浙江大学电子信息类工程硕士专业必修课

**人工智能算法与系统 （课程号：2142001）**

**算法实践报告**

学号： 12021232 姓名：刘柯

1. **算法实践拟解决问题**（描述拟解决的问题，不超过300字）

当下的文学作品越来越多，其中不乏剽窃他人作品的行为。为了鉴定某些存在争议的文学作品的作者，判断文章是否剽窃他人作品，我们可以设计一种文本分类的算法，通过识别文本中的个人言语表达特征，实现根据文本识别其作者的功能。

1. **描述所实现算法内容**（提供如下内容：1、相关算法现状分析及所采用算法的动机；2、算法概貌及画出算法流程图；3、算法流程图中主要模块内容描述；4）算法实验，包括训练数据构成、算法优化训练方法、算法结果呈现和分析。该部分内容不少于800字、不多于1000字）

**背景分析：**对于文本分类问题，主要使用两类模型，一类是使用传统的机器学习模型，如朴素贝叶斯，最大熵，支持向量机（SVM）等等，第二类是是使用神经网络模型，包括CNN和RNN。传统的机器学习模型在分类前先进行特征工程，如把文本转化为词袋，并转化为TF-IDF矩阵，然后再做分类。而使用神经网络模型则可以让模型自己提取特征并进行文本分类，从而获得优于传统模型的效果。

**算法介绍**：在本次从实验中，我使用了神经网络模型，主要的算法是先将原始数据全部转换为词向量的方式，再用LSTM来进行特征提取，然后使用全连接层来进行分类，算法流程图如下：

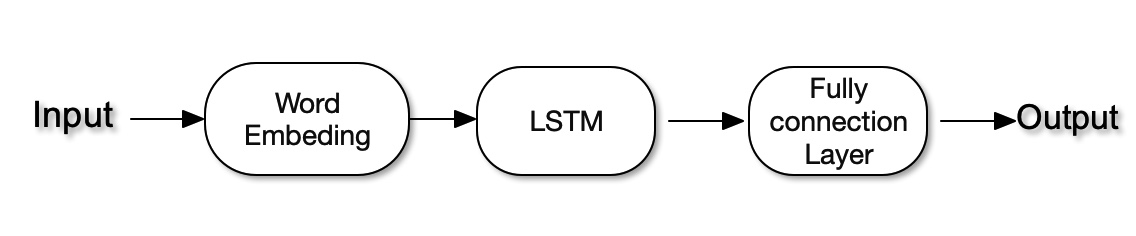


Figure 1：网络总体架构

Word Embedding 层：输入文本的one-hot编码，将这些编码映射到规定的维度上去，word embedding可以考虑到不同词之间的差异，词义词性越靠近，其向量位置越靠近，这一层可以用来降低输入数据的维度，提高计算速率和准确度。

LSTM层：长短期记忆网络，能够处理序列变化的数据，比如某个单词的意思会因为上文提到的内容不同而有不同含义。网络通过门控状态来控制传输状态，记住需要长时间记忆的，忘记不重要的信息。这里我使用了双向LSTM，使得网络能够从上下文中提取特征。

FC 层：从上一层提取完特征之后经过一个FC层，计算出属于5个class的概率。

**实验过程及结果：**

实验数据来自于老师提供数据集，其中包含了8438个经典中国文学作品片段，对应文件分别以作家姓名的首字母大写命名，其中鲁迅片段1500条，莫言片段2219条，钱钟书片段1419条，王小波片段1300条，张爱玲片段2000条。这些数据按照7：3的比例分为训练组和测试组。

几个需要设置的超参：词嵌入的embedding\_dim,LSTM的隐藏层维度，循环神经网络的层数，dropout，bidirectional。 经过测试，embedding\_dim设为500，LSTM隐层维度设为128，dropout设为0.5（单层时无dropout）时效果较好。此时层数和bidirectional的影响如下：

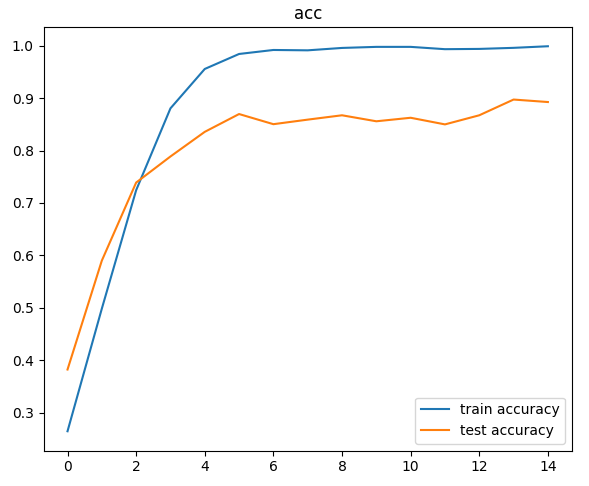
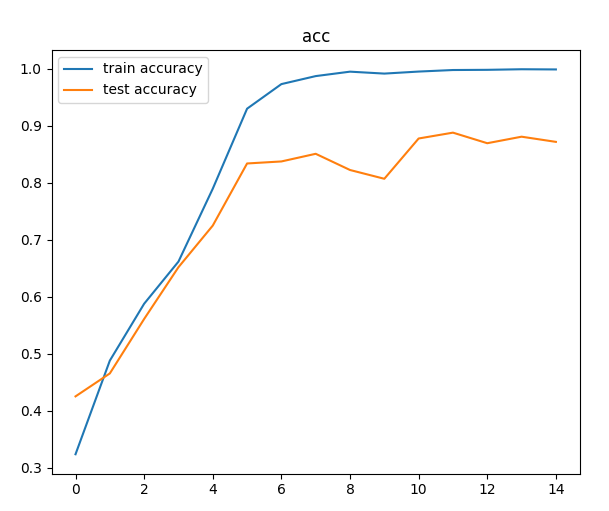
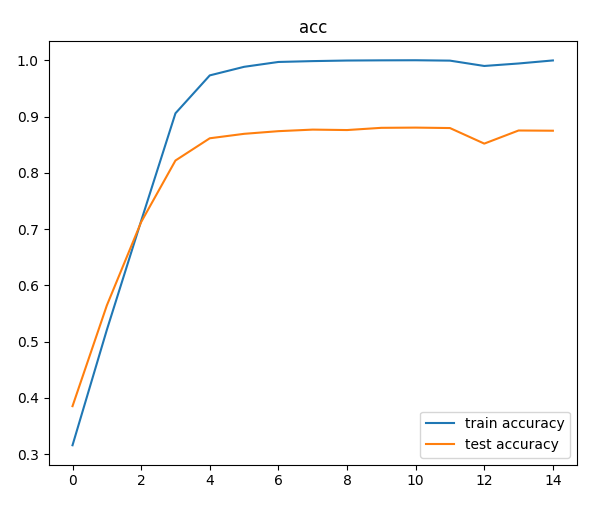
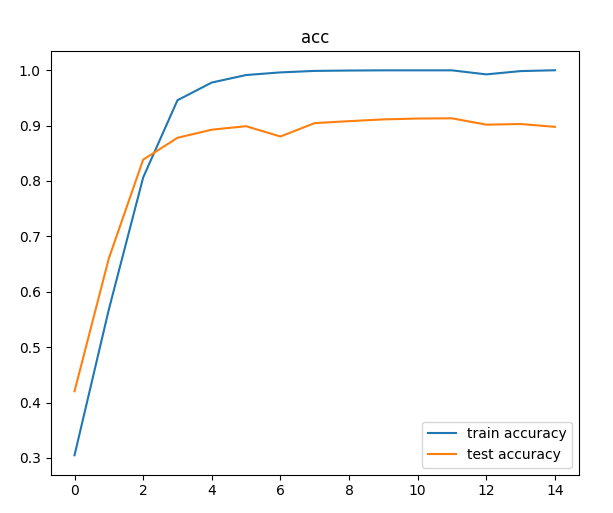


Figure 2：左上图为单向单层单向LSTM，epoch=4时收敛，正确度为90%；右上图为双向单层LSTM，epoch=6时收敛，正确度为88%；左下图为单向双层LSTM，epoch=7时收敛，正确率为88%；右下图为双向双层LSTM，epoch=7时收敛，正确率为88%。

可以看到，对于当前的数据集，LSTM的层数和单双向对于结果的影响不算特别大，但随着层数的增多和单向变为双向，收敛的时间变长了。这一点的原因在于同样隐层维度设置，层数多的网络参数更多，双向的网络比单向更多。

理论上双向LSTM应该比单向LSTM效果更好，于是尝试对双向LSTM的优化器加上了weight\_decay的限制，在经过调参后，发现当weight\_decay=0.001时，虽然收敛的时间变长了，但达到了94%的效果。经过测试，Weight\_decay=0.001的限制条件无法在单层LSTM上取得成功，出现了梯度消失的现象。

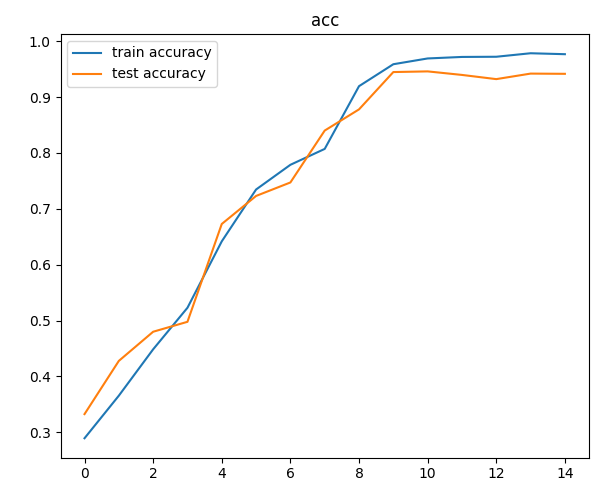


Figure 3：加入weight\_decay参数后，在epoch=10时收敛，双向单层LSTM正确率为94%

1. **算法所涉及内容的未来趋势和挑战分析**（不超过300字）

我所用的算法适用于文本分类问题，在未来文本分类问题还将面临如下几个挑战。第一是数据标注瓶颈，虽然在本次实验中标签的获取十分简单，可以直接获取作者信息，但如情感分类等问题是不易获得高质量的标签数据的。第二是深度学习的可解释性较差，我们在预训练结束后，在参数迁移，特征迁移，针对下游任务进行微调时，我们很难确定该保留哪些特征。第三是跨语种或多语种的文本分类，适用于一种语言的模型不一定适合于其他语言，特别是语法差距较大的小语种。在未来，我们可以继续对传统方法进行优化，将更多更新的模型和理论引入文本分类问题，如图卷积神经网路，知识库，知识图谱等等。

提交形式：doc文档

提交日期：2020年11月23日星期一