

# 大连理工大学

## 毕业设计（论文）任务书

（理工类）

题 目 名 称 基于方面的情感分析系统

学部（学院）软件学院 专 业 软件工程

学 生 姓 名 宋安洋 学 号 201892436

指 导 教 师 张晓彤 职 称 特聘副研究员

下发 2021 年 12 月 12 日 上交 2022 年 6 月 1 日

# 本科生毕业设计（论文）须知

1. 认真学习理解《大连理工大学大学生毕业设计（论文）工作管理办法》。

2. 努力学习、勤于实践、勇于创新，保质保量地完成任务书规定的内容。

3. 独立完成规定的工作任务，严格遵守学术道德、学术规范，严禁抄袭、剽窃、伪造、篡改、买卖、代写等学术不端行为，一经发现将根据《大连理工大学学生违纪处分规定》给予相应处分，并追究相关指导教师指导责任；已授予学位的经查实将依法撤销已授予学位，并注销学位证书。

4. 毕业设计（论文）成果、资料应于答辩结束后及时交给学部（学院）收存，学生不得擅自带离学校。经指导教师推荐可作为论文发表。

5. 爱护仪器设备，节约材料，严格遵守操作规程及实验室有关制度。

6. 毕业设计（论文）完成后，将《大连理工大学毕业设计（论文）任务书》同毕业设计（论文）一同交给指导教师。

## 毕业设计（论文）任务下达表（此表内容可打印）

题目：基于方面的情感分析系统

Title: Aspect Based Sentiment Analysis System

一、题目来源（在合适的项目前划√）

☐ 科研项目、☒ 工程模拟、☐ 实际应用、☐ 自拟课题、☐ 大创项目

其他：

二、设计（论文）要求：设计参数

1. 工程语言：Anaconda 4.10.1 (Python 3.8.8)
2. 操作系统：Windows10
3. 开发环境：Pycharm2020、Jupyter Notebook
4. 实现目标：通过深度学习方法，建立相关的神经网络模型，完成基于方面的情感分析，设计相应界面，完成整个系统的开发。

三、个人重点工作

1. 了解 ABSA 领域的相关概念
2. 复习机器学习、深度学习相关的理论基础
3. 熟悉深度学习框架 PyTorch 的使用
4. 总结 LSTM、Transformer 等神经网络模型架构
5. 熟悉使用 flask web 应用开发，熟悉 html 页面的编写
6. 结合深度学习理论编码实现用于基于方面的情感分析的神经网络模型
7. 完成基于方面的情感分析整个系统的设计与开发

#### 四、各阶段时间安排、应完成的工作量

第 1-2 周: 了解 ABSA 领域的相关概念

第 3-4 周: 完成毕业设计开题报告, 复习机器学习、深度学习相关的理论基础

第 5 周: 完成外刊论文翻译任务

第 6-8 周: 阅读文献, 总结常用的基于方面的情感分析的方法

第 9-10 周: 初步设计相关神经网络模型

第 11-12 周: 编写代码, 完成相关神经网络模型

第 13-14 周: 完善模型, 编写测试程序, 进行实验测试

第 15-16 周: 整理相关数据, 获得结论, 撰写论文

第 17-18 周: 修改论文, 制作答辩 PPT, 准备答辩

#### 五、应阅读的资料及主要参考文献目录

- [1] Zhou J, Huang J X, Chen Q, et al. Deep learning for aspect-level sentiment classification: survey, vision, and challenges[J]. IEEE access, 2019, 7: 78454-78483.
- [2] Tang D, Qin B, Feng X, et al. Effective LSTMs for target-dependent sentiment classification[J]. arXiv preprint arXiv:1512.01100, 2015.
- [3] Wang Y, Huang M, Zhu X, et al. Attention-based LSTM for aspect-level sentiment classification[C]//Proceedings of the 2016 conference on empirical methods in natural language processing. 2016: 606-615.
- [4] Liu J, Zhang Y. Attention modeling for targeted sentiment[C]//Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers. 2017: 572-577.
- [5] Xue W, Li T. Aspect based sentiment analysis with gated convolutional networks[J]. arXiv preprint arXiv:1805.07043, 2018.
- [6] Huang B, Carley K M. Parameterized convolutional neural networks for aspect level sentiment classification[J]. arXiv preprint arXiv:1909.06276, 2019.
- [7] Zhang C, Li Q, Song D. Aspect-based sentiment classification with aspect-specific graph convolutional networks[J]. arXiv preprint arXiv:1909.03477, 2019.
- [8] Devlin J, Chang M W, Lee K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding[J]. arXiv preprint arXiv:1810.04805, 2018.
- [9] Xu H, Liu B, Shu L, et al. BERT post-training for review reading comprehension and aspect-based sentiment analysis[J]. arXiv preprint arXiv:1904.02232, 2019.
- [10] Gao Z, Feng A, Song X, et al. Target-dependent sentiment classification with BERT[J]. IEEE Access, 2019, 7: 154290-154299.
- [11] 孟霞. 基于层次化双向 LSTM 的评论方面级别情感分析研究[D]. 吉林大学, 2019.
- [12] 程康鑫. 基于 LSTM 与 CNN 的中文餐饮评论情感特征提取算法[D]. 北京邮电大学, 2020. DOI:10.26969/d.cnki.gbydu.2020.002649.
- [13] 王昆, 郑毅, 方书雅, 等. 基于文本筛选和改进 BERT 的长文本方面级情感分析[J]. 计算机应用, 2020, 40(10):7.

指导教师 (签字)

张晓彤

2021 年 12 月 9 日

学院院长 (系主任) (签字)

2021 年 12 月 10 日

教学部长 (院长) (签字)

2021 年 12 月 11 日

## 毕业设计（论文）过程检查情况记录（记录内容需手写）

第 1 次	<p>本周工作进展情况（学生填写）</p> <ol style="list-style-type: none"> <li>1. 查阅了相关文献，了解了 ABSA 领域的研究背景、弄清了 aspect、category 等相关概念。</li> <li>2. 了解了常用数据集有 SemEval 的 restaurant 和 laptop 的评论，以及 ACL 的 twitter 的评论。</li> <li>3. 了解了 SemEval 2014 提出的四个子任务。</li> </ol> <p>指导教师检查意见</p> <p style="padding-left: 40px;">继续阅读相关的文献，对需要构建的模型有一定了解。</p> <p style="text-align: right;">指导教师(签字) <u>张晓彤</u>      2021 年 12 月 17 日</p>
第 2 次	<p>本周工作进展情况（学生填写）：</p> <ol style="list-style-type: none"> <li>1. 复习了学习吴恩达《神经网络和深度学习》、《改善深层神经网络：超参数调试、正则化及优化》、《结构化机器学习项目》、《卷积神经网络》、《序列模型》五门课程时所做的课程笔记。</li> <li>2. 复习了学习邱锡鹏教授的《神经网络与深度学习》教材所做的阅读笔记。</li> <li>3. 复习了 nlp beginner 的五个入门任务。</li> </ol> <p>指导教师检查意见</p> <p style="padding-left: 40px;">继续阅读相关的文献和技术资料，需要对构建的模型有一定的了解。</p> <p style="text-align: right;">指导教师(签字) <u>张晓彤</u>      2021 年 12 月 24 日</p>
第 3 次	<p>本周工作进展情况（学生填写）</p> <ol style="list-style-type: none"> <li>1. 配置了本地的 CUDA 和 cudnn 环境，安装了 cu 版的 pytorch 包，安装了 Transformer 包。</li> <li>2. 完成了论文第一章中对 python 语言、类库介绍的撰写。</li> </ol> <p>指导教师检查意见</p> <p style="padding-left: 40px;">任务完成情况良好，论文需注意格式问题。</p> <p style="text-align: right;">指导教师(签字) <u>张晓彤</u>      2021 年 12 月 31 日</p>

## 毕业设计（论文）过程检查情况记录（记录内容需手写）

第 4 次	<p>本周工作进展情况（学生填写）</p> <ol style="list-style-type: none"> <li>1. 尝试使用了 google colab，学习了如何连接 google drive。一方面由于本地 gpu 内存有限，另一方面为提高实验效率，因此之后会考虑使用 google colab 运行一部分实验。</li> <li>2. 尝试学习了使用 tpu，可能 torch_xla 无法很好地支持自定义数据集，运行效率过慢，因此放弃了对 tpu 的使用，实验决定仅使用 gpu 训练。</li> </ol> <p>指导教师检查意见</p> <p style="padding-left: 40px;">需要继续阅读相关的文献。可以开始外文翻译工作。</p> <p style="text-align: right;">指导教师(签字) <u>张晓彤</u>          2022 年 1 月 7 日</p>
第 5 次	<p>本周工作进展情况（学生填写）</p> <ol style="list-style-type: none"> <li>1. 重点阅读了关于 ABSA 研究的详细综述：Deep Learning for Aspect-Level Sentiment Classification: Survey, Vision, and Challenges. 全面了解过去的研究中已尝试的方法，以及未来的挑战。</li> <li>2. 重点完成了外文翻译任务，并在老师的指导下进行了修改。</li> </ol> <p>指导教师检查意见：还需继续了解过去的方法，并考虑算法的设计。</p> <p style="text-align: right;">指导教师(签字) <u>张晓彤</u>          2022 年 1 月 14 日</p>
第 6 次	<p>本周工作进展情况（学生填写）</p> <ol style="list-style-type: none"> <li>1. 全面阅读了主要的参考文献，充分调研了过去的研究者提出的方法。</li> <li>2. 参考了约 20 篇文献，完成了引言中文献综述相关内容的撰写。</li> </ol> <p>指导教师检查意见：任务完成情况良好，论文撰写过程中的格式问题还需注意。</p> <p style="text-align: right;">指导教师(签字) <u>张晓彤</u>          2022 年 3 月 4 日</p>

## 毕业设计（论文）过程检查情况记录（记录内容需手写）

第 7 次	<p>本周工作进展情况（学生填写）</p> <ol style="list-style-type: none"> <li>1. 回顾并总结了神经网络相关的基础理论知识。</li> <li>2. 完成了第一章相关技术介绍中本文主要用到的技术的神经网络基础知识的撰写。</li> </ol> <p>指导教师检查意见：任务完成良好，论文撰写中的格式问题依然需要注意。</p> <p style="text-align: right;">指导教师(签字) <u>张晓彤</u>      2022 年 3 月 11 日</p>
第 8 次	<p>本周工作进展情况（学生填写）</p> <ol style="list-style-type: none"> <li>1. 学习了系统落地中需要使用到的 flask 框架。</li> <li>2. 完成了论文第一章相关技术介绍中有关 flask 介绍的撰写，至此，完成了整个第一章的撰写。</li> </ol> <p>指导教师检查意见：任务完成情况良好。抓紧开始程序编写，可以尝试先做一个简易的交互界面。</p> <p style="text-align: right;">指导教师(签字) <u>张晓彤</u>      2022 年 3 月 18 日</p>
第 9 次	<p>本周工作进展情况（学生填写）</p> <ol style="list-style-type: none"> <li>1. 重点阅读了 TD-LSTM、ATAE-LSTM、LCF-BERT 等方法的文献。</li> <li>2. 完成了对上述方法的程序编写。</li> <li>3. 完成了第二章系统方法中，TD-LSTM、LCF-BERT 的介绍撰写。</li> </ol> <p>指导教师检查意见：可以参考一下相关代码，继续完成程序编写，抓紧开始实验。</p> <p style="text-align: right;">指导教师(签字) <u>张晓彤</u>      2022 年 3 月 25 日</p>

## 毕业设计（论文）过程检查情况记录（记录内容需手写）

第 10 次	<p>本周工作进展情况（学生填写）</p> <ol style="list-style-type: none"> <li>1. 分析并提出了 TD-Transformer、TDLC-Transformer、TAGG-LSTM 三个新方法。</li> <li>2. 完成了对上述三个方法的代码编写。</li> <li>3. 完成了对上述三个方法架构图的绘制。</li> <li>4. 完成了论文第二章系统方法中，对三个方法的描述。</li> </ol> <p>指导教师检查意见：仍需注意论文格式问题。继续完成程序编写，抓紧开始实验。</p> <p style="text-align: right;">指导教师(签字) <u>张晓彤</u>      2022 年 4 月 1 日</p>
第 11 次	<p>本周工作进展情况（学生填写）</p> <ol style="list-style-type: none"> <li>1. 完成了对数据集的分析。</li> <li>2. 完成了训练和验证相关程序的代码编写。</li> <li>3. 基于对数据集的分析完成了论文第三章系统实验的第一节数据集的撰写。</li> </ol> <p>指导教师检查意见：任务完成情况良好，尽快开始实验。</p> <p style="text-align: right;">指导教师(签字) <u>张晓彤</u>      2022 年 4 月 8 日</p>
第 12 次	<p>本周工作进展情况（学生填写）</p> <ol style="list-style-type: none"> <li>1. 基于过去方法对超参数的推荐值，通过五折交叉验证，完成了对超参数的调试与设置。</li> <li>2. 基于调试好的超参数完成了第三章系统实验的第二节超参数设置。</li> </ol> <p>指导教师检查意见：论文格式问题仍需注意。研究实验已进入核心部分，注意加入不同技术方法间的对比分析。</p> <p style="text-align: right;">指导教师(签字) <u>张晓彤</u>      2022 年 4 月 15 日</p>



## 毕业设计（论文）过程检查情况记录（记录内容需手写）

第 13 次	<p>本周工作进展情况（学生填写）</p> <ol style="list-style-type: none"> <li>1. 完成了文中选用的方法的实验，log 记录了实验的训练过程，以及对测试集评估得到的准确率和 F1 分数。</li> <li>2. 基于实验结果完成了论文第三章系统实验第三节方法比较的撰写。</li> </ol> <p>指导教师检查意见</p> <p style="padding-left: 2em;">任务完成情况良好。实验结果的分析比较较为清晰。论文的图表格式还需注意。</p> <p style="text-align: right;">指导教师(签字) <u>张晓彤</u>                      2022 年 4 月 22 日</p>
第 14 次	<p>本周工作进展情况（学生填写）</p> <ol style="list-style-type: none"> <li>1. 完成了领域语料中，话题内和跨话题数据集继续预训练 BERT 的实验，并整理了结果的准确率和 F1 分数</li> <li>2. 基于实验结果完成了论文第三章系统实验第四节领域预料继续预训练 BERT 的撰写。</li> </ol> <p>指导教师检查意见</p> <p style="padding-left: 2em;">实验具有一定的创新性，分析实验结果的语言描述还应再清晰一些，避免表意不清与阅读困难。</p> <p style="text-align: right;">指导教师(签字) <u>张晓彤</u>                      2022 年 4 月 29 日</p>
第 15 次	<p>本周工作进展情况（学生填写）</p> <ol style="list-style-type: none"> <li>1. 通过可视化等手段对不同方法预测结果的案例做了进一步分析。</li> <li>2. 基于分析结果，完成了对第四章案例分析的撰写。</li> </ol> <p>指导教师检查意见</p> <p style="padding-left: 2em;">任务完成情况良好。尽快完成系统应用的编写。</p> <p style="text-align: right;">指导教师(签字) <u>张晓彤</u>                      2022 年 5 月 6 日</p>

## 毕业设计（论文）过程检查情况记录（记录内容需手写）

第 16 次	<p>本周工作进展情况（学生填写）</p> <ol style="list-style-type: none"><li>1. 完成了系统的界面设计，使用 Flask 框架完成了系统落地。</li><li>2. 基于完成的系统，完成了论文的第五章系统落地的撰写。</li></ol> <p>指导教师检查意见</p> <p>任务完成情况良好。第五章内容可以再扩充一下，最好有不同结果界面的对比。</p> <p>指导教师(签字) <u>张晓彤</u> 2022 年 5 月 13 日</p>
第 17 次	<p>本周工作进展情况（学生填写）</p> <ol style="list-style-type: none"><li>1. 在指导教师的修改建议下对论文进行了反复修改。</li><li>2. 完成了论文的中英文摘要、关键字部分，对部分内容进行了扩充，对表述不清的地方进行了改善。</li></ol> <p>指导教师检查意见</p> <p>论文内容比较充实，结构完整，逻辑清晰，图表排版等格式问题还需完善。</p> <p>指导教师(签字) <u>张晓彤</u> 2022 年 5 月 20 日</p>
第 18 次	<p>本周工作进展情况（学生填写）</p> <ol style="list-style-type: none"><li>1. 完成了初稿的论文查重工作，查重率符合学校要求，指导老师同意参加答辩。</li><li>2. 制作了答辩中所需要的 PPT。</li></ol> <p>指导教师检查意见</p> <p>继续修改论文中的格式问题，完善 PPT，准备答辩。</p> <p>指导教师(签字) <u>张晓彤</u> 2022 年 5 月 27 日</p>

## 毕业设计（论文）指导教师评价表（此表可打印）

指导教师评语(结合研究内容进行评价，包括：研究内容、工作态度、综合能力、工作量、质量与水平、创新点、规范化程度、是否同意答辩等；不少于100字)

宋安洋同学在毕业设计期间，学习刻苦努力，按期完成规定的任务；英语基础很好，外文资料的翻译，语言流畅，符合规定的要求。

作者以python为工具，在阅读了大量的中外文文献的基础上，主要完成如下工作：

- 1、介绍了神经网络相关概念和相关技术框架。
- 2、分析了数据集的特性及分布特点。
- 3、研究了过去研究中的方法与设计思路。
- 4、提出了新的方法架构以及对继续预训练BERT的是否可以提高表现的影响因素。
- 5、建立了基于方面的情感分析系统的基础构架，最后具体作了实现。

论文有一定的深度，结构合理，层次分明，叙述到位，内容详实，图表规范，数据正确，叙述简捷。

论文工作表明该同学具有较扎实的理论基础，并能够利用所学的理论知识解决实际问题，有较强创新能力。

综上，同意该生参见本科论文答辩。

指导教师（签字）

张晓彤

2021年5月30日

## 毕业设计（论文）评阅人评价表（此表可打印）

评阅教师评语(结合研究内容进行评价，包括选题价值与意义、工作量、完成质量、创新点、规范化程度，问题与建议、是否同意答辩等；不少于100字)

宋安洋同学的“基于方面的情感分析系统”一文，是情感分析中一项具有挑战的细粒度分析任务。挖掘用户情感对电商改善产品与服务有着重要作用。论文选题有一定理论研究意义与实用价值，研究成果具有一定创新性。

该生的毕业设计中实验充足。在尝试优化过去方法的过程中，提出了三个新的方法。在尝试通过领域内语料继续预训练BERT以提高方法表现的过程中，对用什么样的数据集继续预训练能提高表现，做了进一步的研究，具有一定的创新性。最后又通过flask框架完成了对基于方面的情感分析系统的开发。

该同学的论文的结构完整，层次清楚，工作量充足，达到了本科毕业设计的要求。同意该生参加本科毕业论文答辩。

评阅教师（签字）

刘晗

2021年5月31日

## 毕业设计（论文）答辩情况记录

自述时间：15 分钟	回答问题时间：5 分钟
------------	-------------

三个主要问题与回答记录：

**1. 可以用一句话再简短地概括一下什么是方面吗？**


基于方面的情感分析中的方面可以指显式出现在句子中的一个实体——Aspect Term 或者隐式表达的一个描述事物的角度——Aspect Category。Aspect Term 就是方面词。当然准确来说是方面词或方面词组，因为它可以是长度大于 1 的一个词组。

**2. 你所提出的三个方法中，表现最好的是哪一个？**

结合三个数据集上的整体表现，可以认为 TDLC-Transformer 表现得最好。和过去的方法相比的话，一般有时会分门别类地去比较，TDLC-Transformer 在 LSTM 和注意力模型两个类别中，在 Twitter 数据集上表现得明显要比其他方法好，可惜的是在另外两个数据集上就比较一般了。另外，当然，和 BERT 这种大型语料库上预训练出来的模型肯定没法比就是了。

**3. 你的系统除了英文外还能支持其他语言吗？**

只支持英文。SemEval2014 提出的数据集只有英文。过去的研究中，为了便于互相比较，研究者们选用的基本都是 SemEval2014 的数据集，因此本文也不例外。由于只对英文做了词嵌入，训练只用的英文数据集，因此只支持英文。

记录人（签字） 

2022 年 6 月 10 日

## 毕业设计（论文）答辩委员会评价表（此表可打印）

答辩委员会意见(结合研究内容评价,包括:研究内容、质量与水平、逻辑思维能力、表述能力、回答问题正确性、是否通过答辩等)

宋安洋的学士学位论文《基于方面的情感分析系统》选题有一定的现实意义。作者系统地综述了大量的国内外文献,基本掌握了该领域的国内外研究现状,在此基础上取得了如下创造性研究成果:

提出了三个基于 LSTM 和 Transformer 的用于解决方面级情感分类的神经网络方法架构。证明了 LSTM 和 Transformer 等注意力模型相结合可以很好地优势互补。提出了一种基于方面汇聚的 dropout 方式,有进一步研究价值。通过对方法对比和中性情感极性案例的分析,提出了 Twitter 数据集可能弱依赖于方面词,建议未来研究对 Twitter 数据集单独研究或者尝试提高方法的鲁棒性。通过继续预训练 BERT 的实验证明了领域语料继续预训练 BERT 确实有很大可能提高微调表现,然而却不一定能提高表现,更进一步地说,话题内,甚至任务内数据集继续预训练 BERT 也未必一定能够提高微调表现。实验表明跨话题继续预训练更容易提高表现。根据实验结果提出了继续预训练和微调数据集分布近似达到一定程度时,对微调表现是否提高的主要影响因素可能是继续预训练数据集的分布广度的推测。对 BERT 继续预训练进一步研究的同时取得了最先进的表现,并在最后通过 flask 框架实现了一个简易的系统落地。

论文结构完整,逻辑清晰,撰写规范,结论正确,研究成果有创新。在答辩过程中回答问题圆满,表明作者较好地掌握本门学科的基础理论、专门知识和基本技能,具有从事科学研究工作的初步能力,论文达到了《中华人民共和国学位条例》对学士学位论文的要求,通过答辩。

成绩 96 分

组长(签字)

成员(签字)

陆坤

张晓彤

李静

田

2022 年 6 月 10 日

# 大连理工大学工科专业毕业设计（论文）成绩表

学生姓名：宋安洋      学号： 201892436      专业： 软件工程      班级： 软数 1801      指导老师： 张晓彤

毕业设计（论文）题目：基于方面的情感分析系统

打分项分类	满分	评分标准	得分
① 调查分析 背景、意义、研究现状、文献翻译	20	<div>➤ 使用现代工具获取合适的文献资料；</div> <div>➤ 对设计任务及其背景的分析、调研；</div> <div>➤ 查阅国内外文献能力和国际化视野。</div>	19
② 方案论证 方案对比、技术路线分析	20	<div>➤ 利用数学、自然科学和工程的基本原理，识别、表达和分析复杂工程问题；</div> <div>➤ 合理的研究方案、技术路线、技术经济分析；</div> <div>➤ 研究方案的创新性。</div>	20
③ 研究 设计、计算、试验、优化、分析	30	<div>➤ 选择恰当的科学方法和现代工具进行研究；</div> <div>➤ 合理完整的设计、计算、试验、优化、分析；</div> <div>➤ 获得有效结论。</div>	30
④ 标准与可持续 考虑标准、法规和可持续	10	<div>➤ 考虑社会、健康、安全、文化、环境等因素；</div> <div>➤ 兼顾技术标准、知识产权、法规、工程伦理等。</div>	9
⑤ 设计和表达的规范性 论文和图纸规范性，答辩清晰	20	<div>➤ 论文撰写和图纸输出的规范性、完整性；</div> <div>➤ 答辩时陈述和回答的清晰性、正确性。</div>	18
合计	100		96

注：所有工科专业使用此成绩表，装订时本成绩表附在《毕业设计（论文）任务书》后。