

硕士学位论文

局部多孔质气体静压轴承关键技术的研究

**RESEARCH ON KEY TECHNOLOGIES
OF PARTIAL POROUS EXTERNALLY
PRESSURIZED GAS BEARING**

颜瑶

哈尔滨工业大学

2010 年 12 月

国内图书分类号: TM301.2
国际图书分类号: 62-5

学校代码: 10213
密级: 公开

工学硕士学位论文

局部多孔质气体静压轴承关键技术的研究

硕士研究生: 颜瑶

导 师: 夏晓玲教授

副 导 师: 副导名

联 合 导 师: 联导名

申 请 学 位: 工学硕士

学 科: 机械制造及其自动化

所 在 单 位: 机电工程学院

答 辩 日 期: 2010 年 12 月

授予学位单位: 哈尔滨工业大学

Classified Index: TM301.2

U.D.C: 62-5

Dissertation for the Master's Degree in Engineering

RESEARCH ON KEY TECHNOLOGIES OF PARTIAL POROUS EXTERNALLY PRESSURIZED GAS BEARING

Candidate:	Yu Dongmei
Supervisor:	Prof. XXX
Associate Supervisor:	Prof. Assosuper
Co Supervisor:	Prof. Cosuper
Academic Degree Applied for:	Master of Engineering
Specialty:	Mechanical and Automation
Affiliation:	School of Mechatronics Engineering
Date of Defence:	December, 2010
Degree-Conferring-Institution:	Harbin Institute of Technology

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、图表和非公知公用的符号和术语，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, semi-conductor, 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 XeLaTeX 编译方法的配置.....	1
1.2 模板的使用方法介绍.....	3
1.3 哈尔滨工业大学 L ^A T _E X 技术交流 QQ 群介绍.....	3
第 2 章 文本立场分析相关技术概述	4
2.1 引言	4
2.1.1 文本情感分析相关技术概述	4
2.1.2 插图编排.....	4
2.2 L ^A T _E X 中推荐使用的图片格式	5
2.3 单张图片的插入方法.....	5
2.4 具有子图的图片插入方法	8
第 3 章 基于条件编码长短期记忆社交媒体文本立场分析	11
3.1 引言	11
3.2 条件编码长短期记忆神经网络模型.....	11
3.2.1 基于 GloVe 的词嵌入模型	12
3.2.2 长短期记忆神经网络模型	13
3.2.3 条件编码长短期记忆神经网络模型	15
3.3 基于条件编码长短期记忆的文本立场分析.....	16
3.4 实验结果及分析	19
3.4.1 实验设置.....	20
第 4 章 表格的绘制方法	23
4.1 研究生院的绘表规范.....	23
4.2 普通表格的绘制方法.....	23
4.3 长表格的绘制方法	24
4.4 列宽可调表格的绘制方法	27
4.4.1 表格内某单元格内容过长的情况	27

4.4.2 对物理量符号进行注释的情况.....	28
结 论.....	30
参考文献.....	31
附录 A A 带章节的附录.....	32
A.1 附录节的内容.....	32
附录 B B 这个星球上最好的免费 Windows 软件列表.....	33
攻读硕士学位期间发表的论文及其他成果.....	34
哈尔滨工业大学学位论文原创性声明和使用权限.....	35
致 谢.....	36

第 1 章 绪论

1.1 XeLaTeX 编译方法的配置

模版无法用 XeLaTeX 编译的原因主要是 xeCJK 包不兼容 CJK 包。例如：xeCJK 原来重写了 CJK 包大部分命令，如 \xeCJKcaption。v3.2.11 之后，这个命令也废弃了。

ctexbook 中含有默认的字体设置，texlive2014-full (ubuntu 14.04) 中字体设置在下面这个文件中：

```
/usr/share/texlive/texmf-dist/tex/latex/ctex/fontset/ctex-xecjk-winfonds.def
```

该文件中有些字体默认的设置会导致字体出错，例如无法找到 [SimKai] 之类。解决方法也很简单，要么在这个文件里面设置好字体；要么在 setup/package.tex 中自定义好字体。例如：ctex-xecjk-winfonds.def 文件设置为：

```
% ctex-xecjk-winfonds.def: Windows 的 xeCJK 字体设置，默认为六种中易字体
% vim:ft=tex
```

```
\setCJKmainfont{SimSun}
```

```
\setCJKsansfont{SimHei}
```

```
\setCJKfamilyfont{zhsong}{SimSun}
```

```
\setCJKfamilyfont{zhhei}{SimHei}
```

```
\newcommand*\songti{\CJKfamily{zhsong}} % 宋体
```

```
\newcommand*\heiti{\CJKfamily{zhhei}} % 黑体
```

```
\endinput
```

package.tex 中字体定义为：

```
57 \def\atempxetex{xelatex}\ifx\atempxetex\usewhat%\def\atempxetex{xelatex}
```

main.tex 中已定义；

```
58 \usepackage[xetex,
```

```

59         bookmarksnumbered=true,
60         bookmarksopen=true,
61         colorlinks = false ,
62         pdfborder={0 0 1},
63         citecolor =blue,
64         linkcolor =red,
65         anchorcolor=green,
66         urlcolor =blue,
67         breaklinks =true ,
68         naturalnames %与algorithm2e宏包协调
69     ]{ hyperref}
70 %\usepackage[BoldFont, normalindentfirst ,BoldFont,SlantFont]{xeCJK}
71 \ defaultfontfeatures {Mapping=tex-text}
72 \xeCJKsetemboldenfactor{1}%只对随后定义的CJK字体有效
73 \setCJKfamilyfont{hei}{SimHei}
74 \xeCJKsetemboldenfactor{4}
75 \setCJKfamilyfont{song}{SimSun}
76 \xeCJKsetemboldenfactor{1}
77 \setCJKfamilyfont{fs}{FangSong}
78 \setCJKfamilyfont{kai}{KaiTi}
79 \setCJKfamilyfont{li}{LiSu}
80 \setCJKfamilyfont{xw}{STXinwei}
81 \setCJKmainfont{SimSun}
82 \setmainfont{Times New Roman}
83 \setsansfont{Arial}
84 \newcommand{\hei}{\CJKfamily{hei}}% 黑体 (Windows自带simhei.ttf)
85 \newcommand{\song}{\CJKfamily{song}} % 宋体 (Windows自带simsun.ttf)
86 \newcommand{\fs}{\CJKfamily{fs}} % 仿宋体 (Windows自带simfs.ttf)
87 \newcommand{\kaishu}{\CJKfamily{kai}} % 楷体 (Windows自带simkai.ttf)
88 \newcommand{\li}{\CJKfamily{li}} % 隶书 (Windows自带simli.ttf)
89 \newcommand{\xw}{\CJKfamily{xw}} % 隶书 (Windows自带simli.ttf)
90 \newfontfamily\arial{Arial}
91 \newfontfamily\timesnewroman{Times New Roman}

```

1.2 模板的使用方法介绍

如果您是 \LaTeX 新手,请您在使用此模板之前,先观看一下模板的使用说明 PPT 演示文档《哈工大学位论文 \LaTeX 模板使用方法介绍》,先大致了解一下此模板的使用方法,之后再准备使用此模板撰写学位论文;如果您有一定的 \LaTeX 技术基础,可以跳过此步骤。

1.3 哈尔滨工业大学 \LaTeX 技术交流 QQ 群介绍

《哈工大硕博学位论文 \LaTeX 模板》项目现已加入哈工大研究生“学术桥”活动中,其两个官方 QQ 群分别为

学术桥- \LaTeX 交流群 1: 38872389

学术桥- \LaTeX 交流群 2: 88984107

欢迎大家加入。加入 QQ 群之后,请大家将自己的名字前面加上当前月份标识,如果您没有标识,在 QQ 群人数已满但仍有人要加入此 QQ 群时,我们会将您优先请出 QQ 群,谢谢合作!您可以在 QQ 群中和其他人讨论关于此模板或其它 \LaTeX 技术相关的任何问题,在提问之前,可以先去 CTEX 论坛或其它网站搜索您所要得到的解决方案,然后确定是否要继续提问,从而节省您的宝贵时间。群共享中包含大量的 \LaTeX 技术资料,方便大家下载阅读,同时还包含了最新的《哈工大硕博学位论文 \LaTeX 模板》。

原则上,模板在 Google Code 的 PlutoThesis 项目的下载列表中也进行了同步更新,但是模板维护人员不再解答用户在此网站提出的问题,如有问题请加入上述的两个 QQ 群中再询问,敬请谅解。

第 2 章 文本立场分析相关技术概述

2.1 引言

本章概要介绍立场分析及其相关技术。首先从目前研究相对成熟的文本情感分析入手，分布从传统的基于规则、机器学习、深度学习分别讨论文本情感分析技术。由于本文主要以深度学习的方法解决社交媒体中立场分析，所以单独详细分析深度学习在文本情感分析的研究。随后作为本文的重点研究对象详细介绍了分别基于机器学习和深度学习模型的文本立场分析技术相关研究本章总结了各项研究工作的特点，在分析优缺点的基础上引出本文的后续研究

本章 2.2 节介绍情感分析的相关研究，2.3 节介绍立场分析相关研究，2.4 节着重介绍基于深度学习模型的立场分析研究。

2.1.1 文本情感分析相关技术概述

文本情感分析，指用自然语言处理、文本挖掘以及计算机语言学等方法来识别和提取原素材中的主观信息。通常来说，情感分析的目的是为了找出说话者，作者在某些话题上或者针对一个文本两极的观点的态度。这个态度或许是他或她的个人判断或是评估，也许是他当时的情感状态，或是作者有意向的情感交流。文本的情感分析是自然语音处理的重要研究内容之一，且其具有重要的科研价值与商业实用价值，吸引了大量的研究人员的关注。研究人员从不同的角度和不同方法对文本情感分析展开了研究。本节将从基于情感词典、机器学习和深度学习的三个方向分别概述近年来情感分析的研究进展，且论述的重点主要集中与文本的后续研究相关的深度学习。

2.1.2 插图编排

插图之前，文中必须有关于本插图的提示，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。

2.2 L^AT_EX 中推荐使用的图片格式

在 L^AT_EX 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。

2.3 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 2-1 所示。

其插入图片的代码及其说明如下。

```
\begin{figure}[htbp]
\centering
```

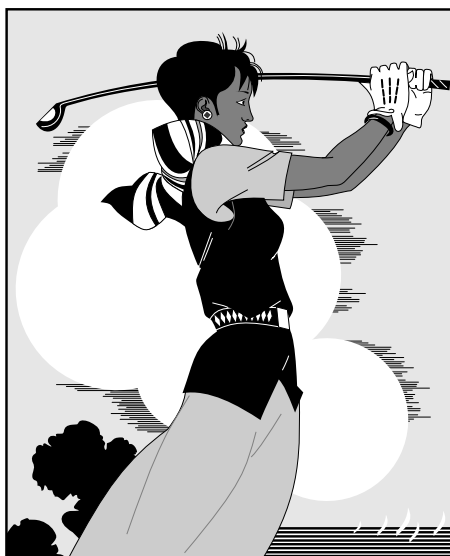


图 2-1 打高尔夫球的人

Fig.2-1 The person playing golf

```
\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**\bcaption**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2-2和图2-3所示。



图 2-2 打高尔夫球的人

Fig.2-2 The person playing golf



图 2-3 打高尔夫球的人

Fig.2-3 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}
\ bcaption [标签名]{}{中文标题}{Fig.$!$}
{English caption}
\end{minipage}
\begin{minipage}{0.4\textwidth}
\ centering
\ includegraphics [width=\textwidth]{文件名}
\ bcaption [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 $(1/n)\text{width}$ 。

2.4 具有子图的图片插入方法

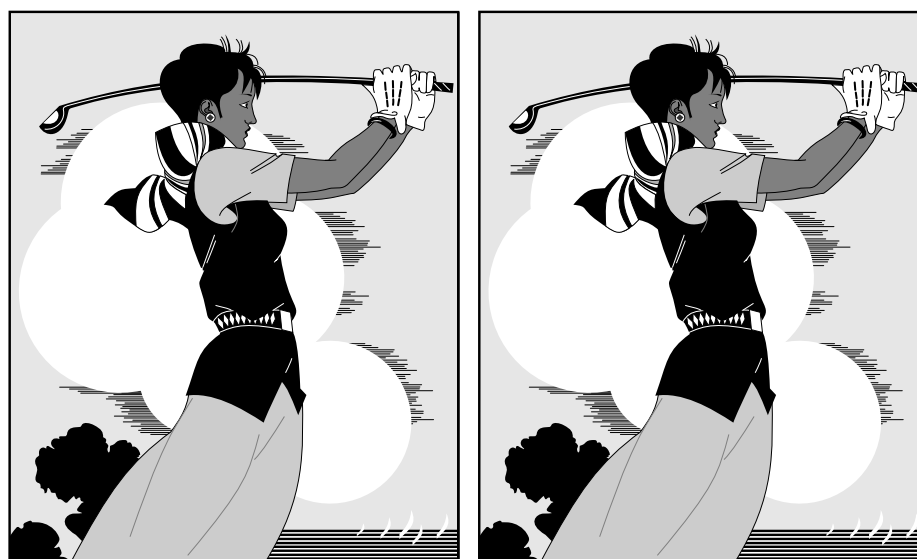
图中若含有子图时，需要调用 subfigure 宏包。博士学位论文规范要求不止总图的标题为中英文形式，其各个子图也应具有中英文形式的标题。然而 ccaption 宏包却无法实现子图的中英文标题功能，这里采用对\subfigure 命令进行嵌套的方法来实现子图的中英文标题功能，如图 2-4 所示。

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

```



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 2-4 打高尔夫球的人

Fig.2-4 The person playing golf

`\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 2-4 a) 和图 2-4 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 L^AT_EX and pdfL^AT_EX 文档。

第 3 章 基于条件编码长短期记忆社交媒体文本立场分析

3.1 引言

当前社交媒体文本立场分析的研究中, 研究人员的研究方向主要集中在如何提取社交媒体文本有效的分类特征。忽略了文本立场分析一种重要的出发点是文本基于某个特定的目标, 若原文本脱离了特定目标, 文本立场分析与情感分析将无差别。基于这个出发点, 本章提出了基于条件编码长短期记忆神经网络的模型, 通过以条件编码的形式引入文本的目标消息, 使立场分析的效果得到显著的提升。本章研究基于条件编码长短期记忆的社交媒体文本立场分析方法, 通过以不同形式给模型接入文本立场的目标消息, 表明了接入目标信息对文本立场分析有明显提升效果。通过在 SemEval2016 英文立场分析数据集和 NLPCC2016 中文立场分析数据集的实验, 论证上述结论。

本章的各节结构如下:3.2 节介绍条件编码长短期记忆神经网络模型;3.3 节介绍基条件编码长短期记忆在文本立场分析;3.4 节为本章实验和结果分析;最后一节为本章小结。

3.2 条件编码长短期记忆神经网络模型

循环神经网络 (Recurrent Neural Networks-RNN) 已经在众多自然语言处理任务上取得了巨大成功。不同于传统过前向反馈神经网络的同层节点无连接层于层之间节点有连接, 循环神经网络引入了定向循环, 可以处理序列数据。RNN 中最大的缺陷是后面时间的节点对于前面节点的感知力下降, 当网络深时训练效果下降。LSTM 可以解决这一问题, 目前 LSTM 是 RNN 中使用最广泛最成功的模型。Rocktaschel[??] 在 2016 年在句子之间的文本蕴含识别 (Recognizing textual entailment RTE) 的研究中提出了条件编码的思想, 其论证了在文本含义识别任务上, 条件编码比单独编码更能抽取两个句子之间的信息。结合文本立场分析的也有文本和目标需要同时考虑特点, 把借鉴条件编码的思想来解决文本立场分析的任务。

3.2.1 基于 GloVe 的词嵌入模型

词的表示是自然语音处理中一个基础且十分重要的任务，大量的研究人员投入到词表示的研究中。早期的有词的表示方式为 **one hot encoding**。每个词独占一个维度，每个词向量有一个维度是 1，其他维度为 0，词向量的维度是所以单词的的长度。**One hot** 编码的特点是假设所有的单词互相独立，这是一个很强的假设，显然在有些任务中并不合适，如词语相似度方面，**dog** 和 **cat** 的相似度应当比 **dog** 和 **not** 高，但是在 **one hot** 编码中他们相似性一样。**one-hot** 编码词表示有一些缺点，如容易受位数灾难的困扰，且不能很好地刻画词与词之间的相似性。词嵌入表示模型能很好的改善 **one-hot** 编码的缺点，词的嵌入模型用稠密且固定长度的向量表示每一个词，而且相似的词具有相似的词向量表示。基于这点 Mikolov[??]2013 提出了 **Word2Vec** 模型，解决了词向量训练速度慢，效率低的缺点。其利用了 **CBOW** (**Continuous Bag-of-words Model**) 和 **Skip-Gram** 的两种语言模型。其中 **CBOW** 的思想是利用词语的上下文词的信息来预测该单词。而 **Skip-Gram** 则采取一种和 **CBOW** 相反的策略，用中间的词的消息来预测上下文的词。**GloVe**(**Global Vectors for Word Representation**) 是斯坦福大学发表的一种词嵌入模型,**GloVe** 尝试借鉴 **NNLM** 和 **word2vec** 的优势来弥补旧方法的劣势，取得了不错的效果。该文发表于 **word2vec** 之后，其方法内核比较朴实和简单，官方实验中，**GloVe** 是略胜 **word2v** 一筹。

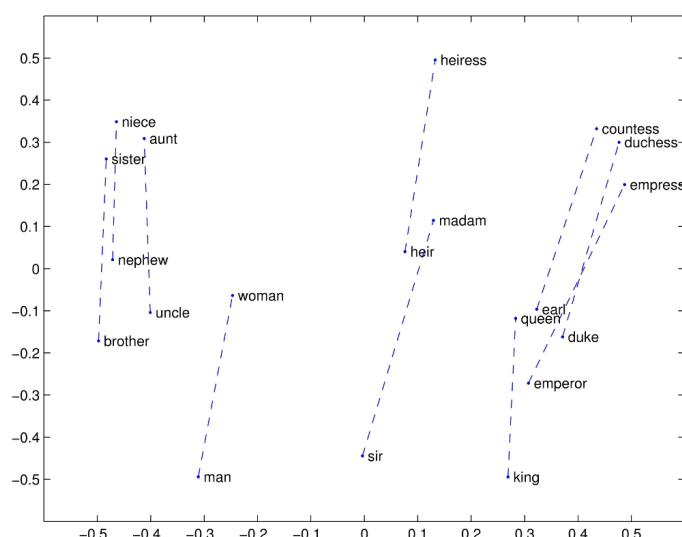


图 3-1 GloVe 可视化词向量

GloVe 结合了基于矩阵分解的词嵌入表示和基于语言模型（例如 **Word2Vec**）的词嵌入表示。基于矩阵分解具有训练快，容易实习的优点，但是生成词向量语义消息十分有限。基于语言模型的词嵌入表示则具有更多语言学层次的支持，在更

多自然语言处理任务上表现更好,但是模型关注的上下文特征较少,忽略了全局的信息。GloVe 发明的初衷,就是想结合两者的长处,建立一个充分利用统计量的更好训练的适用程度更广的词嵌入模型。具体模型建立公式如下

$$F(w_i, w_j, w_k^c) = \frac{P_{ij}}{P_{jk}}$$

其中,取 $word_i$ 的出现次数为 X_i , 定义 $P_{ij} = P(j|i) = \frac{X_{ij}}{X_i}$ 表示在 X_i 的上下文下 $word_i$ 的出现几率, F 则是某一种能够实现我们需求的变换。 w_i, w_j 是实数空间下的 $word_i, word_j$ 的词向量, w_k^c 也是实数空间下的 $word_k$ 的上下文词向量,其作用类似 word2vec 中的上下文向量。为了精简计算引入词向量的线性加减和点乘计算

$$F((w_i - w_j)^T w_k^c) = \frac{F(w_i^T w_k^c)}{F(w_j^T w_k^c)}$$

GloVe 每个词涉及到两个词向量,一个词语本身的向量 w_i , 一个词的 context 向量 w_i^c 。最初这样设计,将词向量和上下文的向量分开,不用一套,是为了在更新参数的时候能够简单地套用 SGD。实验证明两个向量加起来最后起到的效果最好。后面英文的词向量用的是 GloVe 模型在大量的 Twitter 文本上训练的 100 维度的词向量,中文微博词向量是 200 维度的词向量。

3.2.2 长短期记忆神经网络模型

由于文本序列的通常具有较长的长度,导致神经网络的层数较多,而传统的递归神经网络解决序列问题经常会出现梯度消失的问题 (vanishing gradient problem) 与梯度爆炸问题 (gradient exploding problem)。梯度消失问题和梯度爆炸问题一般随着网络层数的增加会变得越来越明显。出现的原因在于对神经网络参数进行链式求导的过程中,输出对于前面递归神经参数的倒数随着累乘激活函数的导数而接近于 0, 以下图的反向传播为例 (假设每一层只有一个神经元且对于每一层 $y_i = \sigma(z_i) = \sigma(w_i x_i + b_i)$, 其中 σ 为 sigmoid 函数)

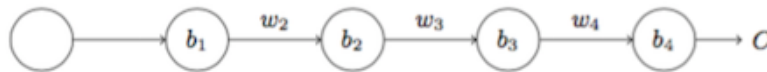


图 3-2 RNN 梯度消失

可以推导出

$$\frac{\partial C}{\partial b_1} = \frac{\partial C}{\partial y_4} \frac{\partial y_4}{\partial z_4} \frac{\partial z_4}{\partial x_4} \frac{\partial x_4}{\partial z_3} \frac{\partial z_3}{\partial x_3} \frac{\partial x_3}{\partial z_2} \frac{\partial z_2}{\partial x_2} \frac{\partial x_2}{\partial z_1} \frac{\partial z_1}{\partial b_1} \quad (3-1)$$

$$= \frac{\partial C}{\partial y_4} \sigma'(z_4) w_4 \sigma'(z_3) w_3 \sigma'(z_2) w_2 \sigma'(z_1) \quad (3-2)$$

一般的非线性激活函数的导数都小于 1（例如 sigmoid 的导数最大值为 $\frac{1}{4}$ ），因此对于上面的链式求导，层数越多，求导结果 $\frac{\partial C}{\partial b_1}$ 越小，因而导致梯度消失的情况出现。长短期记忆（Long Short-Term Memory, LSTM）是一种缓解上述问题的递归神经网络的变种，Hochreiter 在 1997 年首次提出了 LSTM 结构，2000 年 Gers 等人改进 LSTM 模型。

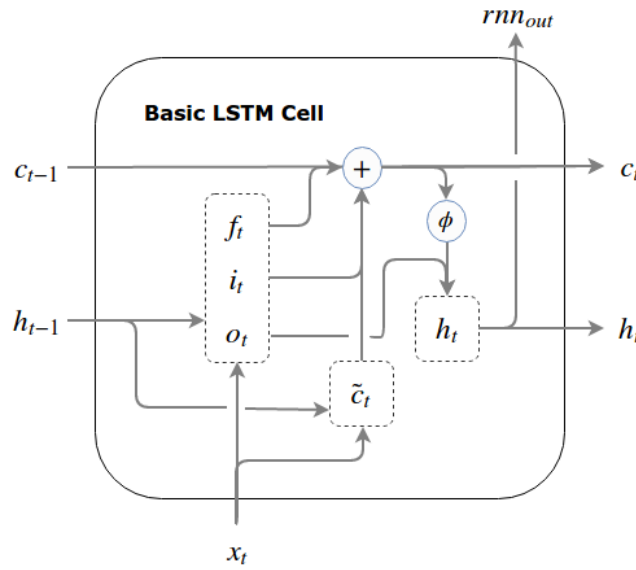


图 3-3 LSTM 单元结构

LSTM 模型提出了记忆存储格（memory cell）的结构，内部包含了遗忘门（forget gate）/ 输入门（input gate）/ 输出门（output gate）。各种门的作用在于调节记忆体在外部输入的情况下应该采取怎么的存储测量，具体门状态和记忆体内部参数的更新公式如下。

$$i_t = \sigma_g(W^i x_t + U_i h_{t-1} + b^i) \quad (3-3)$$

$$f_t = \sigma_g(W^f x_t + U_f h_{t-1} + b^f) \quad (3-4)$$

$$o_t = \sigma_g(W^o x_t + U_o h_{t-1} + b^o) \quad (3-5)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \sigma_c(W_c x_t + U_c h_{t-1} + b_c) \quad (3-6)$$

$$h_t = o_t \odot \sigma_h(c_t) \quad (3-7)$$

其中 σ_g 为 sigmoid 的激活函数, σ_c, σ_h 为 thah 的激活函数, x_t 为输入向量, h_t 为输出向量, c_t 为记忆向量, W, U, b 是矩阵参数和向量参数。各个门的值是保持在 0-1 之间的向量。其中遗忘门向量 f_t 表示上一时刻的记忆体信息需要遗忘多少, 输入门向量 i_t 表示有多少当前时刻输入信息需要加入到记忆体中, 输出门向量 o_t 表示记忆体输出多少信息。

前已经证明, LSTM 是解决长序依赖问题的有效技术, 并且这种技术的普适性非常高, 导致带来的可能性变化非常多。各研究者根据 LSTM 纷纷提出了自己的变量版本, 这就让 LSTM 可以处理千变万化的垂直问题

3.2.3 条件编码长短期记忆神经网络模型

Rocktaschel[] 等在句子之间的文本蕴含识别的研究中提出了条件编码长短期记忆神经网络模型, 文本蕴含定义为一对文本之间的有向推理关系, 其中蕴含前件记作 T(Text), 蕴含后件记作 H(Hypothesis)。如果人们依据自己的常识认为 H 的语义能够由 T 的语义推理得出的话, 那么称 T 蕴含 H, 记作 $T \rightarrow H$, 作者提出的模型的结构是首先有一个 LSTM 模型编码 Text 消息, 另一个不同参数的 LSTM 模型编码 Hypothesis。作者不是简单把两个特征向量拼接在一起, 而是做了如下转换。把第一个编码 Text 信息的 LSTM 模型的记忆状态 (Cell) 保留下来, 作为第二个编码 Hypothesis 的 LSTM 模型记忆状态 (Cell) 的初始值, 此模型建立的了 Text 消息作为条件下的对 Hypothesis 的编码表示。

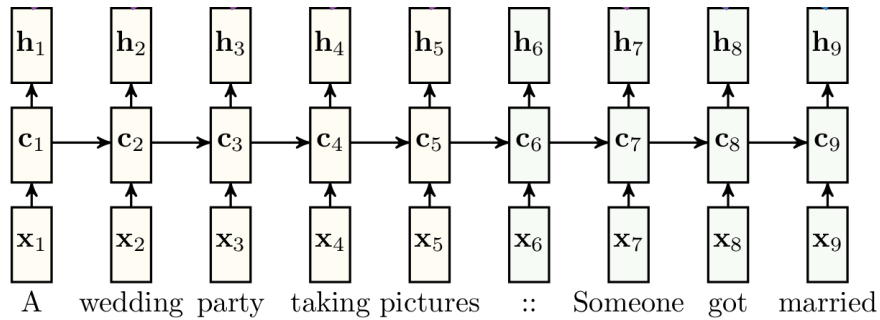


图 3-4 条件编码长短期记忆

如上图图表示所示 “A wedding party taking pictures” 作为我们的 Text 文本, “Someone got married” 作为我们的 Hypothesis, 其中 c_5 作为前一个 LSTM 的记忆体状态被当做编码 Hypothesis 的初始记忆状态。两个 LSTM 具体的状态转移公式如下:

$$[h_1 \ c_1] = LSTM^{Text}(x_1, h_0, c_0) \quad (3-8)$$

...

$$[h_T \ c_T] = LSTM^{Text}(x_1, h_{T-1}, c_{T-1}) \quad (3-9)$$

$$[h_{T+1} \ c_{T+1}] = LSTM^{Hypothesis}(x_1, h_0, c_T) \quad (3-10)$$

...

$$[h_N \ c_N] = LSTM^{Hypothesis}(x_1, h_{N-1}, c_{N-1}) \quad (3-11)$$

$$c = \tanh(Wh_N) \quad (3-12)$$

其中 $(x_1 \dots x_T)$ 为 Text 的序列消息, $(x_{T+1} \dots x_N)$ 为 Hypothesis 的序列信息。 h_0, c_0 为 LSTM 的初始化向量。

实验证明在文本蕴含任务上, 条件 LSTM 模型比单独编码高 3.3% (从 77.6% 提升到 80.9%) 的性能。这种条件编码能使 Text 的信息更好的流向对 Hypothesis 编码的 LSTM 模型, 有了第一个 LSTM 模型传来的记忆状态, 第二个 LSTM 模型能更好的编码 Hypothesis 的消息。

3.3 基于条件编码长短期记忆的文本立场分析

通过 Rocktaschel 在文本蕴含的任务的实验可知, 在处理两个文本序列的编码任务上, 条件编码长短期记忆神经网络比单独独立编码两个文本序列有更好的建模能力。在文本立场分析的任务上, 有文本的信息和目标主题两个文本序列信息, 我们可以借鉴条件编码长短期记忆神经网络在文本蕴含建模的方式, 把文本信息和目标主题信息更好结合起来。在实验部分设计了多种文本序列信息的结合方式, 通过实验证明了以目标主题文本作为条件编码文本信息的模型对文本立场分析有更好的效果。

本文后期实验将在 NLPCC2016 中文微博立场分析数据集和 SemEval2016 英文 Twitter 立场分析数据集, 为较清晰阐述条件编码长短期记忆模型, 以下简短的介绍下两个数据集的样例, 具体的有关数据集的信息将会在下面实验部分做详细介绍。

例 1: 目标主题文本:”深圳禁摩限电” 微博文本:”支持深圳交警。电单车继续治理” 立场分析类标: “Favor” (持支持立场)

目标的文本主题有关“深圳禁摩限电”的主题的, 而从微博文本“支持深圳交

警。电单车继续治理”中，我们可以知道微博的作者首先是赞同了深圳交警的行为，然后叙述了电单车需要得到继续的整治，从两个方法肯定了”深圳禁摩限电”这个主题目标的，因此给出的类标是“Favor”也就是持支持目标主题的立场。

例 1: 目标主题文本:”Hillary Clinton” Twitter 文本:”Hopefully Hillary Clinton gets cancer and dies before she gets the opportunity to embarrass our country any further “，立场分析类标 “Against”（持反对立场）

译文:”真希望希拉里克林顿得癌症然后死去，这样她就不再会有机会再让我们国家蒙羞了。”

目标的文本主题有关“希拉里克林顿”的主题的，这个 Twitter 文本是有关 2016 年美国大选，显然 Twitter 作者一直咒骂希拉里克林顿，希望她得癌症，不让她侮辱国家，可以看出作者有强烈反对主题目标“希拉里克林顿”，因此给出的类标是“Against”，也就是持反对目标主题的立场。

从上述的两个简单的样例可知，立本立场是有两个输入的，一个是立场主题例如“深圳禁摩限电”和“Hillary Clinton”。另外一个立场下的文本“支持深圳交警。电单车继续治理”和” Hopefully Hillary Clinton gets cancer and dies before she gets the opportunity to embarrass our country any further “。在这通过中文微博阐述条件编码长短期记忆模型的建立。首先经过一些数据预处理和分词把主题目标”深圳禁摩限电“和”支持深圳交警。电单车继续治理”转变成”深圳禁摩限电“和”支持深圳交警电单车继续治理“。

在文本立场分析的任务，一般目标主题包含的信息较少，而文本包含了大部分的信息。例如上面两个例子所举例的，目标主题文本分别为” Hillary Clinton “和”深圳禁摩限电“，而 Twitter 和微博文本包含的消息较多，结合立场分析文本的特点，改善了原有的条件编码长短期记忆的网络结构，后续实验论证在多种条件编码长短期记忆的改进方案中，下面所示的网络结构具有更好的实验效果，后续在实验分析其可能的原因。

模型的具体公式如下

$$[h_1 \ c_1] = LSTM^{target}(t_1, h_0, c_0) \quad (3-13)$$

...

$$[h_M \ c_M] = LSTM^{target}(t_M, h_{M-1}, c_{M-1}) \quad (3-14)$$

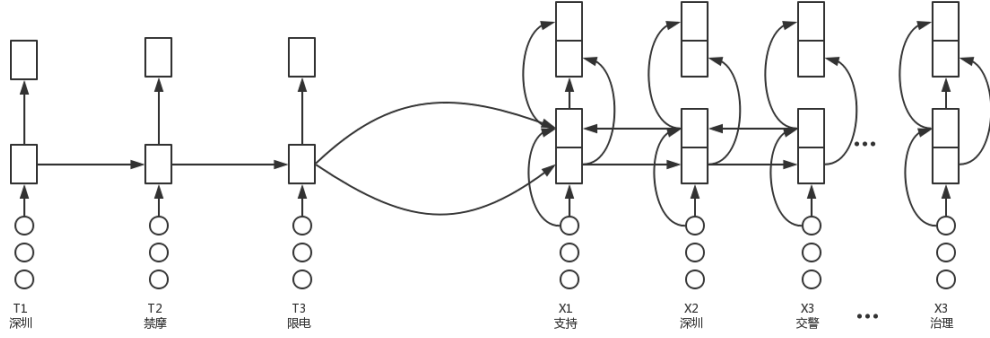


图 3-5 条件编码长短期记忆

$$[h_1^{forward} \ c_1^{forward}] = LSTM^{forward}(x_1, h_0, c_T) \quad (3-15)$$

...

$$[h_N^{forward} \ c_N^{forward}] = LSTM^{forward}(x_n, h_{N-1}^{forward}, c_{N-1}^{forward}) \quad (3-16)$$

$$[h_N^{backward} \ c_N^{backward}] = LSTM^{backward}(x_n, h_0, c_T) \quad (3-17)$$

...

$$[h_1^{backward} \ c_1^{backward}] = LSTM^{backward}(x_1, h_2^{backward}, c_2^{backward}) \quad (3-18)$$

$$c_i = Softmax(W[h_n^{forward} \ h_1^{backward}]) \quad (3-19)$$

其中 M 为主题目标的文本长度, N 为微博 \bigcirc **Twitter** 的文本长度。 $LSTM$ 单向编码主题目标, $LSTM^{forward}$ 为前向编码文本信息, $LSTM^{backward}$ 为后向编码文本信息, h_0 为 $LSTM$ 的初始化向量, c_T 为 $LSTM$ 主题目标编码的最后一个 Cell 状态, $h_1^{backward}$, $h_N^{forward}$ 分别为前向和后向编码的最后一个隐藏状态。

本节以“深圳禁摩限电”为话题目标, 微博文本“支持深圳交警。电单车继续治理”为例, 按本节模型的 4 个层次, 描述基于条件编码长短期记忆的立场分析的过程。

(1) 输入层

先将话题目标和微博文本经过预处理操作, 然后通过分词工具把话题目标和微博文本进行划分, 对于同一个话题目标, 微博文本分词后句子长度有可能不一致, 为了方便后续神经网络框架中的批量的并行计算, 通过统计选择 30 为固定长

度, 长度超过固定长度进行截断操作, 不够的进行补齐词表中规定 <PAD> 关键词。如例微博文本最后转换成”支持深圳交警。电单车继续治理 <PAD> ... <PAD> “

(2) 词向量嵌入层

词向量的嵌入层, 此层的功能是对输入的每一次词检索其词向量 (lookup 操作), 后续实验词向量的预训练由 GloVe 模型在大量无监督语料上训练可得, 预训练的词向量维度为 100, 且把词向量设置为可训练, 随神经网络模型的训练动态调整权重。

(3) 主题目标编码层

通过一个单向的长短期记忆 (LSTM) 模型模型编码主题目标文本, 把分词后的主题目标文本“深圳禁摩限电”通过 lookup 操作取出相应的词向量, 经过一个隐藏层为 64 的单向 LSTM 模型, 且保留最后的记忆体的状态, 如上图所示, 保留 XXX 的单元的信息, 给下一层文本编码做输入。

(4) 文本编码层

由于文本包含大部分信息, 所以采用了双向的 LSTM 模型而不是 Rocktaschel 提出的单向的 LSTM 模型编码文本信息。如上图所示, 文本编码的双向 LSTM 模型的初始 Cell 状态是由上一层的目标主题编码的最后的 Cell 状态填充, 表示主题目标的消息流入到当前的双向 LSTM 模型中参与对文本的编码操作。每一个方向 LSTM 的最后的隐状态 h 作为对整个文本的最终编码表示。

(5) 全连接层

全连接层接受来自文本编码层每个方向最后的隐状态, 拼接两个隐状态信息作为最后的特征向量。全连接层的输出个数为 3, 表示每个立场的预测概率。通过 softmax 激活函数归一化三个立场的概率, 在预测阶段我们选取概率最大的立场当做预测的类标。

3.4 实验结果及分析

本节主要介绍条件编码长短期记忆神经网络模在 2016NLPCC 微博中文语料库和 2016SemEval 英文 Twitter 数据集中的实验结果及分析。并从实验的角度论证条件编码长短期记忆在立场分析问题上的有效性。本节包含两部分: 3.4.1 小节介绍中英文两个数据集中训练和测试文本的分布、实验评价方式以及对比方法; 3.4.2 小结介绍模型训练阶段的性能调优, 以及本章实验与其他方法的比较。

3.4.1 实验设置

为验证条件编码长短期记忆神经网络在立场分析任务上算法的性能表现, 本节采用 2016 NLPCC 中文微博语料和 2016 SemEval 英文的 Twitter 语料两个数据集验证算法的效果。以下分别介绍两数据集的分布。

中文数据集来自 NLPCC2016 立场分析测评任务 [55], 数据集的 5 个话题目标分别为 iPhone SE、春节放鞭炮、俄罗斯在叙利亚的反恐行动、开放二胎政策和深圳禁摩限电。所有语料都来自于新浪微博, 每个微博文本的立场属于“支持”、“反对”和“其他”三者之一。NLPCC 2016 中文微博数据集的训练集、测试集按照 75% 与 25% 的比例划分, 如表 3-1 所示详细介绍每个话题目标下数据的分布。

表 3-1 训练集、测试集话题数量及立场分布比例 (中文数据集)

预置话题分类	训练集数量和立场比例 (%)				测试集数量和立场比例 (%)				文本数量
	数量	支持	反对	其他	数量	支持	反对	其他	
iPhone SE	600	40.8	34.8	24.3	200	37.5	52.0	10.5	800
春节放鞭炮	600	41.7	41.7	16.7	200	44.0	47.0	9.0	800
俄在叙反恐行动	600	41.7	41.7	16.7	200	47.0	43.0	10.0	800
开放二胎政策	600	43.3	33.3	23.3	200	49.5	47.5	3.0	800
深圳禁摩限电	600	26.7	50.0	23.3	200	31.5	55.0	13.5	800
总计	3000	38.8	40.3	20.9	1000	41.9	48.9	9.2	4000

英文数据集来自 SemEval2016 Task6 stance detection[55], 数据集的 5 个话题目标分别为 Atheism(无神论)、Climate Change is a Real Concern(气候变化真实性)、Feminist Movement(女权运动)、Hillary Clinton (希拉里克林顿) 和 Legalization of Abortion(堕胎合法化)。所有语料都来自于英文 Twitter 文本, 每个 Twitter 文本的立场属于“支持”、“反对”和“其他”三者之一。不同于上述中文语料的分布, 每个话题目标英文 Twitter 语料的数量参差不齐, 但总体上训练集和测试集按 70% 与 30% 的比例划分, 如表 3-2 所示详细介绍每个话题目标下数据的分布。

为了和已有方法进行性能比较, 本文在两数据集上都按照分别测评的比例划分出训练集和测试集。其中训练集负责模型的训练和调优, 测试集则进行最后模型的性能的评估。中英文数据集均包含 5 个不同的话题目标, 每个话题目标包含若干的话题文本, 由于每个主题目标关注的内容不同且具有各自独特的语言特点。为了使模型更好的拟合每一种话题的特性, 本文首先按照中英文不同语料集合划分成两个大任务, 然后根据每个语料库在细分成 5 个不同子任务, 分别建立不同的条件

表 3-2 训练集、测试集话题数量及立场分布比例（英文数据集）

预置话题分类	训练集数量和立场比例 (%)				测试集数量和立场比例 (%)				文本数量
	数量	支持	反对	其他	数量	支持	反对	其他	
Atheism	513	17.9	59.3	22.8	220	14.5	72.7	12.7	733
Climate Change	395	53.7	3.8	42.5	169	72.8	6.5	20.7	564
Feminist Movement	664	31.6	49.4	19.0	285	20.4	64.2	15.4	949
Hillary Cliton	689	17.1	57.0	25.8	295	15.3	58.3	26.4	984
Legal of Abortion	653	18.5	54.4	27.1	280	16.4	67.5	16.1	933
总计	2914	25.8	47.9	26.3	1249	24.3	57.3	18.4	4163

编码长短期记忆模型。各子模型预测结束后,统计各个子任务上的性能并汇总预测结果进行最后统一指标的计算。

中英文数据集上的社交媒体文本的立场结果有“支持”,“反对”,“其他”,可以把任务当成一个三分类任务,但是由于三个类在不同的数据集下的不同主题目标的分布有可能很不平均,如果只单独用正确率(Accuracy)作为评测指标则缺失了评价指标的客观性。本文c采取了两个评测任务都使用的”支持“和”反对“的F1指标的微平均(micro-average)作为最后模型的评测指标,但是为了更清晰的评价每个主题目标的性能,每个目标主题也会单独计算微平均评测指标。为了清晰地解释指标的含义,列举以下公式说明

首先定义准确率(Precision, P)、召回率(Recall, R),如公式 3-20 和公式 3-21 所示。

$$P = \frac{TP}{TP + FP} \quad (3-20)$$

$$R = \frac{TP}{TP + FN} \quad (3-21)$$

TP: 正样例预测为正样例的个数

FP: 负样例预测为正样例的个数

FN: 正样例预测为负样例的个数。

精确率计算的是所有”正确被检索的样例(TP)”占有所有”实际被检索到的(TP+FP)样例的比例。召回率计算的是所有”正确被检索的运力(TP)”占有所有”应该检索到的样例(TP+FN)”的比例。如果要同时考虑精确率和召回率,则需要采样两者的调和平均值,也称为F1值,其定义如公式 3-22 所示

$$F1 = \frac{2PR}{P + R} = \frac{2TP}{2TP + FP + FN} \quad (3-22)$$

有关”支持”立场F1值的计算,“支持”类标作为正样本,“反对”后“其他”

作为负样本。因此其计算公式如下所示

$$F1_{favor} = \frac{2P_{favor}R_{favor}}{P_{favor} + R_{favor}} \quad (3-23)$$

同样的”反对“立场 F1 值的计算，“反对”类标作为正样本，“支持”后“其他”作为负样本。因此其计算公式如下所示

$$F1_{against} = \frac{2P_{against}R_{against}}{P_{against} + R_{against}} \quad (3-24)$$

立场分析总的平均指标 F1 的微平均（Micro-average）的计算公式如下

$$F1_{average} = \frac{F1_{favor} + F1_{against}}{2} \quad (3-25)$$

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

Table 4-1 Table in agreement of the standard from graduate school

$D(\text{in})$	$P_u(\text{lbs})$	$u_u(\text{in})$	β	$G_f(\text{psi.in})$
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]
\caption[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
云南省	云/滇	昆明市

表 4-2（续表）

名称	简称	省会或首府
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[1.5 pt] 表头第1个格 & 表头第2个格 & ... & 表头第n个格\midrule[1 pt]
\endfirsthead
\multicolumn{n}{r}{表~\thetable（续表）}\vspace{0.5em}
\toprule[1.5 pt] 表头第1个格 & 表头第2个格 & ... & 表头第n个格\midrule[1 pt]
\endhead
\bottomrule[1.5 pt]
\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。

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

绘制这种表格的代码及其说明如下。

```

\begin{table}[htbp]
\caption[标签名]{{中文标题}{Table$!\$}{English caption}}
\vspace{0.5em}\wuhao
\begin{tabularx}{\textwidth}{1 ... X ... 1}
\toprule[1.5 pt]
表头第1个格 & ... & 表头第X个格 & ... & 表头第n个格 \\
\midrule[1 pt]
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[1.5 pt]
\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 ρ ——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；
 ω_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。

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.

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：

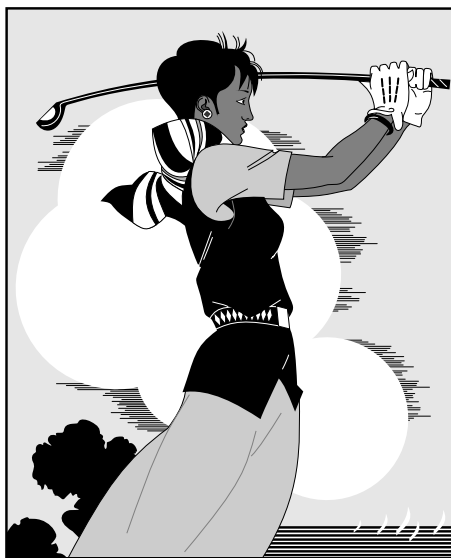


图 1-1 打高尔夫球的人

Fig.1-1 The person playing golf

附录中公式的示例：

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

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

附录 B 这个星球上最好的免费 Windows 软件列表

杀毒软件

avast! 免费杀毒软件——推荐

AVG 杀毒永久免费版——推荐

Avira Free Antivirus（小红伞）

攻读硕士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.

哈尔滨工业大学学位论文原创性声明和使用权限

学位论文原创性声明

本人郑重声明：此处所提交的学位论文《局部多孔质气体静压轴承关键技术的研究》，是本人在导师指导下，在哈尔滨工业大学攻读学位期间独立进行研究工作所取得的成果，且学位论文中除已标注引用文献的部分外不包含他人完成或已发表的研究成果。对本学位论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。

作者签名：

日期： 年 月 日

学位论文使用权限

学位论文是研究生在哈尔滨工业大学攻读学位期间完成的成果，知识产权归属哈尔滨工业大学。学位论文的使用权限如下：

(1) 学校可以采用影印、缩印或其他复制手段保存研究生上交的学位论文，并向国家图书馆报送学位论文；(2) 学校可以将学位论文部分或全部内容编入有关数据库进行检索和提供相应阅览服务；(3) 研究生毕业后发表与此学位论文研究成果相关的学术论文和其他成果时，应征得导师同意，且第一署名单位为哈尔滨工业大学。

保密论文在保密期内遵守有关保密规定，解密后适用于此使用权限规定。

本人知悉学位论文的使用权限，并将遵守有关规定。

作者签名：

日期： 年 月 日

导师签名：

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。

...