, 但对本专业同行有参考价值的资料;
- (e) 某些重要的原始数据、数学推导、计算程序、框图、结构图、注释、统计表、计算机打印输出件等。

2. 附录与正文连续编页码。

3. 每一附录均另页起。

A.8 结尾部分(必要时)

为了将论文迅速存储入电子计算机, 可以提供有关的输入数据。可以编排分类索引、著者索引、关键词索引等。

参考文献

- [1] 徐亮, 周影辉, 韦博成, 2007. 基于 mm 算法的 lad 回归的影响分析. 高校应用数学学报 A 辑, 48–58.
- [2] 王立洪, 王金德, 1999. 时间序列模型 11-估计量的极限分布:平稳自回归模型. 数学学报42.
- [3] 王磊, 1997. \LaTeX 插图指南.
- [4] 中国 CTeX 用户小组, 2007. 一份不太简短的 $\text{\LaTeX} 2_{\epsilon}$ 介绍.
- [5] Adomavicius, G., Tuzhilin, A., 2005. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. Knowledge and Data Engineering, IEEE Transactions on 17, 734–749.
- [6] Al-Osh, M.A., Alzaid, A.A., 1987. First-order integer-valued autoregressive (I-NAR(1)) process. Journal of Time Series Analysis 8, 261–275.
- [7] DoubanFM, . <http://douban.fm>.
- [8] Duffie, D., 2001. Dynamic Asset Pricing Theory. Princeton University Press. 3rd edition.
- [9] Freeland, R., 1998. Statistical analysis of discrete time series with applications to the analysis of workers compensation claims data. Ph.D. thesis. University of British Columbia, Canada.
- [10] Lastfm, . <http://last.fm>.
- [11] Michel Goossens, Frank Mittelbach, A.S., 1997. The LaTeX Companion. Addison Wesley Longman, Inc.. chapter 8 Higher Mathematics. pp. 215–258.
- [12] Oetiker, T., Partl, H., Hyna, I., Schlegl, E., 2008. The Not So Short Introduction to LaTeX2 ϵ (Or LaTeX 2 ϵ in 141 minutes). ctan.org. 4.26 edition.
- [13] Pandora, . <http://pandora.com>.
- [14] Reckdahl, K., 2006. Using imported graphics in \LaTeX and pdf \LaTeX .

- [15] Rentfrow, P.J., Gosling, S.D., 2003. The do re mi's of everyday life: the structure and personality correlates of music preferences. *Journal of personality and social psychology* 84, 1236.
- [16] Scheibehenne, B., Greifeneder, R., Todd, P.M., 2010. Can there ever be too many options? a meta-analytic review of choice overload. *Journal of Consumer Research* 37, 409–425.
- [17] Song, Y., Dixon, S., Pearce, M., 2012. A survey of music recommendation systems and future perspectives, in: 9th international symposium on computer music modelling and retrieval (CMMR 2012), pp. 19–22.
- [18] Xiami, . <http://www.xiami.com>.
- [19] Yang, C.C., Chen, H., Hong, K., 2003. Visualization of large category map for internet browsing. *Decision Support Systems* 35, 89–102.

简历与科研成果

基本情况

杨文博，男，汉族，1986 年 11 月出生，河南省民权县人。

教育背景

2007.9~2010.6	中国科学院研究生院信息安全国家重点实验室	硕士
2003.9~2007.6	南京大学数学系	本科

攻读硕士学位期间完成的学术成果

- [1] Wenbo Yang and Wen Tao Zhu, “Voting-on-Grid Clustering for Secure Localization in Wireless Sensor Networks,” in *Proc. IEEE International Conference on Communications (ICC) 2010*, May. 2010.
- [2] Wenbo Yang and Wen Tao Zhu, “Protecting Source Location Privacy in Wireless Sensor Networks with Data Aggregation,” in *Proc. 7th International Conference on Ubiquitous Intelligence and Computing (UIC) 2010*, Oct. 2010.

攻读硕士学位期间参与的科研课题

1. 国家自然科学基金面上项目“无线传感器网络在知识获取过程中的若干安全问题研究”（课题年限 2010.1~2012.12），负责位置相关安全问题的研究。
2. 中国科学院知识创新工程重要方向项目下属课题“下一代移动通信安全机制研究”（课题年限 2010.1~2010.12），负责 LTE/SAE 认证相关的安全问题研究。

致 谢

首先感谢XXX