Q1: Human Ratings

a. Collection

I gathered the following comments from three of my classmates.

Reviewer A = "My iPhone is still OK to maintain my daily life, but if there are conditions, I will keep changing to the latest iPhone because I pay more attention to freshness than daily phone use."

Reviewer B = "I really like the iOS system on the iPhone, mainly because it is very smooth, with a long time and no lag; sliding the phone to switch the interface interaction effect is very smooth."

Reviewer C = "I hate iPhones. It has many restrictions on installing applications, which can only be downloaded from the Appstore. Some apps not used in the AppStore cannot be downloaded and used directly, which is very inconvenient."

I got three other people to rate the above comments. The rating criteria was positive: 1, negative: -1, and neutral: 0.

Rater A = [-1,1,0]

Rater B = [-1,1,-1]

Rater C = [-1,1,1]

We can see that all three agree with the first comment and the second comment. The three disagree on the third comment, giving three different results.

b. Calculation

1. Data Matrix

```
              Rater_A   Rater_B   Rater_C
Reviewer_A       -1         1         0
Reviewer_B       -1         1        -1
Reviewer_C       -1         1         1
```

2. Cohen's Kappa Score

Score(Rater_A, Rater_B) = 0

Score(Rater_A, Rater_C) = 0.5

Score(Rater_B, Rater_C) = -0.5

Apparently, due to the third comment, all three ratings are categorized to different degrees. The degree of agreement between Rater_A and Rater_C is higher. Between Rater_A and Rater_B, the agreement is the same as expected by chance. The agreement between Rater_ B and Rater_C is lower because Rater_B and Rater_C hold opposing views on the third comment.

3. Pearson Score

Score(Rater_A, Rater_B) = 0.8660254

Score(Rater_A, Rater_C) = 0.8660254

Score(Rater_B, Rater_C) = 0.5

I obtained a strong correlation between Rater_A and Rater_C, consistent with Cohen's Kappa Score values above. However, there is a strong correlation between Rater_A and Rater_B. There is also a moderate correlation between Rater_B and Rater_C. This is different from the above calculation. Due to the difference in absolute values between the three values, the different evaluations of the third review, under the influence of the two main ones, make the Pearson Score values present different expectations.

Q2: Sentiment Lists

Three Sentiment Lists were mentioned.

1. Harvard General Inquirer List

2. MPQA Corpus

3. WordNet: Affective Wordnet

I randomly selected five positive and five negative words from the Harvard General Inquirer List and MPQA Corpus. And I try to find the opposite meaning for them.

a. Harvard General Inquirer List

Positive: Excellent, Glad, Cute, Vital, Yearn.

Word	Positive	Negative	Discuss
Excellent	Your performance is excellent.	In order to stay excellent, he repeatedly made mistakes in the game.	Normally, excellent is a very positive word.
Glad	I am glad to see you.	Resenting such tutelage, I am only too glad to avail myself of it.	Glad is a very positive word.
Cute	This cat is cute.	The way you screwed up this thing is so cute.	In the contemporary sense, cute is sometimes used to describe the comicality of people.
Vital	She is vital and attractive.	this book is far from vital.	Vital is a very positive word.
Yearn	He yearned for home.	In the university, you will miss studying when you are	Yearn is a positive word most of the time, but a

		bent on yearning towards pleasure.	negative one when it means addiction.

Negative: Invalid, Envy, Bitterness, Worry, Harm.

Word	Positive	Negative	Discuss
Invalid	He looks after his invalid brother.	His invalid father is becoming a burden.	Invalid is a strong negative word.
Envy	He envied your trip to Greece.	She was devoured by envy and hatred.	Sometimes envy means "want" or "like" in a positive sense.
Bitterness	If you have tasted the bitterness of gall, you know better the sweetness of honey.	She was full of hatred and bitterness.	Bitterness is a really negative word.
Worry	I found my dog contentedly worrying a bone.	She is always worrying about little things.	Worrying means positive when it means satisfaction.
Harm	There is no harm in trying.	Indulgence in smoking can seriously harm your health.	Harm is a very negative word.

b. MPQA Corpus

Positive: Esteem, Convincing, Wish, Great, Hero.

Word	Postive	Negative	Discuss
Esteem	The people esteemed the president for his unselfishness.	Then you must hold God in very low esteem.	Esteem is a really positive word.
Convincing	We hear she was quite convincing.	The message was far from convincing.	Convincing is a very strong positive word.
Wish	He wishes to journey round the world.	Anyone who wishes to may leave now.	Most of the time wish means positive. In certain contexts wish may mean negative.
Great	A great future is reserved for you.	But it eroded a great deal yesterday.	Great should be like a neutral word. Except for the examples, like a great loss, a good view.
Hero	He was acclaimed hero of the country.	No man is a hero to his valet.	Typically, hero is a very positive word.

Negative: Defile, Brutal, Betrayer, Unlucky, Veto.

Word	Positive	Negative	Disscuss
Defile	They that touch pitch will be defiled.	She felt defiled by the filth.	Defile is a strong negative word.

Brutal	She has to face the brutal reality.	His answer was brutal, cruel and inhumane.	Brutal is a strong negative word.
Betrayer	It's easy to see betrayer in a sentence.	The deepest circle of hell is reserved for betrayers and mutineers.	Betrayer is a totally negative word.
Unlucky	Unlucky, you'll get it in next time.	I was unlucky enough to lose my glasses.	Unlucky is a strong negative word.
Veto	I've used up all my vetoes.	My father vetoed my plan.	Veto is a strong negative word.

In general, ten randomly selected positive and ten negative words were classified more accurately in the two corresponding dictionaries. However, a few words may need to be correctly classified due to non-daily usage. Yearn is a positive word most of the time, but a negative one when it means addiction. Sometimes envy means "want" or "like" in a positive sense. Worry means positive when it means satisfaction. Moreover, Cute is often used to describe cuteness, sometimes with an implied layer of comic relief in its modern meaning. Great should be classified as neutral when describing the number of gaps in positive or negative usage. Since a few words are not common types in everyday use, the overall word classification in both dictionaries makes sense.

Q3: A Program

a. Remove stop-words

Based on the 75 percent training set, the latter was removed stop-words. Most informative features remained the same in both tests. After removing stop-words, the accuracy of the test decreases in terms of decline, and the values of positive precision and negative recall decrease. In terms of improvement, positive recall and negative precision were improved. Since there are quite a few removed stop-words in the data, these are easily classified correctly, causing the former to be more correct than the latter. After removing stop-words, fewer correct data are used for prediction, causing a decrease in the number of correct predictions in the prediction results, resulting in a boost in positive recall and negative precision and a decrease in accuracy.

b. 85% training data

```
Using all words as features                                  Remove stop words
train on 9064 instances, test on 1600 instances             train on 9064 instances, test on 1600 instances
accuracy:0.7812                                              accuracy:0.7719
pos precision:0.7945                                         pos precision:0.7715
pos recall:0.7588                                            pos recall:0.7725
neg precision:0.7691                                         neg precision:0.7722
neg recall:0.8037                                            neg recall:0.7712
Most Informative Features                                    Most Informative Features
         engrossing = True      pos : neg   =     20.3 : 1.0          engrossing = True      pos : neg   =     20.3 : 1.0
              flaws = True      pos : neg   =     14.3 : 1.0               flaws = True      pos : neg   =     14.3 : 1.0
            generic = True      neg : pos   =     14.3 : 1.0             generic = True      neg : pos   =     14.3 : 1.0
           mediocre = True      neg : pos   =     13.7 : 1.0            mediocre = True      neg : pos   =     13.7 : 1.0
         refreshing = True      pos : neg   =     13.7 : 1.0          refreshing = True      pos : neg   =     13.7 : 1.0
          absorbing = True      pos : neg   =     13.0 : 1.0           absorbing = True      pos : neg   =     13.0 : 1.0
          inventive = True      pos : neg   =     13.0 : 1.0           inventive = True      pos : neg   =     13.0 : 1.0
             boring = True      neg : pos   =     12.7 : 1.0              boring = True      neg : pos   =     12.7 : 1.0
               flat = True      neg : pos   =     12.6 : 1.0                flat = True      neg : pos   =     12.6 : 1.0
            routine = True      neg : pos   =     12.3 : 1.0             routine = True      neg : pos   =     12.3 : 1.0
```

After boosting the proportion of training data, the accuracy rate improved in both tests. The overall trend is the same as the one tested above. There is also a new appearance in Most Informative Features, where the last three new singles replace the last three words in the above test.