Product Reviews – An Emotional Roller-Coaster: A Presentation Of AAMER, A Method For Extracting Aspect-Based Emotions From Online Product Reviews

Amardeep Sarang - 20651461
David R. Cheriton School of Computer Science
University of Waterloo
Waterloo, Canada
assarang@uwaterloo.ca

***Abstract*— In this paper, we present our AAMER (Adaptable Aspect-based Multi Emotion Recognition) method for aspect-based emotion analysis of product reviews. AAMER focuses on addressing some shortcomings of previous aspect extraction and emotion analysis works by being adaptable to new datasets and will require no previous training data. The method can extract multiple emotion values towards a product aspect instead of a single polarity value. We describe, test, and analyze results from 12 different variations of our method while using a new dataset of RC car reviews from amazon. We found that AAMER can produce reasonable and interpretable results which allow a user to quickly gain a sense of what emotions reviewers have felt towards aspects of a given product. We found the best variations of AAMER for doing so was the first emotion analysis method using non-zero mean and using our rule-based heuristics method of aspect extraction.**

Keywords—component, aspect extraction, emotion analysis, text mining of product reviews

I. Introduction

Ecommerce has become a key part of our modern society. In 2020, 82% of Canadians shopped online, spending $84.4 billion [9]. Among these e-retailers, Amazon was the biggest [20]. A key component of shopping on these platforms is product user reviews, which are text-based descriptions of the characteristics of a product and experience one may have when using it, these reviews are written by consumers who have previously bought the product and are meant to help others in their purchase decision [8]. These reviews are an important part of the buying process with 74.04% of survey participants stating that they found online reviews are important or very important and 85.57% saying that they read the reviews often when making a purchase online [8]. By reading the positive and negative reviews shoppers hope to get a sense of the good and bad aspects of the product to figure out if the product is right for them. Looking at star ratings can speed this process up, but normally, websites only provide one rating for the whole product. Thus, making it impossible to find aspect-specific information without reading multiple long reviews which, is an activity that survey participants report is a very time-consuming process [8]. Some websites do provide aspect-specific ratings but there are only a few websites that do this, and they are normally reserved for specific aspects of select product categories [10][8].

To help tackle this problem we present a product aspect-based review emotion analysis method called AAMER (Adaptable Aspect-based Multi Emotion Recognition). AAMER is a rule-based text mining approach to extract customer emotions towards aspects (or features, these two words may be used interchangeably) of a product from the product's reviews on Amazon. By presenting this method we hope to give users of the tool a better overview of how certain features of a product impacted the experience of that product more so than they may get from simply looking at a rating or trying to read many long reviews. For example, instead of simply seeing a mostly positive 4/5 rating, our tool will show users more detail about how other reviewers of a phone were positively surprised with the battery life, but also mildly disgusted and angry that the manufacture no longer includes a charger in the box.

As part of this paper, we will first outline previous works that have tackled problems related to aspect-based emotion analysis, and discuss how some shortcomings of these past works have motivated the goals we set forth for our AAMER method. We give a detailed description of the methods we used to create our AAMER method, and the data we used to test it. We then present the test results of several variations of our method, which we evaluate and compare with some statistical analysis. Finally, we conclude the paper with a discussion about the quality of the results overall and how we may further address some shortcomings.

II. Related works

The works related to the problem can be split into two main categories: aspect extraction and emotion analysis. The first problem of aspect extraction has had many different approaches, some used learning-based approaches such as deep convolutional neural networks or RBF kernels [13][16]. Others use grammar rule-based (also may be referred to as unsupervised) approaches similar to the one we present for our method to extract aspects from the texts [7][21]. These methods are often simpler than those that use learning, and perform worse in some test datasets, but can often be used on a variety of data, which is why we will look to follow a similar approach in our method [21]. These rule-based approaches also have tailored their methods to focus on extracting specific types of aspects, in [7] a focus was on extracting multiword aspects, like menu items with multi-word names from restaurant reviews. While [21] focused on extracting implicit and explicit aspects. Many of these systems were tested on prelabelled datasets such as Semeval and Citysearch datasets [23][2] and could thus directly compare performance with one another. It was found that those using learning approaches did slightly outperform rule-based approaches in these datasets[7][21].

The second problem of sentiment and emotion analysis is the process of extracting emotions or sentiment from text. Standard sentiment analysis found in works such as [19][5][17] extracts a polarity (positive, negative, neutral) values. Whereas emotion analysis provides a set of values that show the presents of several emotions in a text[4][3]. Both of these have been applied to multiple levels of granularity: document level, sentence level, and aspect level, where aspect level is the most used among works looking to gain information from product reviews [10]. The main way most methods extract this sentiment or emotion is using sentiment lexicon-based approaches which contain a set of pre-compiled sentiment terms accompanied by labeled sentiment values. One example of such an approach are works concerning Kansei engineering, which uses pre-existing sets of Kansei words and text to quantify emotion towards a given product[19][3].

III. Goal and motivation for AAMER

Based on the above works, we identified several areas that presented an opportunity for improvement. The first of these areas was the product aspect extraction task. We found that most of the models proposed required some degree of training with some form of labeled data[13][11]. Papers such as [16] and [13] which were trained and tested on the Semeval and Citysearch dataset [23][2] achieved F-scores of 86.4% and 86.2% respectively. However, these methods required large amounts of labeled training data. Also, the training data needs to be specific to the types of products being analyzed which in the case of Semeval and Citysearch dataset was laptops and restaurants. We felt that requiring such vast amounts of labeled training data would severely limit any efficiency one could gain when using an automatic system. At the very least, it would greatly limit the system's ability to be deployed quickly across many product categories as one may find on large e-commerce websites like Amazon where there are many product categories relative to the number of reviews in each category. Also, most of the works have used the same few datasets. This does allow them to be easily compared to each other but further raises questions about how adaptable they are to new data. Even the rule-based unsupervised methods which should, in theory, be adaptable limit their testing to these datasets [7][21]. To counter these shortcomings, our method implements and test several variations of an unsupervised rule base aspect noun extraction system that requires no previous training data, and tests it on an entirely new dataset.

Another area where we plan to improve on other works is emotion analysis. Most works such as [19][5][17] only implement sentiment analysis meaning they simply apply a polarity (positive, negative, neutral) value to a product aspect. We implement emotion analysis, to provide a set of values that show the presents of several emotions towards an aspect/feature of the product and how strong the emotions are. This is similar to other works like [4][3]. The emotions we will detect will be fear, anger, anticipation, trust, surprise, generally positive, generally negative, sadness, disgust, and joy [1][6][15]. As part of our method, we will discuss several ways of extracting these emotions using the NRCLex python library by Mark M. Bailey.

In the following sections, we will present a product aspect-based review emotion analysis method, called AAMER (Adaptable Aspect-based Multi Emotion Recognition), that can work on any new set of review data without needing to be pretrained to extract product aspects, and will also be able to identify multiple emotions towards the said aspect, even for previously unseen review datasets. By doing so we try to address some of the shortcomings discussed.

IV. Data used

To simulate using the system on a new product category we created a new dataset, we used customers' reviews from the 2018 Amazon Review Dataset created by Jianmo Ni for their paper: [14]. This set had review data for many product departments of Amazon but for our study, we used product reviews from the "Toys and games" department. We choose this category for its range of products at various prices. Given the large variety of products, it would be likely that a shopper in this category would have difficulty comparing between different products, especially if a shopper is not familiar with the space, such as new parents, or someone looking to purchase a gift for a younger relative. These factors make this category a good and rigorous testbed for the method. Also, with it being one of the available categories with a smaller number of reviews (8 million reviews as opposed to 20 million reviews in other categories made available by Jianmo Ni) it makes data preprocessing and sampling more feasible on the equipment (laptop) available.

To focus our study on a more homogenous set of products we further filter down the products we would analyze to a specific category like products containing either "RC car", "remote control", or "car" in the title and from this subset, we selected the 30 products that had the most reviews. Our final data set had 30 products with 2050 unique reviews. The number of reviews for each product ranged from 41 to 145 reviews.

V. Method

Our AAMER method has 3 main modules: aspect extraction, emotion analysis, and results visualization. In this section, we will describe the task of each module as well as the various techniques we used to implement each of them.

A. *Aspect extraction*

The first stage in our AAMER system is aspect extraction. This module takes in a list of reviews for each product and then generates a list of words where each word corresponds to an aspect/feature of that product. This is done for each of the 30 products. For this study, we explored 3 methods of extraction:

a simple noun frequency method, a rule-based heuristic method, and a manual method which are described below:

1) *Noun frequency method*

In this method, we first extracted all nouns from the reviews of a given product using the part of speech (POS) tagger from the NLTK python library to identify the nouns. All the words identified as a type of noun by having the POS tags "NN", "NNP", "NNS", or "NNPS" are added to a list and the frequency of each noun is counted. The top N most frequent nouns across all reviews for the product will be selected as aspects. For our method, we kept N=10 most frequent nouns. This method works under the theory that if a certain thing (noun) is mentioned in multiple reviews it may be an important feature of the product.

When we tried this in its most basic form, we did encounter some problems. These included having the same word appear on the top 10 list in several forms such as singular words eg. "car" and plural words eg. "cars", and also having common nouns like "work" appearing in the list. To address these issues we lemmatized the nouns before counting them so that words such as "car" and "cars" would both be counted towards the single aspect word "car". We also filtered out common nouns using the stop word list in NLTK.

We also observed that in many cases when testing this method, the top aspect words would feature multiple words from the title, for example, the top ten aspect list of a product titled "Kettler Kettcar Kabrio Cart, Yellow" had the words "kettcar" and "cart" both in the list. These words were used to refer to the products themselves and were blocking aspect-related words from appearing on the top 10 list. To mitigate this issue we added a step to filter all words in the product title from the noun list before counting them to form the top 10.

The final modification we made to this method is what we will refer to as "consolidation". For these processes, we defined a list of words that would all be replaced in the noun list by a single overarching term. For example, we consolidate the words "product", "car", "toy", "car", "item", and "thing" under the single word "product", meaning that they will all count towards a single frequency count and only "product" will appear on the top 10 list (if the frequency count is high enough). The reason for doing so was similar to filtering the title words, but we felt in these cases the words did refer to valid aspects (or the product as a whole), so we wanted to keep the aspect they referred to but only recognize the words as a single aspect. One example of this was seen when words from the set { 'kid', 'son', 'grandson', 'daughter', 'granddaughter', 'nephew', 'niece', 'child', 'husband','wife'} kept appearing on the product top 10 aspect list. Keeping these as aspects were beneficial for our dataset of toy reviews since it described how the recipient (child) of the toy felt about the toy, a key aspect for the consumer (adult) buying the product. For this dataset we consolidated the terms { 'kid', 'son', 'grandson', 'daughter', 'grandaughter', 'nephew', 'niece', 'child', 'husband','wife'} to "product_recipient", as well as, the terms { 'product','car','toy','car','item','thing'} to "product". It should be noted that creating these mapping lists is the only manual dataset-specific task required in the whole system.

Upon applying these filters this relatively simple method for aspect extraction produced decent results as we will discuss in the results and analysis sections.

2) *Rule-based heuristic method*

The second method of aspect extraction we developed and tested was a rule-based heuristic method. For this method, we use a more advanced rule-based approach to extract aspect words. The rules and algorithms used were outlined in Venugopalan and Gupta's paper and can be seen in figure 1 [7].

***Inputs**: Pre-processed data for the respective domain S*
Liu : opinion word dictionary of Bing Liu
***Output** :aspect[][] a two dimensional array where each row aspects[i]corresponds to the aspects in each sentence of S*
LEVEL 1:
for i = all sentences in *S*
for every word in *S[i]*
R1: if word is a noun and preceded by an adjective which is not an opinion word then concatenate the word and its preceding word and append it to aspect[i]

R2: if word is a noun and preceded by another noun then concatenate the word and its preceding word and append it to aspect[i]

R3: if word is a noun and preceded by an adjective which is an opinion word then append the word to aspect[i]

R4: if word is a noun and is in a dobj relationship with a verb in the sentence then append the word in aspect[i]

R5: if word is a noun and is in a nsubj relationship with an adjective in the sentence then append the word in aspect[i]

R6: if the sentence S[i] contains a SUBJECT VERB and if the word has any adverbial or adjective modifier which is an opinion word append the word to aspect[i]

R7: if word is a noun and is in a modifier relationship with a copula verb in the sentence then append the word to aspect[i]

R8: if word is a noun and if the previous word is not a preposition of place or a word corresponding to an NER tag of {'TIME', 'ORDINAL', 'NUMBER', 'DATE', 'PERCENT'} and there is atleast one opinion word in the sentence , then append the word to aspect[i]

Fig. 1. The rule-based algorithm for extracting aspect words as outlined in Venugopalan and Gupta's paper "An Unsupervised Hierarchical Rule-Based Model for Aspect Term Extraction Augmented with Pruning Strategies" [7].

These rules make use of grammatical relations which can be extracted using POS and NER (Named Entity Recognition) tags and the scapy library. For a more detailed description of each grammatical relationship please refer to [12]. These grammatical relationships indicated the presents of opinion words towards a given noun thus making it a likely candidate for being an aspect word. Another advantage of this method is it facilitates the extraction of multi-word aspect phrases like "race car".

Following the extraction of the candidate aspect words using these rules we then apply the same lemmatization, stop word removal, title word removal, and consolidation techniques that we applied in the noun frequency method. Then we rank the filtered candidate aspect words to find the top 10. We experimented with two ways of doing so. The first is the same as in the noun frequency method where we simply rank the candidate aspect words based on their frequency in the reviews and keep the top 10 most frequent aspect words. This approach is referred to as "rule-based" in the results.

The second is what we refer to as a heuristic ("rule-heuristic" in results) approach which takes into account both the frequency and polarity magnitude of a given candidate aspect word. We calculate this heuristic which we call importance for each word w using the equation $importance_w = \frac{count_w{}^2 \times polarity_w}{10}$, where count_w is the frequency of a given candidate aspect word in the reviews. The polarity value is a measure of the average magnitude of emotions tied to each sentence where the given aspect word is found. We compute the polarity by taking each sentence from the reviews containing the aspect word and computing the polarity for each sentence using NLTK's implementation of VADER (Valence Aware Dictionary and sEntiment Reasoner). This function provided both negative, neutral, and positive polarity scores but we kept the magnitude of the largest score among the negative, and positive polarity scores. We then summed up these values and divided by the total number of sentences containing the aspect word ie. for a candidate aspect word $w, polarity_w = \frac{\sum_{s \in S} polarity(s)}{|S|}$, where S is the set of sentences containing the aspect word w. This heuristic allows us to capture the cluster of meaningful aspect words, in the middle of figure 2, that have a medium-high polarity and appear fairly frequently in the reviews. The theory behind this heuristic was that if a word is mentioned a lot and people have strong feelings towards it, then it is likely an important aspect. Our reason for dividing by 10 was that it allowed us to adjust when the frequency would overpower the polarity term and thus, we could fine-tune the denominator to tweak the heuristic based on the number of reviews in the dataset. Once we calculate the $importance_w$ heuristic we then choose and output the top 10 words with the highest value.

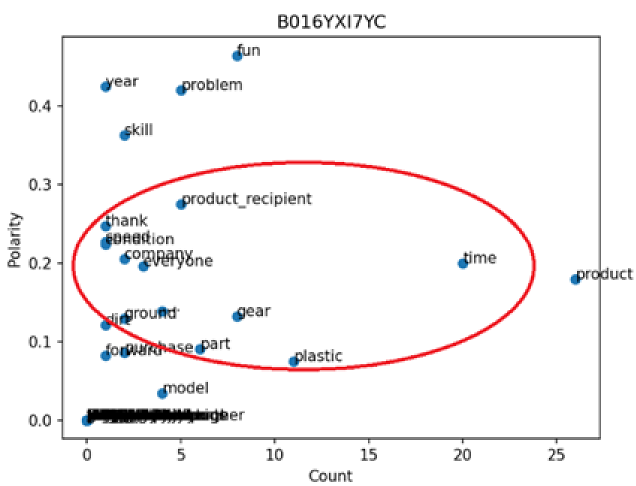

Fig. 2. A scatter plot showing the polarity vs. the count frequency in reviews for candidate aspect words of product B016YXI7YC. The middle part of the chart (circled in red) shows the most plausible candidates that we are looking to extract with the heuristic.

3) Manual method

The final method of aspect extraction that we tested was one of manually assigning aspect words to each product. This method was meant to act as a control and could be useful in the case where the user already knows the important aspects in the product category. For this method, we created a list of aspect words based on our own intuition of what aspects of an RC car would be, the list of words are as follows (note the first 2 are consolidated terms): product, product_recipient, quality, battery, price, box, color, speed, controller. In this method, the same list of aspect words is assigned to all products. Doing so was necessary since reading all reviews to find a unique set for each product would defeat the purpose of using an automated system.

For all aspect extraction methods, the system takes the top 10 (or any N chosen during configuration) aspect words for each product and outputs them to a CSV file where each row contains the ASIN (Amazon product ID number) and the list of aspect words for the product. The aspect words generated by each method can be found in appendix 2.

B. Emotion analysis

The next stage in our AAMER system is emotion analysis. This module takes in the CSV file created by the aspect extraction module which contains the aspect words for each product. The task of this module is to look at each aspect of each product and compute a set of emotion values for that aspect by analyzing the product's reviews. For each aspect word an emotion value is computed for each of the following emotions: fear, anger, anticipation, trust, surprise, generally positive, generally negative, sadness, disgust, and joy [1][6][15]. For our system, the emotions were computed using the NRCLex python library by Mark M. Bailey this library uses a lexicon-based Approach and relies on the NRC Emotion Lexicon created by the National Research Council Canada (NRC). When fed a text string this library produces a value associated with each emotion listed above [1][6][15]. We experimented with two different ways of using the library to see how the results would vary when it was fed long vs. short text strings.

1) Method 1: Sentence based emotional analysis followed by aggregation across reviews

In the first method for each aspect word extracted for a given product we tokenize each review into sentences for that product and iterate over each sentence. If the sentence contains that aspect word, we pass the sentence to the NRCLex library to generate the emotion values. We then average the emotion values of all sentences containing the word to get the set of emotion values for that aspect. This method was used because it only required short pieces of text (sentences) to be passed to the library.

A variation on this method which we refer to as "non-zero mean" was also tested. In this variation when calculating the average of a given emotion value, we only count that subset of sentences that gave a non-zero value for that emotion. Doing so was found to help amplify emotion values since the denominator was smaller when averaging given that only sentences with those emotions were counted. This variation also made sense since just because an emotion towards an aspect didn't appear in a given sentence does not mean it should dilute the value seen in sentences where it does appear.

2) Method 2: Single emotional analysis on aggregated aspect text blob

In the second method when we find that a sentence contains a given aspect word, we concatenate the sentence to a string. Once we have iterated over each sentence of each review, we will have a single large blob of text containing all sentences that include a given aspect word. We then pass this entire text blob to the NRCLex library to generate the emotion values for the given aspect. This method was used because it allows us to test the effects of passing larger strings of text to the library.

C. Results and data visualization:

We tested all combinations and variations of our aspect extraction and emotion analysis methods to get 12 sets of test results. The final output of each of these tests will be a set of 10 emotion values for each aspect of each product. We then visualize the results in a grouped bar chart where each product is a separate chart. In the chart, each group represents an aspect of the product and has 10 bars denoting the emotional values of that aspect. An example resulting chart is shown in the figure below.

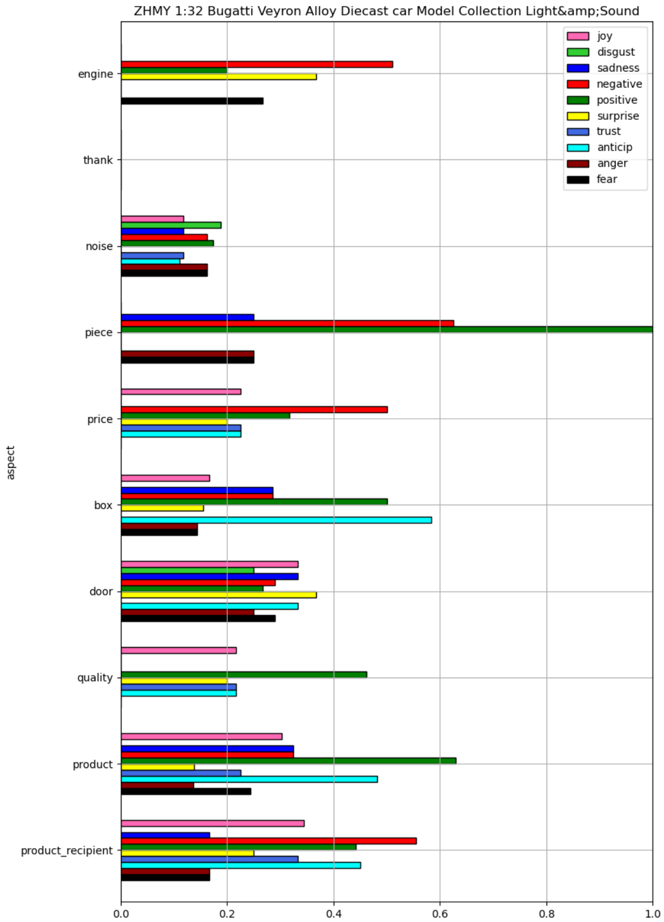

Fig. 3. The chart shows emotion values for each extracted aspect of product B01CA4R218, titled "ZHMY 1:32 Bugatti Veyron Alloy diecast". Aspect words were extracted using noun-based frequency method and emotion analysis was done with method 1 with the non-zero mean for this chart.

From the chart in figure 3, we can get an overview of emotions towards certain aspects. In the example, we see that reviewing customers have significant negative emotions towards the price but have more positive emotions about the color and quality of the product. Significant surprise emotions can also be seen towards the engine aspect of the product, from which a user can quickly infer some deficiency with the engine in the product. However, this example also demonstrates how our system can sometimes generate noise or nonsense results during both aspect extraction and emotion analysis. This can be seen when the system generated the "piece" aspect for the example result which is not an intuitive aspect of an RC car. There is also a fair bit of noise in emotions detected towards certain aspects like for the "door" and "product_recipient" aspects in the examples, where no clear standout emotions can be seen.

We also qualitatively compare results from all tests by laying out all resulting charts for the same product side by side (see appendix 1 for figure). From here a few patterns are obvious. The first is that non-zero mean tends to give higher emotion values across all products. The second is that for the manual method certain aspects do not give any aspect values. This is expected since the manually entered aspects may not appear in all products. Other than these observations, there were no significant visually discernible differences between the tests. To provide more insight into the difference between each test a quantitative analysis will be performed in the next section.

VI. Analysis

The analysis of our AAMER method was the most challenging part of this study. This was largely due to the fact that we decided to create our own data set and wanted to detect multiple emotions. Due to this, unlike in other works that made use of existing datasets such as Semeval, we were unable to directly compare our method's results against other systems [23][13][11]. Also, since the newly created dataset did not have aspect or emotion labels, and creating them was beyond the scope of the project, we also did not have a ground truth for our comparison either. We instead used a set of statistical analysis methods to compare the tests and see which of the test performed the best.

A. *Emotion analysis method difference heat map*

The first analysis we performed compared two emotion analysis methods while using the same aspect extraction method for both. We took the difference in emotion values for all aspects between the two analysis methods and plotted the differences as heat maps. By doing so we could tell that method 1 (with normal averaging) produced a similar value to method 2, with method 2 values only being slightly higher for emotions like joy and positivity. Meaning that the library when passed a large text chunk, like in method 2, performed similarly, to method 1 where smaller chunks were passed and then averaged. This analysis also confirmed that emotion values for method 1 when using non-zero averaging were larger overall than values for method 1 with normal averaging.

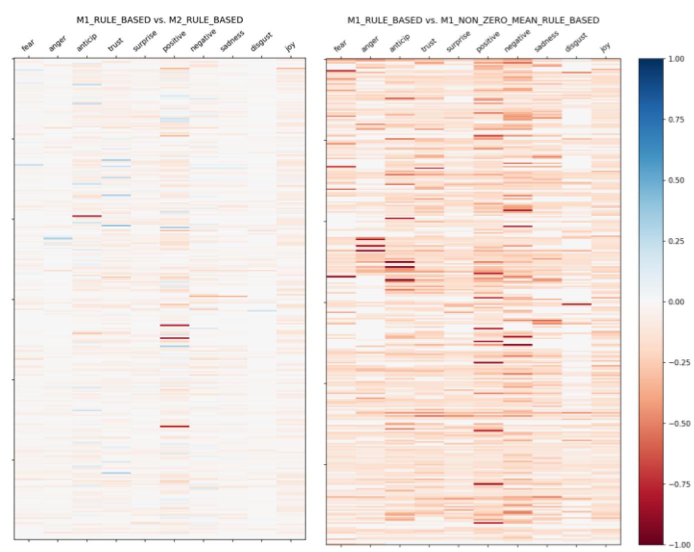

Fig. 4. These two charts visualize differences between emotion values between two emotion analysis method in a heat map. They show differences between

method 1 and method 2 (left), and method 1 when using normal vs. non-zero mean (right). Aspect words extracted using rule-based method for both.

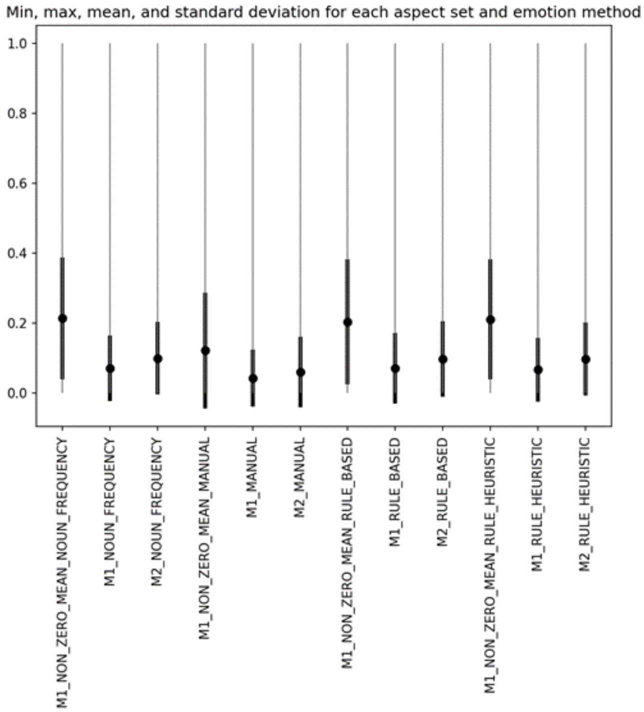

Fig. 5. Chart shows minimum, maximum, mean, and standard deviation in emotion values for each test, across all 12 tests using combinations of all emotion analysis and aspect extraction. The naming scheme for the tests is as follows {emotion analysis method used}_{aspect extraction method used}

B. *Test minimum, maximum, mean , and standard deviation*

In the second statistical analysis, we calculated and plotted the minimum, maximum, mean, and standard deviation in emotion values of all 12 tests. The minimum and maximum values of all tests were 0 and 1 which gave little insight, but by analyzing the standard deviation and mean we could tell the amount of noise vs. interpretable emotion values generated by each test. A large standard deviation would indicate that there is an emotion value far from the mean, therefore that emotion value could be interpretable as being very strong or very weak towards the given aspect. Where a test with a low standard deviation would indicate more noise and the absence of a few strong emotions. This analysis is shown in figure 5. We see that tests using emotion analysis method 1 with non-zero averaging showed the best performance with a larger mean and standard deviation, with tests using method 2 being the runners-up. The difference between the standard deviation of tests with different aspect extraction methods, but the same emotion analysis was negligible.

C. *Correlation with rating*

Our final statistical analysis was used to compare our tests' performance against the star rating values that came with the dataset. We did this by taking the average star rating of each product and calculating its correlation coefficient with the average emotion values of each product (seen in table I). We then show the correlation between the ratings and emotion values for each test in the table below. By doing so we can get a sense of which tests generate emotion values most consistent with the rating. We would intuitively expect the best test to produce strong negative correlation values for more negative emotions like fear, anger, sadness, disgust, and overall negative, and strong positive correlation values for emotions like trust, anticipation, surprise, joy, and overall positive emotions. Using this logic, we observed that the test using method 2 with either the simple (frequency) rule-based or heuristic rule-based approach performed the best in this analysis.

We do acknowledge that there are some flaws with this method of analysis. Most notably that it requires us to average the emotion values of all aspects of a product together and we thereby lose the aspect information. We also acknowledge that in reality, the rating may not always directly correlate with emotions towards a product, as we ourselves argued in the intro of this paper. However, since the given star ratings are the closest analog to ground truth we had for this data set we felt that it was a viable way of comparing the tests to each other in a meaningful manner.

TABLE I. The table shows the correlation coefficients between star rating and each average emotion value across all 12 tests using combinations of all emotion analysis and aspect extraction. The naming scheme for the tests is as follows {emotion analysis method used}_{aspect extraction method used}. The overall column was calculated by taking sum the positive emotion's correlations values added to the negated sum of the negative emotion's correlations values

Test method	**fear**	**anger**	**sadness**	**disgust**	**negative**	**trust**	**anticip**	**surpise**	**joy**	**positive**	**Overall**
M1_NON_ZERO_MEAN_NOUN_FREQUENCY	-0.27	-0.32	-0.07	-0.21	-0.24	-0.18	-0.01	-0.11	0.35	0.52	1.68
M1_NOUN_FREQUENCY	-0.45	-0.59	-0.21	-0.16	-0.73	-0.19	-0.10	0.04	0.34	0.43	2.66
M2_NOUN_FREQUENCY	-0.39	-0.71	-0.27	-0.44	-0.71	0.09	0.14	0.11	0.54	0.66	4.05
M1_NON_ZERO_MEAN_MANUAL	-0.53	-0.50	-0.51	-0.40	-0.44	-0.53	-0.41	-0.42	-0.22	-0.07	0.74
M1_MANUAL	-0.49	-0.56	-0.49	-0.09	-0.59	-0.54	-0.37	-0.42	0.13	0.02	1.04
M2_MANUAL	-0.51	-0.62	-0.54	-0.31	-0.61	-0.49	-0.26	-0.24	0.21	0.28	2.09
M1_NON_ZERO_MEAN_RULE_BASED	-0.30	-0.35	-0.13	-0.29	-0.19	-0.18	-0.09	-0.15	0.26	0.59	1.70
M1_RULE_BASED	-0.37	-0.66	-0.23	-0.21	-0.68	-0.22	-0.08	-0.09	0.18	0.51	2.45
M2_RULE_BASED	-0.39	-0.74	-0.20	-0.42	-0.61	-0.02	0.06	0.05	0.40	0.59	3.43
M1_NON_ZERO_MEAN_RULE_HEURISTIC	-0.23	-0.31	-0.15	-0.28	-0.17	-0.25	-0.07	-0.14	0.26	0.59	1.51
M1_RULE_HEURISTIC	-0.31	-0.37	-0.26	-0.54	-0.55	-0.31	-0.31	-0.19	0.12	0.43	1.77
M2_RULE_HEURISTIC	-0.29	-0.65	-0.24	-0.57	-0.51	-0.03	0.00	0.09	0.39	0.70	3.40

VII. Conclusion and Discussion

For this study, our goal was to create an aspect-based emotion analysis system that can help its user easily extract and interpret reviewers' emotions towards specific aspects of a product. We wanted to create a system that could be easily applied to any product category and would require no training of the system. In this paper, we have shown that this goal is feasible since our simple rule-based methods were able to achieve results, in some examples, that were fairly interpretable and show the system's user a reasonable, and believable, representation of the emotions reviewing customers felt towards given product aspects.

However, despite some promising results like the one exampled above, overall, many of our test results did suffer from noise. The biggest contributor to this issue was the fact that the aspects that our system extracted were not always consistent with aspect terms one would expect from this category. Indeed some aspect terms like "price" and "quality" were believable but other nonsensical aspects like "year" and "bongbong" also appeared in the results despite our efforts to filter them. We expected some of these issues to appear in the simpler noun frequency method but we had hoped that the advancements in the rule-based heuristic method would perform better. However, as the results and analysis have shown, there were similarly noisy aspect results in both methods and the more advanced aspect extraction method showed little improvement in the results.

Our studies statistical analysis also looked at how the emotion analysis method affected our results. It indicated that among the 12 variations in our system those that used the method 1 emotion analysis with non-zero averaging produced results that were most likely to produce stronger standout emotions for each aspect, thus making it the most easily interpretable method. We also saw that method 2 of emotion analysis, although not as interpretable as method 1 with non-zero averaging, did produce emotion values that were the most similar to a given product star rating. Thus, both methods may be usable but more study is required, perhaps on a larger and more diverse dataset, to clearly determine which emotion analysis method is the best. Also, both methods still suffered from a fair bit of noise where many emotions had similar values and no one (or few) emotion was clearly dominant.

To resolve these issues further work is needed. Some simple improvements we can make in future iterations is to split sentences into clauses for more granularity in emotion values [17]. This could help reduce noise when a sentence contains more than one aspect and conflicting emotions towards each aspect. Another improvement that would reduce the noise in emotion values would be to handle word negation before feeding to the emotion library eg. replacing "not happy" with "sad" in a sentence [17].

Our AAMER system provides an easy way to make these improvements as we have designed it to be modular and store and read the data created by each step in CSV files. This modular approach allows us, and future researchers, to improve parts of the system such as the aspect extraction and replace it with a more advanced method, without having to rework other parts of the system. We feel that looking into a new aspect extraction method, perhaps even one that uses a weakly supervised learning method could be a good next step to boost the performance of the system. Looking at other emotion analysis libraries could also help boost performance, and could be easily done with AAMER's modular design.

We believe that these improvements need to be made while keeping in mind the goals that we strived to achieve at the beginning of this paper. There have been other works where unsupervised aspect extraction had been used [7][21], and other works have also used emotion analysis, the more granular form of sentiment analysis with more types of emotions [19][5][11]. We believe that we have been able to contribute to this research topic by creating one of the first aspect-based emotion analysis method that combines the two features. We also wanted to build a system that was more usable in diverse real e-commerce settings like Amazon by normal users. This was only a first attempt at a system with these challenging set of requirements and we strongly encourage more research into this type of product aspect-based review emotion analysis method that not only extracts multiple emotions towards an aspect of the product but that can also be easily used in real-life settings.

References

[1] S. M. Mohammad, "Practical and Ethical Considerations in the Effective use of Emotion and Sentiment Lexicons." 2020.

[2] G. Ganu, N. Elhadad, and A. Marian, "Beyond the stars: improving rating predictions using review text content.," 2009, vol. 9, pp. 1–6.

[3] W. Kim, T. Ko, I. Rhiu, and M. H. Yun, "Mining affective experience for a kansei design study on a recliner," vol. 74, pp. 145–153, 2019, doi: 10.1016/j.apergo.2018.08.014.

[4] H. Binali, C. Wu, and V. Potdar, "Computational approaches for emotion detection in text." IEEE, pp. 172–177, 2010, doi: 10.1109/DEST.2010.5610650.

[5] E. Kauffmann, J. Peral, D. Gil, A. Ferrández, R. Sellers, and H. Mora, "A framework for big data analytics in commercial social networks: A case study on sentiment analysis and fake review detection for marketing decision-making," vol. 90, pp. 523–537, 2020, doi: 10.1016/j.indmarman.2019.08.003.

[6] S. M. Mohammad and P. D. Turney, "Crowdsourcing a Word-Emotion Association Lexicon," vol. 29, no. 3, pp. 436–465, 2013.

[7] M. Venugopalan and D. Gupta, "An unsupervised hierarchical rule based model for aspect term extraction augmented with pruning strategies," vol. 171, pp. 22–31, 2020.

[8] G. Lackermair, D. Kailer, and K. Kanmaz, "Importance of online product reviews from a consumer's perspective," vol. 1, no. 1, pp. 1–5, 2013.

[9] "Online shopping by Canadians in 2020: Results from the Canadian Internet Use Survey," Jun. 22, 2021. https://www150.statcan.gc.ca/n1/pub/11-627-m/11-627-m2021048-eng.htm.

[10] S. Gojali and M. L. Khodra, "Aspect based sentiment analysis for review rating prediction," 2016, pp. 1–6.

[11] J. Bross and H. Ehrig, "Generating a context-aware sentiment lexicon for aspect-based product review mining," 2010, vol. 1, pp. 435–439.

[12] M.-C. De Marneffe and C. D. Manning, "The Stanford typed dependencies representation," 2008, pp. 1–8.

[13] S. Poria, E. Cambria, and A. Gelbukh, "Aspect extraction for opinion mining with a deep convolutional neural network," vol. 108, pp. 42–49, 2016.

[14] J. Ni, J. Li, and J. McAuley, "Justifying recommendations using distantly-labeled reviews and fine-grained aspects," 2019, pp. 188–197.

[15] S. Mohammad and P. Turney, “Emotions Evoked by Common Words and Phrases: Using Mechanical Turk to Create an Emotion Lexicon,” 2010, pp. 26-34, [Online]. Available: https://aclanthology.org/W10-0204.

[16] S. Tulkens and A. van Cranenburgh, “Embarrassingly simple unsupervised aspect extraction,” 2020.

[17] T. U. Haque, N. N. Saber, and F. M. Shah, “Sentiment analysis on large scale Amazon product reviews.” IEEE, pp. 1–6, 2018, doi: 10.1109/ICIRD.2018.8376299.

[18] W. Medhat, A. Hassan, and H. Korashy, “Sentiment analysis algorithms and applications: A survey,” vol. 5, no. 4, pp. 1093–1113, 2014.

[19] W. M. Wang, Z. Li, Z. G. Tian, J. W. Wang, and M. N. Cheng, “Extracting and summarizing affective features and responses from online product descriptions and reviews: A Kansei text mining approach,” vol. 73, pp. 149–162, 2018, doi: 10.1016/j.engappai.2018.05.005.

[20] D. Coppola, “E-commerce in Canada - Statistics & Facts ,” Jul. 05, 2021. https://www.statista.com/topics/2728/e-commerce-in-canada/ (accessed Oct. 01, 2021).

[21] S. Poria, E. Cambria, L.-W. Ku, C. Gui, and A. Gelbukh, “A rule-based approach to aspect extraction from product reviews,” 2014, pp. 28–37.

[22] T. T. Tun, J.-C. Na, and C. S. G. Khoo, “Aspect-based sentiment analysis of movie reviews on discussion boards,” vol. 36, no. 6, pp. 823–848, 2010, doi: 10.1177/0165551510388123.

[23] M. Pontiki et al., “Semeval-2016 task 5: Aspect based sentiment analysis,” 2016, pp. 19–30

APPENDIX 1: MULTI-TEST BAR CHART

Fig. 6. The series of charts shows emotion values for each extracted aspect of product B01CA4R218, titled “ZHMY 1:32 Bugatti Veyron Alloy dicast” across all 12 tests using combinations of all emotion analysis and aspect extraction.

APPENDIX 2: ASPECT EXTRACTION DATA

Table II. This table shows the aspect word extracted for each product using the noun frequency method.

ASIN (product ID)	**Aspect words**									
B00005LBZO	Product	product_recipient	year	time	bike	pedal	seat	wheel	condition	today
B000GPLT68	Product	product_recipient	wheel	time	quality	race	gift	year	piece	christmas
B000Q9KPB4	Product	model	base	quality	wire	box	time	corner	top	price
B002AKKO6S	product_recipient	string	product	year	quality	sound	price	tune	body	instrument
B004KU83C8	Product	product_recipient	set	quality	thomas	track	year	box	engine	fun
B004LH01ZW	product_recipient	product	battery	wheel	light	day	minute	review	fun	color
B006RKH1I6	product_recipient	arm	quality	show	transformer	year	product	form	hour	set
B006RKBK0Q	product_recipient	arm	product	quality	show	time	year	leg	amazon	transformer
B006RKFHBY	product_recipient	product	quality	year	arm	show	transformer	korea	leg	day
B00AM9SZWO	product_recipient	bong	carry	bongbong	quality	year	product	time	lot	picture
B00APTT9SU	product_recipient	year	rogi	quality	time	size	show	door	..	hand
B00E849LS6	Product	price	product_recipient	card	collection	wall	plastic	wheel	star	shipping
B00J97X2YG	Window	fan	product	vehicle	rubber	air	unit	sun	day	heat
B00JPTG45C	Tool	wheel	derby	product	nail	side	year	slot	groove	metal
B00KFBS5QA	Product	product_recipient	communication	quality	seller	price	shipping	day	door	wheel
B00KS6ADQW	Product	paint	collection	awesome	detail	job	product_recipient	box	addition	character
B00S4RM1XK	product_recipient	cute	color	Eye	gift	baby	product	blue	collection	super
B00TSV67XG	Product	Time	quality	easy	piece	sturdy	theme	cute	perfect	stand
B00TWG57CY	product_recipient	product	bruder	door	quality	year	part	vehicle	christmas	officer
B00V8LIPFW	product_recipient	product	bruder	year	door	quality	vehicle	part	price	christmas
B0155MEP1C	Product	product_recipient	price	year	fun	time	speed	wheel	rc	box
B0155MVMFO	Product	product_recipient	time	Rc	year	plastic	controller	speed	fun	wheel
B015SKSQUW	Product	product_recipient	day	costume	skin	year	price	time	zombie	job
B0162HF12Q	Product	product_recipient	time	Day	part	battery	wheel	minute	fun	drive
B016YXI7YC	Product	battery	time	flight	motor	quad	review	size	plastic	money
B017AYXV9U	Truck	time	product_recipient	product	part	fun	minute	speed	body	day
B018VW1X4U	product_recipient	product	year	quality	door	great	seller	price	box	gift
B01B55BPDY	product_recipient	product	k'nex	piece	year	motor	fun	kit	part	race
B01CA4R218	product_recipient	product	quality	Box	door	color	engine	piece	year	price
B01CWSRW62	product_recipient	battery	product	Fun	christmas	minute	day	plug	great	Wheel

Table III. This table shows the aspect word extracted for each product using the rule-based method.

ASIN (product ID)	**Aspect words**									
B00005LBZO	product	product_recipient	wheel	time	bike	pedal	today	condition	year	yr
B000GPLT68	product	product_recipient	race car	wheel	quality	screw	time	gift	plastic	box
B000Q9KPB4	product	display case	quality	model	box	corner	money	price	time	product_recipient
B002AKKO6S	product_recipient	string	sound	product	price	quality	tune	instrument	full size	action
B004KU83C8	product_recipient	product	set	thomas	fun	brio products	wheel	box	box cars	track
B004LH01ZW	product_recipient	product	wheel	battery pack	review	battery	lot	fun	antenna	box
B006RKH1I6	product_recipient	arm	robocar poli	show	quality	transformer	granddaughter	adult	transformation	part
B006RKBK0Q	product_recipient	arm	quality	product	show	robocar poli	amazon	leg	time	youtube
B006RKFHBY	product_recipient	robocar poli	arm	quality	product	show	korea	old son	transformer	leg
B00AM9SZWO	product_recipient	carry	bong bong	quality	bong	red bus	bongbong	top	lot	cartoon
B00APTT9SU	product_recipient	quality	time	rogi	show	lot	..	wheel	friend	sticker
B00E849LS6	product	product_recipient	wheel	price	collection	wall	star	display case	card	shipping
B00J97X2YG	window	product	fan	rubber	vehicle	air	unit	sun	temperature	difference
B00JPTG45C	tool	wheel	pinewood derby	product	nail	derby car	work	side	product_recipient	price
B00KFBS5QA	product	product_recipient	seller	price	great	customer	customer service	door	wheel	shipping
B00KS6ADQW	product	paint	awesome	collection	product_recipient	addition	detail	box	amazon	pre-order
B00S4RM1XK	product_recipient	cute	color	eye	beanie boos	super	product	super cute	beanie boo	baby
B00TSV67XG	easy	product	sturdy	quality	perfect	product_recipient	great	time	birthday party	super
B00TWG57CY	product_recipient	product	door	quality	bruder	other bruder	price	play	bruder trucks	one
B00V8LIPFW	product_recipient	product	door	bruder	quality	price	other bruder	lot	awesome	light
B0155MEP1C	product	product_recipient	price	fun	speed	remote control	wheel	quality	review	rc
B0155MVMFO	product	product_recipient	controller	plastic	wheel	rc car	speed	fun	price	time
B015SKSQUW	product_recipient	product	skin	job	scar tattoos	zombie	look	way	day	sheet
B0162HF12Q	product	product_recipient	part	battery	fun	rc car	seller	wheel	money	time
B016YXI7YC	product	battery	review	motor	controller	video	quad	nothing	great	wind
B017AYXV9U	product_recipient	truck	product	fun	time	price	part	body	rc truck	great
B018VW1X4U	product_recipient	product	door	great	seller	quality	box	battery	price	old son
B01B55BPDY	product_recipient	k'nex	product	fun	motor	race car	piece	instruction	lego	race
B01CA4R218	product_recipient	product	quality	door	box	price	piece	noise	thank	engine
B01CWSRW62	product_recipient	product	battery	christmas	fun	great	wa	money	european plug	wire

Table IV. This table shows the aspect word extracted for each product using the rule heuristic method.

ASIN (product ID)	Aspect words									
B00005LBZO	Product	product_recipient	year	time	highly	condition	pedal	fun	bike	everyone
B000GPLT68	Product	product_recipient	love	easy	great	quality	gift	time	wheel	race car
B000Q9KPB4	Product	model	display case	look	perfect	quality	fit	one	nice	price
B002AKKO6S	Sound	product_recipient	great	string	product	quality	one	price	tune	great guitar
B004KU83C8	product_recipient	product	great	set	quality	excellent	fun	play	awesome	brio train
B004LH01ZW	product_recipient	product	work	battery	light	fun	wheel	review	great	gift
B006RKH1I6	Product	product_recipient	love	great	quality	transformer	show	play	great toy	arm
B006RKBK0Q	Product	product_recipient	love	quality	arm	show	great	time	amazon	play
B006RKFHBY	Product	product_recipient	love	quality	year	play	robocar poli	great	transformer	arm
B00AM9SZWO	Product	product_recipient	bong	great	bong bong	carry	quality	cute	lot	favorite
B00APTT9SU	product_recipient	love	product	great	quality	size	time	rogi	play	month
B00E849LS6	Product	great	perfect	look	product_recipient	price	wheel	hot wheels	hot	display case
B00J97X2YG	Product	window	fan	work	great	hot	rubber	sun	vehicle	day
B00JPTG45C	Tool	product	great	work	wheel	straight	derby	use	nail	pinewood derby
B00KFBS5QA	Product	product_recipient	great	loved it	son loved	..	quality	communication	seller	came
B00KS6ADQW	Great	product	love	awesome	collection	paint	product_recipient	nice	job	detail
B00S4RM1XK	Cute	love	product_recipient	great	super	soft	gift	color	bought	beanie boo
B00TSV67XG	Great	easy	sturdy	cute	product	perfect	quality	look	nice	super
B00TWG57CY	Product	product_recipient	great	bruder	quality	one	back	play	light	part
B00V8LIPFW	Product	product_recipient	great	bruder	quality	year	light	one	back	play
B0155MEP1C	Product	great	product_recipient	fast	one	fun	price	speed	time	day
B0155MVMFO	Product	product_recipient	rc	great	rc car	time	fun	plastic	price	year
B015SKSQUW	Easy	product	look	product_recipient	super	costume	wa	day	review	skin
B0162HF12Q	Product	great	product_recipient	this car	n't	fun	day	time	drive	rc car
B016YXI7YC	Great	product	one	time	battery	flight	quad	motor	fun	review
B017AYXV9U	Product	truck	fun	great	product_recipient	time	part	speed	gift	price
B018VW1X4U	Product	product_recipient	great	nice	quality	gift	..	seller	door	battery
B01B55BPDY	Product	product_recipient	build	k'nex	fun	piece	motor	kit	help	great
B01CA4R218	Product	product_recipient	love	quality	nice	great	awesome	door	came	thank
B01CWSRW62	Product	product_recipient	fun	wa	battery	great	christmas	money	was fun	Waste

APPENDIX 3: EMOTIONAL ANALYSIS SAMPLE DATA

Table V. This table shows a few sample rows of the emotion analysis data extracted using the rule-heuristic method and the 1st emotion analysis method with non-zero mean.

ASIN (product ID)	Aspect word	Emotion values									
		fear	*anger*	*anticip*	*trust*	*surprise*	*positive*	*negative*	*sadness*	*disgust*	*joy*
B00005LBZO	Product	0.49359	0.5	0.484722	0.41456	0.291667	0.611068	0.475962	0.153846	0	0.301465
B00005LBZO	product_recipient	0.205556	0.34	0.326515	0.311364	0.155952	0.436765	0.403333	0.24	0.138889	0.25
B00005LBZO	Year	0.25	0.333333	0.219048	0.269048	0.183333	0.339583	0.416667	0.25	0.5	0.255556
B00005LBZO	Time	0	0.25	0.622917	0.25	0.225	0.29	0.291667	0.333333	0	0.233333
B00005LBZO	Highly	0	0	0.1	0.366667	0.175	0.4125	0.25	0	0	0.275
B00005LBZO	condition	1	0	0.266667	0.2	0.2	0.266667	0.333333	0	0	0.2
B00005LBZO	Pedal	0	0.625	0.25	0	0	0.625	0.25	0	0	0
B00005LBZO	Fun	0.083333	0	0.222222	0.125	0.083333	0.277778	0.083333	0.083333	0.083333	0.277778
B00005LBZO	Bike	0	0.25	0.25	0	0	0.625	0.375	0.5	0	0
B00005LBZO	everyone	0	0	0.25	0.25	0	0.25	0	0	0	0.25
B000GPLT68	product	0.192636	0.165695	0.275366	0.224707	0.204464	0.397535	0.269048	0.132176	0.150449	0.221674
B000GPLT68	product_recipient	0.178182	0.231818	0.335101	0.264069	0.209375	0.439173	0.470644	0.183333	0.1	0.28931
B000GPLT68	love	0	0	0.21875	0.5	0.1875	0.419643	0.375	0.1875	0.5	0.410714
B000GPLT68	easy	0	0	0.333333	0.333333	0	0.583333	0.5	0.5	0	0.388889

We only show a small sample of emotion analysis data from one test, given the full set of emotion data for all 12 tests is approximately 300×12 rows of data. The full raw data set, is available upon request.