

# 浙江工业大学



## 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

## 一、实验目的

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

## 二、实验内容

### 1. 预处理

- a. 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
- b. 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。
- c. 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。