



文本分析与挖掘

上机实验

**计算机科学与技术学院**

**实验八、基于TextCNN的中文情感分析**

1. **实验目的**
2. 理解CNN网络在文本任务中的用法
3. 掌握tensorflow搭建文本分类模型的方法
4. 熟悉中文情感分析过程。

**二、实验内容**

**1.预处理**

1. 加载数据（训练集train.tsv，测试集test.tsv，验证集dev.tsv）并,适当预处理、分词后构建词汇表以及对应索引，留出索引0:padding，1：unknown
2. 把每个句子/评论转成长度统一为max\_len=100的向量，太长的截断，太短的后面补零。
3. 打印两个样本查看结果。

**2.设置模型**

a.设置embedding层输出维度embed\_dim=200；

b.卷积层设置三种卷积核大小kernel\_size=[2,3,4]，每种大小用64个卷积核，即filters=64；

c.倒数第二的全连接dense层输出64维，Relu激活；

d.最后一个输出二分类结果，1个节点，sigmoid激活。

f.打印模型摘要

**3.训练模型**

a.设置损失函数、epoch、优化器、学习率等超参数并训练模型。

b.打印并观察模型的学习曲线，讨论训练情况。

b.基于以上观察适当调整实验设置（超参数、网络结构），对比结果。

**4.测试模型**

a.自己写一条评论文本数据进行测试，输出情感类别及其概率;

b.给定测试集上的准确率。