

基于数据挖掘的情感词分析研究

Data Mining Based on the emotional word analysis

// CONTENT //

01

毕设概述

INTRODUCTION

03

制作过程

MAKING PROCESS

05

总结回顾

REVIEW

02

选题背景

SUMMARY

04

毕设展示

PRODUCT SHOW

06

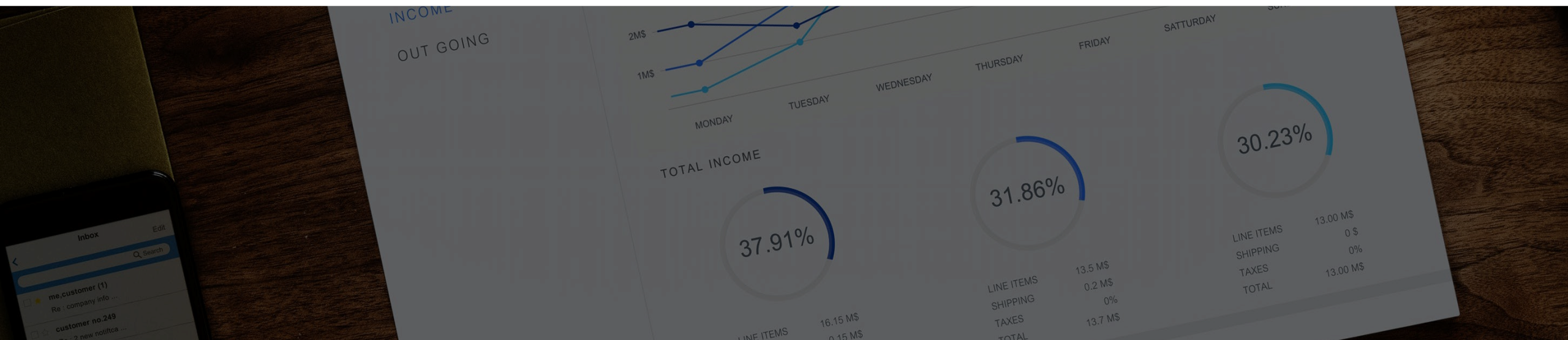
参考文献

REFERENCE

01

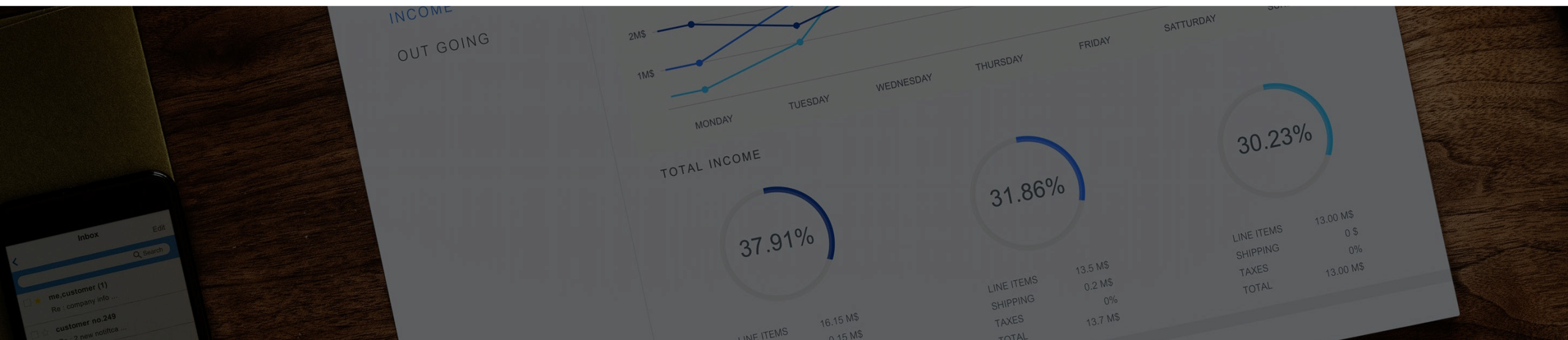
毕设概述

INTRODUCTION



02 选题背景

SUMMERY

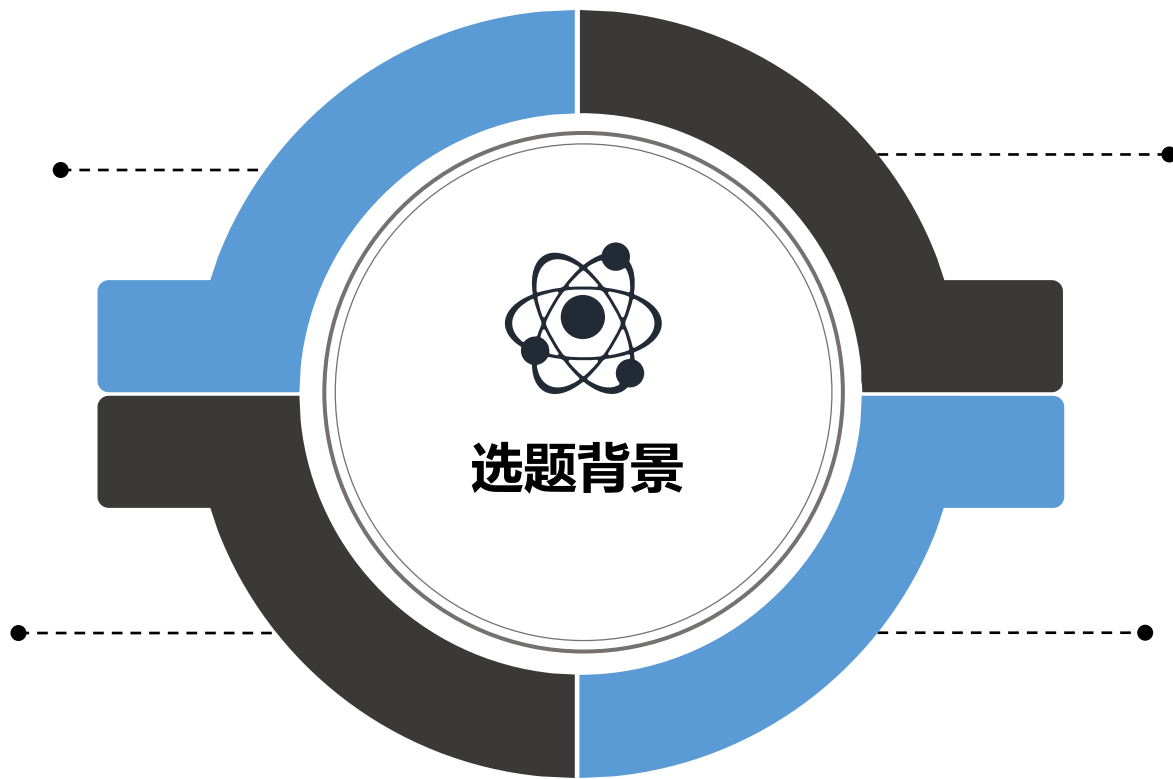


数据挖掘

数据挖掘是数据库知识发现中的一个步骤。数据挖掘通常与计算机科学有关，并通过统计、在线分析处理、情报检索、机器学习、专家系统和模式识别等诸多方法来实现上述目标。

朴素贝叶斯 算法

朴素贝叶斯法是基于贝叶斯定理与特征条件独立假设的分类方法。最为广泛的两种分类模型是决策树模型和朴素贝叶斯模型。



情感分析

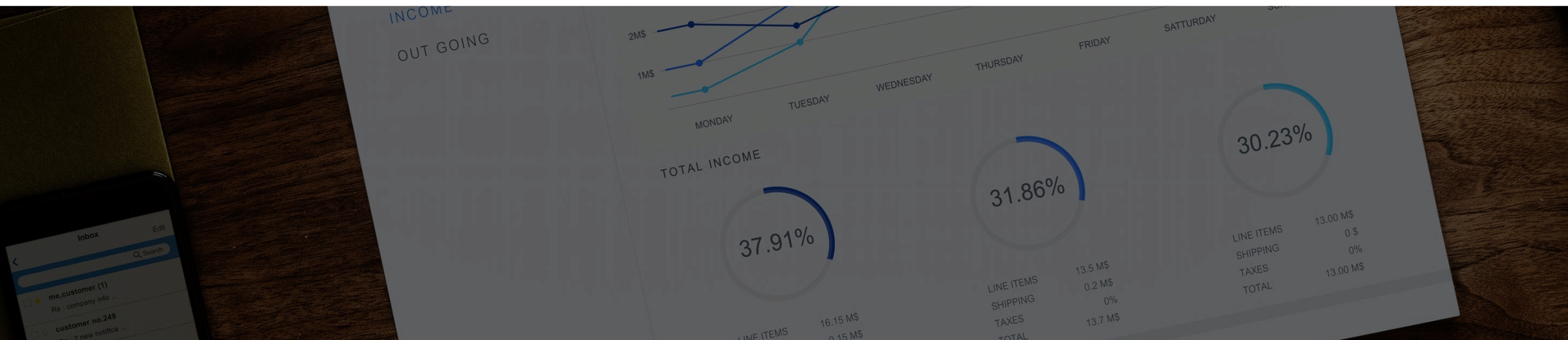
情感分析是对带有情感色彩的主观性文本进行分析、处理、归纳和推理的过程。潜在的用户就可以通过浏览这些主观色彩的评论来了解大众舆论对于某一事件或产品的看法。

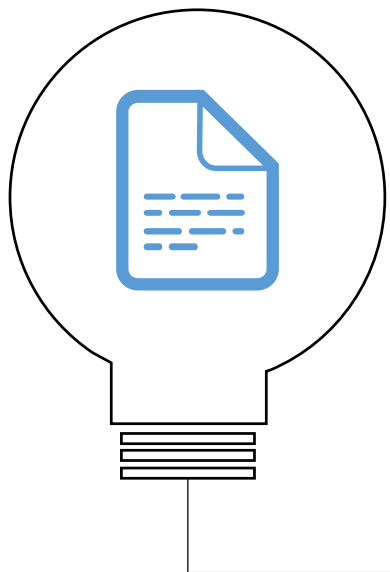
机器学习

机器学习是一门多领域交叉学科，涉及概率论、统计学、逼近论、凸分析、算法复杂度理论等多门学科。是人工智能的核心，是使计算机具有智能的根本途径，主要使用归纳、综合而不是演绎。

03 制作过程

MAKING PROCESS





01.02

开题报告

初步确定了论文的论题及方向
确认了数据分析类论文的研究方案和主要内容
拟定了初拟论文提纲

毕设初稿

对于数据的清洗及训练没有体现
分词部分不够详细
对于数据可视化及动态网站没有具体设计

03.08

03.15

中期检查修改

论文论题与实际毕设研究方案有差异，根据导师的建议修改了论题
论文整体的逻辑不严谨，整体修改并修订了目录，并调整了各章节的比重与内容

毕设答辩

进行最后的调整，包括错别字，
语序用词以及逻辑微调

04.20

04

毕设展示

PRODUCT SHOW



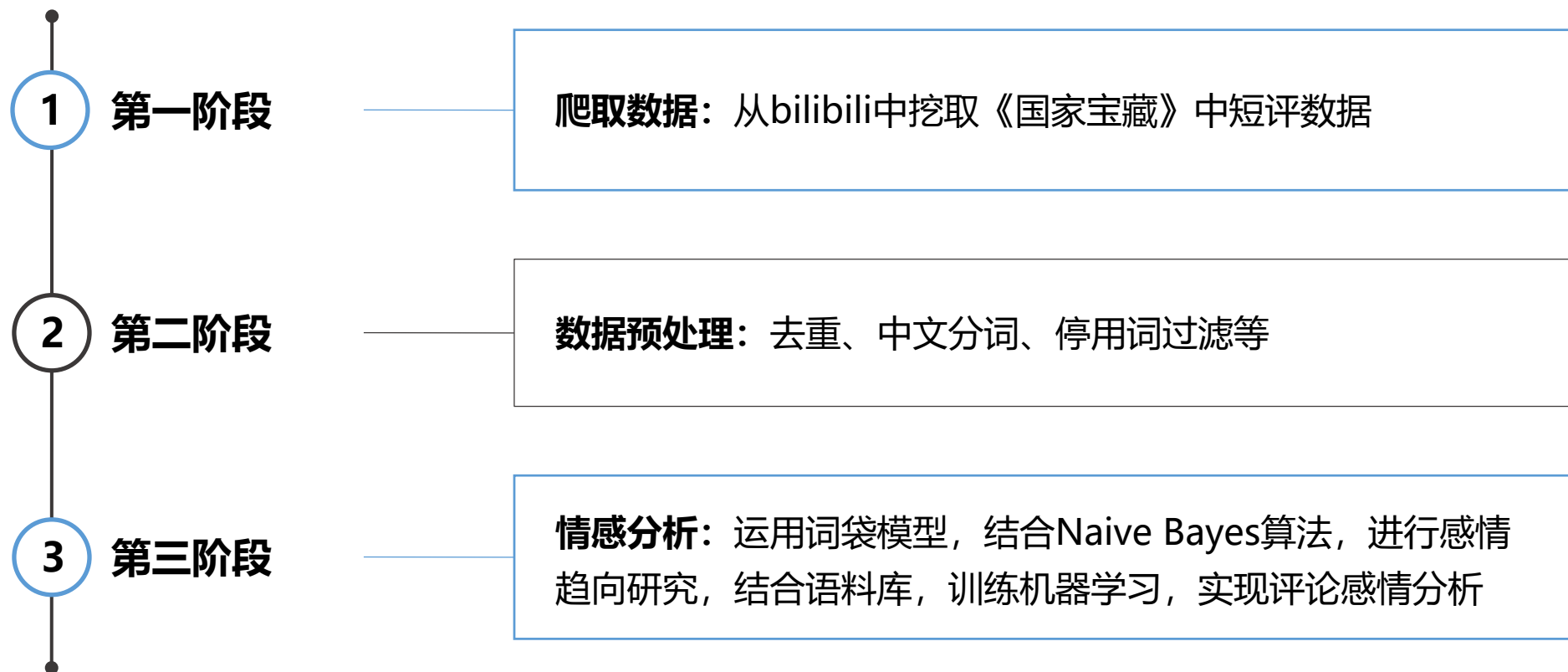
毕设介绍

本文阐述了朴素贝叶斯算法在感情分析中的作用。

通过《国家宝藏》案例分析，提取了本论文发表前视频网站bilibili中该综艺的所有短评及其评分，结合数据进行可视化情感分析。并建立语料库情感分析模型，将朴素贝叶斯的原理应用到机器学习短评分析。进行了感情词分析研究。







1

爬取数据



承古人之创造
开时代之生面

中央广播电视总台
CCTV-9 08:00-20:00

国家宝藏 第二季

历史
人文
社会

总播放 2330.7万	追剧人数 52.3万	弹幕总数 60.9万
----------------	---------------	---------------

9.8
★★★★★
我要点评

2018年12月9日 开播 已完结,全10话

简介: 作为中央广播电视总台创新打造的重要精品项目之一,《国家宝藏》第二季继续由央视和故宫两大文化体强强联手,全新加入河北博物院、山西博物院、山东博物馆、广东省博物馆、四川博物院、云南省博物馆、甘肃省博物馆、新疆维吾尔自治区博物馆等,从第一季的八大博物馆(院)手中接过了讲述中国故事、让国宝活起来的接力棒。观众还将看到熟悉的舞美、熟悉的环节、熟悉的001号讲解员国立叔,听到熟悉的来自那英的《一眼千年》,以及会让很多粉丝单曲回放一整晚的背景音乐。与之同时,“国宝盒子”会以更为震撼的体量 and 视效呈现在观众眼前;节目会...

❤️ 追剧

📱
👤
★
💬
📌

作品详情 长评 (50) **短评 (2819)** 相关视频

短评

默认 ▾

去写短评



西山的馄饨



2018-12-09

快过年了,给春晚留点面子吧

👍 1129



点评动态

🔄 换一换

内容尚可,立意不佳

怎么说呢,这个系列虽然挺有意思,短片和人物介绍也都娓娓道来。但是整体立意并不...

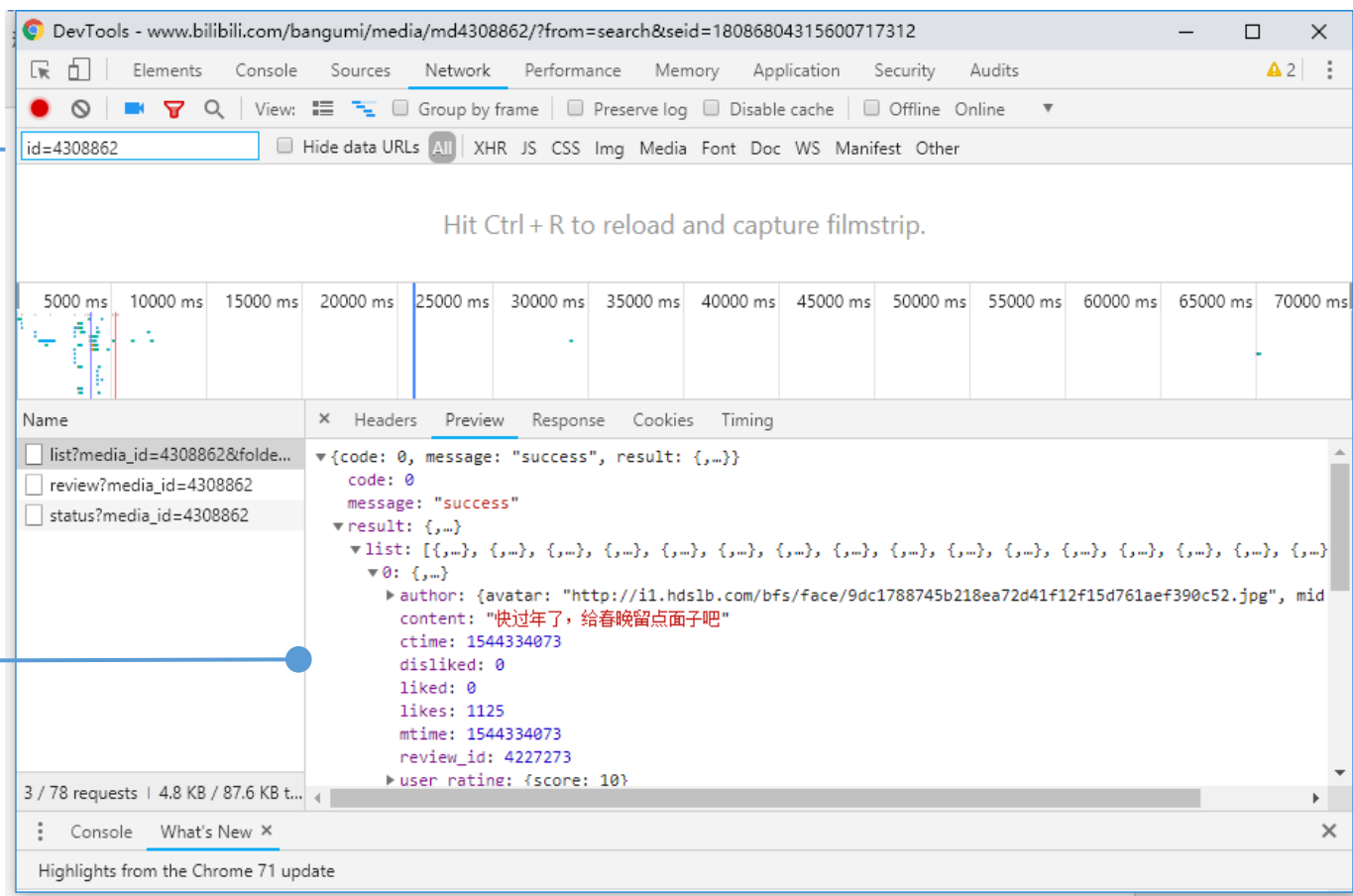


德小白 评 科幻小说预言家 第一季

赞: 3

获取API

获取cursor



DevTools - www.bilibili.com/bangumi/media/md4308862/?from=search&seid=18086804315600717312

Elements Console Sources Network Performance Memory Application Security Audits

id=4308862 Hide data URLs All XHR JS CSS Img Media Font Doc WS Manifest Other

Hit Ctrl + R to reload and capture filmstrip.

5000 ms 10000 ms 15000 ms 20000 ms 25000 ms 30000 ms 35000 ms 40000 ms 45000 ms 50000 ms 55000 ms 60000 ms 65000 ms 70000 ms

Name	Headers	Preview	Response	Cookies	Timing
<input type="checkbox"/> list?media_id=4308862&folde...		<pre>{code: 0, message: "success", result: {, ...}}</pre>			
<input type="checkbox"/> review?media_id=4308862		<pre>code: 0</pre>			
<input type="checkbox"/> status?media_id=4308862		<pre>message: "success"</pre>			

```
▼ result: {, ...}
  ▼ list: [{, ...}, {, ...}, {, ...}, {, ...}, {, ...}, {, ...}, {, ...}, {, ...}, {, ...}, {, ...}, {, ...}, {, ...}, {, ...}, {, ...}, {, ...}, {, ...}, {, ...}
    ▼ 0: {, ...}
      ▶ author: {avatar: "http://i1.hdslb.com/bfs/face/9dc1788745b218ea72d41f12f15d761aef390c52.jpg", mid
        content: "快过年了, 给春晚留点面子吧"
        ctime: 1544334073
        disliked: 0
        liked: 0
        likes: 1125
        mtime: 1544334073
        review_id: 4227273
        user rating: {score: 10}
```

3 / 78 requests | 4.8 KB / 87.6 KB t...

Console What's New x

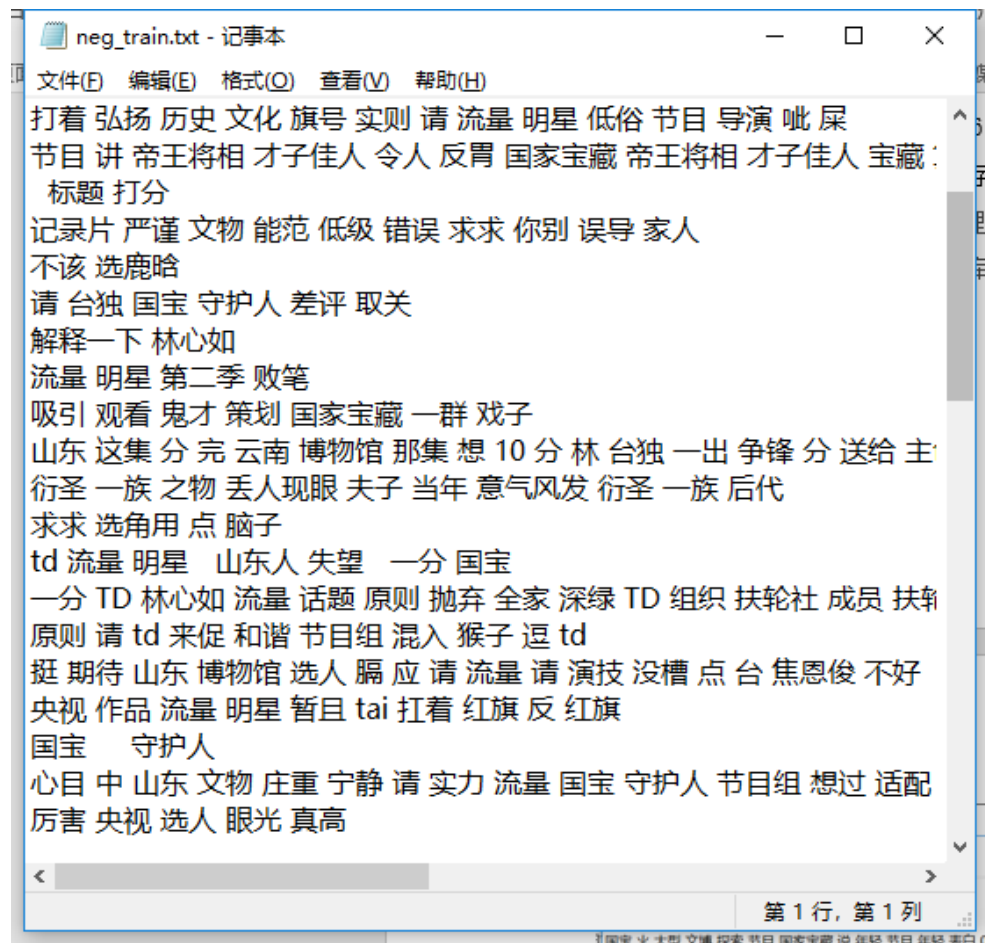
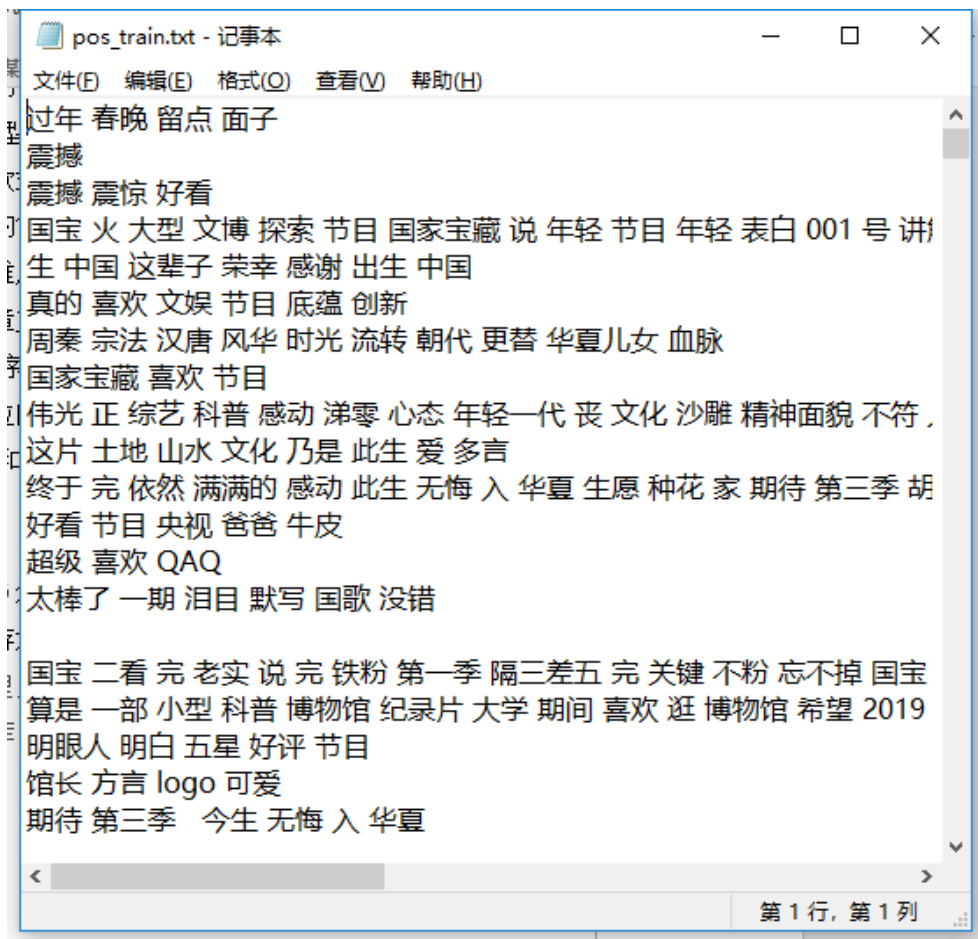
Highlights from the Chrome 71 update

2

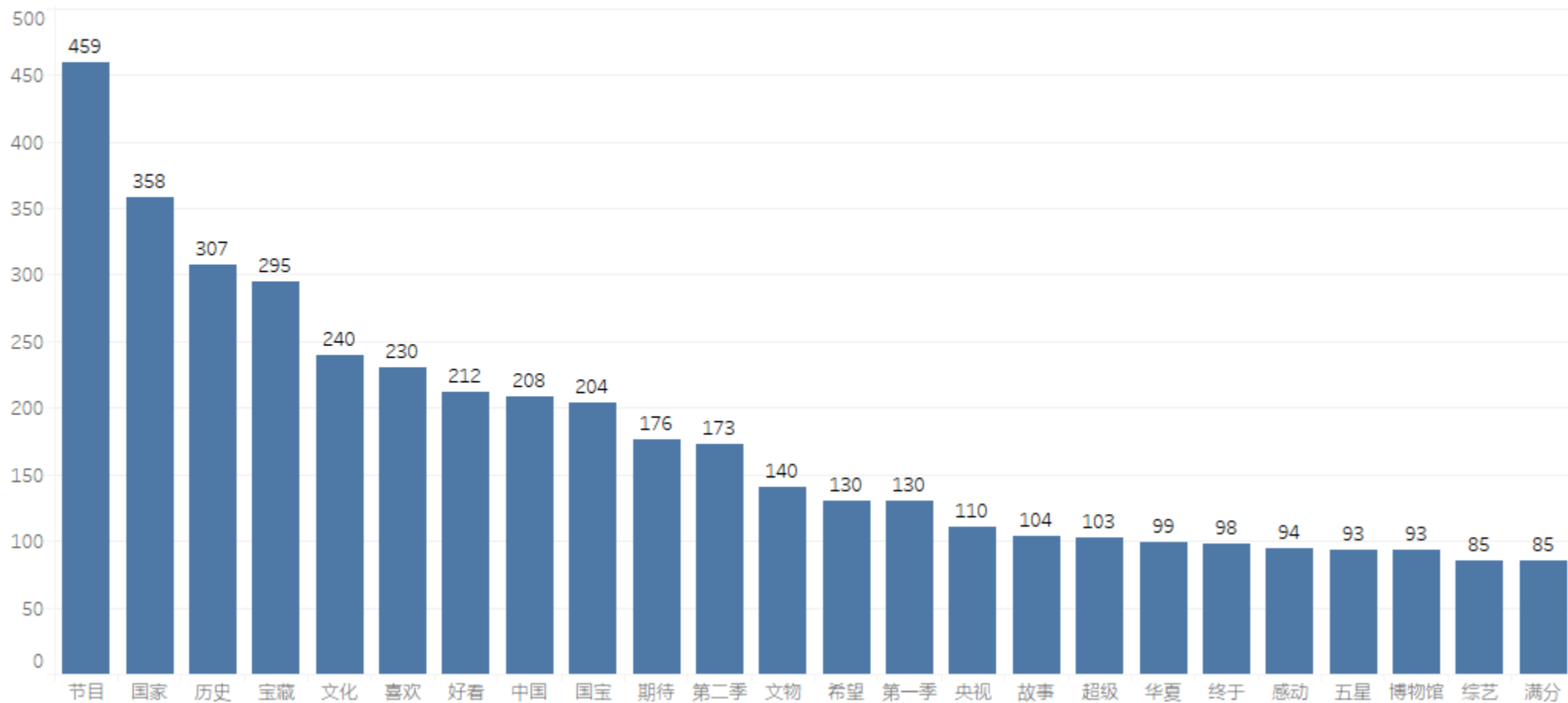
数据预处理

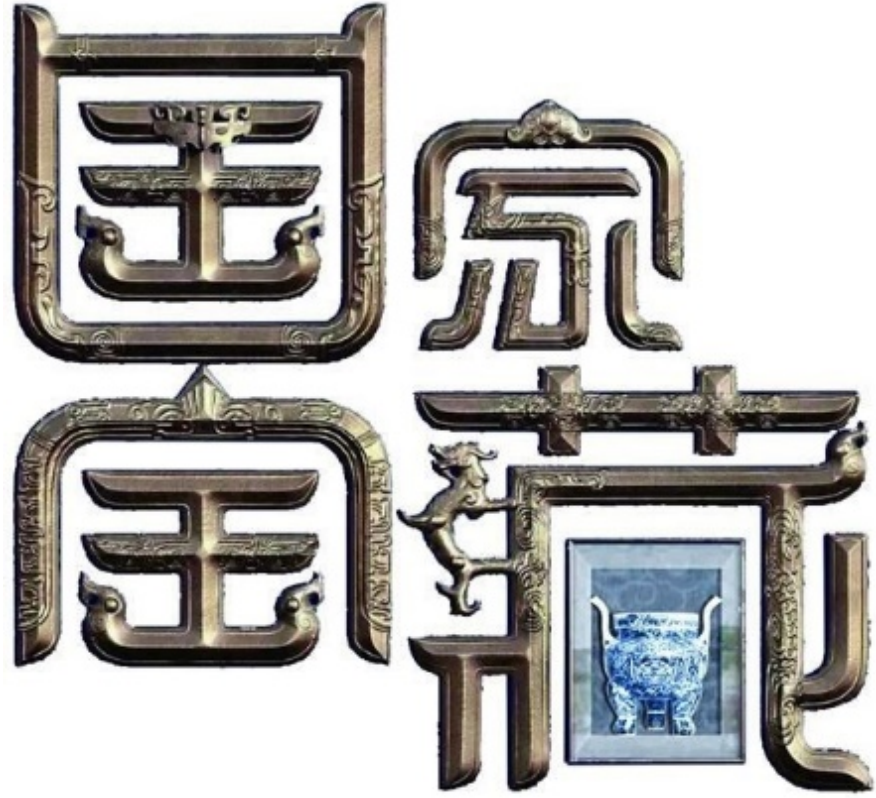
content

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
A1	author	score	disliked	likes	liked	ctime	score	content	last_ep_in	cursor	date					
2	西山的馄饨	10	0	1117	0	1544334073	10	快过年了，给春晚留点面	0	7.70818E+13	2018/12/9 5:41					
3	墨夷潇	10	0	0	0	1551507224	10	震撼	0	7.70818E+13	2019/3/2 6:13					
4	宁湫湫	10	0	0	0	1551493405	10	震撼，震惊，好看	0	7.70818E+13	2019/3/2 2:23					
5	就算再次回	10	0	0	0	1544583469	10	让国宝火起来，这里是	0	7.70818E+13	2018/12/12 2:57					
6	闲暇摸鱼的	10	0	0	0	1551454196	10	生为中国人，是我这辈子	0	7.70818E+13	2019/3/1 15:29					
7	布吉也是星	10	0	0	0	1551451547	10	真的很喜欢这种文艺节目	0	7.70818E+13	2019/3/1 14:45					
8	豆包仓鼠	10	0	0	0	1551323017	10	周秦宗法，汉唐风华，	0	7.70818E+13	2019/2/28 3:03					
9	此人非到炸	10	0	0	0	1551277843	10	国家宝藏是最喜欢的节	0	7.70818E+13	2019/2/27 14:30					
10	岔娇吐你一	10	0	0	0	1551172197	10	被一个伟光正的综艺科普	0	7.70818E+13	2019/2/26 9:09					
11	迷之悲鸣	10	0	0	0	1551170742	10	这片土地上的山水、文化	0	7.70818E+13	2019/2/26 8:45					
12	彦小兰陵	10	0	0	0	1551065529	10	今天终于都看完了！依然	0	7.70818E+13	2019/2/25 3:32					
13	韦高武	10	0	0	0	1551003515	10	很好看的节目，央视爸爸	0	7.70818E+13	2019/2/24 10:18					
14	binちゃん	10	0	0	0	1550985207	10	超级喜欢QAQ！！！！	0	7.70818E+13	2019/2/24 5:13					
15	可爱妍又会	10	0	0	0	1547087935	10	太棒了，每一期都泪目。	0	7.70818E+13	2019/1/10 2:38					
16	_于莫邪	10	0	0	0	1550933535	10	好	0	7.70818E+13	2019/2/23 14:52					
17	永不变更的	10	0	0	0	1550762102	10	国宝二看完了，老实说	0	7.70818E+13	2019/2/21 15:15					
18	空想家Frey	8	0	0	0	1550738864	8	算是一部小型科普的博物	0	7.70818E+13	2019/2/21 8:47					
19	Endless、真	10	0	0	0	1550725075	10	明眼人都明白，五星好评	0	7.70818E+13	2019/2/21 4:57					
20	皮卡丘5元	10	0	0	0	1550684898	10	各个馆长的方言logo好可	0	7.70818E+13	2019/2/20 17:48					
21	Mr嘉诚	10	0	0	0	1550657967	10	期待第三季 今生无悔入	0	7.70818E+13	2019/2/20 10:19					
22	小竹88	10	0	0	0	1550640657	10	不忘初心	0	7.70689E+13	2019/2/20 5:30					
23	Randoll	10	0	0	0	1550605831	10	啥也不说了反手一个满分	0	7.70689E+13	2019/2/19 19:50					
24	RAMEN-	10	0	0	0	1550589723	10	延续了第一季的精彩与诚	0	7.70689E+13	2019/2/19 15:22					



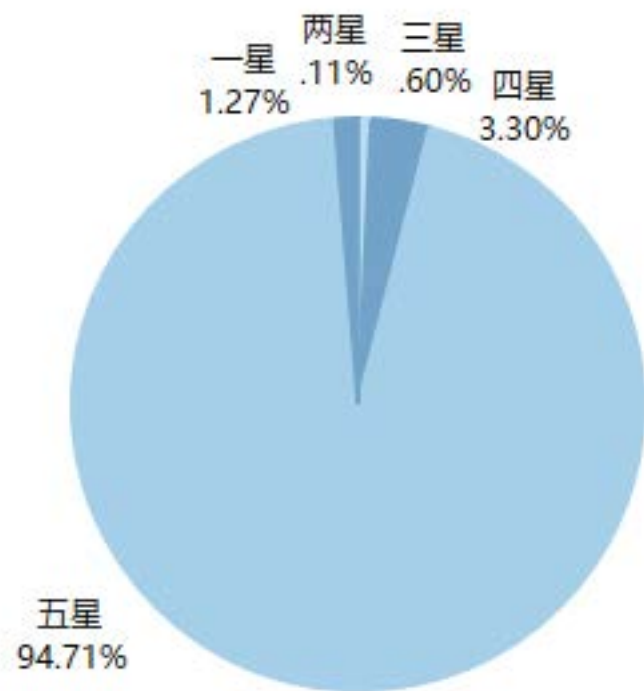
词频



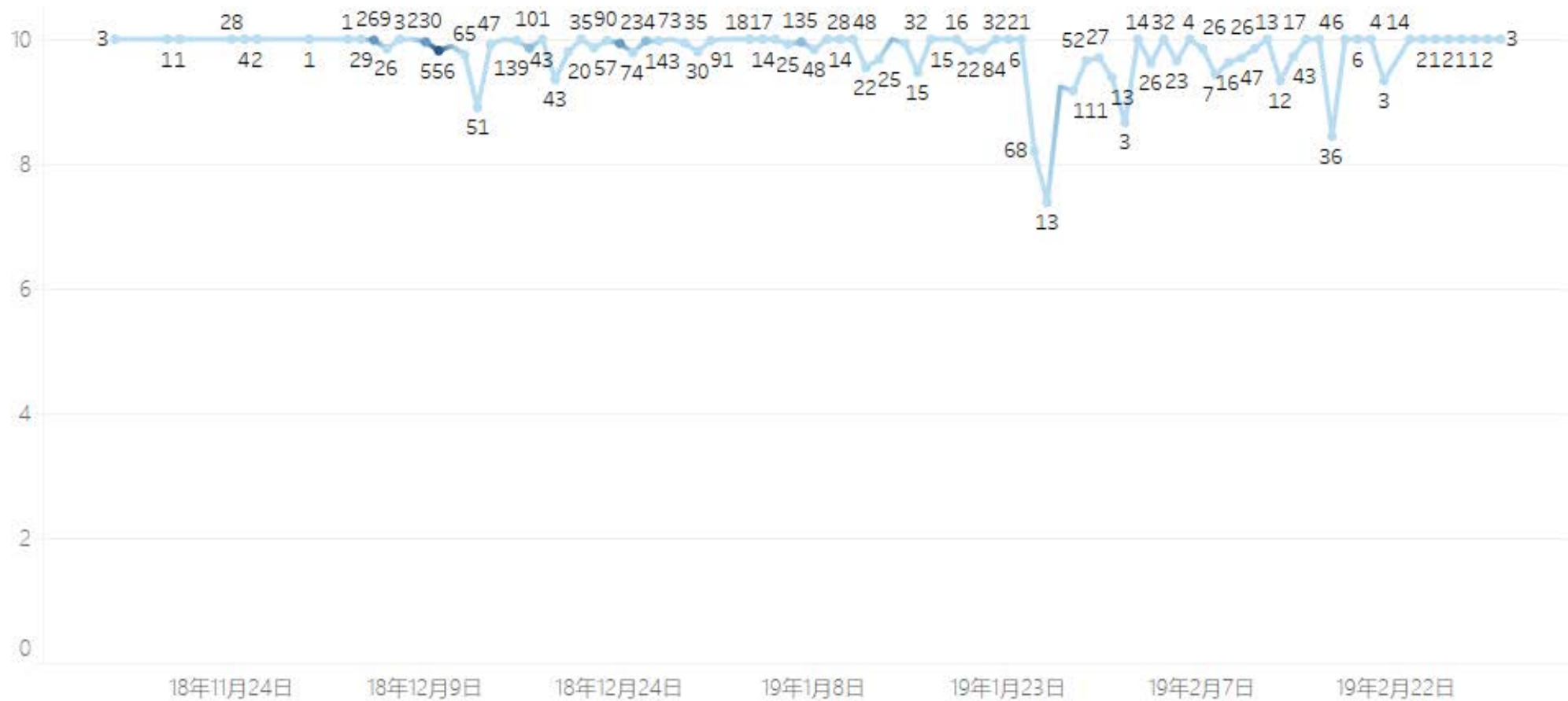


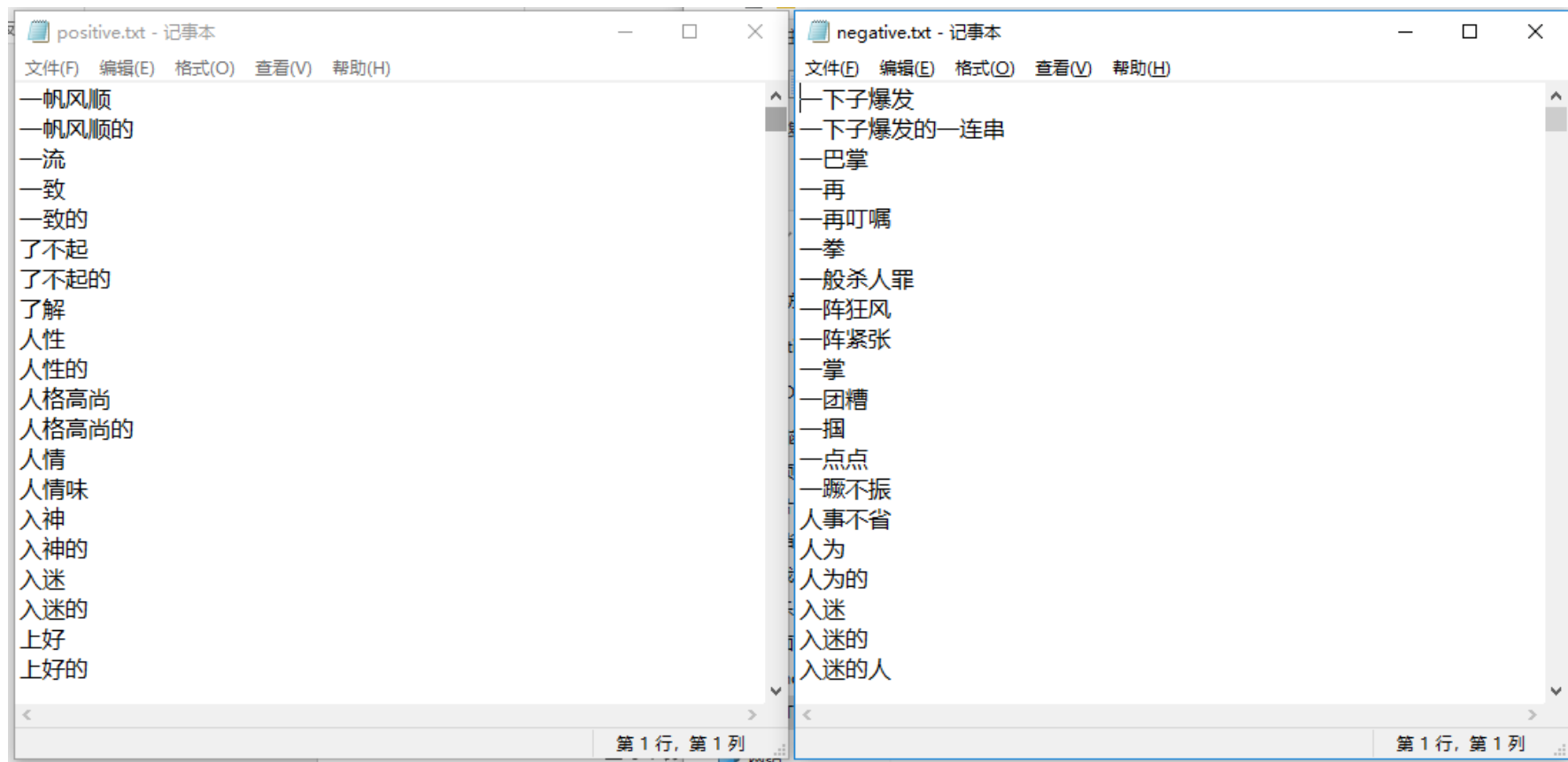
3

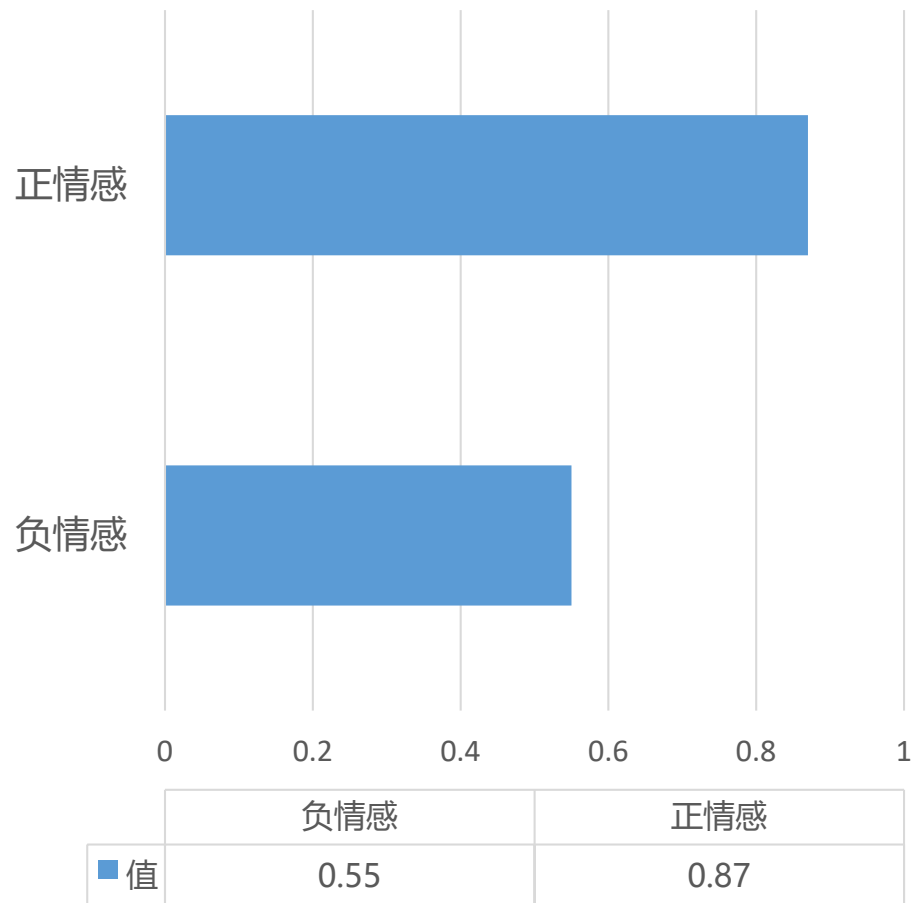
情感分析



评分时间分布







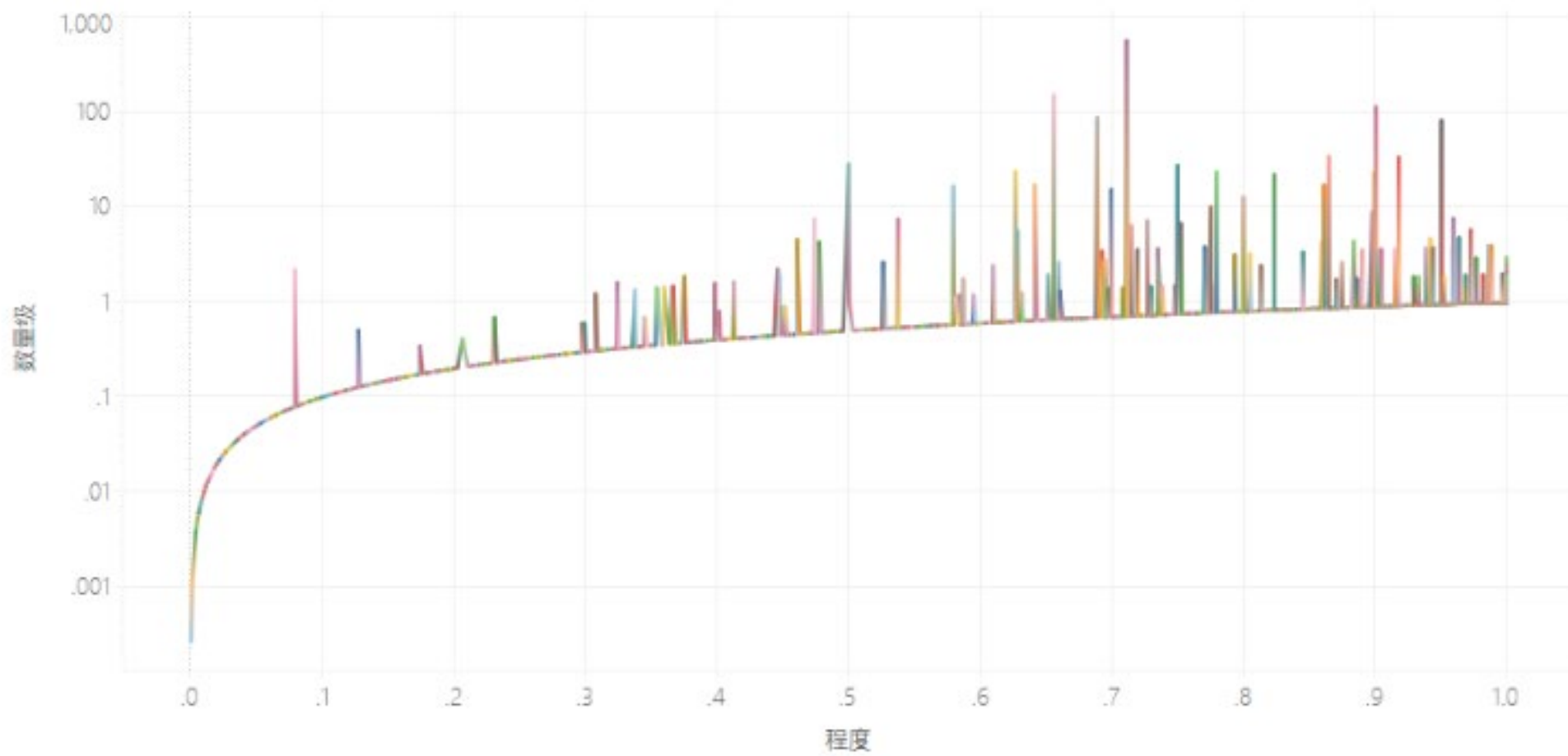
正情感分析精度: 0.8714

p0V : [-7.09589322 -7.09589322 -7.09589322
... -7.7890404 -7.09589322 -7.7890404]

负情感分析精度: 0.5486

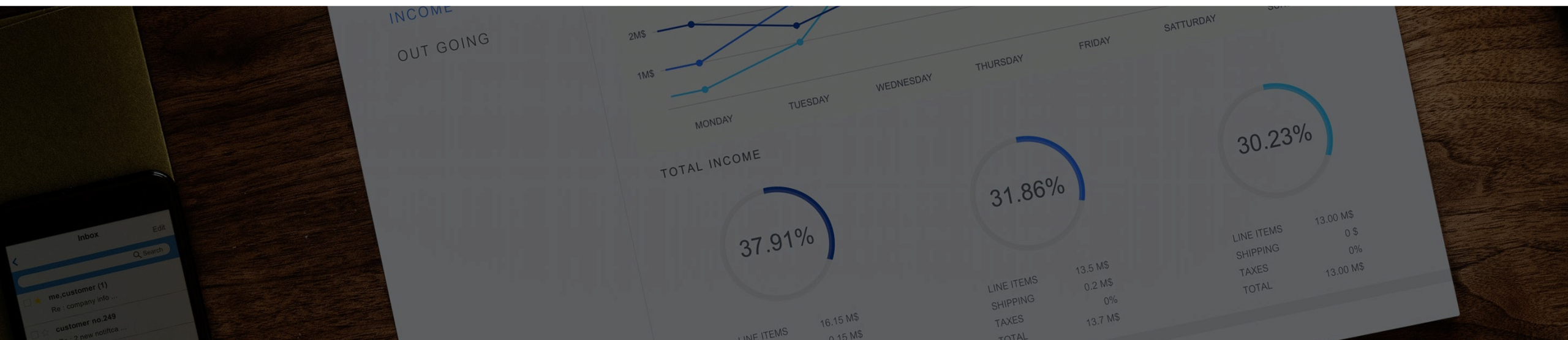
p1V : [-6.20253552 -6.20253552 -6.20253552
... -5.50938834 -6.20253552 -5.50938834]

情感分析



05 总结回顾

REVIEW





前期准备

学习阶段

Python算法
数据分析
数据挖掘
可视化分析
网页抓取



数据获取

实践阶段

网页爬虫
数据处理
完成程序



论文撰写

整合阶段

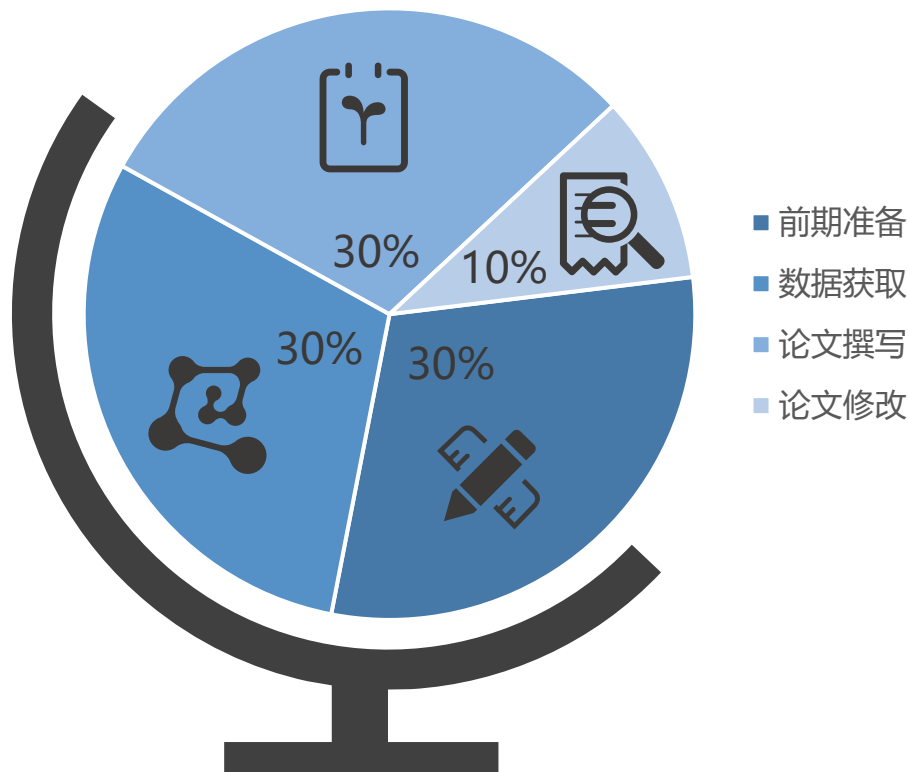
梳理逻辑
修改代码
图表制作



论文修改

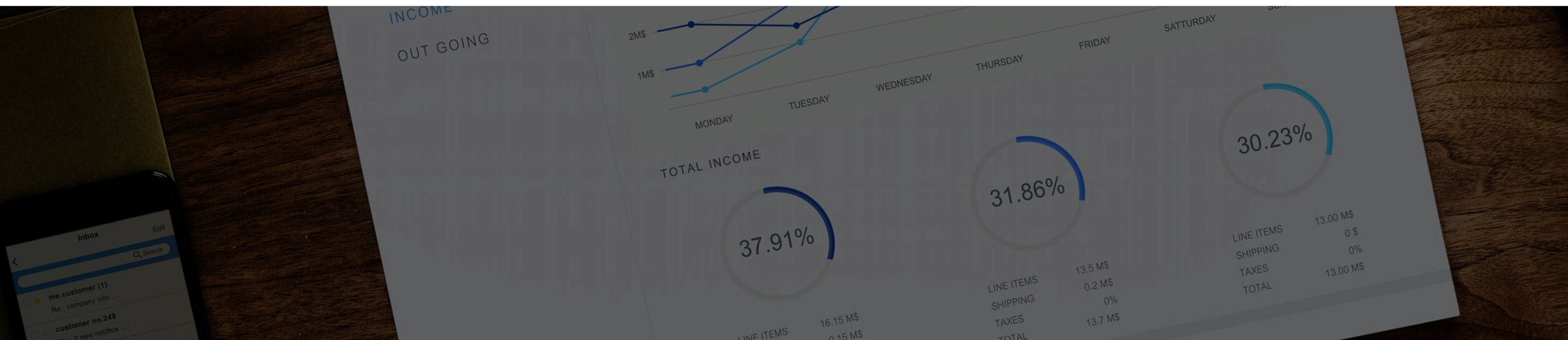
修改阶段

完善逻辑
格式修改
整体调整



06 参考文献

REFERENCE



1	顾君忠	大数据分析[J]	计算机教育, 2014, No.209(5):122-126
2	刘红岩, 陈剑, 陈国青	数据挖掘中的数据分 类算法综述[J]	清华大学学报(自然科学版), 2002, 42(6):727-730
3	Swami A , Jain R	Scikit-learn: Machine Learning in Python[J]	Journal of Machine Learning Research, 2012, 12(10):2825-2830
4	刘信杰, 李艳, 胡学钢	Naive Bayes算法在 垃圾邮件过滤系统 中的应用与改进[J]	潍坊学院学报, 2007, 7(6):26-27
5	周钦强	基于人工智能技术 Naive Bayes文本自 动分类系统研究[D]	广东工业大学, 2005
6	慕春棣, 戴剑 彬, 叶俊	用于数据挖掘的贝叶 斯网络[J]	软件学报, 2000, 11(5):660-666

Morris Charts

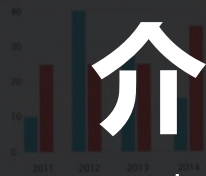
Line Chart



Area Chart



Bar Chart



Donut Chart



介绍完毕 请指正

Introduction is completed, please correct me.

Sparkline Charts

Line Chart



Bar Chart



Pie Chart



Easy Pie Charts

