

# گزارش پروژه چهارم درس مبانی هوش مصنوعی (پردازش زبان طبیعی)

امید ماهیار: ۹۷۳۱۱۰۰

پویان حسابی: ۹۷۳۱۱۲۲

## ۱- تاثیر حذف کلمات پرتکرار و کم تکرار:

همانطور که در فایل result.txt مشاهده می‌کنیم، در حالت کلی اگر رشته‌هایی که زیاد تکرار شدند را حذف کنیم، دقت ما بیشتر می‌شود چرا که:

- از نظر زبانی وقتی کلمه‌ای مثل the را از جمله حذف کنیم تاثیر در معنای آن ندارد.
- از نظر برنامه نویسی و آماری وقتی کلمه‌ای زیاد تکرار شده است باعث می‌شود احتمال کلمات دیگر کم شود.

## ۲- تاثیر مقدار $\lambda$ و $\epsilon$ دقت بدست آمده:

با توجه به تست کیس‌های مختلف مشاهده می‌کنیم بهترین حالت زمانی است که  $\epsilon$  بسیار پایین باشد. به طور کلی وجود  $\epsilon$  برای این است که اگر کلمه‌ای در دیکشنری وجود نداشت احتمال آن صفر نشود.

وقتی  $\lambda 2$  باید از همه مقدار بیشتری داشته باشد که نشان دهنده ی احتمال وجود تک کلمه (کلمه اول) در دیکشنری می‌باشد.

و  $\lambda 3$   $\lambda 1$  باید به ترتیب مقدار خیلی کم و کمی نسبت به  $\lambda 2$  داشته باشند، چرا که به طور کلی  $\lambda 1$  ضربدر اپسیلون می‌شود که از آنجایی که مقدار اپسیلون بسیار کم است پس خوب است که مقدار آن زیاد باشد تا بتوانیم مقدار  $\lambda 3$  را

کمتر کنیم. لاند ۳ احتمالاً وجود آن دو کلمه در کل دیکشنری است که می‌دانیم خیلی احتمال پائینی دارد پس بهتر است ضریب آن هم خیلی کم باشد.

### ۳- مقدار بیشینه:

-> [ $\lambda_1$ : 0.1 , $\lambda_2$ : 0.6 , $\lambda_3$ : 0.3 ,e: 0.0001]

Positive accuracy: 0.73 ,Negative accuracy: 0.71 |[Whole accuracy: 1.44]

در ادامه مجدداً تست کیس های فایل result.txt آورده شده است.

### Cleaned words:

#### Test untrained words bigram:

##### changing e:

[ $\lambda_1$ : 0.33 , $\lambda_2$ : 0.33 , $\lambda_3$ : 0.34 ,e: 0.0001] -> Positive accuracy: 0.74 ,Negative accuracy: 0.68 |[Whole accuracy: 1.42]

[ $\lambda_1$ : 0.33 , $\lambda_2$ : 0.33 , $\lambda_3$ : 0.34 ,e: 0.0005] -> Positive accuracy: 0.73 ,Negative accuracy: 0.67 |[Whole accuracy: 1.4]

[ $\lambda_1$ : 0.33 , $\lambda_2$ : 0.33 , $\lambda_3$ : 0.34 ,e: 0.0010] -> Positive accuracy: 0.71 ,Negative accuracy: 0.67 |[Whole accuracy: 1.38]

[ $\lambda_1$ : 0.33 , $\lambda_2$ : 0.33 , $\lambda_3$ : 0.34 ,e: 0.0050] -> Positive accuracy: 0.68 ,Negative accuracy: 0.64 |[Whole accuracy: 1.32]

[ $\lambda_1$ : 0.33 , $\lambda_2$ : 0.33 , $\lambda_3$ : 0.34 ,e: 0.0100] -> Positive accuracy: 0.68 ,Negative accuracy: 0.65 |[Whole accuracy: 1.33]

[ $\lambda_1$ : 0.33 , $\lambda_2$ : 0.33 , $\lambda_3$ : 0.34 ,e: 0.0500] -> Positive accuracy: 0.66 ,Negative accuracy: 0.64 |[Whole accuracy: 1.3]

[ $\lambda_1$ : 0.33 , $\lambda_2$ : 0.33 , $\lambda_3$ : 0.34 ,e: 0.1000] -> Positive accuracy: 0.63 ,Negative accuracy: 0.63 |[Whole accuracy: 1.26]

[ $\lambda_1$ : 0.33 , $\lambda_2$ : 0.33 , $\lambda_3$ : 0.34 ,e: 0.5000] -> Positive accuracy: 0.64 ,Negative accuracy: 0.56 |[Whole accuracy: 1.20]

[ $\lambda_1$ : 0.33 , $\lambda_2$ : 0.33 , $\lambda_3$ : 0.34 ,e: 0.9000] -> Positive accuracy: 0.64 ,Negative accuracy: 0.54 |[Whole accuracy: 1.18]

##### changing $\lambda_1$ :

[ $\lambda_1$ : 0.2 , $\lambda_2$ : 0.4 , $\lambda_3$ : 0.4 ,e: 0.0001] -> Positive accuracy: 0.74 ,Negative accuracy: 0.68 |[Whole accuracy: 1.42]

[ $\lambda_1$ : 0.4 , $\lambda_2$ : 0.3 , $\lambda_3$ : 0.3 ,e: 0.0001] -> Positive accuracy: 0.74 ,Negative accuracy: 0.68 |[Whole accuracy: 1.42]

[ $\lambda_1$ : 0.6 , $\lambda_2$ : 0.2 , $\lambda_3$ : 0.2 ,e: 0.0001] -> Positive accuracy: 0.74 ,Negative accuracy: 0.68 |[Whole accuracy: 1.42]

[ $\lambda_1$ : 0.8 , $\lambda_2$ : 0.1 , $\lambda_3$ : 0.1 ,e: 0.0001] -> Positive accuracy: 0.74 ,Negative accuracy: 0.68 |[Whole accuracy: 1.42]

##### changing $\lambda_2$ :

[ $\lambda_1$ : 0.4 , $\lambda_2$ : 0.2 , $\lambda_3$ : 0.4 ,e: 0.0001] -> Positive accuracy: 0.73 ,Negative accuracy: 0.68 |[Whole accuracy: 1.41]

[ $\lambda_1$ : 0.3 , $\lambda_2$ : 0.4 , $\lambda_3$ : 0.3 ,e: 0.0001] -> Positive accuracy: 0.74 ,Negative accuracy: 0.70 |[Whole accuracy: 1.44]

[ $\lambda_1$ : 0.2 , $\lambda_2$ : 0.6 , $\lambda_3$ : 0.2 ,e: 0.0001] -> Positive accuracy: 0.71 ,Negative accuracy: 0.71 |[Whole accuracy: 1.42]

[ $\lambda_1$ : 0.1 , $\lambda_2$ : 0.8 , $\lambda_3$ : 0.1 ,e: 0.0001] -> Positive accuracy: 0.68 ,Negative accuracy: 0.69 |[Whole accuracy: 1.37]

**changing  $\lambda_3$ :**

[ $\lambda_1$ : 0.4 , $\lambda_2$ : 0.4 , $\lambda_3$ : 0.2 ,e: 0.0001] -> Positive accuracy: 0.73 ,Negative accuracy: 0.71 |[Whole accuracy: 1.44]

[ $\lambda_1$ : 0.3 , $\lambda_2$ : 0.3 , $\lambda_3$ : 0.4 ,e: 0.0001] -> Positive accuracy: 0.73 ,Negative accuracy: 0.68 |[Whole accuracy: 1.41]

[ $\lambda_1$ : 0.2 , $\lambda_2$ : 0.2 , $\lambda_3$ : 0.6 ,e: 0.0001] -> Positive accuracy: 0.72 ,Negative accuracy: 0.68 |[Whole accuracy: 1.40]

[ $\lambda_1$ : 0.1 , $\lambda_2$ : 0.1 , $\lambda_3$ : 0.8 ,e: 0.0001] -> Positive accuracy: 0.69 ,Negative accuracy: 0.68 |[Whole accuracy: 1.37]

**best:**

**[ $\lambda_1$ : 0.1 , $\lambda_2$ : 0.6 , $\lambda_3$ : 0.3 ,e: 0.0001] -> Positive accuracy: 0.73 ,Negative accuracy: 0.71 |[Whole accuracy: 1.44]**

**Uncleaned words:**

**Test untrained words bigram:**

**changing  $\lambda_1$ :**

[ $\lambda_1$ : 0.3 , $\lambda_2$ : 0.35 , $\lambda_3$ : 0.35 ,e: 0.0001]->Positive accuracy: 0.74 ,Negative accuracy: 0.63 |[Whole accuracy: 1.37]

[ $\lambda_1$ : 0.6 , $\lambda_2$ : 0.2 , $\lambda_3$ : 0.2 ,e: 0.0001] -> Positive accuracy: 0.74 ,Negative accuracy: 0.63 |[Whole accuracy: 1.37]

[ $\lambda_1$ : 0.9 , $\lambda_2$ : 0.05 , $\lambda_3$ : 0.05 ,e: 0.0001]->Positive accuracy: 0.74 ,Negative accuracy: 0.64 |[Whole accuracy: 1.38]

**changing  $\lambda_2$ :**

[ $\lambda_1$ : 0.35 , $\lambda_2$ : 0.3 , $\lambda_3$ : 0.35 ,e: 0.0001]->Positive accuracy: 0.74 ,Negative accuracy: 0.63 |[Whole accuracy: 1.37]

[ $\lambda_1$ : 0.2 , $\lambda_2$ : 0.6 , $\lambda_3$ : 0.2 ,e: 0.0001] -> Positive accuracy: 0.72 ,Negative accuracy: 0.66 |[Whole accuracy: 1.38]

[ $\lambda_1$ : 0.05 , $\lambda_2$ : 0.9 , $\lambda_3$ : 0.05 ,e: 0.0001]->Positive accuracy: 0.71 ,Negative accuracy: 0.64 |[Whole accuracy: 1.35]

**changing  $\lambda_3$ :**

[ $\lambda_1$ : 0.35 , $\lambda_2$ : 0.35 , $\lambda_3$ : 0.3 ,e: 0.0001]->Positive accuracy: 0.73 ,Negative accuracy: 0.64 |[Whole accuracy: 1.37]

[ $\lambda_1$ : 0.2 , $\lambda_2$ : 0.2 , $\lambda_3$ : 0.6 ,e: 0.0001] -> Positive accuracy: 0.74 ,Negative accuracy: 0.64 |[Whole accuracy: 1.38]

[ $\lambda_1$ : 0.05 , $\lambda_2$ : 0.05 , $\lambda_3$ : 0.9 ,e: 0.0001]->Positive accuracy: 0.74 ,Negative accuracy: 0.63 |[Whole accuracy: 1.37]

**best:**

**[ $\lambda_1$ : 0.1 , $\lambda_2$ : 0.6 , $\lambda_3$ : 0.3 ,e: 0.0001] -> Positive accuracy: 0.73 ,Negative accuracy: 0.65 |[Whole accuracy: 1.38]**