

به نام خدا

پروژه تشخیص Unsupervised sentiment درس NLP

داده مورد استفاده: کامنت های مثبت و منفی مربوط به وسایل منزل از سال 2002

90% داده ها را برای آموزش جدا کردیم و همه را با هم در یک فایل قرار دادیم.

ابتدا با استفاده از کلاس splitter کامنت ها را از هم جدا کرده و در یک فایل جدید قرار می دهیم.

(متد tokenizer این کار را انجام می دهد.)

سپس با استفاده از کلاس POSTagger با استفاده از کتابخانه openNLP متن را postag میکنیم و با استفاده از آن phrase های 2 کلمه ای را با قواعد زیر بدست می آوریم:

نکته: کلاس اصلی که ران می شود کلاس polarity است که train و تست در آن انجام می شود.

Dan Jurafsky



## Extract two-word phrases with adjectives

First Word	Second Word	Third Word (not extracted)
JJ	NN or NNS	anything
RB, RBR, RBS	JJ	Not NN nor NNS
JJ	JJ	Not NN or NNS
NN or NNS	JJ	Nor NN nor NNS
RB, RBR, or RBS	VB, VBD, VBN, VBG	anything

حالا با استفاده از 2 کلمه مناسب می خواهیم کامنت ها را به دو دسته مثبت و منفی تقسیم کنیم. این دو کلمه را excellent و poor انتخاب کردیم و با استفاده از این دو polarity مربوط به هر phrase را در کامنت ها به دست آورده و در یک هشتمپ میریزیم.

حالا برای تست هر کامنت را می خوانیم و مجموع Polarity آنرا با توجه به phrase هایی که دارد به دست می آوریم. اگر بزرگتر از صفر بود کامنت را مثبت میگیریم در غیر این صورت منفی

پرسیژن و ریکال هم در نهایت محاسبه می شود و خروجی به شکل زیر است:

```
56 try {
70     if (inputLine.contains(postagger.getPhrases().get(i))) {
71         if (inputLine.contains("excellent"))
72             hitsPhraseNearExcellent++;
73         else if (inputLine.contains("poor"))
74             hitsPhraseNearPoor++;
75     }
76 }
77
78 }
79
80 double polarity = (double) (Math.Log(hitsPhraseNearExcellent * hitPoor))
81     - (double) (Math.Log(hitsPhraseNearPoor * hitExcellent));
82 polarityMap.put(postagger.getPhrases().get(i),polarity);
83
84 }
85 catch (IOException e) {
86     e.printStackTrace();
87 }
```

Run: Polarity

```
, , , : 0.47589629276656387
#of positive comments for pos_test.binclean test is : 99
#of negative comments for pos_test.binclean test is : 95
recall of pos test is : 0.510309
recall of pos test is : 0.523810
#of positive comments for neg_test.binclean test is : 103
#of negative comments for neg_test.binclean test is : 84
prec of neg test is : 0.533679
recall of pos test is : 0.449198
```

Handwritten note: *Prec* with an arrow pointing to the precision values in the output.