Aspect-based Amazon Review Analysis

Repository:

https://drive.google.com/drive/folders/1NjOcMxCGNlnKRNKIeu9U90KjrwW30hzQ?usp=shari

ng

Guanda Jiang, Sida Chen & Fangjie Qi

ISE-540 Text Analytics

Instructor: Professor Mayank Kejriwal

December 1, 2020

Contents

Executive Summary

As product information becomes much easier to acquire, the selling party no longer dominates the sales relationship as before. There are too many similar goods and different platforms to choose. How to attract more loyal consumers in this situation? The answer is catering what the customers need and keep upgrading. In the existing online shopping system, customer review is an important way for both online sellers and potential buyers to know how the consumers feel about the product, service, delivery etc. However, even in the Amazon's review system, the selling party can only see the overall rating and the holistic sentences instead of finding out which factors and how much each factor contributes to the overall rating individually. It is inconvenient for the business owners to know what they did right and what they can improve. Meanwhile, it is not easy for people who want to buy the product to only check what they really concern about. This problem motivates us to come up with a better rating system with several dominant categories and provide aspect-based scores.

We chose the reviews from category "Clothing, shoes, and jewelry" to do analysis. However, there were over 32 million reviews, which was too much to process so that we narrowed it down to only focus on several "Sweatshirt" items. After filtering reviews with more than 100 characters, we had total 7451 reviews to work with. To find the crucial criteria, we attempted two different methods. First, we applied the TF-IDF and K-Means Clustering algorithms that vectorize the corpus first and use the center of each cluster as the category. Second, we adopted the LDA analysis to find the main topics across all the reviews. Finally, we chose the common six categories between these two methods, which were "color", "size", "quality", "comfort", "price" and "material". Then, we manually labeled 1200 reviews with label 1,-1,0 representing good, bad or neural to exploit as the supervised training set and evaluated these six categories by fitting regression models. After determining the categories, we start to solve the aspect-based rating problem. For this, we also tried several different methods and assessed by different metrics. To begin with, we made use of the aspect-based sentiment analysis approach with NLP techniques to separate each review to different parts and estimate each part if it is positive or negative. Then, we run multiple classifiers with TF-IDF vectorization to classify each category to be 1 or -1. Last, we added Word Embedding to the previous classification method and also tried different classifiers to see whether the result can be more accurate with only related words. For all of these three methods, we evaluated them by accuracy score, precision score, recall score and f1 score and not only compared among themselves, but also compared them with the random scenario. At the end, we also used significant test to prove our trained model is statistically significant.

It turns out a surprising consequence. The simplest method, using only TF-IDF vectorization and Random Forest classifier seems to have the best overall performance. After predicting those 6000 more out-sample instances (not manually labeled), we used the prediction as input to run regression models again. Unfortunately, the R square and Mean Absolute Error were not good. There are many reasons and many future work we need to do, we present them in the last section of this report. However, a good news is that we found the importance level of each category is similar from the ground truth and the predictions. "Size " tends to be the most important one, and "color", "price" is not that concernment for the chosen items. From this project, we learned a lot and recognized that framing the problem, building the evaluation, showing the reliability of solution are as important as finding the solution itself. Especially in the very complicated task, like this project, the result may not be that perfect, but the processing is really valuable.

Proposed Approach

Criteria Extraction

The first task we need to do is to determine what aspects of the product can be criteria for product sweatshirt. We tried two different methods.

The first method implements TF-IDF vectorization and K-Means Clustering algorithm. The general idea is to vectorize all reviews and cluster them into 10 groups. We assume that the center of each cluster should be the keyword so that we can use them as the criteria. However, in fact, the centers of many clusters are modified words like “good”, “perfect”, etc. Therefore, we have to manually pick several common important words, which are “size”, “material”, “comfort”, “warm” and “color”.

The second method was the Topic Model (LDA). Before we applied LDA, there were several steps we had done. First, we did tokenization on reviews, which means we split the text into words, lowercase the words, and removed the punctuation. Then we removed words that have fewer than 3 characters. At last, we lemmatized all words and reduced them to their root form. We applied the process above on each cluster we got in the first method. Then on each processed cluster, we used LDA to find 10 keywords.

At last, we combined the results from both two methods and determined six criteria, which are “color”, “size”, “quality”, “comfort”, “price” and “material”.

Aspect-based Sentiment Analysis

First approach we tried is the aspect-based sentiment analysis. In this case, we treat each review as a whole and let the machine decide what are the main terms, which are the modification parts and how are the sentiments.

To catch as much information as we can, before doing the sentiment analysis, we did the Word Embedding and Lemmatize Stemming to collect related and similar words with the six categories. In Word Embedding, we used the cosine similarity larger than 0.6 as the threshold and finally got a list of words in each category. For example, the words like “comfy”, “cozy”, “softer” and “warm” will be transformed to “comfi”, “cozi”, “soft”, “warm” and then be classified to category “comfort” according to the similarity.

Furthermore, we learned the method implementing Natural Language Processing (NLP) techniques from the article *Aspect-based Sentiment Analysis - Everything you wanted to know (Intellica.AI)* to decode the review and rate them separately. Firstly, we globally defined two lists of words including 2007 positive words and 4783 negative words in English from Github open source (Minqing and Bing) to be the standard of sentiment. Secondly, we created a function to analyze each review and apply the sentiment analysis based on the lexicon for positive and negative word dictionaries. The function will iterate each token in the review. If the token is an opinion word, the machine will assign the sentiment to be 1 or -1 and find its parent term to label it. If the vocabulary is a verb, the command will check if there is a direct object. If the token is a noun or compound noun and not in the opinion word list, it will become the key term and the function will label it by analyzing its children tokens. Finally, we can obtain a dictionary with keys as the main terms and values as the scores. For instance, if there is a review “The sizing-chart was confusing also”, after doing the aspect-based sentiment analysis, we can get a result like {‘sizing-chart’: -1}.

Since we did Word Embedding in the first step, we can easily find which categories should these key terms be classified. To be specific, if the key term can be found in the list of words of any category, we can just label the corresponding category with the sentiment score. If it is not covered by any of these six categories, we can just label it as 0, which means neutral, since we will assume this reviewer does not take notice of about those aspects. In the above example, the customer only cares about the sizing-chart, so we can label category “size” as -1 for this feedback. On average, each review may mention about two to three categories, so for the other categories that have not been mentioned, we will put 0 under them. Following this logic, we can label the training set and evaluate by several metrics to see the performance.

Classification with solely TF-IDF

The second method is straightforward. Still, all reviews were treated as a whole, and after being stemmed and lemmatized, the collection of reviews was converted to a matrix of TF-IDF features. Since the criteria set of six factors (“color,” “size,” “quality,” “comfort,” “price” and “material”) had been selected and partial scores (-1, 0 and 1) had been labelled manually, the input matrix was the TF-IDF features and output vector was the partial scores for one specific factor in the criteria set. Thus, the prediction was conducted six times to evaluate the performances of different classification methods for all six factors.

In comparison with various classifiers, decision tree, naïve bayes, stochastic gradient descent, support vector classifier and random forest classifier being selected, multiple scores (accuracy, precision, recall and f1) were used to predict the partial scores and the score matrices will be discussed in the Experimental Results.

Classification with TF-IDF and Word Embedding

After finishing the second method, we started to think about how we can improve it. We focused on the vectorizing part. In the second method, we vectorized all words in all reviews. But sometimes, one review contains the attitude of the reviewer towards more than one criterion. In such a case, those words related to one criterion may affect the output of the classifier that aims at another criterion.

To avoid this situation, we decided to combine the Word Embedding and TF-IDF. First, we applied Word Embedding and Lemmatize Stemming to collect related and similar words with the six categories. We used cosine similarity larger than 0.6 as the threshold and got words related to six criteria. For example, the list of stemmed words related to the criterion ‘color’ is ['array', 'bright', 'color', 'colour', 'depict', 'desrib', 'od', 'pictur', 'picture-', 'pricepoint', 'red', 'satur', 'select', 'shade', 'show', 'vibrant'].

Then in the TF-IDF part, we only vectorized the words related to the category we want to label. In labeling each criterion, we applied several diverse kinds of classifiers and decided to use the Support Vector Classifier since its performance was the best.

Experimental Results

As a whole, we have three parts that should be evaluated. First, how good are the categories we chose? Second, which solution is more effective to implement? Third, how is the performance of overall prediction using the best model we trained?

Evaluation Benchmark

To proceed the evaluation, we need a ground truth to compare with. However, at the beginning, we do not have a given rating for each category from the original dataset, so we decide to first manually label 1200 reviews (without duplicates will be 1191 reviews) base on our subjective judgement as the ground truth. In order to avoid bias as much as possible, we randomly choose 1200 reviews from total 7451 reviews with more than 100 characters. The criterion is that if the review shows positive attitude toward the category, we label it as 1. On the contrary, if it shows negative sentiment, we label it as -1. Otherwise, we label it as 0 which means neural (**Table 1**). Having this dataset makes the work much easier since now it becomes a supervised learning problem. The following table is an example showing the first 11 rows of the manually labeled dataset.

	color	size	qualiti	comfort	price	materi	overall
0	1	1	-1	0	0	0	3
1	0	-1	0	0	0	0	2
2	1	0	0	0	0	1	3
3	0	1	0	0	0	0	5
4	0	1	0	0	0	0	5
5	1	1	0	0	0	1	5
6	0	1	0	0	0	0	5
7	0	-1	0	0	0	0	3
8	0	1	1	0	0	0	5
9	0	0	0	0	0	1	4
10	0	-1	0	0	0	-1	2

Table 1 Manually labelled partial scores over six criteria

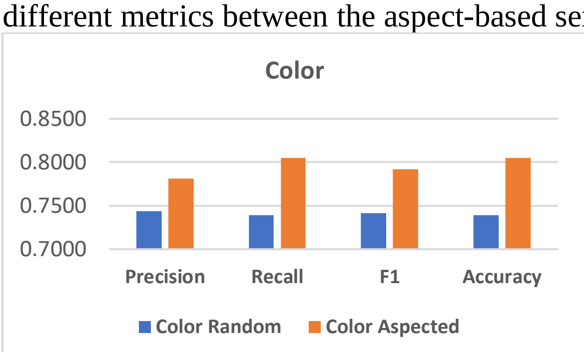
Criteria Evaluation

For the first part, we decide to use R square and Mean Absolute Error (MAE) as the metrics. On one hand, R square shows the percentage of explained variance by the model, if these six categories can explain most of the information, it should be a good choice to use them. On the other hand, MAE can more intuitively reflect the difference between prediction and true target for this rating case. Furthermore, we take manually labeled data as input and overall rating as the response to fit both Linear Regression model and Decision Tree Regression model because they can be evaluated by R square and MAE, and most importantly, both of them can provide coefficients or importance level for ultimate goal. The assessment result is not bad. We achieve the highest R square as 0.64, which means about 64% variance can be explained by these six categories, and MAE as 0.47 from the Decision Tree Regression model. Therefore, it is meaningful to use these categories to do the following work.

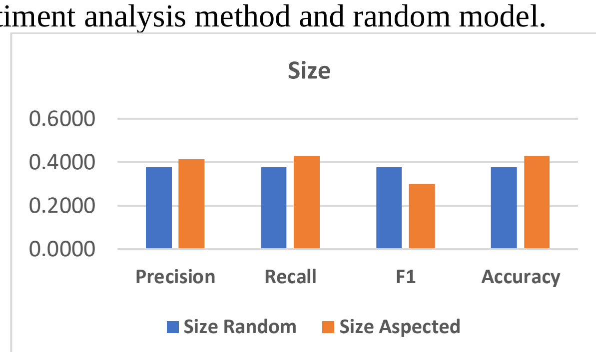
Trained models Evaluation

For the second part, since our methods substantially are solving classification problems, the metrics we select to evaluate the models are accuracy score, precision score, recall score and f1 score. Due to there are three types of criterions in our models which are 1, 0, -1, we decide to use the weighted average score as the final report number in order to make the conclusion clearer. Meanwhile, we create randomly classification model as the baseline to contrast with the accuracy of the models. Moreover, we also did significant test to check if the model is truly statistically significant or just happened by statistical chance.

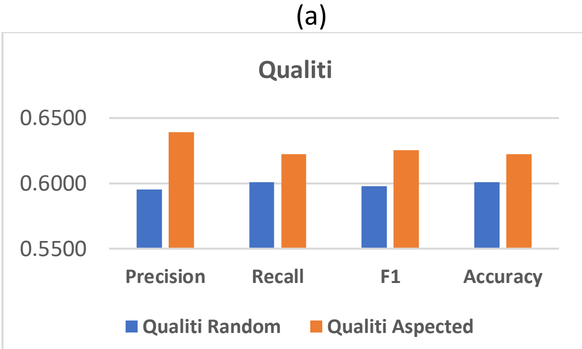
As mentioned in Proposed Approach section, the first method we tried, aspect-based sentiment analysis, will analyze each review and label each category with the corresponding sentiment score. Hence, we can get a same format table as the manually labeled dataset except the “overall” column. Nevertheless, since these metrics in scikit learn module do not support multi-class and multi-output evaluation, we can only assess these six categories one by one. For instance, to evaluate the category “color”, we use the manually labeled rating as true target and labels created by aspect-based sentiment analysis as prediction. Then, run a accuracy-score command provided by scikit learn module to acquire the result. Same procedures were done in different categories and implementing different metrics. After the comparison between ground truth and trained model, we also create a random labeling dataset to see how much more accuracy will the aspect-based sentiment analysis solution better than the random scenario. It is important to notice that we must use the weights of each criterion to build the random model because they are imbalanced. It is rational owing to the average rating of all the 7451 reviews is around 4.4 out of 5, so there are not a lot -1. At the same time, most of customers just mention few aspects in their reviews, so 0 has the dominant portion of labels. Following this logic, we obtain the random model. The evaluation process is the same as we did above, using ground truth as response and random labels as prediction to run the metrics in Python so that we can acquire the accuracy of random model. The following plots (**Fig. 1**) are the evaluation results from different metrics between the aspect-based sentiment analysis method and random model.



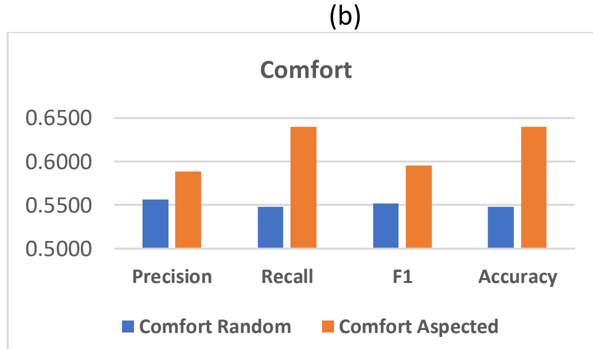
(a)



(b)

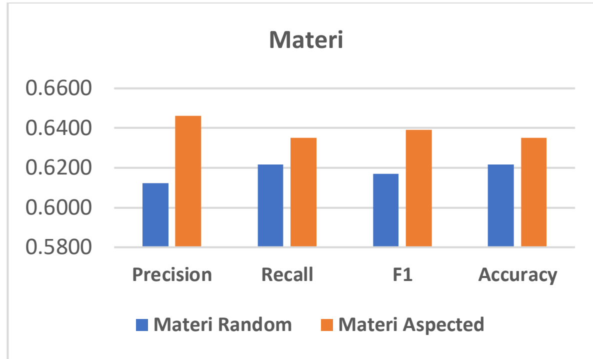


(c)



(d)

(e)

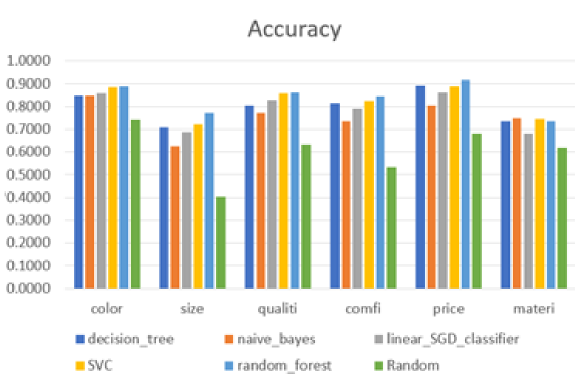


(f)

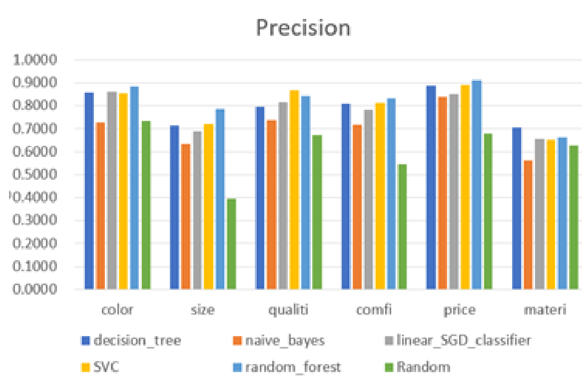
Fig. 1 (a-f) Results of aspect-based sentiment analysis versus random vector

Obviously, our trained model is much better except the category "Size". There are many reasons, and we will explain it later in the Discussion section. Last step, we need to do the significant analysis to prove whether the better performance showed above is statistically significant. Since we use the same method to label the review for six categories, we have six samples from aspect-based sentiment analysis model and six samples from random model for Student's T-test. The null hypothesis is that we assume that there is no significant difference between these two models. We set the confidence interval as 95%. After doing the hypothesis test, we got T-statistic as 7.55 and P-value as 0.0006 for metric accuracy score, which is way smaller than the threshold 0.05. Therefore, we can reject the null hypothesis and show that our first trained model is statistically significant.

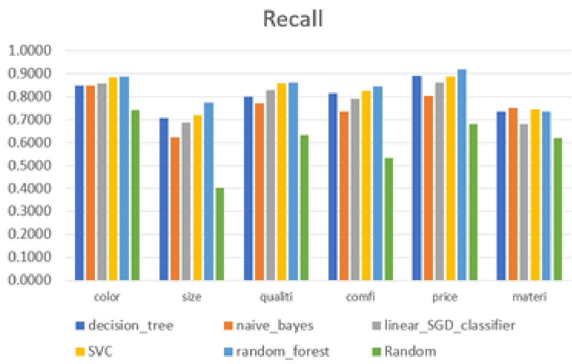
In the second approach, the crucial question here is which classifier is the best one. To provide a baseline for the comparison, the predicted vector was shuffled to generate one vector with random labels but follow the (–1) - 0 - 1 distribution, which can be regarded as one vector that derived from random, as was shown in the green bar in the very right for every bar cluster (**Fig. 2**). All predicting experiments were conducted ten times with the size of the test set being 400 reviews out of 1200 reviews. The chart below is the results of the experiments.



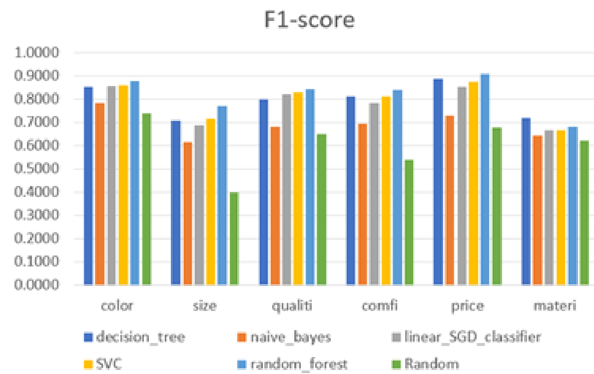
(a)



(b)



(c)



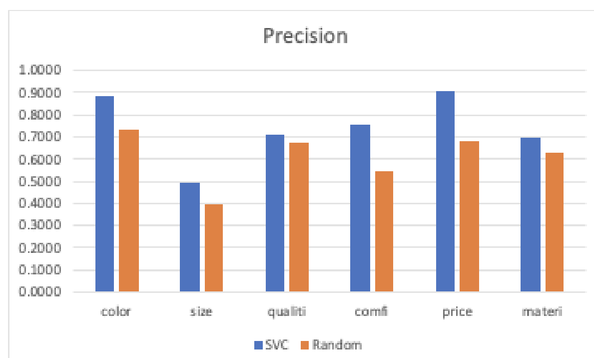
(d)

Fig. 2 (a-d) Results of TF-IDF vectorizing using various classifiers

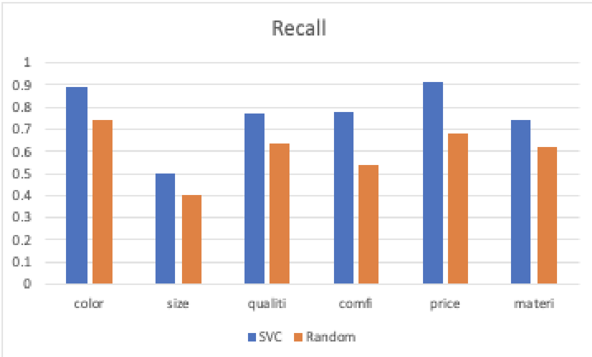
All classifiers performed better than random. Since the team collected evaluation scores from ten experiments, student t test was applied to make the distinction between predicted vectors and random vectors. The confidence interval is 95%. For the comparison between classifiers, although random forest classifier has the highest average for most of the cases, the difference between random forest and other classification methods is correspondingly small. To narrow the scope, random forest was selected as the classifier to compare with the third method (TF-IDF with word-embedding).

For the third method, the same as the second one, we had six different classifiers aimed at six different criteria, so we had to evaluate them separately. We randomly split the ground truth set into a train set and test set for ten times with the size of the test set being 400, then trained six classifiers ten times. Each time, we got the scores of the four metrics we used above. We then create a random labeling set ten times and got corresponding scores of four metrics as our baseline. The following figures show the comparison of four metrics between one out of ten experiments and one random result.

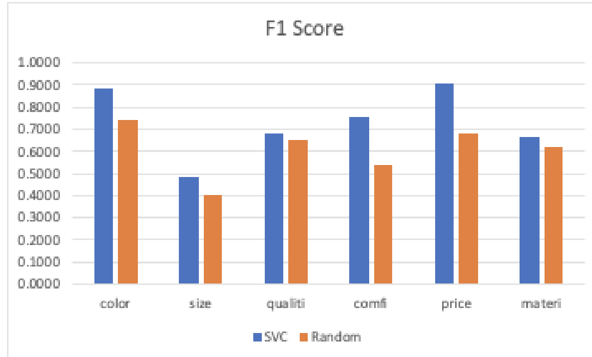
(a)



(b)



(c)



(d)

Fig. 3 (a-d) Results of TF-IDF with word embedding versus random

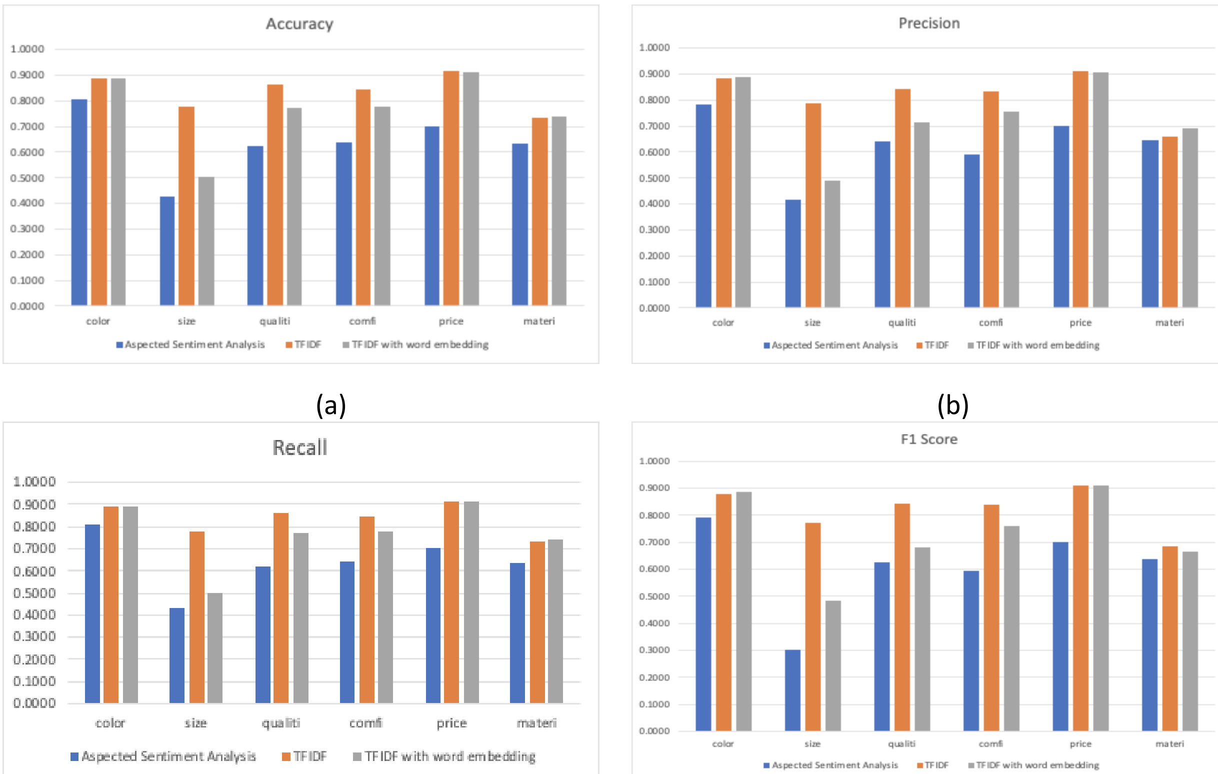
From the plots (**Fig. 3**), the performance of our classifiers seemed better than the baseline. However, we still cannot say it was better. Therefore, we applied the T-test between the scores we got from ten experiments and scores from ten random labeling sets. The null hypothesis is that the mean score of the ten scores from the classifiers is equal to the mean score of ten scores from random experiments, and we set the confidence interval as 95%. Taking the metric "Accuracy" as an example, for each criterion, we applied a one-tail T-test between ten accuracy scores from the classifier and ten accuracy scores from random labeling. If the P-value is less than 0.05, then we can say this classifier is significantly better in accuracy than random labeling. We did the test for every metric and every criterion. The result shows that all P-values are smaller than 0.05, which means the performance of the third method is better than random labeling.

Overall Prediction Evaluation

For the third part, we first used the second method to label all other unlabeled over six thousand reviews and the reason will be contained in the Discussion section. Then we still come to utilize R square and Mean Absolute Error (MAE) as the metrics since eventually we need regression models to provide the weight that each factor contributes. Therefore, it is essential to know how good the regression model fits using predictions. Given the review text, we predict the partial scores and fit regression models with Linear Regression model and Decision Tree Regression model. Linear regression model's R square is 0.186 while MAE is 0.692. For decision tree, R square is 0.294 while MAE is 0.504. The R square shows that our approach can only explain a small proportion of variance of the data, which may not be a good model to use for the Review Analysis task. Nevertheless, we have other findings, and we will cover it in the last session.

Discussion

All three methods were better than the baseline, then we compared these three methods. We first compare their performance in a single experiment by plotting their four metrics. The following figures show the result (**Fig. 4**).



(a) (b)

(c) (d)

Fig. 4 (a-d) A comparison between three methods

Obviously, the performance of the first method is the worst. The difference between the second and the third method is very tiny. We decided to use the same process of T-test that we used to compare the performance between our methods and the baseline to determine which method is better. The result shows that for criterion “color”, there is no significant difference between the performance of the second method and the third method, but for other criteria, the performance of the third method is worse than the second one. This result is contrary to what we expected. We can draw a conclusion that the second method is the best.

For all three methods, the performance on the criteria “size” and “materi” is the worst. There are two possible reasons to explain this phenomenon. The first one is the limit of word embedding in this case. If we want the words related enough, then we may not have enough words or miss many words, but if we want enough words, they may not that related. It is also a possible reason why the third method is worse than the second method. The second explanation is people’s preference. For example, different people may have different attitudes towards the product but with the same expression about the size.

Conclusion, Future Work and Lesson Learned

In this study, the team compared three methods to predict partial scores from six factors of criteria set. The six factors are presented to construct a better rating system for sellers to know what the customers really focus on other than a simple star rating. The solely TF-IDF gives best prediction according to accuracy, precision, recall and f1-score jointly. Although it is a surprising result, sometimes simple method may truly better than more complicated ones. Moreover, the corpus size is not huge enough for training processes to seize every partial word and grab it into the specific bag of words, which may be the main reason that method 1 and method 2 did not perform well.

For the final prediction evaluation part, the mean absolute error scores for both methods are quite high and the R square scores for both are below 0.3, which indicates that our trained model cannot catch most explained variance from the data. However, this does not mean that we are a complete failure.

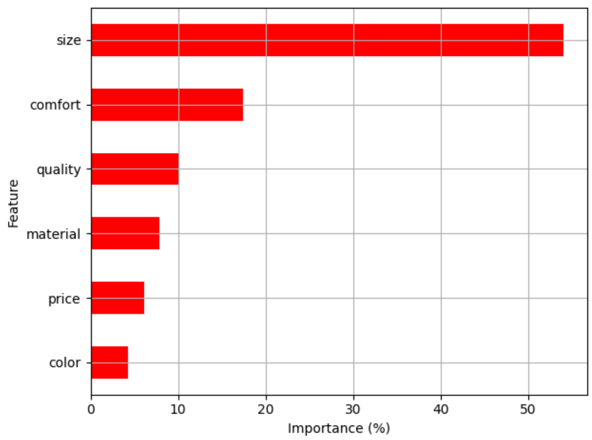


Fig. 5 Importance level of the features using predictions

The regression model aims at giving sellers an importance level of six factors that have been predicted, the results is shown in the histogram above (**Fig. 5**). It is the factor 'size' that contributes to customers' satisfaction (overall ratings) most over any other features, which takes over more than a half of the significance. It is good news because we have similar importance level of these factors using ground truth labels (**Fig. 6**).

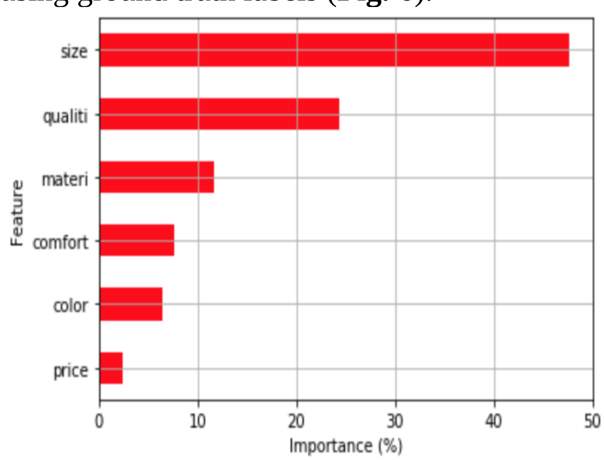


Fig. 6 Importance level of the features using ground truth

These two figures show that “Size” is the most important factor for sellers to pay attention, and “color” or “price” may not be that crucial for these sweatshirt products.

During the project study, the team took advantage of TF-IDF mothed as a solid way to vectorize text reviews into numeric features and we utilized several advanced models that we did not cover in class, like LDA and Word Embedding. Also, the significance analysis provided one practical application in distinction with results from multiple methods. We learned a lot and the future work of this study is listed as following:

- Labeling more reviews with partial scores to enrich training set, as a controlled trial to discover whether TF-IDF with word embedding has positive impacts on prediction.
- Enriching the criterion from –1-0-1 to more specific scores, for example, -1, -0.5, 0, 0.5, 1, to experiment if this helps build a better regression model.
- Improving the Word Embedding model to catch more words, even phrases accurately.
- Since the regression model is not good enough to predict the overall score, it is even much challenging to digging into the significant factors that contribute most to the overall ratings. With the regression model more precise, the significant factors could be provided to the sellers. Though this study aims to demonstrate a comprehensive process from dataset preprocessing to generating testing results, six factors are not enough for genuine business analysis, for better understanding market appetite, more factors/characters are needed for analysis.

References

"Aspect-based Sentiment Analysis – Everything You Wanted to Know." Intellica.AI., 2020. https://medium.com/@Intellica.AI/aspect-based-sentiment-analysis-everything-you-wanted-to-know-1be41572e238

Minqing Hu and Bing Liu. "Mining and Summarizing Customer Reviews." Proceeding of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2004), Aug 22-25, 2004, Seattle, Washington, USA

Appendix

Raw Data

The raw dataset was downloaded from the UCSD's open source that can be found using https://nijianmo.github.io/amazon/index.html, which includes the Amazon review dataset, metadata and other information. We chose the category "Clothes, shoes and Jewelries" and downloaded the review dataset and metadata dataset from the "Complete review data" files. The original dataset is a JSON file, and we read it in Python to transfer to a CSV file. The Data Preprocessing procedures are showed in our coding documents. The general format of row data showed as the following picture.

Clothing_Shoes_and_Jewelry.json — showing 128 MB of 13.17 GB   Open with Microsoft Excel

{"overall": 5.0, "vote": "2", "verified": true, "reviewTime": "05 4, 2014", "reviewerID": "A2IC3NZN488KWK", "asin": "0871167042", "style": {"Format:": " Paperback"}, "reviewerName": "Ruby Tulip", "reviewText": "This book has beautiful photos, good and understandable directions, and many different kinds of jewelry. Wire working and metalsmithing jewelry are covered. Highly recommend this book.", "summary": "Unique designs", "unixReviewTime": 1399161600}
{"overall": 4.0, "verified": true, "reviewTime": "04 26, 2014", "reviewerID": "A3OT9BYASFGU2X", "asin": "0871167042", "style": {"Format:": " Paperback"}, "reviewerName": "Laurie K.", "reviewText": "I love the ideas but have no access sea glass, but will experiment with other stones ... but I do love the book. Great job!", "summary": "makes you want to live by the sea...", "unixReviewTime": 1398470400}
{"overall": 5.0, "vote": "6", "verified": false, "reviewTime": "04 17, 2014", "reviewerID": "A28GK1G2KDXHRP", "asin": "0871167042", "style": {"Format:": " Paperback"}, "reviewerName": "Marie Rhoades", "reviewText": "As someone who has never possessed the manual dexterity or patience for creating jewelry, I think that I could actually make something after reading this new book.\n\nThe book is bursting with exceptional up-close photography which walks you easily through each process. The explanations of basic tools and techniques are written clearly and for those without any jewelry making experience, it all makes sense! Once you learn the how tos of working with different organic elements, the book teaches you numerous beginner and intermediate projects. Again, the photography is exceptional and the instructions easy to understand.\n\nIf you've had the desire to learn jewelry making, especially using sea glass or other natural elements, I highly recommend this book to you!", "summary": "Highly Recommend this Jewelry Making Book", "unixReviewTime": 1397692800}
{"overall": 5.0, "verified": true, "reviewTime": "04 16, 2014", "reviewerID": "A3NFXFEKW8OK0E", "asin": "0871167042", "style": {"Format:": " Paperback"}, "reviewerName": "Bobbie", "reviewText": "beautiful layout and fun pictures. as a collector of beach glass it is great to see other ways to use it beside a mason jar in the window.", "summary": "great book", "unixReviewTime": 1397606400}
{"overall": 5.0, "verified": true, "reviewTime": "04 15, 2014", "reviewerID": "A3I6G5TKBVJEK9", "asin": "0871167042", "style": {"Format:": " Paperback"}, "reviewerName": "mperez", "reviewText": "Some days ago I received my fabulous book: Organic Wire and Metal Jewelry. I'm in love with my book. Love the pictures and the instructions are simple and clear. I loved all the projects and I'm sure everyone who has purchased the book. think like me.", "summary": "Excellent", "unixReviewTime": 1397520000}
{"overall": 5.0, "verified": true, "reviewTime": "03 31, 2014", "reviewerID": "A1A7Y1M8AJWNZ8", "asin": "0871167042", "style": {"Format:": " Paperback"}, "reviewerName": "Laurie", "reviewText": "Excellent! It was just what I was looking for. Great service too! Thank you very much for the great experience!", "summary": "Great!", "unixReviewTime": 1396224000}
{"overall": 5.0, "verified": true, "reviewTime": "03 31, 2014", "reviewerID": "A30FG02C424EJ5", "asin": "0871167042", "style": {"Format:": " Paperback"}, "reviewerName": "NWCancerBaby", "reviewText": "Loved their approach in this book and that it was a \"paperback\". Easy to use and inspiring me to use up my beachcombing finds.", "summary": "Great Book", "unixReviewTime": 1396224000}
{"overall": 5.0, "verified": true, "reviewTime": "03 28, 2014", "reviewerID": "ADQQYU1UCDEWB", "asin": "0871167042", "style": {"Format:": " Paperback"}, "reviewerName": "Marlee", "reviewText": "Well written and beautifully illustrated. Features unique designs using gemstones and/or beach glass. A definite must have for anyone looking for new designs and different mediums to work with.", "summary": "Beach Glass and Wire Jewelry - what a bonus!", "unixReviewTime": 1395964800}
{"overall": 5.0, "vote": "14", "verified": true, "reviewTime": "03 23, 2014", "reviewerID": "A39YL2NXZORK56", "asin": "0871167042", "style": {"Format:": " Paperback"}, "reviewerName": "Susan Coffey", "reviewText": "I highly recommend this book. It's written for those new to the craft of jewelry making and those with more experience who want to expand their skill set. I have some jewelry making experience and consider my skills intermediate to advanced, and I'm pleased to say this book is now on my book shelf. Beth Martin and Eva Sherman are excellent teachers and it comes through in this book. They take you from the fundamentals of tools and material choices,to technique basics of a wrapped loop,to texturing,and even the use of a torch. The instructions are clearly written and easy to follow. The projects are beautiful and will offer a chance for those new to jewelry making to expand on the skills you learn as you move from what they designate as beginner to intermediate. Great purchase!", "summary": "Excellent addition to the jewelry maker's book shelf!", "unixReviewTime": 1395532800}