表Ⅴ—103

**广州大学**

**本科毕业设计任务书**

|  |  |
| --- | --- |
| 设计题目: | [基于深度神经网络的中文阅读理解系统的设计与实现](javascript:f_viewtopic('4a409232674baeba016771a4081e145c');void(0);) |
| 学 院: | 计算机科学与网络工程学院 |
| 专 业: |  |
| 班 级: |  |
| 学 号: |  |
| 学生姓名: |  |
| 指导教师: |  |

下发时间: 年 月 日

**说 明**

1.认真学习《广州大学毕业论文（设计)工作管理规程》；

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任务；

3.遵守纪律，保证出勤，因事、因病离岗，应事先向指导教师请假，否则作为缺席处理。凡随机抽查三次不到，总分降低 10 分。累计缺席时间达到全过程 l／4者，取消答辩资格，成绩按不及格处理；

4.独立完成规定的工作任务，不弄虚作假，不抄袭和拷贝别人的工作内容。否则毕业设计成绩按不及格处理；

5.毕业设计必须符合《广州大学毕业论文（设计）规范化要求》，否则不能取得参加答辩的资格；

6.实验时，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。

7.妥善保存《广州大学毕业设计任务书》。

8.定期打扫卫生，保持良好的学习和工作环境。

9.毕业设计成果、资料按规定要求装订 好后交指导教师。凡涉及到国家机密、知识产权、技术专利、商业利益的成果，学生不得擅自带离学校。如需发表，必须在保守国家秘密的前提下，经指导教师推荐和院领导批准。

**一、题目来源及原始数据资料**

|  |
| --- |
| 题目来源：根据教学要求自拟的课题。  本题目来源于本人曾经接触到的斯坦福问答系统，该系统语言为英文，主要表现为提供一篇英文的文章和根据该文章所提出的问题，要求系统根据文章给出答案。英文的自然语言处理(NLP)发展是不错，中文的NLP还比较落后，主要在于中文语言本身的难度。所以要做一个中文的阅读理解，但难度相对降低，提供一篇文章，在其中的一个词挖空，要求系统根据文章语境进行完形填空。该系统主要用于中文的阅读理解，帮助我们更好地理解语境，那么挖空多个也有可能补全文章，也有可能帮助我们更好地写作，结合语境不易跑题。  原始数据资料：pyhton语言、所有数据均来源于[哈工大讯飞联合实验室（HFL）](http://www.hfl-tek.com/)收集的“儿童读物”阅读材料。 |

**二、毕业设计要求**

|  |
| --- |
| 1、熟悉数据-“儿童读物”阅读材料。  2、熟悉python框架tensorflow  3、熟悉词向量模型word2vec  4、熟悉自然语言处理模型LSTM  5、最终提交：整理资料，完成毕业论文。  6、a首先配置好环境，导入所需要的库  b进行数据的处理，处理训练集、验证集和测试集  c进行模型的训练和验证集的准确率判断  d 模型调参使模型准确率更好  e 进行测试集的预测  f 对于普通提供的童话进行预测 |

**三、进度安排、应完成的工作量**

|  |
| --- |
| 1、准备开发工具和环境，了解项目需求，熟悉python的使用。第1周  2、进行数据的处理，处理训练集、验证集和测试集。第2周  3、进行模型的训练和验证集的准确率判断。第3周  4、模型调参使模型准确率更好。第4周  5、进行测试集的预测。第5周  6、对于普通提供的童话进行预测。第6-7周  7、完善和优化设计。第8周  8、整理设计文档，完成论文。第9-10周 |

**四、主要参考文献**

|  |  |  |
| --- | --- | --- |
| [1] S. Hochreiter and J. Schmidhuber. Long short-term memory. Neural Computation, 9(8):1735–1780, 1997.  [2] A. Graves, M. Liwicki, S. Fernandez, R. Bertolami, H. Bunke, J. Schmidhuber. A Novel Connectionist System for Improved Unconstrained Handwriting Recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 31, no. 5, 2009.  [3] Bengio, Y.; Courville, A.; Vincent, P. Representation Learning: A Review and New Perspectives. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2013, **35** (8): 1798–1828.  [4] 顾健伟,曾诚,邹恩岑,陈扬,沈艺,陆悠,奚雪峰.基于双向注意力流和自注意力结合的机器阅读理解[J].南京大学学报(自然科学),2019.  [5] 李昭昱,艾芊,张宇帆,肖斐.基于attention机制的LSTM神经网络超短期负荷预测方法[J].供用电,2019.  [6] 黄贤英,刘广峰,刘小洋,阳安志.基于word2vec和双向LSTM的情感分类深度模型[J/OL].计算机应用研究:1-7[2019-02-24].  [7] 牛伯浩.循环神经网络实现文本智能预测[J].智能城市,2018.  [8] 孙严伟. 基于改进LSTM单元的循环神经网络中文分词研究[D].南华大学,2018.  [9] 高成亮. 基于LSTM的文本上下文依赖特征的表示方法研究[D].河北科技大学,2019.  [10] 蓝雯飞,徐蔚,汪敦志,潘鹏程.基于LSTM-Attention的中文新闻文本分类[J].中南民族大学学报(自然科学版),2018.  [11] 章登义,胡思,徐爱萍.一种基于双向LSTM的联合学习的中文分词方法[J/OL].计算机应用研究,2019(10):1-7[2019-02-24].  [12] 王乔乐.中文分词和词向量[J].中国新通信,2018.  [13] 冯俐.中文分词技术综述[J].现代计算机(专业版),2018.  [14] 李泽,古超,龙政.基于Python的文本分析方法研究[J].电脑编程技巧与维护,2018.  [15] 陶伟.警务应用中基于双向最大匹配法的中文分词算法实现[J].电子技术与软件工程,2016. | | |
| 指导教师（签名）： | 系（教研室）主任（签名）： | 主管院长（签名）：  （学院章） |