

Assignment 6

Jingyuan Liu

jingyual@andrew.cmu.edu

Deliverables:

Question 3

The training result is as following:

Model Evaluation Metrics:	
Metric	Value
Accuracy	0.6811
Kappa	0.3622

Figure 1. Cross-validation Data Model performances

Question 4

The training result is as following:

Model Evaluation Metrics:	
Metric	Value
Accuracy	0.7095
Kappa	0.419

Figure 2. Development Data Model performances

Question 5

Using the comparing result from LightSide, we can navigate the features to see which feature would have comparatively “big” horizontal absolute difference and vertical absolute differences. The result is as follows:

Act \ Pred	neg	pos
neg	1283	514
pos	518	1238

Feature	Horizontal A...	Vertical Absolute...	Feature Weight
and	0.1118	0.0252	-0.1392
of	0.0555	0.0234	-0.036
i	0.053	0.0208	0.1538
more	0.0523	0.0132	0.2078
that	0.0517	0.0295	-0.0709
but	0.0509	0.0322	0.0936
's	0.0484	0.0737	0.053
movie	0.0431	0.0273	0.2051
as	0.0424	0.024	0.0929
than	0.0423	0.0078	0.0241
an	0.0415	0.0072	-0.1629
on	0.0402	0.0383	0.1623
film	0.0344	0.0264	-0.1726
funny	0.0327	0.0133	-0.6601
for	0.0321	0.003	0.0725
it	0.0317	0.0076	-0.105
n't	0.0297	0.0113	0.2347
is	0.0273	0.0349	-0.0603
like	0.0271	0.0062	0.0916
would	0.0267	0.002	0.3833
family	0.0242	0.0047	-0.7835
have	0.0239	0.0037	0.1491
documentary	0.0239	0.0066	-0.3131
watching	0.0233	0.0204	0.696
director	0.0233	0.016	-0.0047
be	0.0224	0.0136	-0.0107
no	0.0221	0.0202	0.3527
never	0.0213	0.0071	0.2695
still	0.0209	0.0051	-0.7622
with	0.0205	0.0278	-0.0943

Act \ Pred	neg	pos
neg	1283	514
pos	518	1238

Feature	Horizontal A...	Vertical Absolute...	Feature Weight
the	0.0684	0.0575	0.0429
but	0.0511	0.0699	-0.0936
film	0.0469	0.0138	0.1726
of	0.0461	0.0328	0.036
too	0.0456	0.0101	-0.8764
you	0.045	0.0256	0.3426
be	0.0402	0.0042	0.0107
an	0.0392	0.0095	0.1629
a	0.0358	0.0386	0.0203
movie	0.0341	0.0363	-0.2051
bad	0.0339	0.0081	-0.815
even	0.0338	0.0223	0.1362
just	0.0336	0.0023	-0.2384
do	0.0323	0.037	0.1738
what	0.0315	0.0253	0.3761
does	0.0276	0.0169	-0.0368
only	0.0273	0.0033	-0.1965
dull	0.0257	0	-1.6582
's	0.0257	0.0003	-0.053
little	0.0242	0.004	-0.2544
was	0.0234	0.0012	-0.5106
like	0.0231	0.0102	-0.0916
action	0.023	0.0047	-0.5846
there	0.0219	0.0093	-0.1007
thing	0.0207	0.0014	-0.2195
love	0.0198	0.007	0.3019
from	0.0198	0.0145	0.1377
it	0.0191	0.0201	0.105
with	0.0177	0.0066	0.0943
material	0.0175	0.018	-0.1183

Evaluations to Display:
<input checked="" type="checkbox"/> Vertical Absolute Difference
<input type="checkbox"/> Vertical Difference
<input type="checkbox"/> Feature Influence
<input type="checkbox"/> Feature Selection
<input checked="" type="checkbox"/> Feature Weight

Evaluations to Display:
<input checked="" type="checkbox"/> Vertical Absolute Difference
<input type="checkbox"/> Vertical Difference
<input type="checkbox"/> Feature Influence
<input type="checkbox"/> Feature Selection
<input checked="" type="checkbox"/> Feature Weight

We can see from these two tables that the features, like “the”, “but”, and “of” are problematic features. These features are the “stop-words” which appeared a lot but do not have significant meanings.

Question 6

I would propose two features:

1. Removing stop words from input dataset
2. Bigram features

These two features should be better than current features.

1. It removed stop words, thus would not be influenced by the problematic features.
2. Bigram features would contain more information than unigram features.

Question 7

The training result based on the new features:

Model Evaluation Metrics:	
Metric	Value
Accuracy	0.6817
Kappa	0.3633

Figure 3. Cross-validation Data Model performances based on new features

We can it is slightly better than the past model

Question 8

The improvement is insignificant with $p = 0.949$

