

NLP Assignment 3

Name: Bo Li

Date: 4/12//2020

processing techniques:

In this Assignment, it applied **Not_features** for the classification task and the experiment. For the data processor and cleaning task, in the set of amazon reviews, it used **sentence tokenize** to split all reviews into each single sentence. In the set of word_features, I mainly focus on words_list retrieving from **sentence_polarity**. It used **isalpha** as a filter to eliminate non-alphabet tokens, used **stopwords** to eliminate unnecessary tokens. In this task, since it used Not_features, it has to be implemented **negation words** which is defined manually. Furthermore, when processing stopwords, it shall be eliminated stop words rather than defined negation words. The reason to clean up words_list to generate word_features is to increase accuracy on training set & test set.

In this Task, **SL_features** & **NOT_features** are implemented. And they used unigram word_features invoking from sentence_polarity to be trained.

SL_features:

```
train_set2, test_set2 = SL_featuresets[1000:], SL_featuresets[:1000]
classifier2 = nltk.NaiveBayesClassifier.train(train_set2)
nltk.classify.accuracy(classifier2, test_set2)

0.7386981804539486
```

NOT_features:

```
# Accuracy "Two different sets of features"
train_set, test_set = NOT_featuresets[1000:], NOT_featuresets[:1000]
classifier = nltk.NaiveBayesClassifier.train(train_set)
nltk.classify.accuracy(classifier, test_set)

0.7968936409679234
```

Based on the results show above, it seems that NOT_features is more accuracy than SL_features. However, when both sections (SL_features & NOT_features) run multiple times, the results are different. Results cannot determine which is more accurate between SL_features and NON_features. The reason I figure out is that both of them have ways to improve performance and accuracy if possible.

Program Processing:

1. Open and Read original file “clothing_shoes_jewelry.txt”

```
import nltk  
  
path = "/Users/boli/Desktop/nlp/HW3/clothing_shoes_jewelry.txt"  
  
rfile = open(path,"r")  
contents = rfile.read()  
rfile.close()  
print(len(contents))  
  
139641164
```

2. Split contents in “clothing_shoes_jewelry.txt” into lines using ‘splitlines()’

```
lines = contents.splitlines()  
print(len(lines))  
print(lines[:10])  
  
2786318  
['reviewerID:A1KLRMWW2FWPL4', 'asin:0000031887', 'reviewerName:Amazon Customer "cameramom"', 'helpful:[0, 0]', "revie  
wText:This is a great tutu and at a really great price. It doesn't look cheap at all. I'm so glad I looked on Amazon  
and found such an affordable tutu that isn't made poorly. A+++', 'overall:5.0', 'summary:Great tutu- not cheaply mad  
e', 'unixReviewTime:1297468800', 'reviewTime:02 12, 2011', '']
```

3. Extract lines only including “reviewText” to retrieve all review texts.

```
# Remove title "reviewText."  
reviewText = [r[11:] for r in lines if "reviewText" in r] # r[11:] is contents after 'reviewText':  
print(len(reviewText))  
print(reviewText[:3])
```

```
278677  
["This is a great tutu and at a really great price. It doesn't look cheap at all. I'm so glad I looked on Amazon and  
found such an affordable tutu that isn't made poorly. A+++', 'I bought this for my 4 yr old daughter for dance class,  
she wore it today for the first time and the teacher thought it was adorable. I bought this to go with a light blue l  
ong sleeve leotard and was happy the colors matched up great. Price was very good too since some of these go for over  
$15.00 dollars.', 'What can I say... my daughters have it in orange, black, white and pink and I am thinking to buy f  
or they the fuccia one. It is a very good way for exalt a dancer outfit: great colors, comfortable, looks great, easy  
to wear, durables and little girls love it. I think it is a great buy for costumer and play too.']
```

4. Convert list to string using “join”

```
str_join = ",".join(reviewText)  
  
print(len(str_join))  
print(str_join[:300])
```

```
88129771  
This is a great tutu and at a really great price. It doesn't look cheap at all. I'm so glad I looked on Amazon and fo  
und such an affordable tutu that isn't made poorly. A++,I bought this for my 4 yr old daughter for dance class, she w  
ore it today for the first time and the teacher thought it was ado
```

5. Make sentence tokenization to string using “nltk.sent_tokenize()”

```
# sentences tokenize  
textssplit = nltk.sent_tokenize(str_join)  
  
print(len(textssplit))  
print(textssplit[:20])  
  
907576  
['This is a great tutu and at a really great price.', "It doesn't look cheap at all.", "I'm so glad I looked on Amazo  
n and found such an affordable tutu that isn't made poorly.", 'A++,I bought this for my 4 yr old daughter for dance c  
lass, she wore it today for the first time and the teacher thought it was adorable.', 'I bought this to go with a lig  
ht blue long sleeve leotard and was happy the colors matched up great.', 'Price was very good too since some of these  
go for over $15.00 dollars.,What can I say... my daughters have it in orange, black, white and pink and I am thinking
```

6. Defined Negation Words list for Not_ features later

```
##### Set Up Features Starts #####
```

```
# Negation Words
# this list of negation words includes some "approximate negotiators" like hardly and rarely
negationwords = ['no', 'not', 'never', 'none', 'nowhere', 'nothing', 'noone', 'rather', 'hardly', 'scarcely', 'rarely',
'ain', 'aren', 'couldn', 'didn', 'doesn', 'hadn', 'hasn', 'haven', 'isn', 'ma', 'mightn', 'mustn', 'nee
```

7. Implement sentence_polarity as dataset to train model

```
# sentence_polarity Corpus
from nltk.corpus import sentence_polarity
import random
documents = [(sent, cat) for cat in sentence_polarity.categories()
             for sent in sentence_polarity.sents(categories=cat)]
documents[0]
```

```
('simplicity', ',', 'silly', 'and', 'tedious', '.', 'neg')
```

8. Shuffle Documents retrieved from sentence_polarity

```
# Shuffle Documents
random.shuffle(documents)
documents[0]
```

```
('been',
'there',
',
'done',
'that',
')'
```

9. Generate words list from document.

```
# All Words List
all_words_list = [word for (sent,cat) in documents for word in sent]
print(len(all_words_list))
print(all_words_list[:10])
```

```
224073
['been', 'there', ',', 'done', 'that', ',', 'liked', 'it', 'much', 'better']
```

10. Since generated words list, we need to start to clean it up. Filter words list by isalpha()

```
# Alphabets
alpha_words_list = [w for w in all_words_list if w.isalpha()]
print(len(alpha_words_list))
print(alpha_words_list[:10])
```

```
187486
['been', 'there', 'done', 'that', 'liked', 'it', 'much', 'better', 'the', 'first']
```

Starting SL Features

11. Eliminate stopwords in words list

```
# StopWords
stopwords = nltk.corpus.stopwords.words('english')
stop_words_list2 = [word for word in alpha_words_list if word not in stopwords]
len(stop_words_list2)
```

```
105085
```

12. Get most 2000 frequency on words list, and then create word features.

```
# Most Frequency Word with 2000
all_words2 = nltk.FreqDist(stop_words_list2)
word_items2 = all_words2.most_common(2000)
word_features2 = [word for (word,count) in word_items2]
word_features2[:200]

['film',
 'movie',
 'one',
 'like',
 'story',
```

13. Define Subjectivity using a sentiment lexicon with score or counts

```
SLpath = 'subjclueslen1-HLTEMNLP05.tff'
def readSubjectivity(path):
    flexicon = open(path, 'r')
    # initialize an empty dictionary
    sldict = {}
    for line in flexicon:
        fields = line.split() # default is to split on whitespace
        # split each field on the '=' and keep the second part as the value
        strength = fields[0].split("=')[1]
        word = fields[2].split("=')[1]
        posTag = fields[3].split("=')[1]
        stemmed = fields[4].split("=')[1]
        polarity = fields[5].split("=')[1]
        if (stemmed == 'y'):
            isStemmed = True
        else:
            isStemmed = False
        # put a dictionary entry with the word as the keyword
        # and a list of the other values
        sldict[word] = [strength, posTag, isStemmed, polarity]
    return sldict
SL = readSubjectivity(SLpath)
```

14. Define a function SL_features as model to be trained by word_features2 later on

```
def SL_features(document, word_features, SL):
    document_words = set(document)
    features = {}
    for word in word_features:
        features['contains({})'.format(word)] = (word in document_words)
    # count variables for the 4 classes of subjectivity
    weakPos = 0
    strongPos = 0
    weakNeg = 0
    strongNeg = 0
    for word in document_words:
        if word in SL:
            strength, posTag, isStemmed, polarity = SL[word]
            if strength == 'weaksubj' and polarity == 'positive':
                weakPos += 1
            if strength == 'strongsubj' and polarity == 'positive':
                strongPos += 1
            if strength == 'weaksubj' and polarity == 'negative':
                weakNeg += 1
            if strength == 'strongsubj' and polarity == 'negative':
                strongNeg += 1
        features['positivecount'] = weakPos + (2 * strongPos)
        features['negativecount'] = weakNeg + (2 * strongNeg)
    return features
```

15. Invoke SL_features to generate SL_feature sets

```
SL_featuresets = [(SL_features(d, word_features2, SL), c) for (d, c) in documents]
```

16. Split SL_feature sets by 1000. The Accuracy shows blow

```
train_set2, test_set2 = SL_featuresets[1000:], SL_featuresets[:1000]
classifier2 = nltk.NaiveBayesClassifier.train(train_set2)
nltk.classify.accuracy(classifier2, test_set2)
```

0.7386981804539486

17. Display most informative features using classifier

```
classifier2.show_most_informative_features(30)
```

```
Most Informative Features
    contains(dull) = True          neg : pos   = 20.1 : 1.0
    contains(problem) = True       neg : pos   = 16.8 : 1.0
    contains(boring) = True        neg : pos   = 16.8 : 1.0
    contains(jokes) = True         neg : pos   = 14.2 : 1.0
contains(beautiful) = True      pos : neg   = 13.8 : 1.0
    contains(warm) = True         pos : neg   = 13.2 : 1.0
    contains(flat) = True         neg : pos   = 12.8 : 1.0
contains(creates) = True         pos : neg   = 11.1 : 1.0
    contains(gentle) = True       pos : neg   = 11.1 : 1.0
    positivecount = 10            pos : neg   = 10.7 : 1.0
contains(captures) = True       pos : neg   = 10.5 : 1.0
contains(examination) = True    pos : neg   = 10.5 : 1.0
    contains(waste) = True        neg : pos   = 10.2 : 1.0
    contains(mess) = True         neg : pos   = 10.1 : 1.0
contains(inventive) = True      pos : neg   = 9.8 : 1.0
contains(unexpected) = True     pos : neg   = 9.8 : 1.0
    contains(tired) = True        neg : pos   = 9.5 : 1.0
    contains(pacing) = True       neg : pos   = 9.5 : 1.0
contains(touching) = True      pos : neg   = 9.1 : 1.0
contains(gorgeous) = True       pos : neg   = 9.1 : 1.0
    contains(fails) = True        neg : pos   = 8.9 : 1.0
contains(enjoyable) = True      pos : neg   = 8.8 : 1.0
contains(relationships) = True pos : neg   = 8.4 : 1.0
    contains(weak) = True         neg : pos   = 8.2 : 1.0
    ----,----,----
contains(supposed) = True       neg : pos   = 8.2 : 1.0
contains(delicate) = True       pos : neg   = 7.8 : 1.0
    contains(cinema) = True       pos : neg   = 7.8 : 1.0
    contains(bad) = True          neg : pos   = 7.1 : 1.0
contains(realistic) = True      pos : neg   = 7.1 : 1.0
contains(political) = True      pos : neg   = 7.1 : 1.0
```

18. Since the model SL_features has been trained, using it to analyze all reviews, separate them into each list (Positive & Negative).

```
# RESULT
pos2 = []
neg2 = []

for text in textsplit:
    texttokens = nltk.word_tokenize(text)
    inputfeatureset = SL_features(texttokens, word_features2, SL)

    if classifier2.classify(inputfeatureset) == 'pos':
        pos2.append(text)
    elif classifier2.classify(inputfeatureset) == 'neg':
        neg2.append(text)

print(pos2[10:15])
print(neg2[10:15])

["It isn't a full&#34; or high quality skirt, but it is perfect for my daughter to wear over leggings for her little outfits., My daughter liked this, and it with her costume, but she would have liked it to be a bit fuller., For what I paid for two tutus is unbeatable anywhere!", 'I ordered a pink and turquios and they are vibrant and beautiful!', 'Princess style!', 'I paid less than 7 bucks for a tutu I and I feel proud of myself for researching to the point of finding gold! Recommend 2-6 years! My daughter is two!', 'Wears size 4t and this skirt ( one size ) fit perfectly and will probably be able to accommodate her quickly growing waist for some time!, Wonder my niece wears it every single day, yellow is her favorite color right now and this cute little tutu made her da.']
["Thanks for a great product.NEVER BUY FROM DRESS UP DREAMS.....I will buy more as long as I don't buy from &#34;Dress Up Dreams&#34;; I never rec'd or order in FL.", "Only rec'd pink, the purple one was missing.", 'Company is a rip off.', "REFUSES to make good on purchase..... Real creeps., I received this today and I'm not a fan of it but my daughter is. I thought it would be sufficient as it looks in the pic but it's not and the one they sent me is pink underneath
```

19. Create a table to store sample analyzed reviews

```
# Table
import pandas as pd
import os

col_name = ["Pos", "Neg"]
df = pd.DataFrame(columns = col_name)

df["Pos"] = pos2[:100]
df["Neg"] = neg2[:100]

csv_file = "/Users/boli/Desktop/nlp/HW3/SL_table.csv"

if not os.path.isfile(csv_file):
    df.to_csv(csv_file, header=True, index=False, encoding = 'utf-8')

print(df)
```

Pos \

0 This is a great tutu and at a really great price.
1 A++, I bought this for my 4 yr old daughter for...
2 I bought this to go with a light blue long sle...
3 It is a very good way for exalt a dancer outfi...
4 The girls have been wearing them regularly, in...
.. ...
95 Do you think that will work?

20. Save each list (Positive & Negative) into different files.

```
# pos into file
posFile = open('/Users/boli/Desktop/nlp/HW3/SL_feature_pos.txt', 'w')
for r in pos2:
    posFile.write(r + '\n')
posFile.close()

# neg into file
negFile = open('/Users/boli/Desktop/nlp/HW3/SL_feature_neg.txt', 'w')
for r in neg2:
    negFile.write(r + '\n')
negFile.close()
```

Starting NOT Features

21. Create new stop words excluding negation words, clear words list up using new stop words list

```
# StopWords
stopwords = nltk.corpus.stopwords.words('english')
newstopwords = [word for word in stopwords if word not in negationwords]
stop_words_list = [word for word in alpha_words_list if word not in newstopwords]
len(stop_words_list)
```

106321

22. Get most 2000 frequency on words list, and then create word features.

```
# Most Frequency Word with 2000
all_words = nltk.FreqDist(stop_words_list)
word_items = all_words.most_common(2000)
word_features = [word for (word, count) in word_items]
word_features[:200]
```

['film',
 'movie',
 'not',
 'one',
 'like',
 'story',
 'no',
 'much',

23. Define a function NOT_features as model to be trained by word_features later on

```
# Using Not Feature
def NOT_features(document, word_features, negationwords):
    features = {}
    for word in word_features:
        features['contains({})'.format(word)] = False
        features['contains(NOT{})'.format(word)] = False
    # go through document words in order
    for i in range(0, len(document)):
        word = document[i]
        if ((i + 1) < len(document)) and ((word in negationwords) or (word.endswith("n't"))):
            i += 1
            features['contains(NOT{})'.format(document[i])] = (document[i] in word_features)
        else:
            features['contains({})'.format(word)] = (word in word_features)
    return features
```

24. Invoke NOT_features to generate NOT_feature sets

```
# Not Feature Sets
NOT_featuresets = [(NOT_features(d, word_features, negationwords), c) for (d, c) in documents]
NOT_featuresets[0][0]['contains(NOTlike)']
```

False

25. Split NOT feature sets by 1000. The Accuracy shows blow

```
# Accuracy "Two different sets of features"
train_set, test_set = NOT_featuresets[1000:], NOT_featuresets[:1000]
classifier = nltk.NaiveBayesClassifier.train(train_set)
nltk.classify.accuracy(classifier, test_set)
```

0.7968936409679234

26. Display most informative features using classifier

```
# Most informative features
classifier.show_most_informative_features(30)
```

Most Informative Features

contains(engrossing) = True	pos : neg	=	19.5 : 1.0
contains(waste) = True	neg : pos	=	15.8 : 1.0
contains(loud) = True	neg : pos	=	15.8 : 1.0
contains(inventive) = True	pos : neg	=	15.6 : 1.0
contains(mediocre) = True	neg : pos	=	15.1 : 1.0
contains(refreshing) = True	pos : neg	=	14.2 : 1.0
contains(boring) = True	neg : pos	=	14.2 : 1.0
contains(flat) = True	neg : pos	=	13.9 : 1.0
contains(routine) = True	neg : pos	=	13.8 : 1.0
contains(unique) = True	pos : neg	=	13.6 : 1.0
contains(90) = False	neg : pos	=	12.4 : 1.0
contains(dull) = True	neg : pos	=	12.2 : 1.0
contains(wonderful) = True	pos : neg	=	12.1 : 1.0
contains(warm) = True	pos : neg	=	11.7 : 1.0
contains(NOTenough) = True	neg : pos	=	11.1 : 1.0
contains(refreshingly) = True	pos : neg	=	10.9 : 1.0
contains(humanity) = True	pos : neg	=	10.9 : 1.0
contains(provides) = True	pos : neg	=	10.9 : 1.0
contains(stupid) = True	neg : pos	=	10.7 : 1.0
contains(touching) = True	pos : neg	=	10.6 : 1.0
contains(mindless) = True	neg : pos	=	10.4 : 1.0
contains(stale) = True	neg : pos	=	10.4 : 1.0
contains(extraordinary) = True	pos : neg	=	10.3 : 1.0
contains(ages) = True	pos : neg	=	10.3 : 1.0
contains(captures) = True	pos : neg	=	10.1 : 1.0
contains(suffers) = True	neg : pos	=	9.9 : 1.0
contains(save) = True	neg : pos	=	9.7 : 1.0
contains(bears) = True	neg : pos	=	9.7 : 1.0
contains(tender) = True	pos : neg	=	9.6 : 1.0
contains(realistic) = True	pos : neg	=	9.6 : 1.0

27. Since the model NOT_features has been trained, using it to analyze all reviews, separate them into each list (Positive & Negative).

```
# RESULT
pos = []
neg = []

for text in textsplit:
    texttokens = nltk.word_tokenize(text)
    inputfeatureset = NOT_features(texttokens, word_features, negationwords)

    if classifier.classify(inputfeatureset) == 'pos':
        pos.append(text)
    elif classifier.classify(inputfeatureset) == 'neg':
        neg.append(text)

print(pos[10:15])
print(neg[10:15])

['The quality is just fine for the price we paid.', 'I ordered a pink and turquios and they are vibrant and beautifu l!', 'The tutu is very full!', 'Princess style!', 'I paid less than 7 bucks for a tutu I and I feel proud of my self for researching to the point of finding gold!Recommend 2-6 years!My daughter is two !']
['Company is a rip off.', "REFUSES to make good on purchase..... Real creeps.,I received this today and I'm not a fan of it but my daughter is I thought it would be puffier as it looks in the pic but it's not and the one they sent me is pink underneath and the waist band is pink which is not what I wanted due to the fact she already had the sandals she was gonna wear with it now I gotta find another pair of sandals,ima just keep it cuz she likes it.,Bought this as a backup to the regular ballet outfit my daughter has to wear.", 'I was not expecting a designer skirt for this price and got exactly what I paid for.,Great tutu for a great price.', "It isn't a 'full'; or high quality skirt, but it is perfect for my daughter to wear over leggings for her little outfits.,My daughter liked this, and it with her costume, but she would have liked it to be a bit fuller.,For what I paid for two tutus is unbeatable anywhere!", 'Not cheaply made!']
```

28. Create a table to store sample analyzed reviews

```
# Table
import pandas as pd
import os

col_name = ["Pos", "Neg"]
df = pd.DataFrame(columns = col_name)

df["Pos"] = pos[:100]
df["Neg"] = neg[:100]

csv_file = "/Users/boli/Desktop/nlp/HW3/table.csv"

if not os.path.isfile(csv_file):
    df.to_csv(csv_file, header=True, index=False, encoding = 'utf-8')

print(df)
```

29. Save each list (Positive & Negative) into different files.

```
# pos into file
posFile = open('/Users/boli/Desktop/nlp/HW3/pos.txt', 'w')
for r in pos:
    posFile.write(r + '\n')
posFile.close()

# neg into file
negFile = open('/Users/boli/Desktop/nlp/HW3/neg.txt', 'w')
for r in neg:
    negFile.write(r + '\n')
negFile.close()
```

Cross-Validation

30. Using Cross-Validation to obtain Precision, Recall and F measure score on both Positive & Negative

```
reflist = []
testlist = []
for (features, label) in test_set:
    reflist.append(label)
    testlist.append(classifier2.classify(features))

reflist[:30]
testlist[:30]

ref1 = set([i for i,label in enumerate(reflist) if label == 'pos'])
ref2 = set([i for i,label in enumerate(reflist) if label == 'neg'])

test1 = set([i for i,label in enumerate(testlist) if label == 'pos'])
test2 = set([i for i,label in enumerate(testlist) if label == 'neg'])

from nltk.metrics import *

def printmeasures(label, refset, testset):
    print(label, 'precision:', precision(refset, testset))
    print(label, 'recall:', recall(refset, testset))
    print(label, 'F-measure:', f_measure(refset, testset))

printmeasures('pos', ref1, test1)
print()
printmeasures('neg', ref2, test2)

pos precision: 0.7334887334887334
pos recall: 0.7039522744220731
pos F-measure: 0.7184170471841705

neg precision: 0.712005803409503
neg recall: 0.7410343525858815
neg F-measure: 0.7262301146873843
```

Table NOT_Features (Positive & Negative first 100 samples):

Pos	Neg
2 This is a great tutu and at a really great price.	I'm so glad I looked on Amazon and found such an affordable tutu that isn't made poorly.
3 It doesn't look cheap at all.	Price was very good too since some of these go for \$15.00 dollars. What can I say... my daughters have it in orange, black, white and pink and I am thinking to buy for they the fucia one.
4 I just bought this for my 4 yr old daughter for dance class, she wore it today for the first time and the teacher thought it was adorable.	Sorry about that.
5 I bought this to go with a light blue long sleeve leotard and was happy the colors matched up great.	The girls have been wearing them regularly, including out to play, and the tutus have stood up well.
6 It is a very good way for a dancer costume outfit, colors, comfortable, looks great, easy to wear, durables and little girls love it.	Clearly plenty of room to grow.
7 Only on sale when the kids pull off the tutus, the waste band gets twisted, and an adult has to re-tangle.	But this is not difficult. Thank you Habs Heaven great product for Little Girls.
8 Fits like a 3-5 yr old & the 5-6 yr old well.	My Grand Daughters Love these Tutu's.
9 My daughter loves it!	Will buy more from this seller.
10 She loves it!	Thanks for your review NEVER BUY FROM DRESS UP DREAMS..... I will buy more as long as I don't buy from 6834.Dress Up Dreamdutu4. I never rec'd or order in FL.
11 Go, go, who's going to stop her! her Cinderella dreams but I am sure we'll be able to use it for a recital sometime soon.	One day I will, the people one was missing.
12 The quality is just fine for the price we paid.	Company is a rip off.
13 I ordered a pink and turquoise and they are vibrant and beautiful!	REFUSED to make good on purchase.... Real creeps. I received this today and I'm not a fan of it but my daughter is I thought it would be puffer as it looks in the pic but it's not and the one they sent me is pink underneath.
14 The tutu is very full!	I was not expecting a designer skirt for this price and got exactly what I paid for. Great tutu for a great price.
15 Princess style!	It isn't cheap tutu's or high quality skirt, but it is perfect for my daughter to wear over leggings for her little outfitts.
16 I paid less than I looks for a tutu and I feel proud of myself for researching to the point of finding gold!Recommend 2-6 yearstly daughter is two !	My daughter liked this, but she would have liked it to be a bit fuller. For what I paid
17 I paid less than I looks for a tutu and I feel proud of myself for researching to the point of finding gold!Recommend 2-6 yearstly daughter is two !	It's a bit cheaply made but I still like it.
18 It's a pretty outfit!	Obviously someone made these with love and care!
19 I've always been a beauty with many pieces in every color.	It is well built and we hope the piece lasts us out of it. My daughter has worn this skirt almost everyday since she received it and it's even been through the washer along with the other clothes.
20 This tutu is a beautiful purple color that looks just like the picture.	She fits it & it's fit & it's above her knee, and has a little bit of growing room, although I'm not so sure as much as others are saying.
21 It looks just adorable on our little fairy. This was a gift for a two year old and a five year old.	But considering how often she wears it, I'm not worried.
22 The tutu seems to be well made, and we stand up to handling by little princesses.	J.Purchased this tutu for my grandmother's first birthday.
23 I recommend these tutu. My 3yr old loved this tutu in pink!	It is nicely with room to grow.
24 She loves it!	This girl loves it and she wears it all the time, they grabbed the tutu, put them on, and wore them the rest of the day.
25 Its great and easy to clean!	Was hoping to find one in different colors. She had tried this one before this stitching came apart in 2 weeks now it's lying in her closet. Altogether she wore it like 4-5 times for 20 mins or so with the stitching was of better qual
26 I bought this for her for Christmas and she never wanted to take it off.	Can't recommend. Perfect for the grand daughter birthday.
27 It is almost a year later and she still has it!	I bought several more colors. Never GOT this item... but gave a 1 STAR because the replies from the SUPPLIER was GREAT! They tried to send the item more than once. My \$ was refunded in a timely manner too. It was a shal
28 Great color and fit for a 2 year old as well as her aunt who is 30!	I would recommend this for girls under 10 yr old.
29 The elastic waist will expand with her as she grows..The minute I saw this my heart skipped a beat.	It will be too short and small for older girls, bought this for my niece for a party of her fairy outfit.
30 It is soft!	She loves it and wears it at the beach.
31 It is suitable for an adult dress.	Well made for the price. My 3 year old daughter wants to wear this dress up.
32 Day Care Jewelry Box - we put together inside nicely... Love the pink & looks so nice on my vanity. This product is great for anyone with a lot of jewelry my girlfriend has a lot and this gift for her is perfect.	We'll keep it for sure.
33 My granddaughter likes the Shining Image Jewelry Box given to her for her birthday gift from Grandmas.	I would recommend. Our granddaughters are all very gifts, so when the youngest one received this for Christmas, they all wanted it!
34 I, on the other hand, have had lots of practice with jewlery drawers, but also struggle with removing the drawer without spilling the contents. Really nice box, a bit cheap made.	I would recommend this tutu for all little girls.. Prompt delivery, and it is exactly as described.
35 Every inch is useable space for jewelry & very compact, which is a nice feature.	The only reason I did not give it 5 stars is because I haven't washed it yet - so I don't know how it will hold up... Other than that my little girl LOVES her tutus (we got one in light pink also), especially spinning and running in them.
36 While a class environment might work better for some, there is not always time for that either.	What a nice size & so nice collection.
37 The software comes with a headset and microphone and that while you are learning, you also speak into the headset and respond to questions.	I love all the compartments and the layout too.
38 I highly recommend this product. It is far better than the one you guys are selling. I am appreacitive.	If this is a gift for a child, it is a great gift.
39 I own both an iPhone and an iPad and the Rosetta Stone app lets me directly to my account. I could learn and practice whenever I needed to.	It will help even the most fussy ones as I am.
40 If want to continue learning this way I need to purchase a very expensive subscription.	The price is fair if you check around and did.
41 I am very disappointed with Rosetta Stone and Amazon.	Seems well made too.
42 I want to continue to learn on my iPad!	Not meant as a toy.
43 Three months access included with purchase.*	This is for serious jewelry lover's.
44 * There are learning sessions in the software to allow speaking with native French speakers.	I recommended this to any one who is a bit of a jewelry freak about this particular Shining Image jewelry box in brown, so this was my second buy.
45 * You can earn points for the software and can earn them at a formal classroom. Enter "The Simulation in the number of seats".	It's one of the millions of Americans that massively struggles with any language other than English.
46 * Lack of matching supplemental physical material such as index cards with conjugations and grammatical rules.	Obviously someone made these with love and care!
47 * One simple ritual to help with identifying grammatical patterns. Chay, I admit it.	It is well built and we hope the piece lasts us out of it. My daughter has worn this skirt almost everyday since she received it and it's even been through the washer along with the other clothes.
48 I'm one of the millions of Americans that massively struggles with any language other than English.	Can't recommend. Perfect for the grand daughter birthday.
49 The University I'm enrolled in requires two years of a foreign language.	I bought several more colors. Never GOT this item... but gave a 1 STAR because the replies from the SUPPLIER was GREAT! They tried to send the item more than once. My \$ was refunded in a timely manner too. It was a shal
50 I am calling that a huge success.	I would recommend this for girls under 10 yr old.
51 The Rosetta Stone "dynamic immersion" method is different than traditional methods where you see a word in English and then see a word in the new language to memorize.	I am calling that a huge success.
52 You will feel confident to speak after this. La Feminne Bell....	It is made of cardboard with plastic.
53 I think it is a great product.	Although the box is not very pretty, I'm not so sure how it will stand up to my daughter's use.
54 They really did not have a job making learning fun and easy.	That was nice to know if you are able to learn the software, before investing even more money into it.
55 It was the wearing syndrome holds over my ear...J	Since total immersion in another country to learn is not always an option, Rosetta Stone is the next best thing.
56 I got the french lesson, I don't want to have to install the application and then a language pack.	Sometimes there's just enough time on the weekends, or in the evening after work to get a few hours of studying in.
57 The software should be smart enough to install both the app and language pack together to make the process even easier.	Or maybe you just don't want to sit in a school classroom, trying to learn from an instructor.
58 Learn as quickly, important phrases, practice some everyday: use it, use it, use it - hear it say it and if you can, practice with a native speaker- you really learn that way, and you get over the nervousness of it.	That's where Rosetta Stone comes in.Rosetta Stone is to be more than just a computer program.
59 Even so, the course of I was able to access were very enjoyable and enhanced by French skills.	Instead it's interactive, and a cross between software and an in-home classroom Spanish lesson.
60 Version 4 comes with the learning lessons, headset with mic, app to use with your iPhone or iPod Touch for the go learning and access to a virtual community online for sessions, games and chat	It is software based, so it is not as good as a book.
61 I certainly would have bought it for each of my children throughout their schooling.	Version 4 comes with the learning lessons, headset with mic, app to use with your iPhone or iPod Touch for the go learning and access to a virtual community online for sessions, games and chat
62 It also teaches you the four areas of language - speaking, listening, reading and writing.	Version 4 comes with the learning lessons, headset with mic, app to use with your iPhone or iPod Touch for the go learning and access to a virtual community online for sessions, games and chat
63 Bilingual. What do you call a person who speaks one language?	Rosetta Stone lets a person learn at the pace they feel at first, so don't be afraid to miss anything along the way.
64 I considered the other romance languages (Spanish, Portuguese, Italian,	I'm more comfortable learning from the program as well, than in front of a group of people. While going to France and living there for a while is probably the easiest and fastest way to learn French, Rosetta Stone is the ne
65 Some languages are dying out in Germany.	It makes it easy to learn French, and a cross between software and an in-home classroom Spanish lesson.
66 I am learning to speak French in these languages, no one is question whether I really know it.	It is software based, so it is not as good as a book.
67 Neither do I if you really need to learn a new language, and I am a fact, Rosetta Stone is the way.	Version 4 comes with the learning lessons, headset with mic, app to use with your iPhone or iPod Touch for the go learning and access to a virtual community online for sessions, games and chat
68 Very little of the repeat when she stuff from normal classes.	Version 4 comes with the learning lessons, headset with mic, app to use with your iPhone or iPod Touch for the go learning and access to a virtual community online for sessions, games and chat
69 Actual fun lessons.	It is software based, so it is not as good as a book.
70 They have the basics covered.Oh, what I don't like - the complexity of the install with all the verifications and such.	Although the box is not very pretty, I'm not so sure how it will stand up to my daughter's use.
71 This means you learn more effectively, as your brain is forced to make certain connections on its own.	That was nice to know if you are able to learn the software, before investing even more money into it.
72 I thought Rosetta Stone did a great job bringing up concepts and reminding me of the concepts later on.	Since total immersion in another country to learn is not always an option, Rosetta Stone is the next best thing.
73 I am learning to speak French in these languages, no one is question whether I really know it.	Sometimes there's just enough time on the weekends, or in the evening after work to get a few hours of studying in.
74 I am learning to speak French in these languages, no one is question whether I really know it.	Or maybe you just don't want to sit in a school classroom, trying to learn from an instructor.
75 I am learning to speak French in these languages, no one is question whether I really know it.	That's where Rosetta Stone comes in.Rosetta Stone is to be more than just a computer program.
76 I am learning to speak French in these languages, no one is question whether I really know it.	Instead it's interactive, and a cross between software and an in-home classroom Spanish lesson.
77 I am learning to speak French in these languages, no one is question whether I really know it.	It is software based, so it is not as good as a book.
78 I spoke some dialect, but conjugation issues!	Version 4 comes with the learning lessons, headset with mic, app to use with your iPhone or iPod Touch for the go learning and access to a virtual community online for sessions, games and chat
79 The way the teaching of the language is set up, is, I believe, the closest to learning it naturally with native speakers.	Version 4 comes with the learning lessons, headset with mic, app to use with your iPhone or iPod Touch for the go learning and access to a virtual community online for sessions, games and chat
80 I highly recommend this software...gratit,I've always found learning a new language to be a huge challenge.	It is software based, so it is not as good as a book.
81 It's a very comfortable way to learn the language, and I am learning to speak French in these languages, no one is question whether I really know it.	Version 4 comes with the learning lessons, headset with mic, app to use with your iPhone or iPod Touch for the go learning and access to a virtual community online for sessions, games and chat
82 I am learning to speak French in these languages, no one is question whether I really know it.	It is software based, so it is not as good as a book.
83 It's a nice thing about becoming fluent in Italian or looking for more than just the ability to "say a", "new" language, then you'll want to get the advanced versions (unfortunately, at a much higher cost).It is a	Version 4 comes with the learning lessons, headset with mic, app to use with your iPhone or iPod Touch for the go learning and access to a virtual community online for sessions, games and chat
84 The simple approach to learning a language as a child does was one terrific idea! agree with other posters that to become fluent you will need to purchase more than just this level 1 program.	It is software based, so it is not as good as a book.
85 I'm enjoying it immensely! love Rosetta Stone is to easy to use in a few quick lessons you are speaking Italian.	Version 4 comes with the learning lessons, headset with mic, app to use with your iPhone or iPod Touch for the go learning and access to a virtual community online for sessions, games and chat
86 I like how the lessons are short and easy to finish! like the pictures I can't wait to go to Italy!Install Mediawire Intel Mar.2 0.6 GHz Intel Core 2 Duo, 4 GB DDR3 RAM, OS 10.6.8Intel was smooth and easy.	It is software based, so it is not as good as a book.
87 Total install time with level 1 language pack was maybe 10 minutes or less (there is a lot of data copied to the PC).Launching the software was straightforward and launching into the first lesson was easy.	Version 4 comes with the learning lessons, headset with mic, app to use with your iPhone or iPod Touch for the go learning and access to a virtual community online for sessions, games and chat
88 This is a better way to learn to make the association of the foreign word to the object instead of translating it in your head through your native tongue first.Maybe being level 1 is a easier on me, but I found it is	It is software based, so it is not as good as a book.
89 However, how much would it cost to take a foreign language class at a local college?	Version 4 comes with the learning lessons, headset with mic, app to use with your iPhone or iPod Touch for the go learning and access to a virtual community online for sessions, games and chat
90 Great for those who don't have time to take an instructor led course. Rosetta Stone Italiano Level 1 Rosetta Stone Italiano Level 1 is easy to install and use.	It is software based, so it is not as good as a book.
91 The included instructions are simple and straightforward.	Version 4 comes with the learning lessons, headset with mic, app to use with your iPhone or iPod Touch for the go learning and access to a virtual community online for sessions, games and chat
92 Rosetta Stone relies on photographs to help you out, which is quite useful but I could not get a good grasp of the language to actually use it in normal speech.	It is software based, so it is not as good as a book.
93 Unfortunately, like other users, the phrases are too simple for use in everyday conversation. The Level 1 it includes a USB headset with a microphone to allow you to speak the phrases using the built-in speech.	Version 4 comes with the learning lessons, headset with mic, app to use with your iPhone or iPod Touch for the go learning and access to a virtual community online for sessions, games and chat
94 This helps with pronunciation, although I found that I will dismiss my main audio and I have to switch back and forth.	It is software based, so it is not as good as a book.
95 This isn't a bad deal for most people, and there's no easy workaround. The box advertises that many Government agencies use Rosetta Stone, like the US Army and State Department.	Version 4 comes with the learning lessons, headset with mic, app to use with your iPhone or iPod Touch for the go learning and access to a virtual community online for sessions, games and chat
96 Then vocabulary, spoken and written Italian is better than what I have learned in school.	It is software based, so it is not as good as a book.
97 So your learning is reinforced in these ways: you hear the phrase spoken, you see the phrase written, and you see a photo illustrating each phrase (a boy running, boys running).	Version 4 comes with the learning lessons, headset with mic, app to use with your iPhone or iPod Touch for the go learning and access to a virtual community online for sessions, games and chat
98 In later lessons you dive into conversations with full sentences. Your Main Menu shows you which lessons you've completed.	It is software based, so it is not as good as a book.
99	
100	
101	

Table SL_Features (Positive & Negative first 100 samples):

SL_habla	Habla
1	It's a great tutu and at a really great price.
2	A+ I bought this for my 4 yr old daughter for dance class, she wore it today for the first time and the teacher thought it was adorable.
3	I bought this to go with a light blue long sleeve leotard and was happy the colors matched up great.
4	It is a very good way for exalt a dancer outfit great colors, comfortable, looks great, easy to wear, durables and little girls love it.
5	The girls have been wearing them regularly, including out to play, and the tutus have stood up well.
6	Meets well and cuts on the girls.
7	So far, she's using it to play her Cinderella dreams but I am sure we'll be able to use it for a recital sometime soon.
8	The quality is just for the price we paid.
9	I was not expecting a designer skirt for this price and got exactly what I paid for.Great tutu for a great price.
10	It isn't a #454fullskirt or high quality skirt, but it is perfect for my daughter to wear over leggings for her little outfit..My daughter liked this, and it with her costume, but she would have liked it to be a bit fuller.
11	Ordered a pink and tan one and they are vibrant and beautiful!
12	Princess style!
13	I just love this tutu! I ordered a pink one and it is perfect and will probably be able to accommodate her quickly growing waist for some time!Wonder my niece wears it every single day, yellow is her favorite color right now!
14	Meets well and cuts on the girls.
15	It looks great and I think it's a great value for the price.
16	Love this tutu! It looks great and fits perfectly. I am so happy with it!
17	It's an amazing quality!
18	I purchased this tutu for my granddaughter's first birthday.
19	It fit nicely with room to grow.
20	I very pleased with it and will purchase more in varying colors in the future.Full and well stitched.
21	This tutu is a beautiful purple color that looks just like the picture.
22	It is a great tutu. My 3yr old loves this tutu skirt in pink!
23	Doesn't recommend.
24	Beautiful vibrant colors.
25	Fits great and easy to clean!
26	Had several more items.Never GOT this item - but gave a 1 STAR because the replies from the SUPPLIER was GREAT.They tried to send the item more than once.My \$ was refunded in a timely manner too.
27	It is almost a year later and she still has it!
28	Great color and fits her well as well as her waist is 30!
29	The elastic waistband expand with her as she grows..The minute I saw this my heart skipped a beat.
30	She loves it and I am sure she will continue to love it.
31	She will love this tutu too.
32	The price is fair if you check around as I did.
33	It is suitable for an adult though.
34	Buy it...Love this Jewelry Box, so well put together holds plenty... Love the peek & looks so nice on my vanity,This products is great for anyone with a lot of jewelry my girlfriend has a lot and this gift for her will be a hit.
35	It arrived with some of the leather scratches,Even though I liked it was the first time opening the box.
36	My granddaughter likes Shining Image Jewelry Box given to her for a birthday gift from Grandmas.
37	It arrived the day before Christmas and she loves it.
38	It is a great gift for anyone who loves jewelry.
39	My granddaughter is five years old ,as having the background knowledge of knowing when to stop the action of pulling the drawer from the base requires practice.
40	On the other hand, he has lots of years of practice with peeling drawers, but also struggle with removing the drawer without spilling the contents..Really nice box, a bit cheaply made.
41	Although the organization for all my jewelry is great, I'm not so sure how it will stand up to my heaven daughter's use.
42	Every inch is useable space for jewelry & very compact, which is a nice feature..Since learning a language from software is not exactly the same as learning a language from going to live in the country itself, I recommend this item.
43	That way you know if you are able to learn from the software, before investing even more money into it.
44	While a classroom environment might work better for some, there's not always time for that either.
45	The software helps you to practice the words and pronunciations and you can learn to help retain it.
46	Under a classroom environment, you can work at this at your own pace, which I appreciate.
47	I'm more comfortable learning from the program as well, than in front of a group of strangers. While going to France and living there for a while is probably the easiest and fastest way to learn French, Rosetta Stone is the best.
48	So when I heard that there is "An App for that", I was thrilled.
49	I own both an iPhone and an iPad so I've used the Rosetta Stone apps that link directly to my account I could learn and practice whenever I needed to.
50	For three months I made great progress whether at home, in the backpack or at work or my lunch hour.
51	If I want to continue learning this way I need to purchase a yearly subscription.
52	I've used it on my iPhone and my iPad.
53	I ended up putting the iPad on my laptop, especially to use the program in other locations.
54	Installation was easy enough though the program did stuck a couple of times along the way.
55	The voice recognition software is impressive.
56	There are plenty of exercises and repetitions, and this will surely help in many ways including pronunciation.
57	There are learning sessions in this software to allow speaking with native French speakers.
58	I feel the need for this and can source it out myself, but again I prefer to learn outside a formal classroom.
59	Lack of simple visual to help with identifying grammatical patterns...Okay, I admit it.
60	The one of the millions of Americans that measure struggles with any language other than English.
61	The software is great and requires few years of a foreign language.
62	I'm calling that a huge success.
63	The Rosetta Stone "dynamic immersion" method is different than traditional methods where you see a word in English and then see a word in the new language to memorize.
64	I'd say it's more like du immersion et raccourcissement amusant et facile.I've never had french, learned french, or know anything about French.
65	I guess we're "ready" interested, it's worth it.For it's purpose, it is almost perfect software.
66	With months of use I made great progress whether at home, in the backpack or at work or my lunch hour.
67	Under a classroom environment, you can work at this at your own pace, which I appreciate.
68	This was my first experience with Rosetta Stone software and I like it.
69	They really do a heck of a job making learning fun and easy.
70	I would definitely try other languages after seeing just how easy it is to use this stuff.What didn't I like?For starters the headphones and microphone that came with it are NOT quality, they are not comfortable at all.
71	The microphone seemed to work well, my only gripe here is the comfort of wearing it.
72	I would have liked a faster, easier program to install.
73	I got the french lesson, I don't want to have to install the application and then a language pack.
74	The software needs to be updated to include both the app and language pack together to make the process even easier.And yes, I am a software developer so when I review software I look at it from a completely objective point of view.
75	Learn the quick, important phrases, practice everyday use. If, use it, hear it, say it and if you can, practice with a native speaker you really learn that way, and you get over the nervousness of it.
76	It offers activities that have you practice a combination of skills reading and writing, as well as lessons that focus on just one of the skills.
77	I am excited to try French Level 1 and learn even more!AHHH! She asked her if she had tried any other language learning products that she could compare it to and she said she had not, but she was very happy.
78	The friends I made had a quite they liked to use: "What do you call a person who speaks three languages?"
79	Bilingual: What do you call a person who speaks one language?
80	Americans I have wanted to improve my language status for many years.
81	My High School Latin teacher said that I learned Latin, the other romance languages would be easy.
82	After a few months using this for four years studying Latin made French grammar easier.
83	Soon I remembered my years living in Germany.
84	Those who do not speak English (maybe one percent of those I encountered) used a different dialect, and I was back to using my German book and pointing to words.
85	The lesson for me is that now I must spend some time in France if I really want to learn the language.
86	Oh well, I can read French as long as people do not get technical.Recently I thought of a possible answer to the dilemma I have with my European friends.
87	Then I will tell them I am "quadri linguis".
88	Do you think that will work?
89	Neither do I.If you ready to learn a new language, and learn it fast, Rosetta Stone is the way.
90	But by far, the easiest way to a new language - and outside of taking one as a kid in public school - the cheapest way, too, is Rosetta Stone.No, I don't work for them or even get to date that spokeswoman in the ad.
91	But I do love learning languages and I ready the RS.
92	And actual fun lessons.
93	It doesn't look cheap at all.
94	I'm glad I looked on Amazon and found such an affordable tutu that isn't made poorly.
95	Price was very good too since some of these go for over \$15.00 dollars.What can I say... my daughters have it in orange, black, white and pink and I am thinking to buy for they the fucca one.
96	I think it is a great buy for costume and play too, I bought several tutus at once, and they are high reviews.
97	Sturdy and seemingly well-made.
98	Only cost is that when she puts pull off the tutu, the waist band gets twisted, and an adult has to untangle.
99	But this is not difficult. Thank you Haha Heaven great product for Little Girls.
100	My Great Grand Daughters Love these Tutu's.
101	Will buy more from this seller.
102	Thanks for a great product NEVER BUY FROM DRESS UP DREAMS.....I will buy more as long as I don't buy from #834.Dress Up Dreams#834; I never rec'd or order in FL.
103	Only rec'd pink, the purple one was missing.
104	Company is a rip off.
105	Highly recommend good on purchases... Real creeps.I received this today and I'm not a fan of it but my daughter is I thought it would be puffer as it looks in the pic but it's not and the one they sent me is pink underneath.
106	The tutu is a bit tight.
107	Not cheeply made!
108	Not cheap material.
109	Obviously someone made these with love and care.
110	It is well built and she loves it lots of wear of it..My daughter has worn this skirt almost every day since she received it and it's even through the washer along with the other clothes.
111	She fits it 4T and it's just above her knees, and has a little bit of room, although I'm not so sure as much as others are saying.
112	But consider how often we wash it. I'm not.
113	It looks just like the picture we saw.
114	It's a pink tutu.
115	It's a pink tutu.
116	It's a pink tutu.
117	It's a pink tutu.
118	It's a pink tutu.
119	It's a pink tutu.
120	It's a pink tutu.
121	It's a pink tutu.
122	It's a pink tutu.
123	It's a pink tutu.
124	It's a pink tutu.
125	It's a pink tutu.
126	It's a pink tutu.
127	It's a pink tutu.
128	It's a pink tutu.
129	It's a pink tutu.
130	It's a pink tutu.
131	It's a pink tutu.
132	It's a pink tutu.
133	It's a pink tutu.
134	It's a pink tutu.
135	It's a pink tutu.
136	It's a pink tutu.
137	It's a pink tutu.
138	It's a pink tutu.
139	It's a pink tutu.
140	It's a pink tutu.
141	It's a pink tutu.
142	It's a pink tutu.
143	It's a pink tutu.
144	It's a pink tutu.
145	It's a pink tutu.
146	It's a pink tutu.
147	It's a pink tutu.
148	It's a pink tutu.
149	It's a pink tutu.
150	It's a pink tutu.
151	It's a pink tutu.
152	It's a pink tutu.
153	It's a pink tutu.
154	It's a pink tutu.
155	It's a pink tutu.
156	It's a pink tutu.
157	It's a pink tutu.
158	It's a pink tutu.
159	It's a pink tutu.
160	It's a pink tutu.
161	It's a pink tutu.
162	It's a pink tutu.
163	It's a pink tutu.
164	It's a pink tutu.
165	It's a pink tutu.
166	It's a pink tutu.
167	It's a pink tutu.
168	It's a pink tutu.
169	It's a pink tutu.
170	It's a pink tutu.
171	It's a pink tutu.
172	It's a pink tutu.
173	It's a pink tutu.
174	It's a pink tutu.
175	It's a pink tutu.
176	It's a pink tutu.
177	It's a pink tutu.
178	It's a pink tutu.
179	It's a pink tutu.
180	It's a pink tutu.
181	It's a pink tutu.
182	It's a pink tutu.
183	It's a pink tutu.
184	It's a pink tutu.
185	It's a pink tutu.
186	It's a pink tutu.
187	It's a pink tutu.
188	It's a pink tutu.
189	It's a pink tutu.
190	It's a pink tutu.
191	It's a pink tutu.
192	It's a pink tutu.
193	It's a pink tutu.
194	It's a pink tutu.
195	It's a pink tutu.
196	It's a pink tutu.
197	It's a pink tutu.
198	It's a pink tutu.
199	It's a pink tutu.
200	It's a pink tutu.

Ways to Improve:

- In this project, the accuracy of NOT_feature is only around 79% using NOT_features and clean up most of necessary tokens/stopwords. However, it still cannot be perfectly analyzed sentences accurately. Here're some ways could improve the accuracy:
 1. **Add more negation words into list.** Because if negation list is large enough, the trained model could analyze sentences efficiently.
 2. **Using other word_features.** Word_features is extremely important in this task. A good word_features could make analyzing more precisely. Thus, it's good idea to use some other corpus instead of sentence_polarity to make comparing.
 3. **The number of Train Set & Test Set.** In this task, train set and test set are spitted by 1000, this could more accurate. But there might be some other proportions between train set and test set which could be better choice.
- In the SL_feature, the accuracy is only around 76.2%. In this case, for cleaning up words list, it also used stopwords and isalpha(). However, the method using stopwords is a bit different as NOT_feature's. It just cleans up regular stop words as cleaning. Here are some ways to improve the performance:
 1. **The number of Train Set & Test Set.** Set up different amount of train set and test set to see which one is most accurate.
 2. **Change the number of the most frequency unigram.** It could be changed smaller amount of number as the most frequency. The original used 2000 as the most frequency unigrams, but if change it to be smaller, it'll be probably more accurate.

Summary:

In this task, most of thing are exactly same as Lab 10. Most techniques are provided to train the model. The results could show two lists with Positive & Negative. As later process, it's better to find out natural sentences, which is more precise to analyze sentiment.

SKLearn Program (Bonus Credit):

In this project, it used jieba as a feature and SKLearn as classifier. It analyzed labelled data (Positive & Negative) using SKLearn. NLTK plays a role to retrieve word_features, and some feature options. Whereas, SKLearn is mainly used to sort algorithms, get some statistics by given data.

Program Processing

1. Make jieba as a feature

```
import jieba
|
def read(file):
    stop_words = [line.strip() for line in open('stopword.txt', 'r', encoding='utf-8').readlines()]
    f = open(file, 'r')
    line = f.readline()
    result = []
    while line:
        s = line.split('\t')
        fenci = jieba.cut(s[0], HMM=True)
        result.append(list(set(fenci)-set(stop_words)))
        line = f.readline()
    return result
```

2. In the jieba_feature, the parameter “number” is the feature of the most data retrieved (chi-squared statistics)

```
from nltk.probability import FreqDist, ConditionalFreqDist
from nltk.metrics import BigramAssocMeasures

def jieba_feature(number):
    posWords = []
    negWords = []
    for items in read('/Users/boli/Desktop/nlp/HW3/pos.txt'):
        for item in items:
            posWords.append(item)
    for items in read('/Users/boli/Desktop/nlp/HW3/neg.txt'):
        for item in items:
            negWords.append(item)

    word_fd = FreqDist()
    cond_word_fd = ConditionalFreqDist()

    for word in posWords:
        word_fd[word] += 1
        cond_word_fd['pos'][word] += 1

    for word in negWords:
        word_fd[word] += 1
        cond_word_fd['neg'][word] += 1

    pos_word_count = cond_word_fd['pos'].N()
    neg_word_count = cond_word_fd['neg'].N()
    total_word_count = pos_word_count + neg_word_count

    word_scores = {}

    for word, freq in word_fd.items():#word_fd={'word':count}
        pos_score = BigramAssocMeasures.chi_sq(cond_word_fd['pos'][word], (freq, pos_word_count), total_word_count)
        neg_score = BigramAssocMeasures.chi_sq(cond_word_fd['neg'][word], (freq, neg_word_count), total_word_count)
        word_scores[word] = pos_score + neg_score

    best_vals = sorted(word_scores.items(), key=lambda item:item[1], reverse=True)[:number]
    best_words = set([w for w,s in best_vals])

    return dict([(word, True) for word in best_words])
```

3. Adjust the Setting, Generate build features based on jieba feature.

```
def build_features():
    feature = jieba_feature(300)

    posFeatures = []
    for items in read('/Users/boli/Desktop/nlp/HW3/pos.txt'):
        a = {}
        for item in items:
            if item in feature.keys():
                a[item] = 'True'
        posWords = [a, 'pos']
        posFeatures.append(posWords)

    negFeatures = []
    for items in read('/Users/boli/Desktop/nlp/HW3/neg.txt'):
        a = {}
        for item in items:
            if item in feature.keys():
                a[item] = 'True'
        negWords = [a, 'neg']
        negFeatures.append(negWords)

    return posFeatures, negFeatures
```

4. Analyze the trained data.

```
posFeatures, negFeatures = build_features()

from random import shuffle
import sklearn
from nltk.classify.scikitlearn import SklearnClassifier
from sklearn.svm import SVC, LinearSVC, NuSVC
from sklearn.naive_bayes import MultinomialNB, BernoulliNB
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score

shuffle(posFeatures)
shuffle(negFeatures)

train = posFeatures[300:] + negFeatures[300:]
test = posFeatures[:300] + negFeatures[:300]

data, tag = zip(*test)

def score(classifier):
    classifier = SklearnClassifier(classifier)
    classifier.train(train)
    pred = classifier.classify_many(data)
    n = 0
    s = len(pred)
    for i in range(0, s):
        if pred[i] == tag[i]:
            n = n+1
    return n/s |
```

5. Show the result of BernoulliNB using SklearnClassifier

```
print('BernoulliNB` s accuracy is %f' %score(BernoulliNB()))

BernoulliNB` s accuracy is 0.761667
```

6. Show the result of MultinomialNB using SklearnClassifier

```
print('MultinomialNB` s accuracy is %f' %score(MultinomialNB()))

MultinomialNB` s accuracy is 0.730000
```

7. Show the result of LogisticRegression(solver='lbfgs') using SklearnClassifier

```
print('LogisticRegression` s accuracy is %f' %score(LogisticRegression(solver='lbfgs')))

LogisticRegression` s accuracy is 0.768333
```

8. Show the result of LinearSVC using SklearnClassifier

```
print('LinearSVC` s accuracy is %f' %score(LinearSVC()))  
LinearSVC` s accuracy is 0.750000
```

Summary:

In this Task, we mainly focus on jieba feature and SKLearn classifier. And based on the data give (Positive & Negative) retrieved from reviews, we analyze several fields of statistics s.t BernoulliNB, MultinomialNBusing, LogisticRegression, and LinearSVC.