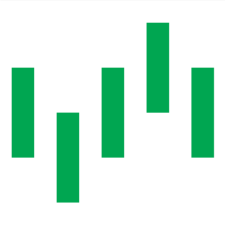
**北京理工大学研究生课程**

**学习报告**

****

厚德 明理 慎独 求是

|  |  |
| --- | --- |
| 论文名称： | 大数据分析典型案例解析-互联网新闻情感分析预测 |
| 学生姓名： | 丘绎楦 |
| 学生学号： | 1820221050 |
| 学生学科： | 计算机科学与技术 |
| 学生学院： | 信息与电子学院 |
| 课程名称： | 0500091-大数据分析 |
| 任课教师： | 罗森林 |
| 开课单位： | 信息系统及安全对抗实验中心（工信部） |

目 录

[1 “大数据分析”课程主要内容 3](#_Toc473208701)

[2 案例名称及简介 3](#_Toc473208702)

[3 案例的架构及其计算模式详解 3](#_Toc473208703)

[3.1 功能性能 3](#_Toc473208704)

[3.2 核心架构 3](#_Toc473208705)

[3.3 计算模式 3](#_Toc473208706)

[3.4 应用方案及效果 3](#_Toc473208707)

[4 案例的优势和不足分析 3](#_Toc473208708)

[4.1 案例的优势分析 3](#_Toc473208709)

[4.2 案例的不足分析 3](#_Toc473208710)

[5 参考文献 3](#_Toc473208711)

[6 附录 3](#_Toc473208712)

# “大数据分析”课程主要内容

* 大数据分析核心架构
* 大数据分析计算模式
* 虚拟机的使用
* linux基础
* 大数据分析的计算方法
  + 深度学习
  + 监督学习
  + 无监督学习
  + 集成学习
  + 机器学习
  + 强化学习
* 大数据分析计算平台
  + spark平台
  + Hadoop平台
  + flink平台
* 实验实践
  + 互联网感情分析预测
  + 西尼罗河病毒预测
  + 乘用车细分市场销量预测

# 案例名称及简介

互联网新闻情感分析预测

该案例通过深度学习技术，结合自然语言处理（NLP）和情感分析算法，对互联网新闻进行情感倾向的预测。利用已标注的训练数据集，模型学习了新闻标题和内容的情感特征，然后通过训练好的模型对未标注的新闻进行情感分析，判断其情感倾向是正面、负面还是中性。

# 案例的架构及其计算模式详解

## 功能性能

该案例的主要功能是对互联网新闻进行情感分析，提供对新闻情感倾向的预测。功能性能主要体现在模型的准确性和泛化能力，以及对大规模数据的高效处理和分析。

## 核心架构

案例的核心架构采用了深度学习技术，使用了Keras库来构建神经网络模型。具体架构包括了Embedding层、卷积层（Conv1D）、全局最大池化层（GlobalMaxPooling1D）、Dense层等。模型利用预训练的Word2Vec模型对文本进行词向量表示，然后通过卷积神经网络（CNN）对文本特征进行提取，最后通过Dense层输出情感分类结果。

## 计算模式

该案例采用了监督学习的计算模式。通过已标注的训练数据集，模型进行有监督的学习，优化损失函数，提高对新闻情感的预测准确性。训练过程中使用了Adam优化器和交叉熵损失函数。

## 应用方案及效果

主要应用于互联网新闻情感分析领域，可用于舆情监测、品牌声誉管理等。通过对新闻情感的预测，企业或机构可以及时了解社会舆论，做出相应决策。

通过训练和验证集的性能评估，案例展现了对互联网新闻情感的较好预测能力。在实际应用中，可以根据F1分数等指标来评估模型的效果。

# 案例的优势和不足分析

## 案例的优势分析

使用了预训练的Word2Vec模型：通过预训练的词向量模型，提高了模型对文本语义的理解能力。

采用了卷积神经网络：CNN在文本分类任务中表现出色，能够有效提取局部特征，帮助模型更好地理解文本。

使用了Embedding层和全局最大池化层：这些层有助于提高模型的训练效率和泛化能力。

## 案例的不足分析

数据清洗限制： 模型在进行情感分析前进行了文本清洗，但清洗规则可能并不完善，会有一些信息丢失。

模型训练数据量： 模型的性能受到训练数据量的限制，如果数据量不够大，可能会影响模型的泛化能力。

超参数调优： 对于神经网络模型，超参数的选择对性能有很大影响。是否进行了充分的超参数调优没有在代码中明确展示。

# 参考文献

# 附录