



FEATURE BASED OPINION MINING AND RANKING

Magdalini Eirinaki, Shamita Pisal, Japinder Singh

Computer Engineering Department, San Jose State University,
One Washington Square, San Jose, CA 95192, United States



ABOUT THE PROJECT

- The project aims to make use of the highly unstructured text reviews about any product, extract the potential features about the product, assign scores to them and then classify the reviews as either positive or negative.
- The features of the product play a crucial role in the decision making process of the potential customer. It is these features that distinguish one product from other similar products from different brands. Most of the companies focus on a specific feature as their selling point.

Example Dataset – Canon Powershot G3

```
1  ### excellent picture quality / color
2  i recently purchased the canon powershot g3 and am extremely satisfied with the purchase . the camera is very easy to use , in fact on a recent trip this
   past week i was asked to take a picture of a vacationing elderly group . after i took their picture with their camera , they offered to take a picture of
   us . i just told them , press halfway , wait for the box to turn green and press the rest of the way . they fired away and the picture turned out quite
   nicely . ( as all of my pictures have thusfar ) . a few of my work constituents owned the g2 and highly recommended the canon for picture quality . i 'm
   easily enlarging pictures to 8 1/2 x 11 with no visable loss in picture quality and not even using the best possible setting as yet ( super fine ) .
   ensure you get a larger flash , 128 or 256 , some are selling with the larger flash , 32mb will do in a pinch but you 'll quickly want a larger flash
   card as with any of the 4mp cameras . bottom line , well made camera , easy to use , very flexible and powerful features to include the ability to use
   external flash and lense / filters choices . i 'd highly recommend this camera for anyone who is looking for excellent quality pictures and a combination
   of ease of use and the flexibility to get advanced with many options to adjust if you like . great job canon !
3  ### cool toy
4  yep . this is my first digital camera , and what a ' toy ' it is ! i am a software engineer and am very keen into technical details of everything i buy ,
   i spend around 3 months before buying the digital camera ; and i must say , g3 worth every single cent i spent on it . just a little overview , powershot
   g3 is the flagship of canon 's powershot series and its an slr-like camera , its 4 megapixel and ( alsmost ) full manual control gives the pictures a
   touch of brilliance . whether you are a novice or an expert , its ease of use and functionality goes together . + you can have different kind of lens if
   you want + flashes , etc. as its 4mp , you might need bigger storage to store high quality images and recording movies ( you can record 3 minutes of
   video ) . i am using kingston 512mb cf which works great and is very fast . a good choice could be cf type ii , microdrives that can store around 1 gb of
   images . some things that i did n't like ( but hey . nothing is perfect ) . * main dial is not backlit . * lens visible in optical viewfinder . ( well
   thats only for old-school die-hard optical viewfinder fans ) use lcd instead which is brilliant and you can twist around too . overall i 'm happy with my
   toy .
5  ### canon g3 ? incredible camera
6  i did extensive research comparing different 4 and 5 megapixel cameras . a couple of online camera review sites were extremely helpful , as they provided
   complete exhaustive reviews of nearly every camera made , plus provided a great number of high resolution sample images to compare amongst the models .
   following my research , i decided on the g3 . in a word , " awesome " is how i would describe this camera . i wo n't go into the details here , as the
   online sites ( stevesdigicams , dpreviews ) go into great depth in their reviews , but i am very pleased with the features , the speed , the picture
   quality , the flexibility , the automation , and the functional conveniences this camera offers . with the automatic settings , i really have n't taken a
   bad picture yet . and with the panoramic " stitch " mode , it guides you through stitching together multiple pictures to build a seamless panoramic image
   . i am absolutely in awe of this camera , canon put enormous quality into the g3 . do n't walk , run and buy this camera if you have the budget , i
   guarantee you will not be disappointed .
7  ### great camera
8  i bought my canon g3 about a month ago and i have to say i am very satisfied . i have taken hundreds of photos with it and i continue to be amazed by
   their quality . the g3 is loaded with many useful features , and unlike many smaller digital cameras , it is easy to hold steady when using slower
   shutter speeds . flaws ? the lens is visible in the viewfinder when the lens is set to the wide angle , but since i use the lcd most of the time , this
   is not really much of a bother to me . still i am a little supriised that canon did not correct this design flaw before releasing the camera . despite
   this minor disappointment , i highly recommend the canon g3 anyone who is serious about digital photography .
9  ### have n't had practice but i 'm already in love with it
10 i have only had this camera for one full day and i have to say that it is wonderful . the photo quality is amazing and i know i 'm going to have fun with
   all the features . i 've figured out most of its features on my own already ( because i have used digital cameras before , so i know things like symbols
   ) . i 'm one of those people that can only concentrate on one screen at a time so i 've been having trouble paying attention to the little display panel
   on top , but that does n't stop me from loving this camera . the reason i rated it a four is because of that darn diopter adjustment dial. its very small
   and hard to turn so you can 't get an accurate adjustment ( for those of you who do n't know what a diopter adjustment is , it is to adjust the focus of
```





Output using the algorithm for Canon G3



Topics in Data Mining project

SEARCH

☐ Search The Web



Features of interest:

functionality	4.0
range	4.0
battery life	3.6
design	3.4
detail	3.4
zoom	2.8
size	-0.37
lens cover	-0.57

Reviews:

Great camera

What else can you say about a camera that works for both of you .I make photographs at work and so wanted a camera good enough to compare with what i use professionally .At the same time i wanted my wife to not be intimidated by knobs and buttons .I recieved the camera inserted a larger compact flash_card charged the battery and handed it to my wife .I showed her how to turn it on where the lens zoom lever is and she loves it .This camera has canon 's great colorimetry plus what you see in the lcd is what you get .The prints are beautiful and you get about 120 images on a 256mb card at highest quality .I tried out some other brands in the stores and was disappointed by the battery_life of the other company plus what you see in the lcd (no optical finder) is n't what you get not even for color the output was less than i expected .Although canon 's batteries are proprietary they last a really long time recharge fairly quickly in the camera plus if you want more power you can even find a knockoff charger and spare batteries right here on amazon .

Upgraded from g2 and not looking back

The powershot g3 is a great camera .Why because it can help a so-so photographer take spectacular photos .4 megapixels is plenty for all but pros and you will not be disappointed with the results of this camera .It 's slightly lighter than the g2 and packed with even more features .make sure you get a big compactflash_card (it comes with a 32mb) the 512mb cards are pretty [inexpensive] here on amazon.com and can hold about 500 photos at a time .

Ok for a digital camera

Maybe it is my lack of experience but i found shots with this camera very disappointing .It was (i sold it after a few months) too expensive for the shots that i got .Sure it had all the features but when i tried to shoot a girl 's basketball game it just wa n't up to the task .I do n't think i was asking too much .Any film zoom p and s would have given superior results with a minimal effort and with the 500 you save you could print the results for the life of the camera .I found that low light situations combined with any sort of action left this camera in the dust .I wanted it to shoot concerts as well that meant even less light and just as

Opinion Mining and Ranking Algorithms

- Basically, two algorithms are used sequentially for our purpose:
 - The 1st algorithm ("HAC algorithm") identifies and extracts the potential features from the reviews of the product.
 - The 2nd algorithm ("MOS algorithm") takes these potential features as input, assigns scores to them and finally helps in classifying every review as either positive or negative.

ASSUMPTION OF THE HAC ALGORITHM

- The HAC algorithm extracts the good potential features only when the reviews in the dataset follows the assumption that:
 - Users comment on the features they care about using adjectives.
 - The more often a noun is in the vicinity of an adjective, the more likely it is to be a representative feature of the item reviewed.
 - Example:
 - If the review is "*This camera has very good image quality*", then the adjective 'good' is more likely to describe the noun "*image quality*" than the noun "*camera*".

1. The High Adjective Count Algorithm (HAC)

- Instead of using frequency of keywords, the algorithm starts identifying adjectives and nouns in every review.
- For every review, if each adjective is associated with a noun to which it is closest, this adjective is more likely to describe the noun.
- If an adjective is found to be describing a noun, then the noun_score is incremented by 1. After processing all the reviews, we will have a score associated with each noun, which we call them as opinion scores.
- So nouns with high score have more adjectives to describe them. Then we can have a threshold and nouns having score more than threshold are considered as potential features.
- The adjectives are even extracted from this step .

2. Max Opinion Score Algorithm (MOS)

- The MOS algorithm takes 3 arguments as inputs:
 - The first argument is the list of adjectives which are used to express opinions (obtained from the 1st algorithm). These adjectives are given scores between $[-4, +4]$ in such a way that a high score indicates a stronger opinion than lower score.
 - Example:
 - The adjective 'excellent' gets a higher positive opinion score than the adjective 'good'.

MOS Algorithm (cont)

- The second argument is the list of inversion words like '**not**' or '**never**' which inverts the opinion orientation, so when these words occur in the left context of opinion words, they change the opinion sense.
 - Example:
 - If the review is "*The image quality is not good*", then image quality must be given a negative score. So, these inversion words are kept track of.
- The third argument is the list of potential features obtained by using HAC algorithm.

MOS Algorithm (cont)

- For each sentence in the review, we look at the opinion words and identify features closest to it. The score of feature is summation of scores of opinion words associated with it.
- The scores of features are further summed up to calculate score of review. So, for each review, a score is obtained and reviews are classified and ranked based on this score.
- A negative score for the review implies a negative review. The reviews of a particular category (+ve / -ve) are ranked based on their scores.
- For each review, we score the features separately for title and body.

$$Reviewscore = \frac{\alpha \cdot Titlescore + Bodyscore}{\alpha + 1} \quad (\alpha \text{ is the hyper parameter})$$

Output using the algorithms for Canon G3



Topics in Data Mining project

SEARCH

Search The Web

Features of interest:

functionality	4.0
range	4.0
battery life	3.6
design	3.4
detail	3.4
zoom	2.8
size	-0.37
lens cover	-0.57

Reviews:

Great camera

What else can you say about a camera that works for both of you .I make photographs at work and so wanted a camera good enough to compare with what i use professionally .At the same time i wanted my wife to not be intimidated by knobs and buttons .I recieved the camera inserted a larger compact flash_card charged the battery and handed it to my wife .I showed her how to turn it on where the lens zoom lever is and she loves it .This camera has canon 's great colorimetry plus what you see in the lcd is what you get .The prints are beautiful and you get about 120 images on a 256mb card at highest quality .I tried out some other brands in the stores and was disappointed by the battery_life of the other company plus what you see in the lcd (no optical finder) is n't what you get not even for color the output was less than i expected .Although canon 's batteries are proprietary they last a really long time recharge fairly quickly in the camera plus if you want more power you can even find a knockoff charger and spare batteries right here on amazon .

Upgraded from g2 and not looking back

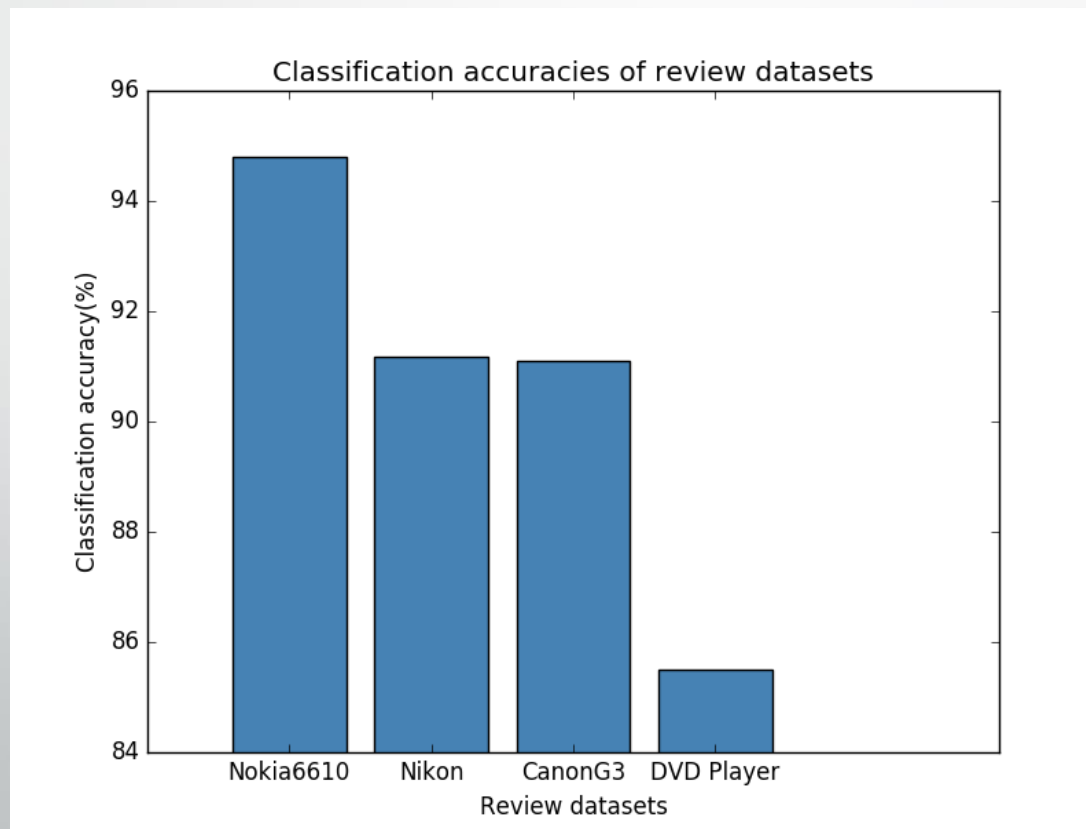
The powershot g3 is a great camera .Why because it can help a so-so photographer take spectacular photos .4 megapixels is plenty for all but pros and you will not be disappointed with the results of this camera .It 's slightly lighter than the g2 and packed with even more features .make sure you get a big compactflash_card (it comes with a 32mb) the 512mb cards are pretty [inexpensive] here on amazon.com and can hold about 500 photos at a time .

Ok for a digital camera

Maybe it is my lack of experience but i found shots with this camera very disappointing .It was (i sold it after a few months) too expensive for the shots that i got .Sure it had all the features but when i tried to shoot a girl 's basketball game it just wa n't up to the task .I do n't think i was asking too much .Any film zoom p and s would have given superior results with a minimal effort and with the 500 you save you could print the results for the life of the camera .I found that low light situations combined with any sort of action left this camera in the dust .I wanted it to shoot concerts as well that meant even less light and just as

REVIEWS CLASSIFICATION ACCURACY

- Using the above algorithms, the classification accuracy for the Canon Powershot G3 dataset was found to be 91.1%.
- Also, the precision for the extracted features was found to be around 41% whereas the recall was around 49%.





FEATURE EXTRACTION USING OTHER METHODS

- The type of algorithm suitable for extracting the potential features of a product can be chosen based on the review dataset.
- If the assumption of the HAC algorithm fails in the dataset, TF / TF-IDF was found to generally perform better in those cases although now-a-days rule-based approaches and semantic dependencies are used for extracting the features from the reviews.

TECHNIQUES FOR EXTRACTING N-GRAM FEATURES

- The procedure to extract the n-gram potential features was not discussed in the paper. This is one of the problems which highly affects the accuracy of the final result.
- Some techniques that were used:
 - Association Rule Mining – The limitation of this method is that it does not extract the unpopular n-gram features.
 - Manual noun phrase extraction – Extract the consecutive noun phrases from the review dataset.
 - TextBlob noun phrase extractor – Use the noun phrase extractor to extract all the potential features from the review dataset. The limitation of this method is that it gives many false positives.

LIMITATION OF MOS ALGORITHM

- The inversion words were used as opinion orientation inverters in the paper which is generally not only the case. This is one of the fundamental limitations of the MOS algorithm.
 - Example:
 - If the review is "*This camera has a major design flaw*", then the camera is expected to have negative opinion score because of the word "flaw". But the NLP libraries fail to find the sentiment for the words like "flaw". So, these negative opinions are not captured in the analysis.

WORK DONE AFTER THE PRESENTATION

- The files (python codes) were well-organized such that running a single file (main.py) iteratively calls all the algorithms and stores the results (for positive reviews, negative reviews and feature scores) onto separate files.
- The algorithms were run on various datasets and a report was created which contains:
 - A graph plot of variation of classification accuracy for various datasets (bar graph).
 - A table giving some information about the classification results.
 - Graphs showing the variation of classification accuracy with hyper-parameter(α) for various datasets.



THANK YOU