

数据挖掘大作业——TexTCNN对TNEWS数据分类

邹世奇

17123143

778377698@qq.com

摘要

本文档包含准备提交CCL-2020或已被其论文集接受发表的论文的格式说明。此文档本身也遵循自己的格式说明，因此是您论文真实的排版格式。这些说明适用于提交审稿的论文和最终版接受的论文。请所有作者遵守此文档中所说明的要求。

1 引言

下列说明为CCL-2020论文投稿和论文发表的作者提供指导。所有论文都需遵从此说明。作者需要提供PDF版本的论文。**此论文集设计版面为A4纸。**

论文必须为单栏格式，**行间距为单行**。全部页面紧接上边距开始。具体请参阅后面关于首页的格式化准则。论文的长度不得超过第2节中所述的最大页面限制。不要为页面设置编号。

1.1 可用资源

强烈建议您使用L^AT_EX以及官方的CCL-2020格式文件(ccl2020-zh.sty)和文献引用格式文件(ccl.bst)准备您的PDF文件。这些文件包含在ccl2020.zip中，可以从<http://www.cips-cl.org/>下载。文件ccl2020.zip中，也包括ccl2020-zh.pdf文件和它的L^AT_EX源码文件(ccl2020-zh.tex)。

1.2 电子文稿的格式

为了制作电子源文档，您可以使用Adobe的PDF工具。PDF文件通常是在L^AT_EX中使用`pdflatex`生成。如果您的L^AT_EX版本生成的是Postscript文件，您可以使用`ps2pdf`或`dvipdf`将它们转换为PDF文件。在Windows上，您也可以使用Adobe Distiller生成PDF文件。

请确保您的PDF文件包含所有必需的字体（尤其是非拉丁字符的树形图、符号和字体）。当您打印或创建PDF文件时，通常打印机设置中会有一个选项：无、全部或仅非标准字体。请确保您选择以下选项：包括所有字体。**在递交之前，通过不同的计算机打印来测试您的PDF文件格式。**此外，某些文字处理程序可能会生成非常大的PDF文件，其中每个页面都呈现为图像。这样的图像可能清晰度不佳。在这种情况下，请尝试其他方法来获取PDF文件。在某些系统上，一种方法是安装打印到文件驱动程序，将文档发送到打印机，并指定“输出到文件”，然后将文件转换为PDF。

最重要的是指定**A4纸张格式**（21厘米x 29.7厘米）。比如，使用`dvips`时，应该指定`-ta4`。

如果您不能按上述要求提交电子版文件，请尽快联系出版主席。

1.3 排版

使用此说明文档展示的格式，将单列的稿件格式化为一个页面。A4纸页面上的确切尺寸是：

- 左右边距: 2.5 cm
- 上边距: 2.5 cm

- 下边距: 2.5 cm
- 宽度: 16.0 cm
- 高度: 24.7 cm

不得以任何其他纸张尺寸提交论文。如果您不能按以上要求提交电子版文件，请尽快联系出版主席。

1.4 字体

为了统一起见，请使用**宋体**（L^AT_EX2e的默认值）字体。

文本类型	字体	风格
论文标题	15 pt	粗
作者姓名	12 pt	粗
作者地址	12 pt	
标题“摘要”	12 pt	粗
章节标题	12 pt	粗
正文内容	11 pt	
图表标题	11 pt	
图表子标题	9 pt	
摘要正文	11 pt	
参考文献	10 pt	
页脚	9 pt	

Table 1: 字体说明

1.5 首页

将标题、作者姓名和工作关系页面居中，不要将工作关系放在脚注中。不要包含提交过程中分配的论文ID号。在提交供审稿的版本中，请勿包含作者的姓名或工作关系。

标题: 请将标题居中放置在首页顶部，并使用15pt粗体。(关于字体和风格的完整说明，请参照表 1)。长标题可以分成两行，中间不要留有空行。标题位于大约从页面顶部开始2.5厘米处，然后是空白行，接着是作者姓名，以及下一行的工作关系。工作关系应包含作者的完整地址，如果可能的话，还有一个电子邮件地址。从第一页的顶部开始7.5厘米处开始正文页。

为了保持作者信息在所有会议出版物中的一致性，标题、作者姓名和地址应该和论文投稿网站上输入的内容完全相同。如果它们不同，则出版主席可以在不与您协商的情况下做出选择。所以请仔细检查该信息是否一致。

摘要: 在地址和正文之间输入摘要。摘要文字的宽度比正文每侧小约0.6厘米。**摘要**一词以12pt粗体居中放于摘要正文上方。摘要应简明扼要概述论文内容和结论，不超过200个字。摘要文字应为11pt字体。

正文: 摘要之后即是正文。请遵守单列格式，如本文所示，不要包括页码。

缩进: 在开始新段落时，正文和小节标题使用11pt字体，章节标题为12pt，论文标题为15pt。

许可: 将许可声明作为未标记（未编号）脚注放置在最终版论文的首页。有关详细信息和动机，请参见下面的1.9。

1.6 章节

标题: 按本文档中显示的样式，设置章节和小结标题的类型和标签。使用阿拉伯数字编号以方便交叉引用。小节编号紧跟章节编号之后，使用点号分隔。

引用: 文本中的引用出现在括号中,像 (?)或者，如果作者的名字本身出现在文本中，例如Gusfield (?)。年份上附加小写字母以防止歧义。将双作者写为 (?)，但当两个以上作者参与其中时写为 (?)。合并多个引用写为 (?; ?)。也不要使用完整的引文作为句子的组成部分。我们建议不要使用

“(?)表明...”

而请使用

“Gusfield (?)表明...”

如果您使用提供的L^AT_EX和BibT_EX样式文件，则您可以使用命令\newcite获得“作者（年份）”的引用。

由于审核将是双盲的，因此论文的提交版本不应包含作者的姓名和工作关系。此外，包括揭示作者身份的自引用，例如：

“我们之前的研究表明(?) ...”

应该避免。而是使用诸如

“Gusfield (?) 之前的研究表明... ”

请不要使用匿名引用 提交论文进行审阅时，不要包含以下信息：致谢、项目名称、授权号，以及最近3周才公开或即将公开的资源或工具的名称或URL，以免损害提交内容的匿名性。不符合这些要求的论文可能会被拒绝而无需审阅。但是，这些详细信息可以包含在出版的最终论文中。

参考文献：收集全部参考资料放置在标题**参考文献**下面。将本节放在任何附录之前，除非它们包含引用。按字母顺序排列参考文献，由第一作者而非论文中出现的先后决定次序。使用一致的格式提供尽可能完整的引用。

附录：附录（如果有）直接跟随正文和参考文献（但请参见上文）。用字母依次编号并提供有效标题：**附录A.附录标题**。

1.7 脚注

脚注：将脚注放在页面底部，并使用9pt字体。可以用星号、其他编号或引用符号。⁰ 脚注和正文之间应有分隔线。¹

1.8 图形

插图：尽可能将图形、表格和图片放在最初讨论它们的地方，而不是放在最后。不建议使用彩色插图，除非您确认用黑色墨水打印时它们是可以被理解的。

标题：为每个插图提供标题；顺序编号为：“图1.图的标题”、“表1.表格标题”。标题放置在插图或表格内容下方，使用11pt字体。较窄的图形和单列格式可能会导致很大的空白处，参见表 1两侧的宽边距。如果您有多个内容相关的图形，可能的话最好将它们组合成一个图形。您可以使用以下子标题来标识子图形：子图形编号为(a)，(b)，(c)等，并使用9pt字体。L^AT_EX包wrapfig、subfig、subtable和/或subcaption可能会有用。

1.9 许可声明

我们要求作者按照《Creative Commons Attribution 4.0 International Licence (CC-BY)》许可其最终版论文。这意味着作者（版权所有人）保留版权，但授予所有人修改和重新分发论文的权利，只要他们署名原文作者并列修改内容。换句话说，该许可使研究人员可以在没有法律问题的情况下使用研究论文进行研究。请参阅<http://creativecommons.org/licenses/by/4.0/>许可条款。

- 本作品已根据《Creative Commons Attribution 4.0 International Licence》获得许可。许可证详细信息：<http://creativecommons.org/licenses/by/4.0/>.

我们强烈建议您将论文许可为以上CC许可。（请注意，本许可声明仅与接受论文的最终版本有关。提交审稿的论文不需要。）

⁰这是脚注的显示方式。

¹注意将脚注与正文分开的分隔线。

2 提交论文长度

提交论文最大长度为10页（A4），外加无限制的参考页数。最终版本将为被接受论文的作者提供额外的空间以便整合审稿人的意见。

不符合指定长度和格式要求的论文，可能会被拒稿。

致谢

致谢应在参考文献之前进行。不要给致谢章节进行编号。在提交论文进行审稿时，不要包括此部分。

参考文献

大规模文本分类网络TextCNN介绍. url: <https://blog.csdn.net/u012762419/article/details/79561441>

pytorch实现textCNN. url: https://doc.flyai.com/blog/textcnn_pytorch.html

文本分类实战（二）—— textCNN 模型. url: <https://www.cnblogs.com/jiangxinyang/p/10207482.html>

NLP文本分类入门学习及TextCnn实践笔记（一）. url: <https://blog.csdn.net/wangyueshu/article/details/106493048>