

汽车行业用户观点主题及情感识别

——数据科学导论期末报告

何秦兴 PB16110299

吴雨菲 PB15020525

张劲曦 PB16111485

December 5, 2018

Contents

1	题目描述：	3
1.1	题目背景	3
1.2	题目任务	3
1.3	数据说明	3
1.4	评测标准	3
2	题目分析与方案设计：	3
2.1	对数据的基本特征分析和设计思路	3
2.2	特征提取与模型选择	3
3	方法说明：	3
3.1	分词	3
3.2	特征提取：	3
3.2.1	隐含狄利克雷主题模型 (LDA)	3
3.2.2	词袋模型 (BOW)	3
3.2.3	TFIDF	3
3.2.4	输入 LSTM 的词向量矩阵	3
3.3	学习方法：	3
3.3.1	逻辑回归模型 (LR)	3
3.3.2	朴素贝叶斯模型 (NB)	3
3.3.3	多层感知机模型 (MLP)	3
3.3.4	随机森林模型 (RF)	3
3.3.5	集成模型 (Ensemble)	3
3.4	提升和改进的思路：	3
4	试验结果：	3
4.1	线上比赛结果	3
4.2	各种方法的结果与分析	3
4.3	没有解决的问题和主要困难	3
5	附录	3
5.1	个人收获	3
5.2	分工说明	4

1 题目描述：

1.1 题目背景

1.2 题目任务

1.3 数据说明

1.4 评测标准

2 题目分析与方案设计：

2.1 对数据的基本特征分析和设计思路

2.2 特征提取与模型选择

3 方法说明：

3.1 分词

3.2 特征提取：

3.2.1 隐含狄利克雷主题模型 (LDA)

3.2.2 词袋模型 (BOW)

3.2.3 TFIDF

3.2.4 输入 LSTM 的词向量矩阵

3.3 学习方法：

3.3.1 逻辑回归模型 (LR)

3.3.2 朴素贝叶斯模型 (NB)

3.3.3 多层感知机模型 (MLP)

3.3.4 随机森林模型 (RF)

3.3.5 集成模型 (Ensemble)

3.4 提升和改进的思路：

4 试验结果：

4.1 线上比赛结果

4.2 各种方法的结果与分析

4.3 没有解决的问题和主要困难

5 附录

5.1 个人收获

3

吴雨菲：

何秦兴：

张劲墩：

5.2 分工说明

吴雨菲：

何秦兴：

张劲墩：