目标：

1.小论文撰写

2.课外阅读：每天半小时

3.cs224n

4.英语

5.leetcode

计算机基础课程学习：网络编程、软件工程、编译原理 shell编程，

犯困：leetcode

收尾：阅读，琐事，课外书

茫然：cs224n

英语学习：口语纠正45min 阅读30min

Leetcode

1，看懂题目

2，分析，推导解法

3，将思路转换为代码。

从某个类型开始 25min 写不出来直接题解 ＋笔记记录

9~10

完成开题报告撰写，掌握研究方向情况，由2~3篇相关论文，创新点有思路。有进行简单的实验。

8.30~9.3：截至周六

紧急重要：1.找几篇nlp相关安全论文 完成阅读

紧急不重要：整理下周计划，项目内容填充

重要不紧急:英语，运动，

不重要不紧急：课外阅读，leetcode 题

9.6~9.10：截至周六

紧急重要：1.related work中相关论文 5篇，方向包括威胁情报来源信任分析 和非结构化识别

紧急不重要：整理下周计划，做表单，做表格，画ppt

重要不紧急：英语，运动

不重要不紧急：课外阅读，leetcode

9.13~9.18

紧急重要：1.按标准模板书写开题报告格式，开始整理内容 2.看完几篇综述

紧急不重要：整理下周计划，核酸检测，补牙

重要不紧急：英语，运动

不重要不紧急：课外阅读，leetcode

9.22~9.26

紧急重要：1.顶会三年内相关遍历 2.确定论文主要内容 数据集有无 相关性

紧急不重要：整理下周计划

不重要不紧急：课外阅读，leetcode

9.29~10.10

紧急重要：1.针对社交媒体的有毒内容检测 和恶意软件分析 进行评估。

紧急不重要：整理下周计划

不重要不紧急：课外阅读，leetcode，深水证

10.11~10.17

紧急重要：1.社交媒体检测 2.专利转论文

重要不紧急：1.运动 2.口语

紧急不重要：整理下周计划

不重要不紧急：课外阅读，leetcode

10.11：洗衣服、买衣服，内存卡，记账。

运动：了解一下饮食，健身房计划，姿势纠正资料。

课外阅读：心流

论文：委婉语检测 两部分1.委婉语检测 2.委婉语鉴别

10.12：买外套（牛仔外套，普通外套），记账、羽绒服（双十一）、内衣当垫

外套：levis牛仔外套（想凑单）

论文：已记录

10.13：普通外套，记账、羽绒服（双十一）、写微博

10.14 预约牙医，论文转换完毕，健身房，信息会议，

10.15 自由泳教程 找完related work

委婉语含义：没有明确研究的 只有词义发现【60】【61】 词义消歧【62】~【65】

地下论坛黑话识别：

已记录

10.17 自由泳 论文 表达和求助

地下论坛黑话检测

已记录

情绪分析来寻找委婉语

已记录

## 10.18~10.24

紧急重要：1.相关论文阅读 2.论文内容填充

重要不紧急：1.运动 2.口语

紧急不重要：整理下周计划

不重要不紧急：课外阅读，leetcode

10.18情绪分析寻找委婉语

记录

10.19 1.找李老师 2.看自由泳基础 x 3.专利补充

断网消失

10.22 ：论文转英语两章；两篇论文阅读；确定蛋糕；

10.25 论文一章

## 10.26 ~10.31，，

紧急重要：1. 论文ppt 2. 创新点 3.整理完之前看的三篇论文

重要不紧急：1.运动 2.口语

紧急不重要：整理下周计划

不重要不紧急：课外阅读，leetcode

10.26 day1

10.27 day2 论文改成了模板

10.28 精读地下黑市论文，读完另一篇地下黑市工具论文

10.29 做好ppt

10.30

地下市场分析工具：

已记录

识别地下市场产品：识别地下市场产品：

使用词嵌入网络确定仇恨发言中的编码词

已记录

## 11.1~11.7

紧急重要：1. 项目任务2. 创新点 3.ai顶会上的相关论文 以及找到的剩余论文

重要不紧急：1.运动 2.口语 3.周末整理一遍内容

紧急不重要：整理下周计划

不重要不紧急：课外阅读，leetcode

11.1：项目格式改动，讨论班，看完kdes（没啥感想）

11.2

项目书再检查一遍，仔细考察一遍需要提交内容，确定进度安排，是否需要将论文填入。

课题一、二不要。四只需要基于区块链的安全分发与溯源功能模块

子课题任务书（就是整体的任务书截取你们研究这一部分），课题研究方案

四月份要开始原型，七月份要交报告。

论文：找新的年份的委婉语检测的论文，可以重点关注一下含义鉴别。

已记录

游泳：自由泳踢腿姿势练习20min左右

英语口语：半小时 x

中午阅读：半小时

11.3

逼着自己输出

项目：图改完 任务书写完

论文：相关：词级语义相似度的无监督模型

两篇检测论文 浅读即可

英语口音

阅读：

王德峰心学

Transforms 进行识别+towards codewords

11.4

项目、论文、阅读、英语

算法总结：

11.5

索要数据信：

11.6

比赛：如何根据字形和同义进行预处理

论文：精读核心论文主要思想（实验内容）、写要数据的信、俚语识别相关论文（与论文场景的不同之处）

游泳：看两个小节去实践

比赛思考：1.同形、同拼音角度检查 2.针对其他类型错误，联系鲁棒性高的解决方案 3.增加数据集数目 4.模型融合 。先自己探索为主，再使用已有方案。

论文：

已记录

11.7

论文：立项依据、研究内容、研究方案、课题特色

## 11.8~11.16

11.9

王德峰：

开题报告：要最后一份数据、观察数据情况、观察中文语境下的情况、明天约老师的时间

代码：

健身：需要锻炼肱三头肌

11.10：

王德峰：

论文：下午确定创新点，再约李老师时间。对无监督方法进行更细粒度的考察，再看中文语境下的telegram论文。

代码：notebook 如何调用其他py文件里面的函数

### 11.11

论文：用bert做自监督的相关论文，俚语辨识相关论文，Zero-shot Knowledge Transfer

Bert的mlm是否能替换为其他掩码模型

代码：写出qa 模型，最好能试试训练

方向：查询工程方向

王德峰：

### 11.12

代码：qa调试、leetcode一道

论文：俚语鉴别、自监督学习、

实验：筛选有禁用词的句子、

开题报告：缺少鉴别一部分的思路，参考词义发现的文章和之前的几种方法

工程方向 r&d d

### 11.13

1.整理一下目前的想法

2.试着补之前没做好的内容 Python 类基础知识 、正则表达式 俚语鉴别和自监督学习论文

3.完成代码等 qa调试

4.leetcode

Python:z字符串前缀 f 相当于format函数 r停止python解析器，防止将/当成转义字符，b 字符串byte类型 ；u 字符串以unicode 编码

### 11.14

Adamw

整理一下收藏夹

## 11.16 ~11.22

紧急重要： 1.项目任务 2.pytorch的优化方法 3.github练手项目寻找 4.开题的框架和报告撰写

5.阅读论文：风格识别、半监督

重要不紧急：1.运动 2.周末整理一遍内容 3.每天记录内容，并进行输出，

4 优化方法复习

紧急不重要：1 整理下周计划 2 leetcode

不重要不紧急：心流阅读完毕，

* pytorch优化方法：swa 四类 fine-tune 7类 efficiently 3类 加速5类

github练手项目寻找：中文nlp框架、比赛

* 心流阅读：51% 一天 10%

理出一个简单的流程来

* Leetcode：
* 开题的框架和报告撰写
  + 模型框架确定，可行性预估
  + 确定工作内容、创新点包装
* 阅读论文：风格识别、半监督
* 优化方法复习：

### 11.16

修笔，拿咖啡

Leetcode：

王德峰：金刚经、

优化方法：swa、question

框架：

待办不紧急：Paddle 上能不能用pytorch 怎么用、有道云笔记 怎么默认python

、有道云笔记内容格式调整、怎么合理规划任务时间、docker如何使用

paddle 上能不能用pytorch 怎么用：不能

怎么合理规划任务时间：

优化方法：

### 11.17

|  |  |  |  |
| --- | --- | --- | --- |
| 任务排序 | 意愿得分 |  | 番茄 |
| 框架撰写，开题报告重叙 | 4 | 5 | 5.5 |
| 论文阅读 | 5 | 7 |  |
| 优化方法复习 | 1 | 4 | 4 |
| pytorch 优化方法 | 6 | 10 |  |
| leetcode | 3 | 8 |  |
| 王德峰，心流 | 2 | 8 |  |

新增待做：Adadelta深究、adam详解、牛顿法、费曼学习法作用、

苏剑林博客有关bert

风格识别做prompt

向上需求：与人交互、理想自我的投射（热爱、专注）、认识自我、

情绪需求：琐碎、倾诉、

优化方法：

GD 优缺点：简单、更新次数多

动量优化法：

### 11.18

|  |  |  |  |
| --- | --- | --- | --- |
| 任务排序 | 意愿得分 |  | 番茄 |
| 框架撰写，开题报告重叙 | 5 | 6 | 1 |
| 论文阅读 | 6 | 7 |  |
| 竞赛 | 1 | 4 | 3个番茄 |
| pytorch 优化方法 | 2 | 6 | 4 |
| leetcode | 3 | 8 |  |
| 王德峰，心流 | 4 | 10 |  |

代码：

王德峰：

佛性：

数据情况：

### 11.19

|  |  |  |  |
| --- | --- | --- | --- |
| 任务排序 | 意愿得分 |  | 番茄 |
| 竞赛 | 3 | 4 | 1 |
| 框架撰写，开题报告重叙 | 4 | 6 |  |
| Pytorch优化 | 2 | 5 |  |
| 海森堡 | 1 | 5 |  |
| leetcode | 5 | 10 |  |
| 王德峰，心流 | 6 | 11 |  |

滴答清单 子任务 roc voc

MarginRankingLoss

### 11.20~21

优化方法

晚上写总结

## 11.22~11.28

紧急重要： 1.开题ppt撰写 2.项目

重要不紧急： 1.竞赛 2.运动 3.leetcode 4.pytorch优化方法

紧急不重要：1.周末整理一遍内容 2.每天记录内容，并进行输出

不重要不紧急：心流重新看一遍并整理

### 11.22

|  |  |  |  |
| --- | --- | --- | --- |
| 任务排序 | 意愿得分 |  | 番茄 |
| 开题报告 | 2 | 3 |  |
| 项目 | 3 | 5 |  |
| 竞赛 | **1** | 4 |  |
| Pytorch 优化方法 | 4 | 8 |  |
| 王德峰,阅读 | 5 | 10 |  |
| leetcode | 6 | 12 |  |

Leetcode刷题：

Codetop 企业题库

闲暇时间用于思考刷题

书籍 算法4 底层算法原理实现

目标增加：算法

方法：按tag刷、按使用频率刷、最后时间用周赛模拟面试

 第一遍：可以先思考，之后看参考答案刷，结合其他人的题解刷。思考、总结并掌握本题的类型，思考方式，最优题解。

 第二遍：先思考，回忆最优解法，并与之前自己写过的解答作比对，总结问题和方法。

 第三遍：提升刷题速度，拿出一个题，就能够知道其考察重点，解题方法，在短时间内写出解答。

基础list ：basic leetcode 前200 easy题

Most frequent / Top 100 linked

进阶list：premium /second frequent 稍微更难 但仍然出现频率很大的题目  
首先解决基础和进阶，确保里面80%以及掌握情况再刷一遍FLAG，都了解之后再面试什么公司着重刷该公司的。

开会时间可以干什么

1.阅读比赛信息

阅读说明、评价、时间线、奖品和规则，可能有些比赛可能有更多的细节，所以养成阅读所有提供的信息和标签的习惯。

2. 验证数据文件格式

阅读数据描述并查看数据文件：train、test、sample\_submission。检查所有字段与描述是否匹配。

3. 设置您的环境

将数据集下载到您的本地机器或使用免费的代码资源，如Kaggle Notebooks和Google Colab。

4. 探索数据

探索数据。了解数据。在阅读其他人的分享之前，自己探索数据通常是个好主意。

5. 阅读比赛论坛

比赛论坛有很多有用的信息和有趣的讨论，你应该关注论坛，或者不时检查新帖子和评论。

6. 阅读比赛Notebook

阅读公开的EDA和baseline Notebook，这是开始自己的代码的好地方，也可以跟随开源的进度。

7. 建立一个基线模型

拥有一个非常基本的端到端模型，不一定需要是机器学习模型。

8.提交到排行榜

使用模型对测试数据进行预测，并将其提交给 Kaggle排行榜。

9. 提出问题

根据上述操作，你或许有一些疑问，可以尝试在比赛论坛提出或者记录下自己的疑问。

**1.想法**

**验证想尝试想法，验证方法次数越多，越接近正确答案。**

**2. 数据清洗**

**对原始数据集处理为最干净的形式，不同的数据集需要不同类型的预处理和转换，有时需要不同的模型。**

**3. 特征工程**

**新功能可以显着提高模型性能。不同类型的特征可能适用于不同的模型。 去野外。努力尝试，尝试尽可能多的功能和想法。**

**4. 错误分析**

**确定模型能够很好预测和失败原因，考虑结果的原因以及你可以做些什么。**

**1.改进可视化**

**增加对数据的可视化和演示，研究高票数的 EDA Notebook并学习如何构建出色的可视化。**

**2. 发布你的EDA Notebook**

**使Kaggle成为如此出色平台的原因是Kagglers和社区，你也可以做一份贡献。**

**3. 分享见解**

**在论坛发帖，或者写一些评论。你分享的越多，你学到的就越多，也会有更多的人帮助你。**

**1.探索模型**

**这是您试验和构建大量模型以找出最有效的模型的机会，阅读并理解不同模型的内部工作原理很重要。**

**2. 集成并组合模型**

**使用blend或者stacking将多个模型组合起来，并为每个基础模型搜索得到各自的特征空间和超参数。**

### 11.24

上午：论文问题并整理

下午：外勤

晚上：kaggle

方向word该怎么使用

滴答清单如何使用

车上应该干啥：

买吹风机

Github推送

Torch.view(-1) 拉伸到一维

Detach 切断反向传播

np.concatenate(a,b) 将a b进行拼接 （不扩展维度）

np.mean(final\_preds, axis=0) axis=0 压缩行 对列求平均

当我们试图评判一个人的时候，应该试着多去了解一下，避免陷入自以为是的境地。

王德峰 心即理 第一讲

### 11.25