



北京大學

硕士研究生学位论文

题目：基于 XXXX 的 XXXX 系
统设计与实现

姓名：扎克·施耐德

学号：180XXXXXXXX

院系：XXXXXX 学院

专业：XXXX

研究方向：XXXX

导师：XXX 教授

YYY 教授

二〇二一年六月

版权声明

任何收存和保管本论文各种版本的单位和个人，未经本论文作者同意，不得将本论文转借他人，亦不得随意复制、抄录、拍照或以其他方式传播。否则，引起有碍作者著作权之问题，将可能承担法律责任。



摘要

中文摘要部分...

关键词：A, B, C, D

Design and implementation of a XXXXX system based on XXXX

Zack Snyder (XXXX)

Directed by Prof. XXX and Prof. YYY

ABSTRACT

英文摘要部分...

KEY WORDS: A,B,C,D

目录

第一章 引论	1
1.1 封面及 pkuthssinfo 相关	1
1.2 字体设定	2
1.3 版权声明与原创页	2
1.4 摘要部分	2
1.5 目录	3
1.6 主要符号对照表.....	3
1.7 图表相关	3
1.7.1 表格样例	3
1.7.2 图片样例	4
1.8 公式	5
1.9 参考文献	6
1.10 其他	6
1.11 与 pkuthss v1.9.0 异同	6
第二章 理论基础	7
2.1 Overleaf 适配存在的问题.....	7
2.2 适配细节	7
2.2.1 C _T E _X 字库特点	7
2.2.2 fontset 设定.....	8
2.2.3 编码方式	9
2.2.4 粗体设定	9
第三章 Model 1	11
第四章 Model 2	13
第五章 Model 3	15
第六章 总结和展望	17
参考文献	19
攻读硕士期间发表的论文及其他成果	21

致谢	23
北京大学学位论文原创性声明和使用授权说明	25

主要符号对照表

x, y, m, n, t	标量, 通常为变量
K, L, D, M, N, T	标量, 通常为超参数
$x \in \mathbb{R}^D$	D 维列向量
(x_1, \dots, x_D)	D 维行向量
$(x_1, \dots, x_D)^T$ or $(x_1; \dots; x_D)^T$	D 维行向量
$A \in \mathbb{R}^{K \times D}$	大小为 $K \times D$ 的矩阵
$x \in \mathbb{R}^{KD}$	(KD) 维的向量
\mathbb{M}_i or $\mathbb{M}_i(\mathbf{x})$	第 i 列为 $\mathbf{1}$ (或者 \mathbf{x}), 其余为 $\mathbf{0}$ 的矩阵
$diag(\mathbf{x})$	对角矩阵, 其对角元素为 \mathbf{x}
\mathbf{I}_N or I	$(N \times N)$ 的单位阵
$diag(\mathbf{A})$	列向量, 其元素为 \mathbf{A} 的对角元素
$\mathbf{A} \in \mathbb{R}^{D_1 \times D_2 \times \dots \times D_K}$	大小为 $D_1 \times D_2 \times \dots \times D_K$ 的张量
$\{\mathbf{x}^{(n)}\}_{n=1}^N$	集合
$\{(\mathbf{x}^{(n)}, \mathbf{y}^{(n)})\}_{n=1}^N$	数据集
$\mathcal{N}(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$	变量 x 服从均值为 $\boldsymbol{\mu}$, 方差为 $\boldsymbol{\Sigma}$ 的高斯分布

① 本符号对照表内容选自邱锡鹏老师的《神经网络与深度学习》^[1]一书。

第一章 引论

本章为 `iofu728/pkuthss`^① 文档类的示例文档，对硕士学位论文写作过程中常见用法和问题进行介绍说明。

1.1 封面及 `pkuthssinfo` 相关

`pkuthssinfo` 相关配置与 `pkuthss` 原文档类完全一致，均为参数配置，可参考 `pkuthss` 文档^②。目前 `iofu728/pkuthss` 基于 `pkuthss v1.9.0` 进行开发。

封面部分参考《北京大学研究生学位论文写作指南 V2.0-2019》(以下简称写作指南) 第 1.1 节中描述和《硕士论文模板 2020》进行修改。

```
\pkuthss{}info{
  cthesisname = {硕士学位论文},
  thesiscover = {硕士研究生学位论文},
  ethesisname = {Master Thesis},
  ctitle = {基于 XXXX 的 XXXX 系统设计与实现},
  etitle = {Design and implementation of a XXXXX system based on XXXX},
  cauthor = {扎克·施耐德}, eauthor = {Zack Snyder},
  studentid = {180XXXXXXXX},
  % 具体时间以教务为准，初稿 3 月，送审 4 月，答辩 5 月，最终 6 月。
  date = {\zhdigits{2021}\ \ 年\ \ \zhnumber{6}\ \ 月},
  school = {XXXXXX 学院},
  cmajor = {XXXX}, emajor = {XXXX},
  direction = {XXXX},
  mentorlines = {2}, % 导师个数
  % 副教授 A.P. 讲师 Lec.
  cmentor = {XXX\ \ 教授\ \YYY\ \ 教授}, ementor = {Prof.\ XXX and Prof.\ YYY},
  ckeywords = {A, B, C, D},
  ekeywords = {A,B,C,D},
  % 盲审模式参数，需在 documentclass 增加 blind
  blindid = {XXXXXXXXXX}, discipline = {XXXX}
}
```

① `iofu728/pkuthss` 是基于 `pkuthss`^[2] 针对硕士学位论文要求进行适配的 L^AT_EX 文档类，符合北京大学硕士学位论文写作规范，并能通过图书馆审查。其官方仓库为 <https://github.com/iofu728/pkuthss>，当前版本 v1.0.0。

② <https://bbs.pku.edu.cn/attach/c8/3e/c83e980c93b838a3/pkuthss-bootstrap-0.1.7.pdf>

1.2 字体设定

本文档类提供五种默认字体设定分别为`windows`, `windows@overleaf`, `mac`, `ubuntu`, `fandol`。其中，默认情况，Overleaf 平台下仅可使用 `fandol`, `ubuntu` 两种模式，**推荐使用 `fandol` 模式 (默认)**。如遇字符不显示问题，可使用 `ubuntu` 模式，或者自行收集上传所需的字体文件 `simsun.ttf`, `simhei.ttf`, `simfang.ttf`, `simkai.ttf` 至根目录并使用 `windows@overleaf` 模式 (注意文件名称和版权问题)，亦或者下载至本地 `windows` 环境使用 `windows` 模式。详细可见第二章文档内容。

1.3 版权声明与原创页

写作指南中要求各学生从校内门户或者从研究生院网站下载相应的文件进行签名后扫描替换或者直接替换。**需要注意的是**，门户生成的 PDF 文件未嵌入字体，其指定字体为华文宋体 (即 STSong)。由于各个 PDF 浏览器和操作系统预设的字体不同，导致呈现效果差异较大 (据不完全统计，Mac 端 PDF Expert 显示的是苹方简体，Chrome 显示的是方正粗宋，而 Acrobat 显示的是 AdobeSongStd，Windows 端 Edge 显示的是方正悠黑，在这之中 Acrobat 字体效果最为接近)。请务必使用 Acrobat 进行打印，或者使用本包生成文件进行打印^①。

此外，写作指南中申明目录中需要保留原创页^②。本文使用 `textblock` 和 `colorbox` 宏包通过遮掩和覆盖的方式实现保留目录前提下的的 PDF 文件插入，其中参数按照 PDF 格式为 A4 进行调校，如有需要可以进行微调。

```
\begin{textblock}{1}{(-0.8,-0.08)
\colorbox{white}{
  \includegraphics[height = 1.2448\textheight]{文件路径}
}
\end{textblock}
```

此外在论文电子版制作过程中，需要替换原创页签名扫描件，使用相同方式即可。

1.4 摘要部分

在 `\begin{cabstract}` 和 `\begin{eabstract}` 环境中进行书写。如论文工作受到基金资助，需要在中文摘要第一页的页脚处标注：本研究得到某某基金 (编号: xxx) 资助。

① 由于样例中的 PDF 被编辑过，最终生成文件在不同 PDF 浏览器下仍然显示不同，使用门户下载文件生成则可以保持在不同 PDF 浏览器效果一致。

② 对于这点部分人解读写作指南为目录中不需要，实际上写作指南第 1.10 节中声明，致谢 (后记或说明)、学位论文原创性声明和授权使用说明是论文的最后两项内容，目录中和章平级。

1.5 目录

按照写作指南和《硕士论文模板 2020》进行调整, 1) 对章级增加点线; 2) 对间距和字体进行调整。

1.6 主要符号对照表

参考chap/deno.tex 即可, 在`\begin{denotation}` 环境下, 使用`\item[X]` Y 分别表示符号及其说明。

已知问题: 符号处不能输入中括号 [,]。

1.7 图表相关

1.7.1 表格样例

一般学术论文需要使用三线表(如表 1.2), 需要依赖宏包booktabs, 使用`\toprule`, `\midrule`, `\bottomrule` 控制三线。此外表序和表名位于表格的上方。如果需要对表格内进行脚注, 可通过 `minipage` 中嵌套 `tabular` 来实现, 具体可参考 Stack Overflow^①。

表 1.1 续表样列表。

年龄 (岁)	性别	cp 胸痛型	静息血压 毫米汞柱	chol 胆固醇	空腹血糖 > 120 mg/dl	restecg 静息状态	thalachh 最大心率
63	1	3	145	233	1	0	150
37	1	2	130	250	0	1	187
41	0	1	130	204	0	0	172
56	1	1	120	236	0	1	178
57	0	0	120	354	0	1	163
57	1	0	140	192	0	1	148
56	0	1	140	294	0	0	153
44	1	1	120	263	0	1	173
52	1	2	172	199	1	1	162
57	1	2	150	168	0	1	174
54	1	0	140	239	0	1	160
48	0	2	130	275	0	1	139

续下页

^① <https://stackoverflow.com/questions/2888817/footnotes-for-tables-in-latex>

续表 1.1 续表样列表。

年龄 (岁)	性别	cp 胸痛型	静息血压 毫米汞柱	chol 胆固醇	空腹血糖 > 120 mg/dl	restecg 静息状态	thalachh 最大心率
49	1	1	130	266	0	1	171
64	1	3	110	211	0	0	144

注：数据来源于 Kaggle Heart Attack Analysis & Prediction Data Set。

如需要注明表格中数据来源，则可使用类似的方式，见表 1.3。

表 1.2 表格脚注样列表。表名可通过中括号添加缩略名。

	X	Y	Z	N	M
默认	99.99	99.99	99.99	99.99 ^①	99.99
w/o X	99.99	99.99	99.99	99.99	99.99
w/o Y	99.99	99.99	99.99	99.99	99.99
w/o Z	99.99 ^②	99.99	99.99	99.99	99.99
w/o N	99.99	99.99	99.99	99.99	99.99
w/o M	99.99	99.99	99.99	99.99	99.99

① 表格中的脚注 1

② 表格中的脚注 2

表 1.3 表格数据来源注释样列表。

Model	数据集 A			数据集 B		
	指标 a(%)	指标 b(%)	指标 c	指标 a (%)	指标 b(%)	指标 c
Devlin et al. ^[3]	99.99	99.99	99.99	99.99	99.99	99.99
Yang et al. ^[4]	99.99	99.99	99.99	99.99	99.99	99.99

注：数据来源 XXXXXX。

当表格较大，不能在一页内打印时，可以“续表”的形式另页打印，可使用宏包 `longtable` 实现，如表 1.1。

1.7.2 图片样例

当需要插入多个子图的时候，可以选用宏包 `subfloat`，不推荐使用 `subfigure` 和 `subtable`。



(a) 北京大学校徽



(b) 北京大学中文校名，依照北京大学标识管理办公室出具的北大标识使用基本规范进行使用

图 1.1 包含子图形的大图形

若使用继承于`subfigure`的宏包，例如`subfloat`、`subfigure`等，则可直接使用引用`\ref{sfig:xxxx}`引用子图 label，如图 1.1(a)。否则需要引用主图，再单独标注子图序号，以便符合学位论文要求。

此外，与表格相反，图序和图名需要位于图片的下方。如果含有子图，每个子图需要具有相应的子图名。

如果需要并排使用两个独立的图形，分别编排图序，则可使用`minipage`，如图 1.2和图 1.3。



图 1.2 北京大学校徽



图 1.3 北京大学中文校名，依照北京大学标识管理办公室出具的北大标识使用基本规范进行使用

1.8 公式

公式部分考虑到写作指南中无关于公式页的说明，并未做改动，使用通用`LATEX`规范即可。对于复杂公式需求，可使用`amsmath`宏包结合`Mathpix`^①等自动化识别工具。

$$\int_a^b \left\{ \int_a^b [f(x)^2 g(y)^2 + f(y)^2 g(x)^2] - 2f(x)g(x)f(y)g(y) dx \right\} dy$$

$$= \int_a^b \left\{ g(y)^2 \int_a^b f^2 + f(y)^2 \int_a^b g^2 - 2f(y)g(y) \int_a^b fg \right\} dy$$

上述公式来源于刘宝碁的《不确定规划》刘宝碁^[5]。

① <https://mathpix.com/>

1.9 参考文献

参考文献根据写作指南使用 `gb7714-2015bibstyle` 进行管理，具体引用命令与日常使用类似，`\cite{}`，`\citet{}`，`\citeauthor{}`，具体用法见相应文档^②。

例如 `\cite{devlin2018bert}=[3]`，`\citeauthor{gut2013probability}=Gut, ...` 相对于的 `bib` 文件的书写基本上直接用 Google Scholar 拷贝的 BibTeX 即可，部分属性按提示进行微调。

```
\usepackage[backend=biber,bibstyle=gb7714-2015,citestyle=gb7714-2015]{biblatex}
```

1.10 其他

正文不建议使用四级目录 `\subsubsection{}`。

本示例文档参考写作指南，《硕士论文模板 2020》，《清华大学学位论文 L^AT_EX 模板使用示例文档》和《pkuthss 使用说明》进行书写。遵循 L^AT_EX Project Public License 和 Attribution 4.0 International (CC BY 4.0) 开源协议。

1.11 与 pkuthss v1.9.0 异同

格式方面：

1. “关键词” + “KEY WORDS” 非粗体
2. “题目” key 字号 2 号，value 字号 1 号
3. “姓名” key 字号小 3
4. 隐藏超链接
5. 目录字体、样式 (点线)、间距

功能方面：

1. 增加主要符号对照表
2. 脚注从当前页开始标注
3. 表格内脚注样式
4. 子图引用格式
5. 字体模式
6. 简化 `blind` 模式下用户设定
7. Windows 下中易宋体的粗体用假粗体替代

^② <https://github.com/hushidong/biblatex-gb7714-2015>

第二章 理论基础

本章对 pkuthss 适配 Overleaf 平台的工作进行阐述和整理，希望对有兴趣了解和有修改相应需求的同学有所帮助。本人才疏学浅，如有纰漏欢迎交流指正。

2.1 Overleaf 适配存在的问题

Overleaf 是一个开源在线实时协作 L^AT_EX 编辑器，其 www.overleaf.com 托管版本是基于 Ubuntu 构建的。Overleaf 适配过程中遇到的问题不是 T_EX Live 版本^①，也不是 T_EX 排班引擎/方式的问题^②，主要问题在于如何在 Overleaf 平台设置满足学位论文要求的字体配置。

2.2 适配细节

pkuthss 适配 Overleaf 问题可追溯到 Casper/pkuthss 的 Issue#28^③。事实上当时 @lianze 已经给出了正确的解决方案。

2.2.1 C_TE_X 字库特点

pkuthss 使用 C_TE_X 宏集进行中文排版，具体为文档类 `ctexbook`。而默认情况下，C_TE_X 宏集根据编译方式和操作系统指定相应字库。表 2.1 中归纳了默认状态下各个操作系统和编译方式对应的字库使用策略。

由于本文选用了 X_ƎL^AT_EX 编译方式进行适配，而原始文档类 pkuthss 使用的是 Windows 下 `ctex-fontset` 设定，Overleaf 托管版本基于 Linux 系统，两者区别在于原始文档类 pkuthss 使用商用字体库 (例如中易黑体、微软雅黑等)，而 Overleaf 托管版本使用开源中文字体库 Fandol、思源字体等。

原先 pkuthss 在 `ctex-fontset-pkuthss.def` 定义了使用的字体依赖，具体情况如下：

- `\songti`，中易宋体，作为默认中文字体使用，衬线字体，`\textrm`；
- `\heiti`，中易黑体，无衬线字体，`\bfseries`，`\textbf` 和 `\textsf`；
- `\fangsong`，中易仿宋，等宽字体，`\texttt`；
- `\kaishu`，中易楷体，`\textit`；

① Overleaf 托管版本目前支持 2014-2020 共 8 种 T_EX Live 版本。

② Overleaf 托管版本目前支持 L^AT_EX，pdfL^AT_EX，LuaL^AT_EX，X_ƎL^AT_EX 四种编译方式。而 pkuthss 是支持 X_ƎL^AT_EX、L^AT_EX+DVIPDFMx 和 pdfL^AT_EX 三种模式，本文选择 X_ƎL^AT_EX 编译方式进行适配。

③ <https://github.com/CasperVector/pkuthss/issues/28>

- `\lishu`, `\youyuan` 这两者在 Linux 下不存在, 在 windows、founder、macnew 字库中才存在。原 `pkuthss` 文档类中仅申明未使用。

 表 2.1 C_TE_X 宏集自动配置字体策略

	macOS Old ^①	macOS New ^②	Windows ^③	其他 ^④
X _Y L _A T _E X	xeCJK 华文字库	xeCJK 华文字库 + 苹方	xeCJK 中易字库 + 微软雅黑	xeCJK Fandol 字库 ^⑤
LuaL _A T _E X ^⑥	LuaTeX-ja 华文字库	LuaTeX-ja 华文字库 + 苹方	LuaTeX-ja 中易字库 + 微软雅黑	LuaTeX-ja Fandol 字库
pdfL _A T _E X	不可用	不可用	CJK + zhmetrics 中易字库 + 微软雅黑 ^⑦	不可用
L _A T _E X + DVIPDFM _x	不可用	CJK + zhmetrics 华文字库 + 苹方	CJK + zhmetrics 中易字库 + 微软雅黑 ^⑦	CJK + zhmetrics Fandol 字库
upL _A T _E X + DVIPDFM _x	不可用	zhmetrics-uptex 华文字库 + 苹方	zhmetrics-uptex 中易字库 + 微软雅黑	zhmetrics-uptex Fandol 字库

注: 数据来源与 C_TE_X 宏集手册 v2.5.6。

① Yosemite (10.10) 及以前的 macOS 系统。

② El Capitan (10.11) 及以后的 macOS 系统。

③ 仅支持 Windows Vista 及以后的 Windows 操作系统。

④ C_TE_X 将其他系统统一归为 Linux。

⑤ 由马起园、苏杰、黄晨成等人开发的开源中文字体, 参见: <https://www.ctan.org/pkg/fandol>。

⑥ LuaL_AT_EX 编译时使用 LuaTeX-ja 宏包。

⑦ 微软雅黑字体并不总是有效, 与选项 `zhmap` 的取值有关。

2.2.2 fontset 设定

`iofu728/pkuthss` 在 `documentclass` 中增加 `fontset` 参数, 并预设了 5 种字体模式 (另外还有 `auto,nono` 两种自适应模式), 将字体设置能力暴露给用户, 从而解决跨端的字体适配问题。

查阅写作指南, 其中说明中文字体使用宋体、仿宋和黑体, 英文字体使用 Times New Roman 和 Arial, 共五种字体, 对楷体、隶书、幼圆等字体并未提及。

由于 Overleaf 托管版本只拥有开源字体版权^①, 无商用字体版权 (例如中易字集)。

本文档类申明的 5 种字体模式 `fontset` 对应使用字体情况如表 2.2 所示。其中, `windows` 和 `mac` 模式为商业字体。默认情况下, 不能在 Overleaf 平台上使用。而 `fandol` 和 `ubuntu` 模式为开源字体, 可在 Overleaf 平台上使用。

而对于中文字体, 不同种类的字体实现细节存在细微差异, 所对应的字符数也不同, 对于极生僻词开源字体可能会出现显示异常的情况, 如 `𠄎` yáo, `fandol` 模式下不可

① 具体清单详见 Overleaf 字体说明文档 https://www.overleaf.com/learn/latex/Questions/Which_OTF_or_TTF_fonts_are_supported_via_fontspeg%3F

表 2.2 iofu728/pkuthss 预设的五种字体模式

	windows ^①	mac	ubuntu	fandol
宋体	中易宋体	华文宋体	思源宋体	Fandol 宋体
黑体	中易黑体	华文黑体	思源黑体	Fandol 黑体
仿宋	中易仿宋	华文仿宋	Fandol 仿宋	Fandol 仿宋

① 与 windows@overleaf 模式相同

见。但考虑到版权问题和字体效果，**Overleaf** 平台中仍然推荐使用 **fandol 模式**。如在使用中发现较多字符不显示，可考虑使用 **ubuntu** 模式，或者自行收集上传 **simsun.ttf**, **simhei.ttf**, **simfang.ttf**, **simkai.ttf** 至根目录并使用 **windows@overleaf** 模式 (注意文件名称和版权问题)，亦或者下载至本地 **windows** 环境使用 **windows** 模式。

为支持文档类的 **fontset** 设定能力, 使用宏包 **kvsetkeys**, **kvdefinekeys**, **kvoptions** 对 **key** 的默认值, 可选范围进行管理。

2.2.3 编码方式

由于使用 **X_YLaTeX** 编译时, **CT_EX** 宏集强制使用 **UTF-8** 编码, 相对应的包括 **.def** 文件、所有 ***.tex** 文件均应该使用 **UTF-8** 编码方式进行编码。

2.2.4 粗体设定

由于 **Windows** 下中易宋体无粗体实现, 原文档类 **pkuthss** 对于宋体的粗体 **\textbf{}** 使用中易黑体替代, 本文档类 **iofu728/pkuthss** 则依照 **Windows** 惯常设定, 使用对应字体的假粗体作为粗体 **\textbf{}**。

第三章 Model 1

第三章部分...

第四章 Model 2

第四章部分...

第五章 Model 3

第五章部分...

第六章 总结和展望

第六章部分...

参考文献

- [1] 邱锡鹏. 神经网络与深度学习[M/OL]. 北京: 机械工业出版社, 2020. <https://nndl.github.io/>.
- [2] VECTOR C T. Pkuthss: LaTeX template for dissertations in Peking University[EB/OL]. 2011 [2011-06-26]. <https://gitea.com/CasperVector/pkuthss>.
- [3] DEVLIN J, CHANG M W, LEE K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding[J]. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), 2019.
- [4] YANG Z, DAI Z, YANG Y, et al. XLNet: Generalized Autoregressive Pretraining for Language Understanding[C/OL]//WALLACH H, LAROCHELLE H, BEYGELZIMER A, et al. Advances in Neural Information Processing Systems: vol. 32. [S.l.]: Curran Associates, Inc., 2019. <https://proceedings.neurips.cc/paper/2019/file/dc6a7e655d7e5840e66733e9ee67cc69-Paper.pdf>.
- [5] 刘宝碇. 不确定规划及应用[M]. [出版地不详]: 清华大学出版社有限公司, 2003.
- [6] GUT A. Probability: a graduate course[M]. [S.l.]: Springer Science & Business Media, 2013.

攻读硕士期间发表的论文及其他成果

- [1] 扎克·施耐德, XXX XXX, et al. XXXX Title[C/OL]/III H D, SINGH A. Proceedings of Machine Learning Research: Proceedings of the 37th International Conference on Machine Learning: vol. 119. [S.l.]: PMLR, 2020: XXXX-XXXX. <http://proceedings.mlr.press/XXXX.html>.(一作, CCF-A)

致谢

致谢部分...

