

Movie Review Sentiment Analysis

Sentiment Analysis is a Big Data problem which seeks to determine the general attitude of a writer given some text they have written. For instance, we would like to have a program that could look at the text “The film was a breath of fresh air” and realize that it was a positive statement while “It made me want to poke out my eye balls” is negative.

One algorithm that we can use for this is to assign a numeric value to any given word based on how positive or negative that word is and then score the statement based on the values of the words. But, how do we come up with our word scores in the first place?

That’s part of the problem that we’ll solve in this assignment. You are going to search through a file containing movie reviews from the Rotten Tomatoes website which have both a numeric score as well as text. You’ll use this to learn which words are positive and which are negative. The data file looks like this:

```
1 A series of escapades demonstrating the adage that what is good for the go
4 This quiet , introspective and entertaining independent is worth seeking .
1 Even fans of Ismail Merchant 's work , I suspect , would have a hard time s
3 A positively thrilling combination of ethnography and all the intrigue , be
1 Aggressive self-glorification and a manipulative whitewash .
4 A comedy-drama of nearly epic proportions rooted in a sincere performance b
1 Narratively , Trouble Every Day is a plodding mess .
3 The Importance of Being Earnest , so thick with wit it plays like a reading
1 But it does n't leave you with much .
1 You could hate it for the same reason .
1 There 's little to recommend Snow Dogs , unless one considers cliched dialo
1 Kung Pow is Oedekerk 's realization of his childhood dream to be in a marti
4 The performances are an absolute joy .
3 Fresnadillo has something serious to say about the ways in which extravagar
3 I still like Moonlight Mile , better judgment be damned .
3 A welcome relief from baseball movies that try too hard to be mythic , this
3 a bilingual charmer , just like the woman who inspired it
2 Like a less dizzily gorgeous companion to Mr. Wong 's In the Mood for Love
1 As inept as big-screen remakes of The Avengers and The Wild Wild West .
2 It 's everything you 'd expect -- but nothing more .
```

Note that each review starts with a number 0 through 4 with the following meaning:

- 0 : negative
- 1 : somewhat negative
- 2 : neutral
- 3 : somewhat positive
- 4 : positive

For each word in a new review, you will search every movie review (**movieReviews.txt**) for that word. If you find it, add the score for that review to the word’s running score total. You also will need to keep track of how many appearances the word made so that you can report the average score of reviews containing that word back to the user.

The program will require that you

1. Read in a review
2. Assign each word in the review the score attributed to the review

3. Enter a `WordEntry` object (consisting of the word, total score, and number of occurrences) into a hash table. if word already exists in the hash table, update the score and number of occurrences to the record
4. Repeat Step 1 until all data is entered

Some code in C++ is provided for you. Specifically, the `main.cpp` program that does this is provided for you. For your information, specification of the aforementioned `HashTable` and `WordEntry`, i.e. `HashTable.cpp` and `WordEntry.cpp` as well as their `.h` files, are also provided. By the way if you wonder, yes, **you'll write program in Java** and not C++ ☺

Note that you do NOT have to follow these C++ specifications. They are provided for your convenience only.

Your **Java** program should prompt the user to input a movie review, and automatically score the review based on the average score of the words in the review.

Example output:

```
enter a review -- Press return to exit:
A weak script that ends with a quick and boring finale
The review has an average value of 1.79128
Negative Sentiment
```

```
enter a review -- Press return to exit:
Loved every minute of it
The review has an average value of 2.39219
Positive Sentiment
```

Note in this case, the positive-negative cutoff is at 2.0. Anything less than 2.0 is negative sentiment.