

# 中文词语相似度计算

周奇安

院（系）：计算机科学与技术学院 专 业：计算机科学与技术

学 号： 1130310402 指导教师： 孙大烈

2017 年 5 月

哈爾濱工業大學

# 畢業設計（論文）

題 目 中文詞語

相似度計算

專 業 計算機科學與技術

學 號 1130310402

學 生 周奇安

指 導 教 師 孫大烈

答 辯 日 期 2017 年 5 月

## 摘 要

摘要是论文内容的高度概括，应具有独立性和自含性，即不阅读论文的全文，就能获得必要的信息。摘要应包括本论文的目的、主要研究内容、研究方法、创造性成果及其理论与实际意义。摘要中不宜使用公式、化学结构式、图表和非公用符号和术语，不标注引用文献编号。避免将摘要写成目录式的内容介绍。

**关键词：**关键词 1；关键词 2；关键词 3；……；关键词 6（关键词总共 3 — 6 个，最后一个关键词后面没有标点符号）

## Abstract

Externally pressurized gas bearing has been widely used in the field of aviation, semiconductor, weave, and measurement apparatus because of its advantage of high accuracy, little friction, low heat distortion, long life-span, and no pollution. In this thesis, based on the domestic and overseas researching.....

**Keywords:** keyword 1, keyword 2, keyword 3, ....., keyword 6 (no punctuation at the end) 英文摘要与中文摘要的内容应一致，在语法、用词上应准确无误。

# 目 录

摘 要.....	I
ABSTRACT .....	II
第 1 章 绪论 .....	1
1.1 问题背景 .....	1
1.2 相关工作 .....	1
1.3 数据来源与评价指标.....	2
1.4 本课题研究方法 .....	3
1.4.1 深度学习方法.....	3
1.4.2 集成学习方法.....	3
1.4.3 基于词典或知识库的方法的改进 .....	4
第 2 章 深度学习的二分类方法 .....	5
2.1 原理介绍 .....	5
2.1.1 训练数据生成.....	5
2.1.2 深度学习模型.....	6
第 3 章 图片的插入方法 .....	7
3.1 研究生院的插图规范.....	7
3.1.1 图题及图中说明.....	7
3.1.2 插图编排.....	7
3.2 L <sup>A</sup> T <sub>E</sub> X 中推荐使用的图片格式 .....	8
3.3 单张图片的插入方法.....	8
3.4 具有子图的图片插入方法 .....	11
第 4 章 表格的绘制方法 .....	14
4.1 研究生院的绘表规范.....	14
4.2 普通表格的绘制方法.....	14
4.3 长表格的绘制方法 .....	15
4.4 列宽可调表格的绘制方法 .....	18
4.4.1 表格内某单元格内容过长的情况 .....	18
4.4.2 对物理量符号进行注释的情况.....	19

第 5 章 数学公式的输入方法 .....	21
5.1 研究生院的公式规范 .....	21
5.2 生成 L <sup>A</sup> T <sub>E</sub> X 数学公式的两种方法 .....	21
5.2.1 基于 MathType 软件的数学公式生成方法 .....	21
5.2.2 基于 MATLAB 软件的数学公式生成方法 .....	22
5.3 数学字体 .....	23
5.4 行内公式 .....	24
5.5 行间公式 .....	24
5.6 可自动调整大小的定界符 .....	25
5.7 数学重音符号 .....	26
第 6 章 模板的其它说明 .....	27
6.1 中英文封面的相关信息 .....	27
6.2 B <sup>I</sup> B <sub>T</sub> E <sub>X</sub> 文献文件的写法 .....	27
6.3 参考文献的引用 .....	28
6.4 单层罗列环境 .....	29
6.5 算法 .....	30
6.6 定理定义 .....	31
结 论 .....	32
参考文献 .....	33
附录 A 带章节的附录 .....	34
A.1 附录节的内容 .....	34
攻读本科学位期间发表的论文及其他成果 .....	35
哈尔滨工业大学本科毕业设计（论文）原创性声明 .....	36
致 谢 .....	37

# 第 1 章 绪论

## 1.1 问题背景

词语的相似度计算是自然语言处理中的一项基本任务。它的目的是使用自动化的方法给出一个量化的指标，来衡量一对词语之间的相似程度。词语相似度计算对于很多高层次的自然语言处理任务有着重要的意义，例如问答、信息检索、改写检测与文本蕴含。

为了评价一种词语相似度计算方法的好坏，可以使用内部任务评价或外部任务评价。内部任务评价直接使用一个标准的词语相似度标注数据集，并使用一些评价指标来衡量机器给出的结果与标准数据集之间的差异；而外部任务评价则利用词语相似度计算的结果解决其他的高层次机器学习问题，通过衡量不同计算结果在这些问题上的性能来简介衡量词语相似度计算方法的好坏。

什么样的词语可以被称为“相似”的，这本身并不是一个定义良好的问题。词语的相似度非常依赖于人的主观判断。（写写每个数据集都是怎么搞的。）

本课题取自NLPCC-ICCPOL 2016会议中的“中文词语相似度计算”开放任务。该开放任务使用内部任务评价的方式来衡量各参赛队提交的结果。其组织者发布了一个PKU-500数据集，其中包含500个词语对，以及对每个词语对之间的相似度标注。共有49个参赛队伍报名参加了此次开放任务，而最终由21个参赛队伍提交了总计24个系统。

## 1.2 相关工作

在英文词语相似度任务中，已经有多个数据集可以作为基准测试。其中，使用最广泛的是WordSim-353数据集。这个数据集包含353个词语对，并由16名标注者对其相似度进行了标注。

在NLPCC-ICCPOL 2016会议举办之前，Semeval-2012会议上也举行了一次中文词语相似度的比赛，比赛所使用的词语对是经过翻译的WordSim-353数据集，并重新进行了人工标注。遗憾的是，这次比赛只有两个队伍参加。

本次开放任务中的参赛队伍使用的方法大致分为三类：基于辞典或知识库的方法、基于语料的方法与基于信息检索的方法。

基于词典或知识库的方法利用了一些现有的语言学资源，例如知网

(HowNet) 和同义词词林。这些词典或知识库的作者往往同时给出了用于计算词语相似度的规则。有些参赛队伍直接使用了这些规则，还有一些对这些规则进行了修改。这些方法依赖于词典或知识库的组织结构，并且无法处理未登录词。此外，人工制定的规则是否合理也是影响模型性能的重要因素。

基于语料的方法主要是根据上下文信息进行词嵌入，这类方法假设“相似的词语应该出现在相似的上下文中”。典型的词嵌入方法，例如word2vec，具有成熟的工具包，因此在各参赛队伍中非常受到欢迎。除了词嵌入方法以外，来自大连理工大学的参赛队使用了深度学习的方法来识别描述词语类比的句子，例如“寂寞和孤单的区别是什么”。如果这类句子出现在语料中，很可能意味着这两个词语是相关的。基于语料的方法需要大规模的语料库和漫长的训练时间，由于这些模型普遍使用无需人工标注的生语料库，因此语料的获取来源非常广泛。而训练时间的缩短只能通过使用运算能力更强的硬件，因此实验可能受到硬件条件的制约。针对中文语料而言，绝大部分的模型是以词为单位进行处理，这样分词的准确度也会对模型最终性能产生一定影响。

基于信息检索的方法利用搜索引擎查询包含目标词语的页面，根据同时包含两个词语的页面数、仅包含一个词语的页面数利用公式计算目标词语的相似度。这种方法的实现非常的简单，只需爬取搜索引擎的结果页面，或调用搜索引擎提供的API。有的队伍将这种方法作为多种评价方法之一，也有的队伍用信息检索得到的结果作为弱监督学习的指标。这种方法的一大问题是词语相似度和词语共同出现的频率并不一致，例如一些事物的别名经常在科普类的文章中被介绍（例如“西红柿”和“番茄”），而因口语习惯而导致不同的同义词（例如“拖后腿”和“拉后腿”）就很难在一篇文章中同时出现。

### 1.3 数据生成与评价指标

PKU-500数据集的生成过程分为三个步骤，分别为词语选择、词语对生成和相似度标注。

PKU-500数据集的词语选择过程遵循以下的条件：

领域 同时涉及传统书面语言与近期的网络语言。

频率 应同时包含高频、中频与低频的词语。

词性 应同时包含名词、动词与形容词。除上述实词外还应包含虚词（例如副词和连词）。

字数 应包含1-4个字的词语。



词义 应包含部分多义词。

极性 应同时包含正面词语与负面词语。

根据上述条件，词语的选择过程如下：

1. 数据来源于两个领域：三个月的《人民日报》新闻与一大批新浪微博数据。
2. 数据经过开源的ansj工具进行分词和词性标注。
3. 从两个领域中分别根据其频率、词性与字数抽取词语。
4. 根据词义与极性手动挑选自动抽取的词语，最终从《人民日报》新闻与微博数据中分别得到了514和202个词语。

## 1.4 本课题研究方法

本课题的主要研究内容是深度学习方法和集成学习方法。上述两种方法将直接用于词语相似度计算，或用于改进与结合现有的方法。（这里要改）

### 1.4.1 深度学习方法

本课题中使用的深度学习方法是一种基于语料的方法。本课题利用循环神经网络及其变种构造分类器。语料库中的所有句子经过一定的处理之后作为输入，其中将句子不加变化地输入循环神经网络作为正例，而将每个句子中挑选一个词语替换成其它的词语作为负例。分类器的训练目标是尽可能准确地识别被修改的句子。最终评价词语相似度时，将目标词语所在的句子中的目标词语替换成词语对中的另一个词语，并利用上文中的分类器进行识别，使用分类器输出的结果和置信度作为词语相似度的评价标准。

此外，课题还将使用深度学习方法辅助词嵌入的训练。这种方法训练一个带有嵌入矩阵的双向循环神经网络，用来解决词预测问题。令神经网络通过词语两侧的上下文信息补全每一个空缺词。为了解决需要在整个单词表的范围内做规范化的问题，可以使用噪声对比估计的方式。

### 1.4.2 集成学习方法

词语相似度的计算具有多种方法，为了结合这些方法的优势，提高模型的准确率与稳定性，需要利用集成学习的方法。集成学习的方法是通过一些结合策略，将多个个体学习器的输出结合成最终的输出结果。集成学习中最常见的方法是简

单平均法和加权平均法。除此之外，在其他参赛选手的系统中还出现了很多人为制定的基于规则的结合策略。这些方法虽然简单朴素，甚至有些看起来缺乏理论根据，但是在词语相似度计算的任务中取得了相当优秀的成绩。因此作为对比，本课题也会对这些方法进行尝试。另外一种比较高级的集成学习方法是学习法，这种方法使用另一个学习器来进行结合。这个学习器以每个初级学习期的输出作为输入，而输出一个单个的结果。这样就可以使用初级学习器在测试集上的输出和测试集的标注训练次级学习器。由于官方给出的测试集规模较小，因此不能使用基于深度学习的方法构造次级学习器。本课题将会选用较为简单的模型用于集成学习。由于可供测试的数据只有宝贵的500个词语对，本课题将使用交叉验证甚至留一法的方式训练次级学习器。这样所有的数据既能参与训练，又能用于评测。而且，进行交叉验证所训练出的学习器还可以再次使用简单平均法集成，形成一个最终的模型。

### 1.4.3 基于词典或知识库的方法的改进

基于词典或知识库的方法往往利用词典和知识库作者提供的相似度计算规则。本课题将会利用词典或知识库的结构提取出特征，然后用机器学习的方式计算相似度。由于这里也需要使用测试集来训练，因此同样需要选择更为简单的模型。由于基于词典或知识库的学习器和集成学习的次级学习器相近，因此也可以用统一的学习器，输入深度学习模型的输出与词典或知识库中的特征，输出最终的结果。

## 第2章 深度学习的二分类方法

### 2.1 原理介绍

为了更清晰的介绍深度学习的二分类方法，本文将先给出一些形式化的定义。

在深度学习的二分类方法中使用的语料库 $C$ 是一个句子的多重集合，其中每一个句子是一个词语的序列，即：

$$C \subset \mathcal{V}^+ \quad (2-1)$$

其中 $\mathcal{V}$ 是所有词语的集合。

替换函数 $f_r$ 可以将一个句子 $s \in C$ 中的某个词语 $s_i$ 替换为一个不同的词语 $v \neq s_i$ ，即：

$$f_r(s, i, v) = (s_1, \dots, s_{i-1}, v, s_{i+1}, \dots, s_{|s|}) \quad (2-2)$$

对语料库中的句子应用替换函数 $f_r$ 得到的句子称为替换句：

$$C_r = \{f_r(s, i, v) \mid s \in C, v \in \mathcal{V}\} \quad (2-3)$$

一般来说， $C_r \cap C \approx \emptyset$ 。本课题希望得到一个二分类器 $M_C$ ，满足：

$$M_C(s) = \begin{cases} 1 & s \in C, \\ 0 & s \in C_r \end{cases} \quad (2-4)$$

使得 $M$ 可以区分原句与替换句。

利用这样的二分类器，我们就可以定义词语对 $(v_a, v_b)$ 的相似度：

$$\text{sim}(w_a, w_b) = E\left(M_C(f_r(s, i, v_b)) \mid s \in C, s_i = v_a\right) \quad (2-5)$$

其中 $E$ 为数学期望。其中的思想是，利用包含词语 $v_a$ 的句子以 $v_b$ 替换得到的替换句测试二分类器 $M_C$ ，若二分类器容易将这样的替换句误判为原句，则说明词语对 $(v_a, v_b)$ 相似。为了提高准确性，应同时考虑从 $v_a$ 到 $v_b$ 和从 $v_b$ 到 $v_a$ 的替换。

#### 2.1.1 训练数据生成

为了训练二分类器 $M_C$ ，我们需要一些带有标注的句子 $(x, y) \in \mathcal{V}^+ \times \{0, 1\}$ 作为训练数据。其中：

$$y = \begin{cases} 1 & x \in C, \\ 0 & x \in C_r \end{cases} \quad (2-6)$$

其中，原句是很容易生成的，只需将 $M_C$ 中全部的句子标注为1即可。而替换句的生成需要一些额外的步骤。

利用随机采样的方法，我们可以生成一些采样句。采样句定义如下：

$$C_s = \{f_r(s, i, v) \mid s \in C, i \sim \mathcal{U}\{1, |s|\}, v \sim \mathcal{D}_v\} \quad (2-7)$$

其中 $\mathcal{U}$ 为均匀分布， $\mathcal{D}_v$ 为词语在自然语言中出现的概率分布。也就是说，采样句中的被替换词语位置 $i$ 是随机选择的，而替换词语 $v$ 亦是从词语分布中采样得到的。

显然， $C_s \subset C_r$ ，而随机采样对于计算机来说是轻而易举的任务。这样，我们就可以通过生成采样句的方式获得用于训练的替换句。

### 2.1.2 深度学习模型

上文提到，本课题需要训练一个二分类器 $M_C$ 用于区分原句和替换句。本课题使用了深度学习的模型来完成此任务。

## 第3章 图片的插入方法

### 3.1 研究生院的插图规范

图应有自明性。插图应与文字紧密配合，文图相符，内容正确。选图要力求精练，插图、照片应完整清晰。图中文字和数字等字号用宋体5号字。

机械工程图：采用第一角投影法，严格按照 GB4457—GB131-83《机械制图》标准规定。

数据流程图、程序流程图、系统流程图等按 GB1526-89 标准规定。

电气图：图形符号、文字符号等应符合附录3所列有关标准的规定。

流程图：必须采用结构化程序并正确运用流程框图。

对无规定符号的图形应采用该行业的常用画法。

坐标图的坐标线均用细实线，粗细不得超过图中曲线，有数字标注的坐标图，必须注明坐标单位。

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

引用文献图表必须标注出处。

#### 3.1.1 图题及图中说明

每个图均应有图题（由图序和图名组成），图名在图序之后空一格排写。图序按章编排，如第1章第一个插图的图号为“图1-1”等。图题置于图下，硕士论文可只用中文书写，博士论文用中、英文两种文字居中书写，中文在上，要求中文用宋体5号字，英文用 Times New Roman 5号字。有图注或其它说明时应置于图题之上。引用图应注明出处，在图题右上角加引用文献号。图中若有分图时，分图题置于分图之下或图题之下，分图号用 a)、b) 等表示。

图中各部分说明应采用中文（引用的外文图除外）或数字项号，各项文字说明置于图题之上（有分图题者，置于分图题之上）。

#### 3.1.2 插图编排

插图之前，文中必须有本插图的提示，如“见图1-1”、“如图1-1所示”等。插图与其图题为一个整体，不得拆开排写于两页。插图处的该页空白不够排

写该图整体时，则可将其后文字部分提前排写，将图移到次页。

## 3.2 L<sup>A</sup>T<sub>E</sub>X 中推荐使用的图片格式

在 L<sup>A</sup>T<sub>E</sub>X 中应用最多的图片格式是 EPS（Encapsulated PostScript）格式，它是一种专用的打印机描述语言，常用于印刷或打印输出。EPS 格式图片可通过多种方式生成，这里介绍一款功能强大的免费图片处理软件——ImageMagick，360 软件管家也提供此软件的下载。此软件可将其它格式图片转换为 EPS 格式图片，同时还可以锐化图片，使图片的局部清晰一些。

此软件对图片的格式转换操作都是在命令提示符（cmd.exe）中实现的，可以通过“开始→运行→输入 cmd→回车”或“开始→程序→附件→命令提示符”找到它。在命令提示符下，首先采用“盘符命令”或“cd 命令”将当前目录改为待处理图片所在的目录，在此目录下就可通过 convert 命令将图片转换为 EPS 格式，其命令的语法格式为

**convert** [可选参数] 原文件名.原扩展名 新文件名.eps

若 convert 命令中无可选参数，则将原来的图片格式直接转换为 EPS 格式，对图片不进行任何处理，这也是最常用的方法。也可以选用可选参数，可选参数有很多选择，但最常用的有如下两个：

**-sharpen radius{xsigma}**——此参数用来锐化图片，一般用在图片像素不高，需要提高图片清晰度的情况下。其中 radius 只能为整数，它用来确定转换命令采取哪一种锐化算法，我们可以只取 radius 为 0；sigma 为所采取算法的锐化度，它的取值为 0.1–3 之间的任意一个浮点数，数值越大，锐化程度也越大，通常取为 0.5–1 之间；x 在参数中为分隔符。

**-resize geometry**——此参数用来改变图片的大小，若图片的存储空间过大，可通过此命令缩小图片尺寸，但同时也将导致图片像素降低，其具体用法请参见 -resize geometry 的官方说明。

除此之外，一些文字处理软件和科学计算软件也支持生成 EPS 格式的文件，请使用“另存为”功能查看某款软件是否能够将图片以 EPS 格式的形式保存。

## 3.3 单张图片的插入方法

单张图片独自占一行的插入形式如图 3-1 所示。

其插入图片的代码及其说明如下。

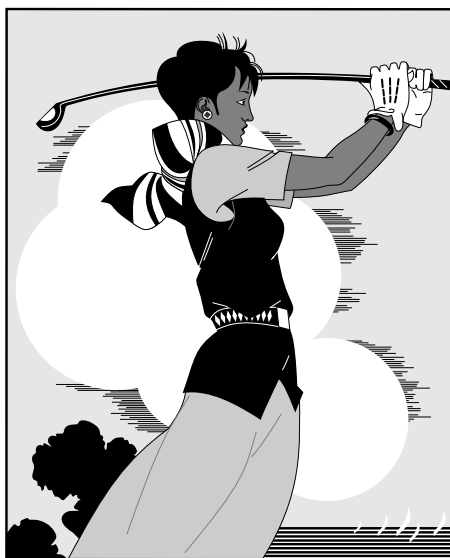


图 3-1 打高尔夫球的人

Fig.3-1 The person playing golf

```
\begin{figure}[htbp]
\centering
\includegraphics[width=0.4\textwidth]{文件名(.eps)}
\bicaption[标签名(英文)]{}{中文标题}{Fig.$\!$}
      {English caption (首字母大写)}\vspace{-1em}
\end{figure}
```

**figure**环境的可选参数[htbp]表示浮动图形所放置的位置，**h** (**here**)表示当前位置，**t** (**top**)表示页芯顶部，**b** (**bottom**)表示页芯底部，**p** (**page**)表示单独一页。在**word**等软件中，图片通常插入到当前位置，如果当前页的剩余空间不够，图片将被移动到下一页，当前页就会出现很大的空白，其人工调整工作非常不便。由**LaTeX**提供的浮动图片功能，总是会按**h**->**t**->**b**->**p**的次序处理选项中的字母，自动调整图片的位置，大大减轻了工作量。

**\centering**命令将后续内容转换成每行皆居中的格式。

“**\includegraphics**”的可选参数用来设置图片插入文中的水平宽度，一般表示为正文宽度(**\textwidth**)的倍数。

**\bicaption**命令的使用需要调用**ccaption**宏包，它可以为图片或表格插入双语标题（博士学位论文要求），可选参数“标签名”为英文形式，一般不以图片或表格的数字顺序作为标签，而应包含一定的图片或表格信息，以便于文中引用（若图片、表格、公式、章节和参考文献等在文中出现的先后顺序发生了变化，其标注

序号及其文中引用序号也会跟着发生变化，这一点是word等软件所不能做到的）。第4个参数中的“ $\$!\$$ ”表示 $-1/6$ 个空铅宽度，这样可以缩小Fig.和Table与后面数字序号之间的水平距离。另外，图题或表题并不会因为分页而与图片或表格体分置于两页，章节等各级标题也不会置于某页的最底部，LaTeX系统会自动调整它们在正文中的位置，这也是word等软件所无法匹敌的。

注：硕士学位论文的图表只需要插入中文标题，因此需将**\bicaption**一句命令替换为如下两条命令（下同）：

**\caption{中文标题}**

**\label{标签名(英文)}**

**\vspace**将产生一定高度的竖直空白，必选参数为负值表示将后续文字位置向上提升，参数值可自行调整。**em**为长度单位，相当于大写字母**M**的宽度。

引用方法：“见图~**\ref{标签名(英文)}**”、“如图~**\ref{标签名(英文)}**~所示”等。

若需要将2张及以上的图片并排插入到一行中，则需要采用**minipage**环境，如图3-2和图3-3所示。



图 3-2 打高尔夫球的人

Fig.3-2 The person playing golf



图 3-3 打高尔夫球的人

Fig.3-3 The person playing golf

其代码如下所示。

**\begin{figure}[htbp]**

**\centering**



---

```

\begin{minipage}{0.4\textwidth}
\centering
\includegraphics[width=\textwidth]{文件名}
\bicaption[标签名]{}{中文标题}{Fig.$\!$}
        {English caption}
\end{minipage}
\begin{minipage}{0.4\textwidth}
\centering
\includegraphics[width=\textwidth]{文件名}
\bicaption[标签名]{}{中文标题}{Fig.$\!$}
        {English caption}
\end{minipage}\vspace{-1em}
\end{figure}

```

---

`minipage`环境的必选参数用来设置小页的宽度，若需要在一行中插入 $n$ 个等宽图片，则每个小页的宽度应略小于 $(1/n)\text{width}$ 。

---

### 3.4 具有子图的图片插入方法

图中若含有子图时，需要调用 `subfigure` 宏包。博士学位论文规范要求不止总图的标题为中英文形式，其各个子图也应具有中英文形式的标题。然而 `ccaption` 宏包却无法实现子图的中英文标题功能，这里采用对 `\subfigure` 命令进行嵌套的方法来实现子图的中英文标题功能，如图 3-4 所示。

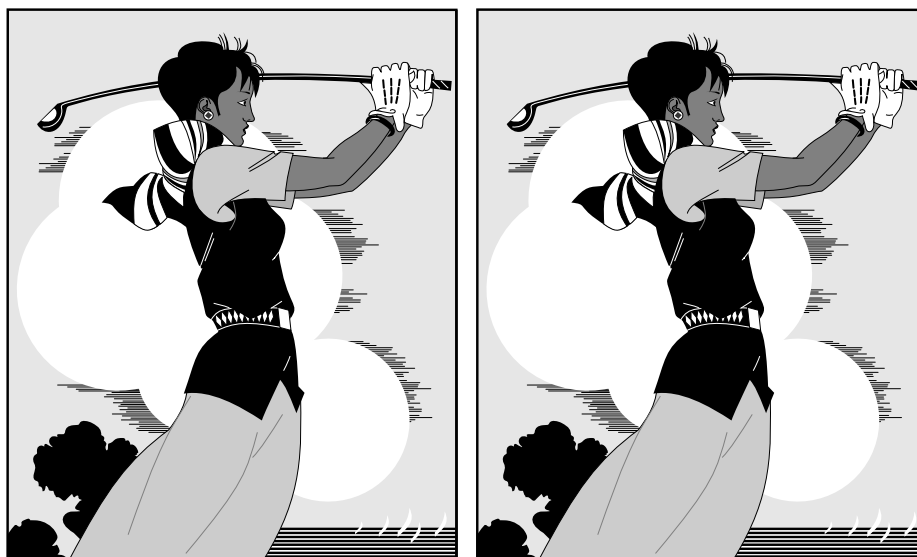
其代码及其说明如下。

---

```

\begin{figure}[htbp]
\centering
\subfigure{\label{第1个子图标签名}}\addtocounter{subfigure}{-2}
\subfigure[The 1st subfigure caption]{\subfigure[第1个子图标题]
        {\includegraphics[width=0.4\textwidth]{文件名}}}
\subfigure{\label{第2个子图标签名}}\addtocounter{subfigure}{-2}
\subfigure[The 2nd subfigure caption]{\subfigure[第2个子图标题]
        {\includegraphics[width=0.4\textwidth]{文件名}}}

```



a) 打高尔夫球的人 1                      b) 打高尔夫球的人 2  
a) The person playing golf                  b) The person playing golf

图 3-4 打高尔夫球的人

Fig.3-4 The person playing golf

```
\bicaption[总标签名]{}{中文总标题}{Fig.$!$}{The total caption}
\vspace{-1em}
\end{figure}
```

`\addtocounter`把指定的值加到计数器上，这里是对`subfigure`计数器进行减2操作。这是因为每插入1个子图，就调用3次`\subfigure`命令，第1次调用`\subfigure`命令用来生成紧随其后所插入子图的标签，而之后的双层嵌套调用`\subfigure`命令用来插入子图并生成该子图的中英文标题。因此，每插入1张子图，`subfigure`计数器的值就自动加3，为了使得子图的序号能每次加1，则需在每插入1张子图前手动把`subfigure`计数器的值减2。

`\subfigure`命令的双层嵌套使用可用来生成中英文标题，其内层`\subfigure`命令用来插入子图并生成中文标题，外层`\subfigure`命令将插入的子图和中文标题作为一个整体，生成这个整体的英文标题，因此英文标题会置于中文标题的下面。

硕士学位论文只需要中文标题，其代码如下：

```
\begin{figure}[htbp]
\centering
\subfigure[第1个子图标题\label{第1个子图标签名}]
{\includegraphics[width=0.4\textwidth]{文件名}}
```

```
\subfigure[第2个子图标题\label{第2个子图标签名}]
    {\includegraphics[width=0.4\textwidth]{文件名}}
\caption{中文总标题}\label{总标签名}
\vspace{-1em}
\end{figure}
```

引用方法：总图的引用方法同本章第1节，子图的引用方法用`\ref{第n个子图标签名}`来代替。

---

子图的引用示例：如图 3-4 a) 和图 3-4 b) 所示。

若想获得插图方法的更多信息，请参见网络上的 Using Imported Graphics in  $\LaTeX$  and pdf $\LaTeX$  文档。

## 第 4 章 表格的绘制方法

### 4.1 研究生院的绘表规范

表应有自明性。表格不加左、右边线。表的编排建议采用国际通行的三线表。表中文字用宋体 5 号字。

每个表格均应有表题（由表序和表名组成）。表序一般按章编排，如第 1 章第一个插表的序号为“表 1-1”等。表序与表名之间空一格，表名中不允许使用标点符号，表名后不加标点。表题置于表上，硕士学位论文只用中文，博士学位论文用中、英文两种文字居中排写，中文在上，要求中文用宋体 5 号字，英文用新罗马字体 5 号字。

表头设计应简单明了，尽量不用斜线。表头中可采用化学符号或物理量符号。

全表如用同一单位，则将单位符号移至表头右上角，加圆括号。表中数据应准确无误，书写清楚。数字空缺的格内加横线“—”（占 2 个数字宽度）。表内文字或数字上、下或左、右相同时，采用通栏处理方式，不允许用“//”、“同上”之类的写法。

表内文字说明，起行空一格、转行顶格、句末不加标点。

如某个表需要转页接排，在随后的各页上应重复表的编号。编号后加“（续表）”，表题可省略。续表应重复表头。

### 4.2 普通表格的绘制方法

表格应具有三线表格式，因此需要调用 booktabs 宏包，其标准格式如表 4-1 所示。

表 4-1 符合研究生院绘图规范的表格

Table4-1 Table in agreement of the standard from graduate school

$D(\text{in})$	$P_u(\text{lbs})$	$u_u(\text{in})$	$\beta$	$G_f(\text{psi.in})$
5	269.8	0.000674	1.79	0.04089
10	421.0	0.001035	3.59	0.04089
20	640.2	0.001565	7.18	0.04089

其绘制表格的代码及其说明如下。

---

```

\begin{table}[htbp]
\bicaption[标签名]{}{中文标题}{Table$!\$}{English caption}
\vspace{0.5em}\centering\wuhao
\begin{tabular}{cc...c}
\toprule[1.5pt]
表头第1个格    & 表头第2个格    & ... & 表头第n个格    \\
\midrule[1pt]
表中数据(1,1) & 表中数据(1,2) & ... & 表中数据(1,n)\\
表中数据(2,1) & 表中数据(2,2) & ... & 表中数据(2,n)\\
.....\\
表中数据(m,1) & 表中数据(m,2) & ... & 表中数据(m,n)\\
\bottomrule[1.5pt]
\end{tabular}
\end{table}

```

---

**table**环境是一个将表格嵌入文本的浮动环境。

**\wuhao**命令将表格的字号设置为五号字（10.5pt），在绘制表格结束退出时，不需要将字号再改回为**\xiaosi**，正文字号默认为小四号字（12pt）。

**tabular**环境的必选参数由每列对应一个格式字符所组成：**c**表示居中，**l**表示左对齐，**r**表示右对齐，其总个数应与表的列数相同。此外，**@{文本}**可以出现在任意两个上述的列格式之间，其中的文本将被插入每一行的同一位置。表格的各行以**\\**分隔，同一行的各列则以**&**分隔。

**\toprule**、**\midrule**和**\bottomrule**三个命令是由**booktabs**宏包提供的，其中**\toprule**和**\bottomrule**分别用来绘制表格的第一条（表格最顶部）和第三条（表格最底部）水平线，**\midrule**用来绘制第二条（表头之下）水平线，且第一条和第三条水平线的线宽为**1.5pt**，第二条水平线的线宽为**1pt**。

引用方法：“如表~**\ref{标签名}**~所示”。

---

## 4.3 长表格的绘制方法

长表格是当表格在当前页排不下而需要转页接排的情况下所采用的一种表格环境。若长表格仍按照普通表格的绘制方法来获得，其所使用的**table**浮动环境无

法实现表格的换页接排功能，表格下方过长部分会排在表格第1页的页脚以下。为了能够实现长表格的转页接排功能，需要调用 `longtable` 宏包，由于长表格是跨页的文本内容，因此只需要单独的`longtable`环境，所绘制的长表格的格式如表 4-2 所示。

此长表格 4-2 第 2 页的标题“编号（续表）”和表头是通过代码自动添加上去的，无需人工添加，若表格在页面中的竖直位置发生了变化，长表格在第 2 页及之后各页的标题和表头位置能够始终处于各页的最顶部，也无需人工调整，`LaTeX` 系统的这一优点是 `word` 等软件所无法比拟的。

表 4-2 中国省级行政单位一览

Table4-2 Overview of the provincial administrative unit of China

名称	简称	省会或首府
北京市	京	北京
天津市	津	天津
河北省	冀	石家庄市
山西省	晋	太原市
内蒙古自治区	蒙	呼和浩特市
辽宁省	辽	沈阳市
吉林省	吉	长春市
黑龙江省	黑	哈尔滨市
上海市	沪/申	上海
江苏省	苏	南京市
浙江省	浙	杭州市
安徽省	皖	合肥市
福建省	闽	福州市
江西省	赣	南昌市
山东省	鲁	济南市
河南省	豫	郑州市
湖北省	鄂	武汉市
湖南省	湘	长沙市
广东省	粤	广州市
广西壮族自治区	桂	南宁市
海南省	琼	海口市
重庆市	渝	重庆
四川省	川/蜀	成都市
贵州省	黔/贵	贵阳市

表 4-2（续表）

名称	简称	省会或首府
云南省	云/滇	昆明市
西藏自治区	藏	拉萨市
陕西省	陕/秦	西安市
甘肃省	甘/陇	兰州市
青海省	青	西宁市
宁夏回族自治区	宁	银川市
新疆维吾尔自治区	新	乌鲁木齐市
香港特别行政区	港	香港
澳门特别行政区	澳	澳门
台湾省	台	台北市

绘制长表格的代码及其说明如下。

```

\wuhao\begin{longtable}{cc...c}
\longbionenumcaption{{中文标题\label{标签名}}{Table$!\$}
{{English caption}\vspace{0.5em}}\
\toprule[1.5pt] 表 头 第1个 格 & 表 头 第2个 格 & ... & 表 头 第n个
格\ \midrule[1pt]
\endfirsthead
\multicolumn{n}{r}{表~\thetable（续表）}\vspace{0.5em}\
\toprule[1.5pt] 表 头 第1个 格 & 表 头 第2个 格 & ... & 表 头 第n个
格\ \midrule[1pt]
\endhead
\bottomrule[1.5pt]
\endfoot
表中数据(1,1) & 表中数据(1,2) & ... & 表中数据(1,n)\
表中数据(2,1) & 表中数据(2,2) & ... & 表中数据(2,n)\
.....\
表中数据(m,1) & 表中数据(m,2) & ... & 表中数据(m,n)\
\end{longtable}\xiaosi

```

在绘制长表格的前面留出一个空白行，并在第2行的一开始全局定义长表格的字号为五号字，这样能够保证长表格之前段落的行距保持不变。在绘制长表格结束后，需要\xiaosi命令重新将字号改为小四号字。

长表格的中英文标题是通过ccaption宏包的\longbionenumcaption命令得到的。

\endhead之前的文字描述的是第2页及其之后各页的标题或表头；\endfirsthead之前的文字描述的是第1页的标题和表头，若无此命令，则第1页的表头和标题由\endhead命令确定；同理，\endfoot之前的文字描述的是除最后一页之外每页的表格底部内容；\endlastfoot之前的文字描述的是最后一页的表格底部内容，若无此命令，则最后一页的表格底部内容由\endfoot命令确定；由于规范中长表格每页底部内容均相同（水平粗线），因此模板中没有用到\endlastfoot命令。

注：硕士学位论文的长表格只需要插入中文标题，因此需将\longbionenumcaption一句命令替换为如下两条命令：

\caption{中文标题}

\label{标签名}

## 4.4 列宽可调表格的绘制方法

论文中能用到列宽可调表格的情况共有两种，一种是当插入的表格某一单元格内容过长以至于一行放不下的情况，另一种是当对公式中首次出现的物理量符号进行注释的情况，这两种情况都需要调用 tabularx 宏包。下面将分别对这两种情况下可调表格的绘制方法进行阐述。

### 4.4.1 表格内某单元格内容过长的情况

首先给出这种情况下的一个例子如表 4-3 所示。

绘制这种表格的代码及其说明如下。

```
\begin{table}[htbp]
\bicaption[标签名]{{中文标题}}{Table$!\$}{English caption}
\vspace{0.5em}\wuhao
\begin{tabularx}{{\textwidth}}{1...X...1}
\toprule[1.5pt]
```



表 4-3 最小的三个正整数的英文表示法

Table4-3 The English construction of the smallest three positive integral numbers

Value	Name	Alternate names, and names for sets of the given size
1	One	ace, single, singleton, unary, unit, unity
2	Two	binary, brace, couple, couplet, distich, deuce, double, doubleton, duad, duality, duet, duo, dyad, pair, snake eyes, span, twain, twosome, yoke
3	Three	deuce-ace, leash, set, tercet, ternary, ternion, terzetto, threesome, tierce, trey, triad, trine, trinity, trio, triplet, troika, hat-trick

```

表头第1个格    & ... & 表头第X个格    & ... & 表头第n个格    \\
\midrule[1pt]
表中数据(1,1) & ... & 表中数据(1,X) & ... & 表中数据(1,n)\\
表中数据(2,1) & ... & 表中数据(2,X) & ... & 表中数据(2,n)\\
.....\\
表中数据(m,1) & ... & 表中数据(m,X) & ... & 表中数据(m,n)\\
\bottomrule[1.5pt]
\end{tabularx}
\end{table}

```

**tabularx**环境共有两个必选参数：第1个参数用来确定表格的总宽度，这里取为排版表格能达到的最大宽度——正文宽度`\textwidth`；第2个参数用来确定每列格式，其中标为**X**的项表示该列的宽度可调，其宽度值由表格总宽度确定。

标为**X**的列一般选为单元格内容过长而无法置于一行的列，这样使得该列内容能够根据表格总宽度自动分行。若列格式中存在不止一个**X**项，则这些标为**X**的列的列宽相同，因此，一般不将内容较短的列设为**X**。

标为**X**的列均为左对齐，因此其余列一般选为**l**（左对齐），这样可使得表格美观，但也可以选为**c**或**r**。

#### 4.4.2 对物理量符号进行注释的情况

为使得对公式中物理量符号注释的转行与破折号“——”后第一个字对齐，此处最好采用表格环境。此表格无任何线条，左对齐，且在破折号处对齐，一共有“式中”二字、物理量符号和注释三列，表格的总宽度可选为文本宽度，因此应该采用**tabularx**环境。由**tabularx**环境生成的对公式中物理量符号进行注释的公式如式(4-1)所示。

$$\ddot{\rho} - \frac{\mu}{R_t^3} \left( 3\mathbf{R}_t \frac{\mathbf{R}_t \rho}{R_t^2} - \rho \right) = \mathbf{a} \quad (4-1)$$

式中  $\rho$ ——追踪飞行器与目标飞行器之间的相对位置矢量；  
 $\ddot{\rho}$ ——追踪飞行器与目标飞行器之间的相对加速度；  
 $\mathbf{a}$ ——推力所产生的加速度；  
 $\mathbf{R}_t$ ——目标飞行器在惯性坐标系中的位置矢量；  
 $\omega_t$ ——目标飞行器的轨道角速度；  
 $\mathbf{g}$ ——重力加速度， $= \frac{\mu}{R_t^3} \left( 3\mathbf{R}_t \frac{\mathbf{R}_t \rho}{R_t^2} - \rho \right) = \omega_t^2 \frac{R_t}{p} \left( 3\mathbf{R}_t \frac{\mathbf{R}_t \rho}{R_t^2} - \rho \right)$ ，这里  $p$  是目标飞行器的轨道半通径。

其中生成注释部分的代码及其说明如下。

---

```
\begin{tabularx}{\textwidth}{@{}l@{\quad}r@{— — —}X@{}}
式中 & symbol-1 & symbol-1的注释内容； \\
      & symbol-2 & symbol-2的注释内容； \\
      & ..... & .....； \\
      & symbol-m & symbol-m的注释内容。
\end{tabularx}\vspace{\wordsep}
```

---

`tabularx`环境的第1个参数选为正文宽度，第2个参数里面各个符号的意义为：

第1个`@{}`表示在“式中”二字左侧不插入任何文本，“式中”二字能够在正文中左对齐，若无此项，则“式中”二字左侧会留出一定的空白；

`@{\quad}`表示在“式中”和物理量符号间插入一个空铅宽度的空白；

`@{— — —}`实现插入破折号的功能，它由三个1/2的中文破折号构成；

第2个`@{}`表示在注释内容靠近正文右边界的地方能够实现右对齐。

---

由此方法生成的注释内容应紧邻待注释公式并置于其下方，因此不能将代码放入`table`浮动环境中。但此方法不能实现自动转页接排，可能会在当前页剩余空间不够时，全部移动到下一页而导致当前页出现很大空白。因此在需要转页处理时，还请您手动将需要转页的代码放入一个新的`tabularx`环境中，将原来的一个`tabularx`环境拆分为两个`tabularx`环境。

若想获得绘制表格的更多信息，请参见网络上的 `Tables in LATEX 2ε: Packages and Methods` 文档。

## 第 5 章 数学公式的输入方法

### 5.1 研究生院的公式规范

论文中的公式应另起行，原则上应居中书写，与周围文字留有足够的空间区分开。若公式前有文字（如“解”、“假定”等），文字空两格写，公式仍居中写。公式末不加标点。

公式应标注序号，并将序号置于括号内。公式序号按章编排，如第 1 章第一个公式序号为“(1-1)”。公式的序号右端对齐。

公式较长时最好在等号“=”处转行，如难实现，则可在 +、-、 $\times$ 、 $\div$  运算符号处转行，转行时运算符号仅书写于转行式前，不重复书写。

文中引用公式时，一般用“见式 (1-1)”或“由公式 (1-1)”。

公式中用斜线表示“除”的关系时应采用括号，以免含糊不清，如  $a/(b \cos x)$ 。通常“乘”的关系在前，如  $a \cos x/b$  而不写成  $(a/b) \cos x$ 。

不能用文字形式表示等式，如：刚度 =  $\frac{\text{受力}}{\text{受力方向的位移}}$ 。

对于数学公式的输入方法，网络上有一个比较全面权威的文档 **Math mode** 请大家事先大概浏览一下。下面将对学位论文中主要用到的数学公式排版形式进行阐述。

### 5.2 生成 L<sup>A</sup>T<sub>E</sub>X 数学公式的两种方法

对于先前没有接触过 L<sup>A</sup>T<sub>E</sub>X 的人来说，编写 L<sup>A</sup>T<sub>E</sub>X 数学公式是一件很繁琐的事，尤其是对复杂的数学公式来说，更可以说是一件难以完成的任务。实际上，生成 L<sup>A</sup>T<sub>E</sub>X 数学公式有两种较为简便的方法，一种是基于 MathType 数学公式编辑器的方法，另一种是基于 MATLAB 商业数学软件的方法，下面将分别对这两种数学公式的生成方法作一下简单介绍。

#### 5.2.1 基于 MathType 软件的数学公式生成方法

MathType 是一款功能强大的数学公式编辑器软件，能够用来在文本环境中插入 Windows OLE 图形格式的复杂数学公式，所以应用比较普遍。但此软件只有 30 天的试用期，之后若再继续使用则需要付费购买才行。网络上有很多破解版的

MathType 软件可供下载免费使用，笔者推荐下载安装版本号在 6.5 之上的中文破解版。

在安装好 MathType 之后，若在输入窗口中编写数学公式，复制到剪贴板上的仍然是图形格式的对象。若希望得到可插入到  $\text{\LaTeX}$  编辑器中的文本格式对象，则需要对 MathType 软件做一下简单的设置：在 MathType 最上排的按钮中依次选择“参数选项 → 转换”，在弹出的对话框中选中“转换到其它语言（文字）：”，在转换下拉框中选择“Tex -- LaTeX 2.09 and later”，并将对话框最下方的两个复选框全部勾掉，点击确定，这样，再从输入窗口中复制出来的对象就是文本格式的了，就可以直接将其粘贴到  $\text{\LaTeX}$  编辑器中了。按照这种方法生成的数学公式两端分别有标记\[[和标记\]]，在这两个标记之间才是真正的数学公式代码。

若希望从 MathType 输入窗口中复制出来的对象为图形格式，则只需再选中“公示对象（Windows OLE 图形）”即可。

## 5.2.2 基于 MATLAB 软件的数学公式生成方法

MATLAB 是矩阵实验室（Matrix Laboratory）的简称，是美国 MathWorks 公司出品的商业数学软件。它是当今科研领域最常用的应用软件之一，具有强大的矩阵计算、符号运算和数据可视化功能，是一种简单易用、可扩展的系统开发环境和平台。

MATLAB 中提供了一个 latex 函数，它可将符号表达式转化为  $\text{\LaTeX}$  数学公式的形式。其语法形式为 latex(s)，其中，s 为符号表达式，之后再将 latex 函数的运算结果直接粘贴到  $\text{\LaTeX}$  编辑器中。从  $\text{\LaTeX}$  数学公式中可以发现，其中可能包含如下符号组合：

$\backslash\text{quad}$ =两个空铅(quad)宽度

$\backslash\text{quad}$ =一个空铅宽度

$\backslash;$ =5/18空铅宽度

$\backslash:$ =4/18空铅宽度

$\backslash,$ =3/18空铅宽度

$\backslash!$ =-3/18空铅宽度

$\backslash_$ =一个空格

所以最好将上述符号组合从数学公式中删除，从而使数学公式显得匀称美观。

对于 word 等软件的使用者来说，在我们通过 MATLAB 运算得到符号表达式形式的运算结果时，在 word 中插入运算结果需要借助于 MathType 软件，通

过在 MathType 中输入和 MATLAB 运算结果相对应的数学表达形式，之后再 MathType 数学表达式转换为图形格式粘贴到 word 中。实际上，也可以将 MATLAB 中采用 latex 函数运行的结果直接粘贴到 MathType 中，再继续上述步骤，这样可以大大节省输入公式所需要的时间。此方法在 MathType 6.5c 上验证通过，若您粘入到 MathType 中的仍然为从 MATLAB 中导入的代码，请您更新 MathType 软件。

### 5.3 数学字体

在数学模式下，常用的数学字体命令有如下几种：

`\mathnormal` 或无命令 用数学字体打印文本；  
`\mathit` 用斜体 (`\itshape`) 打印文本；  
`\mathbf` 用粗体 (`\bfseries`) 打印文本；  
`\mathrm` 用罗马体 (`\rmfamily`) 打印文本；  
`\mathsf` 用无衬线字体 (`\sffamily`) 打印文本；  
`\mathtt` 用打印机字体 (`\ttfamily`) 打印文本；  
`\mathcal` 用书写体打印文本。

在学位论文撰写中，只需要用到上面提到的 `\mathit`、`\mathbf` 和 `\mathrm` 命令。若要得到 Times New Roman 的数学字体，则需要调用 txfonts 宏包（此宏包实际上采用的是 Nimbus Roman No9 L 字体，它是开源系统中使用的免费字体，其字符字体与 Times New Roman 字体几乎完全相同）。表 5-1 中分别列出了得到阿拉伯数字、拉丁字母和希腊字母各种数学字体的命令。

表 5-1 常用数学字体命令一览

Table 5-1 Summary of common commands for setting math fonts

	阿拉伯数字&大写希腊字母	大小写拉丁字母	小写希腊字母
斜体	<code>\mathit{}</code>	无命令	无命令
粗斜体	<code>\boldsymbol{\mathit{}}</code>	<code>\boldsymbol{}</code>	<code>\boldsymbol{}</code>
直立体	无命令	<code>\mathrm{}</code>	字母后加up
粗体	<code>\mathbf{}</code> 或 <code>\boldsymbol{}</code>	<code>\mathbf{}</code>	<code>\boldsymbol{}</code> 字母后加up

下面列出了一些应采用直立数学字体的数学常数和数学符号。

d、D、p ——微分算子      e ——自然对数之底数  
 i、j ——虚数单位           $\pi$  ——圆周率

## 5.4 行内公式

出现在正文一行之内的公式称为行内公式，例如  $f(x) = \int_a^b \frac{\sin x}{x} dx$ 。对于非矩阵和非多行形式的行内公式，一般不会使得行距发生变化，而 word 等软件却会根据行内公式的竖直距离而自动调节行距，如图 5-1 所示。这三幅图分别为 L<sup>A</sup>T<sub>E</sub>X

a) 由 L<sup>A</sup>T<sub>E</sub>X 系统生成的行内公式

a) Inline mode equation derived from L<sup>A</sup>T<sub>E</sub>X system

b) 由 word 软件生成的 .doc 格式行内公式

b) Inline mode equation displayed as .doc format file  
derived from word software

c) 由 word 软件生成的 .pdf 格式行内公式

c) Inline mode equation displayed as .pdf format file  
derived from word software

图 5-1 由 L<sup>A</sup>T<sub>E</sub>X 和 word 生成的 3 种行内公式屏显效果

Fig.5-1 Three kinds of inline mode equation displayed effects derived from L<sup>A</sup>T<sub>E</sub>X and word

和 word 生成的行内公式屏显效果，从图中可看出，在 L<sup>A</sup>T<sub>E</sub>X 文本含有公式的行内，在正文与公式之间对接工整，行距不变；而在 word 文本含有公式的行内，在正文与公式之间对接不齐，行距变大。因此从这一点来说，L<sup>A</sup>T<sub>E</sub>X 系统在数学公式的排版上具有很大优势。

L<sup>A</sup>T<sub>E</sub>X 提供的行内公式最简单、最有效的方法是采用 T<sub>E</sub>X 本来的标记——开始和结束标记都写作 \$，例如本节开始的例子可由下面的输入得到。

$f(x)=\int_a^b \frac{\sin x}{x} \mathrm{d}x$

## 5.5 行间公式

位于两行之间的公式称为行间公式，每个公式都是一个单独的段落，例如

$$\int_a^b f(x) dx = \lim_{\|\Delta x_i\| \rightarrow 0} \sum_i f(\xi_i) \Delta x_i$$

除人工编号外， $\text{\LaTeX}$  各种类型行间公式的标记见表 5-2。另外，在自动编号的某

表 5-2 各种类型行间公式的标记

Table 5-2 Tags for several kinds of displaymath mode equations

	无编号	自动编号
单行公式	<code>\begin {displaymath}.....</code> <code>\end {displaymath}</code> 或 <code>\begin {equation} .....</code> <code>\end {equation}</code> <code>\[...\]</code>	<code>\begin {equation} .....</code> <code>\end {equation}</code>
多行公式	<code>\begin {eqnarray*} .....</code> <code>\end {eqnarray*}</code>	<code>\begin {eqnarray} .....</code> <code>\end {eqnarray}</code>

行公式行尾添加标签 `\nonumber`，可将该行转换为无编号形式。

行间多行公式需采用 `eqnarray` 或 `eqnarray*` 环境，它默认是一个列格式为 `rc1` 的 3 列矩阵，并且中间列的字号要小一些，因此通常只将需要对齐的运算符号（通常为等号“=”）置于中间列。

## 5.6 可自动调整大小的定界符

若在左右两个定界符之前分别添加命令 `\left` 和 `\right`，则定界符可根据所包围公式大小自动调整其尺寸，这可从式(5-1)和式(5-2)中看出。

$$\left(\sum_{k=\frac{1}{2}}^{N^2}\right) \quad (5-1)$$

$$\left(\sum_{k=\frac{1}{2}}^{N^2}\right) \quad (5-2)$$

式(5-1)和式(5-2)是在  $\text{\LaTeX}$  中分别输入如下代码得到的。

```
(\sum_{k=\frac{1}{2}}^{N^2})
\left(\sum_{k=\frac{1}{2}}^{N^2}\right)
```

`\left` 和 `\right` 总是成对出现的，若只需在公式一侧有可自动调整大小的定界符，则只要用“.”代替另一侧那个无需打印出来的定界符即可。

若想获得关于此部分内容的更多信息，可参见 `Math mode` 文档的第 8 章“Brackets, braces and parentheses”。

## 5.7 数学重音符号

数学重音符号通常用来区分同一字母表示的不同变量，输入方法如下（需要调用 `amsmath` 宏包）：

<code>\acute</code>	$\acute{a}$	<code>\mathring</code>	$\mathring{a}$	<code>\underbrace</code>	$\underbrace{a}$
<code>\bar</code>	$\bar{a}$	<code>\overbrace</code>	$\overbrace{a}$	<code>\underleftarrow</code>	$\underleftarrow{a}$
<code>\breve</code>	$\breve{a}$	<code>\overleftarrow</code>	$\overleftarrow{a}$	<code>\underlefterightarrow</code>	$\underlefterightarrow{a}$
<code>\check</code>	$\check{a}$	<code>\overleftrightarrow</code>	$\overleftrightarrow{a}$	<code>\underline</code>	$\underline{a}$
<code>\ddot</code>	$\ddot{a}$	<code>\overline</code>	$\overline{a}$	<code>\underrightarrow</code>	$\underrightarrow{a}$
<code>\dot</code>	$\dot{a}$	<code>\overrightarrow</code>	$\overrightarrow{a}$	<code>\vec</code>	$\vec{a}$
<code>\grave</code>	$\grave{a}$	<code>\tilde</code>	$\tilde{a}$	<code>\widehat</code>	$\widehat{a}$
<code>\hat</code>	$\hat{a}$	<code>\underbar</code>	$\underbar{a}$	<code>\widetilde</code>	$\widetilde{a}$

当需要在字母  $i$  和  $j$  的上方添加重音符号时，为了去掉这两个字母顶上的小点，这两个字母应该分别改用 `\imath` 和 `\jmath`。

如果遇到某些符号不知道该采用什么命令能输出它时，则可通过 [Detexify<sup>2</sup>](#) 网站来获取符号命令。若用鼠标左键在此网页的方框区域内画出你所要找的符号形状，则会在网页右方列出和你所画符号形状相近的 5 个符号及其相对应的  $\text{\LaTeX}$  输入命令。若所列出的符号中不包括你所要找的符号，还可通过点击“[Select from the complete list!](#)”的链接以得分从低到高的顺序列出所有符号及其相对应的  $\text{\LaTeX}$  输入命令。

最后，笔者建议大家还是要以 **Math mode** 这篇 pdf 文档作为主要参考。若要获得最为标准、美观的数学公式排版形式，可以查查文档中是否有和你所要的排版形式相同或相近的代码段，通过修改代码段以获得你所要的数学公式排版形式。



## 第 6 章 模板的其它说明

### 6.1 中英文封面的相关信息

国内图书分类号（Classified Index）的查询网址：

<http://www.ztflh.com/>——中国图书馆分类法

国际图书分类法（U.D.C）的查询网址：

<http://www.udcc.org/udcsummary/php/index.php>——UDC Summary

学校代码查询网址：

<http://www.marry360.com.cn/Tools/UniversityCodeList.aspx>——高校代码查询

哈尔滨工业大学的学校代码为 10213。

英文封面下方的学位论文相关信息可以采用 `tabular` 和 `tabularx` 两种表格环境，具体使用哪一种环境和具体的相关信息有关。若信息内容不太长，不会引起信息内容分行时，则应该采用 `tabular` 环境；若信息内容过长，会引起信息内容分行时，则应该采用 `tabularx` 环境。具体用法请见 `format.tex` 文件的相应代码。

### 6.2 BibTeX 文献文件的写法

用在  $\text{\LaTeX}$  中的 BibTeX 文献文件的扩展名为 `bib`，此模板中，该文件即为 `reference.bib`。`bibtex.exe` 命令根据 `GBT7714-2005NLang-HIT.bst` 文件定义的文献格式，将 `reference.bib` 中的文献数据转换为输出文档中的文献列表。`GBT7714-2005NLang-HIT.bst` 文件是在 `GBT7714-2005NLang-UTF8.bst` 文件的基础上修改得到的，所做的唯一一处改动是将姓氏字母全部大写的英文作者名改为只首字母大写，以保证和《研究生学位论文撰写规范》及其《研究生学位论文书写范例》相一致。

`bib` 文件的编写方法可参考模板中已给出的例子，也可参考 `GBT7714-2005.bst` 说明文档 20060919 中所给出的例子。

中文文献需要添加一个额外的 `language` 域，并使得域值非空，这样 `bst` 文件就能够判断此文献为中文文献，进而能正确地生成参考文献格式。

`GBT7714-2005.bst` 对于国标 GB/T 7714-2005 的文献分类如表 6-1 所示。对于每种文献类型的缺省类型，已经设置好相应的文献标识码，因此不需要输入相应

的文献标识码。扩展类型的文献则应再添加一个 `TypeofLit` 域，并需要将其域值改为相应的文献标识码。

表 6-1 GBT7714-2005.bst 的分类方式  
Table6-1 Classification method of GBT7714-2005.bst

文献类型	缺省类型	扩展类型（需要手工加入文献标识码）	主要特征
article	文章[J]	报纸中的析出文献[N] 在线文章[J/OL]	年,卷(期):页码
book	书[M]	论文集、会议录[C] 在线书[M/OL] 汇编[G]	
inbook	书的某几页[M]		
incollection	书中析出的文章[M]//	汇编的析出文献[G]// 标准的析出文献[S]//	析出文献[文献标识码]//
proceedings			
inproceedings	论文集、会议录中的析出文献[C]//	在线论文集、会议录[C/OL]//	析出文献[文献标识码]//
/conference			
mastersthesis	毕业论文[D]		类似book类
phdthesis	毕业论文[D]		类似book类
techreport	科技报告[R]		类似book类
misc		杂项[], 例如:专利[P] 网上专利[P/OL] 网上电子公告[EB/OL] 磁盘[CP/DK]	此类一般是网上文件, 按照国标规定顺序 编码制时不输出年份

《研究生学位论文撰写规范》及《研究生学位论文书写范例》中所列英文参考文献例子中的文章名的每个实词首字母都大写，因此需要将英文参考文献的 `title` 域手动修改为每个实词首字母大写。

英文参考文献在 `author` 域中的作者名需要将姓置前，名置后。

### 6.3 参考文献的引用

需要将 `main.tex` 文件中的语句 `\nocite{*}` 屏蔽掉，这样，文中未引用的参考文献就不会出现在文后的参考文献列表中。文中参考文献的引用方法：

- 行文引用请使用命令 `\cite{引用词}`，引用效果为 “[1]”；
- 上标引用请使用命令 `\citeup{引用词}`，引用效果为 “[1]”。

其中，上标引用命令 `\citeup{}` 为本模板自定义的命令，其定义为

```
\newcommand{\citeup}[1]{\textsuperscript{\cite{#1}}}
```

## 6.4 单层罗列环境

哈工大学位论文一般可采用两种罗列环境：一种是并列条目有同样标签的 `itemize` 罗列环境，另一种是具有自动排序编号符号的 `enumerate` 罗列环境。这两种罗列环境的样式参数可参考图 6-1。通过调用 `enumitem` 宏包可以很方便地控

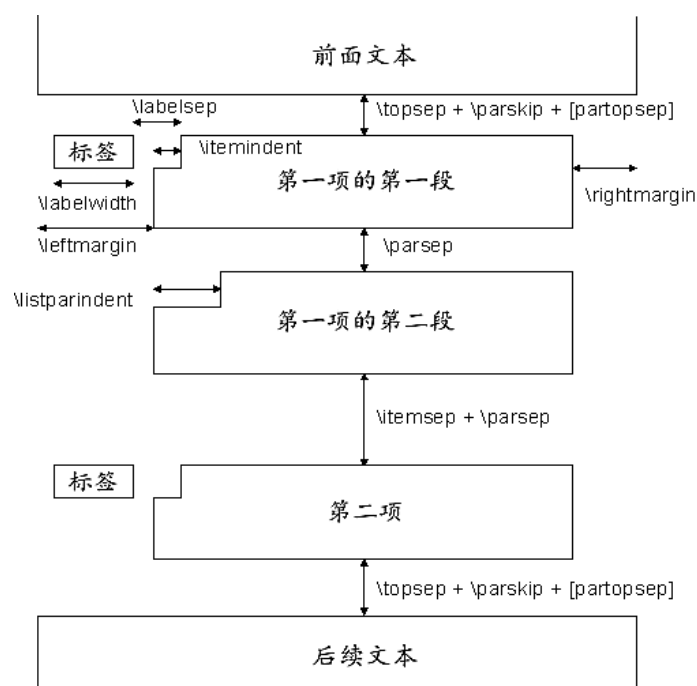


图 6-1 罗列环境参数示意图

Fig.6-1 Schematic diagram of list environments

制罗列环境的布局，其 `format.tex` 文件中的 `\setitemize` 和 `\setenumerate` 命令分别用来设置 `itemize` 和 `enumerate` 环境的样式参数。采用 `itemize` 单层罗列环境的排版形式如下：

- 第一个条目文本内容
- 第二个条目文本内容
- 第三个条目文本内容

其代码如下

```
\begin{itemize}
  \item 第一个条目文本内容
```

```
\item 第二个条目文本内容
...
\item 第三个条目文本内容
\end{itemize}
```

采用 `enumerate` 单层罗列环境的排版形式如下：

1. 第一个条目文本内容
2. 第二个条目文本内容
3. 第三个条目文本内容

其代码如下

```
\begin{enumerate}
\item 第一个条目文本内容
\item 第二个条目文本内容
...
\item 第三个条目文本内容
\end{enumerate}
```

## 6.5 算法

这是一个算法的例子，来自 `worldguy@lilacbbs`。建议将算法放在 `minipage` 环境中，避免算法出现在页面版心之外。

**Input:** training samples,  $(d_i, d_j)_q$ ;  $\mathbf{q}_i, \mathbf{q}_j \in C, q \in \mathbf{Q}$

**Output:** parameter setting  $\lambda^T$

```
1 for  $t=1$  to  $T$  do
2    $\lambda_n^{t+1} = \lambda_n^t + \eta(f_n(q, c, d_i) - f_n(q, c, d_j))$ 
3 end
```

算法环境中右侧空白比较多，若想把右侧的空白框减小，可以采用 `minipage` 环境实现。把 `algorithm` 环境放到 `minipage` 环境里面，并且加上选项[H]禁止算法浮动，下面给出一个例子。需要说明的是，一般不需要进行这种处理。算法标题可有可无，若有中英文标题，请使用 `\AlgoBiCaption{中文标题}{英文标题}`。下面给出两个有标题的例子。需要说明的是，算法的标题是自动换行，没有必要手动换行。

**Input:** training samples,  $(d_i, d_j)_q$ ;  $\mathbf{q}_i, \mathbf{q}_j \in C, q \in \mathbf{Q}$

**Output:** parameter setting  $\lambda^T$

```
1 for  $t=1$  to  $T$  do
2    $\lambda_n^{t+1} = \lambda_n^t + \eta(f_n(q, c, d_i) - f_n(q, c, d_j))$ 
3 end
```

**算法 6-1** 这是一个简短的算法中文图题

Algo. 6-1 This is the English caption of the algorithm

**Input:** training samples,  $(d_i, d_j)_q$ ;  $\mathbf{q}_i, \mathbf{q}_j \in C, q \in \mathbf{Q}$

**Output:** parameter setting  $\lambda^T$

```
1 for  $t=1$  to  $T$  do
2    $\lambda_n^{t+1} = \lambda_n^t + \eta(f_n(q, c, d_i) - f_n(q, c, d_j))$ 
3 end
```

**算法 6-2** 这是一个算法的较长的中文图题，需要换行，这里采用自动换行，如果手动换行会造成算法目录中同样出现断行

Algo. 6-2 This is a long English caption of the algorithm, a new line required, and this a new line

## 6.6 定理定义

若需要书写定理定义等内容，而且带有顺序编号，需要采用如下环境。除了 `proof` 环境之外，其余 9 个环境都可以有一个可选参数作为附加标题。

定理	<code>theorem</code> 环境	定义	<code>definition</code> 环境
例	<code>example</code> 环境	算法	<code>algo</code> 环境
公理	<code>axiom</code> 环境	命题	<code>proposition</code> 环境
引理	<code>lemma</code> 环境	推论	<code>corollary</code> 环境
注解	<code>remark</code> 环境	证明	<code>proof</code> 环境

## 结 论

学位论文的结论作为论文正文的最后一章单独排写，但不加章标题序号。

结论应是作者在学位论文研究过程中所取得的创新性成果的概要总结，不能与摘要混为一谈。博士学位论文结论应包括论文的主要结果、创新点、展望三部分，在结论中应概括论文的核心观点，明确、客观地指出本研究内容的创新性成果（含新见解、新观点、方法创新、技术创新、理论创新），并指出今后进一步在本研究方向进行研究工作的展望与设想。对所取得的创新性成果应注意从定性和定量两方面给出科学、准确的评价，分（1）、（2）、（3）…条列出，宜用“提出了”、“建立了”等词叙述。

## 参考文献

- [1] 林来兴. 空间控制技术[M]. 北京: 宇航出版社, 1992:25–42.
- [2] 辛希孟, 中国科学院文献信息中心, 孟广均,等. 信息技术与信息服务国际研讨会论文集[C]. 北京: 中国社会科学出版社, 1994:45–49.
- [3] 赵耀东. 新时代的工业工程师[M/OL]. 台北: 天下文化出版社, 1998[1998-09-26].  
[http://www.ie.nthu.edu.tw/info/ie.newie.htm\(Big5\)](http://www.ie.nthu.edu.tw/info/ie.newie.htm(Big5)).
- [4] 湛颖. 空间最优交会控制理论与方法研究[D]. 哈尔滨: 哈尔滨工业大学, 1992:8–13.
- [5] Kanamori H. Shaking without Quaking[J]. Science, 1998, 279(5359):2063–2064.
- [6] Mlot C. Plant Biology in the Genome Era[J/OL]. Science, 1998, 281(5375):331–332[1998-09-23]. <http://www.sciencemag.org/cgi/collection/anatmorp>.
- [7] Lamport L. Paxos made simple[J]. ACM Sigact News, 2001, 32(4):18–25.

## 附录 A 带章节的附录

完整的附录内容，包含章节，公式，图表等

### A.1 附录节的内容

这是附录的节的内容

附录中图的示例：



图 A-1 打高尔夫球的人

Fig.A-1 The person playing golf

附录中公式的示例：

$$a = b \times c \quad (\text{A-1})$$

$$E = mc^2 \quad (\text{A-2})$$



## 攻读本科学位期间发表的论文及其他成果

### （一）发表的学术论文

- [1] XXX, XXX. Static Oxidation Model of Al-Mg/C Dissipation Thermal Protection Materials[J]. Rare Metal Materials and Engineering, 2010, 39(Suppl. 1): 520-524. (SCI 收录, IDS号为 669JS, IF=0.16)
- [2] XXX, XXX. 精密超声振动切削单晶铜的计算机仿真研究[J]. 系统仿真学报, 2007, 19 (4): 738-741, 753. (EI 收录号: 20071310514841)
- [3] XXX, XXX. 局部多孔质气体静压轴向轴承静态特性的数值求解[J]. 摩擦学学报, 2007 (1): 68-72. (EI 收录号: 20071510544816)
- [4] XXX, XXX. 硬脆光学晶体材料超精密切削理论研究综述[J]. 机械工程学报, 2003, 39 (8): 15-22. (EI 收录号: 2004088028875)
- [5] XXX, XXX. 基于遗传算法的超精密切削加工表面粗糙度预测模型的参数辨识以及切削参数优化[J]. 机械工程学报, 2005, 41 (11): 158-162. (EI 收录号: 2006039650087)
- [6] XXX, XXX. Discrete Sliding Mode Control with Fuzzy Adaptive Reaching Law on 6-PEES Parallel Robot[C]. Intelligent System Design and Applications, Jinan, 2006: 649-652. (EI 收录号: 20073210746529)

### （二）申请及已获得的专利（无专利时此项不必列出）

- [1] XXX, XXX. 一种温热外敷药制备方案: 中国, 88105607.3[P]. 1989-07-26.

### （三）参与的科研项目及获奖情况

- [1] XXX, XXX. XX 气体静压轴承技术研究, XX 省自然科学基金项目. 课题编号: XXXX.
- [2] XXX, XXX. XX 静载下预应力混凝土房屋结构设计统一理论. 黑龙江省科学技术二等奖, 2007.

## 哈尔滨工业大学本科毕业设计（论文）原创性声明

本人郑重声明：在哈尔滨工业大学攻读学士学位期间，所提交的毕业设计（论文）《中文词语相似度计算》，是本人在导师指导下独立进行研究工作所取得的成果。对本文的研究工作做出重要贡献的个人和集体，均已在文中以明确方式注明，其它未注明部分不包含他人已发表或撰写过的研究成果，不存在购买、由他人代写、剽窃和伪造数据等作假行为。

本人愿为此声明承担法律责任。

作者签名：

日期： 年 月 日

## 致 谢

衷心感谢导师 XXX 教授对本人的精心指导。他的言传身教将使我终生受益。

感谢 XXX 教授，以及实验室全体老师和同窗们的热情帮助和支持！

本课题承蒙 XXXX 基金资助，特此致谢。

...