Natural Language Processing

Assignment 2 – Report

Tommy Yu

Apr. 30, 2018

1.1 Unigram Binary Features – Perceptron loss

Best Accuracy on DEV: **64.73%**
Best Accuracy on DEVTEST: **58.26%**

1.2 Unigram Binary Features – Hinge loss

Displays slightly better performance than perceptron loss on DEVTEST.
Best Accuracy on DEV: **67.09%**
Best Accuracy on DEVTEST: **58.62%**

1.3 Feature Weights

"lacks" and "pleasant" indeed top the list for negative and positive labels.
For negative and positive labels, top words (those with largest weights) are mostly adjectives or adverbs, with mostly unambiguous sentiment. Bottom words show a similar yet weaker pattern.
For neutral labels, top words are mostly prepositions and nouns, especially proper nouns.

```
Label 0:
  Top 10 features:
        lacks: 1.39
        lacking: 1.27
        tiresome: 1.24
        listless: 1.21
        lousy: 1.18
        worst: 1.17
        unfunny: 1.17
        uneven: 1.17
        sloppy: 1.17
        depressing: 1.17
  Bottom 10 features:
        wonderful: -0.79
        enjoyable: -0.76
        nice: -0.75
        impressive: -0.72
        -RRB-: -0.71
        cute: -0.70
        likable: -0.69
        beautifully: -0.66
        charming: -0.66
        hilarious: -0.66
Label 1:
  Top 10 features:
        means: 0.36
        Mr: 0.35
        While: 0.34
        nor: 0.33
        Big: 0.33
        Cletis: 0.32
        If: 0.30
```

```
            word: 0.29
            entire: 0.29
            York: 0.28
    Bottom 10 features:
            enjoy: -0.43
            sweet: -0.43
            care: -0.42
            perfect: -0.42
            best: -0.41
            endearing: -0.40
            good: -0.40
            terrific: -0.40
            exciting: -0.39
            flawed: -0.39
Label 2:
    Top 10 features:
            pleasant: 1.29
            treat: 1.22
            thought-provoking: 1.22
            wonderfully: 1.20
            touching: 1.17
            vividly: 1.16
            delight: 1.14
            wonderful: 1.13
            funniest: 1.10
            feel-good: 1.10
    Bottom 10 features:
            n't: -1.07
            lacks: -1.06
            lacking: -1.00
            hardly: -0.93
            lack: -0.92
            no: -0.92
            worst: -0.92
            devoid: -0.90
            failure: -0.90
            nor: -0.87
```

1.4 Error Analysis

1   Sentence: With the exception of some fleetingly amusing improvisations by Cedric the Entertainer as Perry 's boss , there is n't a redeeming moment here .

Gold Standard: 0

Prediction: 2

Category: multiple sentiments with most important sentiment at the end; "fleetingly amusing" contributes to a positive sentiment more than "is n't a redeeming moment" does to a negative sentiment.

2   Sentence: Though only 60 minutes long , the film is packed with information and impressions .

Gold Standard: 2

Prediction: 0

Category: negation; plus, no word from "packed with information and impressions" has a clear indication of positive sentiment.

3 Sentence: A great ensemble cast ca n't lift this heartfelt enterprise out of the familiar .

Gold Standard: 0

Prediction: 2

Category: negation; a few clearly positive words.

4 Sentence: There 's just no currency in deriding James Bond for being a clichéd , doddering , misogynistic boy 's club .

Gold Standard: 1

Prediction: 0

Category: negation; a few clearly negative words; "currency" is also used in a unique way here.

5 Sentence: Under 15 ?

Gold Standard: 1

Prediction: 2

Category: sentence with very limited information; the question mark may have played too much of a role here if most of the training examples with a question mark are categorized as positive.

6 Sentence: To say this was done better in Wilder 's Some Like It Hot is like saying the sun rises in the east .

Gold Standard: 0

Prediction: 2

Category: metaphor/sarcasm; positive words like "better", "Like", and (maybe) "hot" contribute to the prediction.

7 Sentence: The ga-zillionth airhead movie about a wife in distress who resorts to desperate measures .

Gold Standard: 1

Prediction: 0

Category: questionable gold standard assignment/ambiguity of the sentence; description itself is neutral, but the content leans towards negative.

8 Sentence: A gripping , searing portrait of a lost soul trying to find her way through life .

Gold Standard: 1

Prediction: 2

Category: questionable gold standard assignment/ambiguity of the sentence; description itself is neutral, but the content leans towards positive.

9 Sentence: As unseemly as its title suggests .

Gold Standard: 2

Prediction: 0

Category: negation; the model doesn't get that "as… as it suggests" leads negation later in the sentence.

10 Sentence: Filmmakers who can deftly change moods are treasures and even marvels .

Gold Standard: 2

Prediction: 0

Category: unknown word "treasure"; "deft" may be used in a negative sense in the training set.

11 Sentence: On the bright side , it contains Jesse Ventura 's best work since the XFL .

Gold Standard: 1

Prediction: 2

Category: questionable gold standard assignment/ambiguity of the sentence; the sentence seems positive enough to me.

12 Sentence: ` They ' begins and ends with scenes so terrifying I 'm still stunned .

Gold Standard: 1

Prediction: 0

Category: questionable gold standard assignment/ambiguity of the sentence; the sentence seems negative enough to me.

13 Sentence: Impostor has a handful of thrilling moments and a couple of good performances , but the movie does n't quite fly .

Gold Standard: 0

Prediction: 2

Category: multiple sentiments with most important sentiment at the end; "thrilling" and "good" are clearly positive, while "fly" is used metaphorically here.

14 Sentence: It 's mighty tedious for the viewer who has to contend with unpleasant characters , hit-and-miss performances and awkwardly staged scenes .

Gold Standard: 1

Prediction: 0

Category: questionable gold standard assignment/ambiguity of the sentence; the sentence seems negative enough to me.

15 Sentence: A difficult , absorbing film that manages to convey more substance despite its repetitions and inconsistencies than do most films than are far more pointed and clear .

Gold Standard: 2

Prediction: 0

Category: negation; a complex structure; plus, "difficult" is used in a positive way here.

16 Sentence: Though Catch Me If You Can is n't badly made , the fun slowly leaks out of the movie .

Gold Standard: 1

Prediction: 0

Category: questionable gold standard assignment/ambiguity of the sentence; I think it is a negative comment on the movie.

17 Sentence: It will not appeal to the impatient , but those who like long books and movies will admire the way it accumulates power and depth .

Gold Standard: 1

Prediction: 2

Category: questionable gold standard assignment/ambiguity of the sentence; I think it is a positive comment on the movie.

18 Sentence: The only excitement comes when the credits finally roll and you get to leave the theater .

Gold Standard: 0

Prediction: 2

Category: metaphor/sarcasm; the model thinks "excitement" tends to be good.

19 Sentence: Oh come on .

Gold Standard: 0

Prediction: 1

Category: sentence with very limited information; the model doesn't get the negative tone; none of the tokens has a clear indication of negative sentiment on its own.

20 Sentence: Let 's hope -- shall we ?

Gold Standard: 1

Prediction: 2

Category: ambiguity of the sentence; implies a sense of positivity to me; "hope" is surely a keyword for positivity in the model.

1.5 Other features
1.5.1 Bigram binary features
Instead of looking at one word/token, look at two words/tokens in a row.
Worse performance on DEVTEST than unigram binary features.

Best Accuracy on DEV: **64.73%**
Best Accuracy on DEVTEST: **54.62%**

Top features:

```
Label 0:
  Top 10 features:
        ('lacks', 'the'): 1.31
        ('never', 'quite'): 1.26
        ("'s", 'hardly'): 1.12
        ('a', 'failure'): 1.11
        ('lacking', 'in'): 1.08
        ('a', 'mess'): 1.08
        ('most', 'offensive'): 1.05
        ('bad', 'movie'): 1.05
        ('pretty', 'mediocre'): 1.04
        ('terrible', 'movie'): 1.03
  Bottom 10 features:
        ('never', 'dull'): -0.65
        ('not', 'too'): -0.61
        ('a', 'masterpiece'): -0.58
        ('certainly', 'does'): -0.58
        ('funny', 'and'): -0.58
        ('I', 'liked'): -0.56
        ('not', 'without'): -0.56
        ('so', 'well'): -0.54
        ('good', 'time'): -0.53
        ("n't", 'feel'): -0.52
Label 1:
  Top 10 features:
        ('Mr', '.'): 0.64
```

```
           ('most', 'part'): 0.54
           ('an', 'exploration'): 0.50
           ('elements', '.'): 0.47
           ('All', 'Fears'): 0.43
           ("'s", 'time'): 0.43
           ('anything', 'else'): 0.42
           ('Chill', "'"): 0.41
           ("'", 'reunion'): 0.41
           ('Bartlett', "'s"): 0.40
  Bottom 10 features:
           ('too', 'much'): -0.37
           ('could', "n't"): -0.36
           ('too', 'long'): -0.36
           ("'s", 'best'): -0.35
           ('as', 'bad'): -0.35
           ('of', 'hope'): -0.35
           ('rich', 'and'): -0.35
           ('see', 'this'): -0.35
           ('Do', "n't"): -0.34
           ('care', 'about'): -0.34
Label 2:
  Top 10 features:
           ('very', 'best'): 1.18
           ("n't", 'disappoint'): 1.15
           ('very', 'funny'): 1.08
           ('powerful', 'and'): 1.08
           ('to', 'behold'): 1.07
           (',', 'amusing'): 1.07
           ('never', 'dull'): 1.06
           ('smart', ','): 1.05
           ('of', 'laughs'): 1.05
           ('funny', 'stuff'): 1.04
  Bottom 10 features:
           ('lacks', 'the'): -0.94
           ('devoid', 'of'): -0.83
           ('fails', 'to'): -0.78
           ('never', 'rises'): -0.78
           ('supposed', 'to'): -0.69
           ('not', 'very'): -0.65
           ('may', 'not'): -0.63
           ('impossible', 'to'): -0.61
           ("n't", 'much'): -0.61
           ('painful', 'to'): -0.61
```

1.5.2 Normalized features

Unigram binary features after transforming to lowercase, lemmatizing, and removing punctuations and stop-words.
1 percentage point better performance on DEVTEST than unigram binary features.
Note that the lemmatization is of very limited quality partly due to incorrect POS tagging (in part due to the original format of the corpus, e.g., "ca n't" instead of "can't"); but this improvement hints at the potential benefits of removing common stop-words and punctuations, e.g., "the", "and", ".".

Best Accuracy on DEV: 66.18%
Best Accuracy on DEVTEST: 59.71%

Top features:

Label 0:
  Top 10 features:
          listless: 1.55
          disappointment: 1.51
          lousy: 1.46
          unfunny: 1.44
          lack: 1.43
          pointless: 1.41
          fails: 1.39
          failure: 1.38
          lackluster: 1.35
          insult: 1.35
  Bottom 10 features:
          refresh: -1.03
          goodnatured: -0.90
          powerful: -0.88
          likable: -0.85
          fascinate: -0.84
          impressive: -0.82
          wonderful: -0.82
          entertaining: -0.76
          lovely: -0.75
          beautifully: -0.74
Label 1:
  Top 10 features:
          greek: 0.72
          left: 0.67
          partner: 0.54
          solondz: 0.53
          50s: 0.46
          warren: 0.43
          1950s: 0.42
          murdock: 0.41
          harvard: 0.41
          deuce: 0.41
  Bottom 10 features:
          flaw: -0.46
          hole: -0.44
          poignant: -0.44
          playful: -0.43
          uneven: -0.43
          attractive: -0.42
          enjoy: -0.42
          perfect: -0.42
          absurd: -0.41
          excite: -0.40
Label 2:
  Top 10 features:
          wonderfully: 1.60
          vividly: 1.59
          delight: 1.42
          pleasant: 1.37
          miracle: 1.30

```
        heartfelt: 1.29
        proud: 1.26
        elegant: 1.25
        embrace: 1.24
        thoughtprovoking: 1.23
Bottom 10 features:
        neither: -1.65
        lack: -1.52
        devoid: -1.28
        none: -1.22
        failure: -1.21
        flat: -1.21
        hardly: -1.19
        waste: -1.14
        fails: -1.13
        bore: -1.10
```