Product Ranking based on Customer Reviews for a Particular Aspect

Anamitra Maji
MT19112
anamitra19112@iiitd.ac.in

Pragya Dara
MT19126
pragya19126@iiitd.ac.in

Swati Verma
MT19073
swati19073@iiitd.ac.in

1 Problem Definition

Nowadays, e-commerce websites contain thousands of products, and each product includes a large number of reviews due to which it becomes quite difficult for the customers to select the product matching their expectations. This paper presents an aspect-based ranking system that can recognize important aspects based on customer reviews and rank the products by taking into account the prevalence and customer sentiment on frequent aspects. To solve the ranking problem, the Learning to Rank technique is used, which uses supervised machine learning regression models to rank the products. Normalized Discounted Cumulative Gain or NDCG is used to evaluate the quality of results obtained. In the end, the customer gets all the relevant products that satisfy their needs and helps them to choose better products.

2 Background

The rapid escalation of internet connectivity over the last few years has caused an unprecedented growth in the field of e-commerce.(1) E-stores allow customers to directly purchase goods and services from a seller over the web store, which saves a lot of time. But due to thousands of products present on the e-commerce websites, it becomes difficult for customers to choose which one to buy without the help of any salesperson. So, some shopping sites allow customers to add reviews corresponding to the product they purchase. But for a particular product, there exists a number of reviews which again creates confusion while choosing products. The main objective of this paper is to rank the products with respect to a particular aspect based on customer reviews.(2)

An item may contain a number of aspects, but a user can easily make buying decisions by seeing only important aspects. In this paper, this problem of identification of important aspects is solved by taking frequently occurring aspects in the customer's review. Classifying reviews as useful or useless is an easy task. However, ranking them according to their usefulness, is very difficult. To address ranking problems, learning to rank techniques has been used, which is one of the applications of machine learning techniques used for building ranking models.(3) Learning to rank focuses more on the relative ordering of the products rather than value of the relevance score of each product. Three machine learning regression models, Linear Regression, Logistic Regression, and Gradient Boosting Regressor, which is an ensemble regressor, are used for predicting the ranking of products. Performance evaluation is done by comparing the ranking obtained by models and the ground truth ranking. The evaluation matrix used here is Normalized Discounted Cumulative Gain or NDCG. Out of three models, Gradient boosting is outperforming with minimum error and highest NDCG values.

3 Dataset used

The dataset used for ranking is taken from Amazon customer reviews on mobile phones, which contains reviews for the top 10 mobile brands: Apple, Sony, Google, Samsung, Motorola, OnePlus, HUAWEI, Nokia, ASUS and Xiaomi. There are 720 different products and out of all the reviews, there are 4471 positive, 775 negative and 999 neutral reviews, which are shown in the graph below.

https://www.kaggle.com/grikomsn/amazon-cell-phones-reviews#20191226-reviews.csv

There are two different datasets one having information about the items like item id and their total number of reviews and another having different

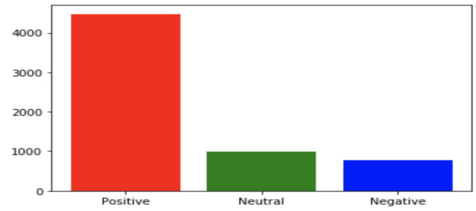

Figure 1: Distribution of reviews in the dataset

reviews corresponding to each item. The fields required for ranking are Product id, Product name, the total number of reviews for each product, and reviews for each product.

4 Literature Review

In the area of product ranking, previously various product ranking frameworks have been proposed where online customer reviews for various products have been taken, and different aspects of the product were identified automatically based on the observations that a large number of people comment on essential aspects and more reviews on important aspects significantly influence the overall rating of the product. Customer sentiments have been then analyzed using sentiment classifiers for all the aspects. A probabilistic ranking algorithm has been applied to compute the overall ranking of the product based on the aspect frequency, and customer's opinion on the important aspects making choices can become easy for customers and producers.

One of the supervised machine learning methods was demonstrated by Pang et al., where the movie reviews are classified on the basis of the overall sentiment of the review. The model classifies the reviews as positive or negative. Three supervised learning models Naive Bayes, SVM and Maximum Entropy were used, and it was found that SVM performs better than Naive Bayes and Maximum entropy by giving maximum accuracy but still none of the methods was good enough for sentiment analysis.

In other supervised techniques, an aspect extractor model was developed, which identifies important aspects in newly given reviews. For this work, Naive Bayes, Hidden Markov Model, and Maximum Entropy supervised models were used.(4) The aspect extractor was learned using the Hidden Markov Model, and it was found that even though supervised learning methods give useful results, it takes a significant amount of time while training reviews.

In contrast to supervised learning, there is unsupervised learning in which the model is not trained.(5) Association rule mining was used by Hu and Liu, which was based on the Apriori algorithm, which was used to extract frequently occurring items. These items are considered as important aspects. But as the position of words is not taken into account in association rule mining, so to eliminate non-relevant aspects, two kinds of pruning was done based on redundancy and compactness.(6) This method was less time consuming because no training is needed but has different drawbacks like the frequently occurring aspects may not be the aspects of the item. Also, the position of words is not considered, so sentiments cannot be obtained for the reviews properly.

Another unsupervised learning was used by Wu et al., which uses dependency parser to extract nouns from the reviews. Then the language model was applied to filter out the non-relevant aspects.(7) However, it was found that this method does not filter out the noise very effectively.

As most of the proposed solutions mainly focus on aspect extraction and not much on the ranking so there is a need to build a model(4) which can take essential aspects as well as rank the reviews based on aspect provided as input

5 Proposed solution sketch

The goal is to rank products on the basis of customer reviews for a particular aspect. In order to do so, we have implemented the Learning to Rank technique. Learning to Rank (LeToR) is one of the techniques of Machine Learning in which the model is trained in order to predict the ranking of the products. A LeTor Model accepts ¡query, document¿ pairs as the input and the relevance score of the document for the given pair as the label or the dependant feature (8). The model is trained on a training set in the above-mentioned format. In order to predict the ranking, the model accepts a query as the input of the testing phase and generates the ranking of the items based on the predicted relevance scores of the testing data.(9)

5.1 Queries

The first step is to generate the various queries for which the model would be trained. For ranking the products based on customer reviews for a particular aspect, the queries would be various aspects. These

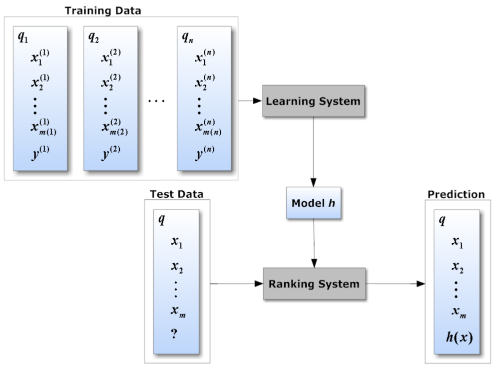

Figure 2: Learning to Rank Model

aspects were extracted from the set of reviews available. Frequently occurring nouns in the reviews have been considered as the aspects. NLTK POS tagging(10)is used to identify the nouns. In the model implemented by this paper we have chosen a set of 9 frequent nouns as aspects. These are, 'Screen', 'Battery', 'Cost', 'Volume', 'Bluetooth', 'Storage', 'Processor', 'Storage' and 'Overall'.

5.2 Documents

The reviews for the items have been treated as documents. Reviews are represented in the form of Term Frequency (Tf)- Inverse Document Frequency (Idf) vectors. (11)

5.3 Relevance Score

The ground truth is the relevance of review with respect to the queries. This paper implements the sentiment of the review with respect to the aspect of the query as the relevance score for the query, document pair(12). First, we found the aspects present in the review. For these aspects, we then found the adjectives associated using NLTK POS tagging. The sentiment of the noun-adjective pair was scaled to a value in the range[1,5] and is treated as the relevance score.

5.4 Training the Model

Learning to Rank problem can be seen as a classification or a regression problem. The task of the classification technique is to predict one of the labels 1,2,3,4,5 which would be the relevance score. The task of regression is to find a real value in the range [1,5]. This paper implements the following regression models.

1. Linear Regression: It tries to fit a linear line between all the points to find the relation between input and output features so as to generate the minimum error.

2. Logistic Regression: It tries to fit a sigmoid function between the points to find the relation between the input and output.

3. Gradient Boosting Regressor: It uses boosting to improve convert weak learners to strong learners. In order to do so, a gradient loss function is used.

5.5 Testing the Model

A query is accepted from the user. For the input query, the score is calculated for the available data records. Since the dataset has various reviews for a single product. The score is the average score for all the reviews of the item. The products are ranked on the basis of the score predicted by the model.

5.6 Evaluation of the Model

Since we are concerned with the ranking of the items Normalised Discounted Cumulative Gain (NDCG) is used as the evaluation criteria of the model. IDCG is the ideal DCG.

$$DCG = (2 * *rel)/log(i + 1, 2) \tag{1}$$

$$IDCG = max(DCG) \tag{2}$$

$$NDCG = DCG/IDCG \tag{3}$$

6 Baseline created

A set of 6000 reviews have been considered to train the model.

1. Identify nouns in the reviews. These nouns are considered as our aspects of the product. We have used a Parts-of-Speech tagger to do so.

2. Using the POS tagger we have identified various noun-adjective pairs in the review.

3. For each aspect, a score has been calculated for the review based on the sentiment of the aspect in the review. This is considered as the ground truth.

4. Various machine learning models are trained for various query document pairs.

5. For a given query,score is predicted by the model for the products.

6. Products are ranked in decreasing order of the average score predicted for the product.

7. NDGC is used to evaluate the model.

7 Results

Different NDCG scores obtained for the aspects battery and screen is shown in the table given below: It can be seen that Logistic Regression is performing better than the other two models with an NDCG score of 0.93073 and 0.9760 for battery and screen, respectively.

	Linear Regression	Logistic Regression	Gradient Boosting Regressor
NDCG for Battery	0.8907528938271746	0.9307339833042494	0.9072936299939266
NDGC for Screen	0.945602945035116	0.9760712615565023	0.9594440693849062

Figure 3: NDCG scores for battery and screen obtained by different models

A graph is plotted between NDCG scores of four different aspects: battery, screen, volume, and cost of the cell phone obtained by the three models: Linear Regression, Logistic Regression, and Gradient Boosting Regression. It can be seen that the NDCG score is coming minimum for battery and maximum for screen. One of the reasons can be that the screen might be coming in more reviews, and battery might be coming in less number of reviews, and hence the model is trained better for the screen than for battery resulting in greater NDGC score for the screen.

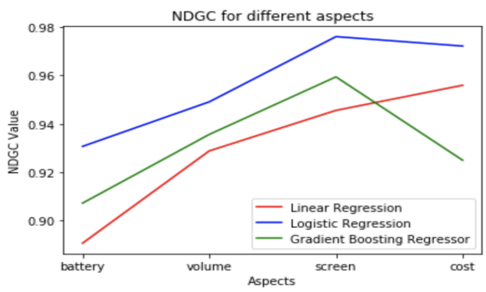

Figure 4: NDCG scores obtained for different aspects

A graph show below is plotted between NDCG scores and different number of ranked products. It can be observed that with increase in number of products obtained the NDCG value is increasing.

Mean squared error and mean absolute error is also calculated for all the models. The errors are

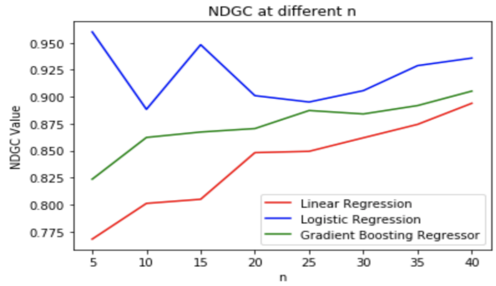

Figure 5: NDCG for different values of n

coming least for Gradient Boosting Regression because this technique uses a number of weak techniques to predict final output.Hence, the model is trained well in this regression technique and output obtained contains less error.

	Linear Regression	Logistic Regression	Gradient Boosting Regressor
Mean absolute error	1.1751485161130026	1.3601992705275332	1.1363485443243067
Mean squared error	1.9214322741448808	3.0180588915576907	1.8312280893921933

Figure 6: Mean absolute error and mean squared error for different models

8 Conclusion and Future Work

A product aspect ranking system has been successfully implemented which retrieves the top k products with respect to a particular aspect taken as input from the user. Learning to Rank techniques were used using different regression models. For now, the model was trained on a set of 9 aspects for 6000 reviews. The model had an NDCG score in the range of 80-90 for most of the aspects considered.

We would further increase the number of reviews and types of products considered in the data set if possible and perform more refined pre-processing along with more refined analysis of the reviews. The products considered so far were mobile phones only. The model can be further extended to include different types of products with different set of aspects. Different classifiers can also be used to implement Learning to Rank technique as a classification technique.

.

References

[1] N. R. Bhamre and N. N. Patil, "Aspect rating analysis based product ranking," *Proceedings - International Conference on Global Trends in Signal Processing, Information Computing and Communication, ICGTSPICC 2016*, pp. 197–202, 2017.

[2] D. Xishuang, C. Xiaodong, G. Yi, X. Zhiming, and L. Sheng, "An overview of learning to rank for information retrieval," *2009 WRI World Congress on Computer Science and Information Engineering, CSIE 2009*, vol. 3, no. January, pp. 600–606, 2009.

[3] Z. Cao, T. Qin, T. Y. Liu, M. F. Tsai, and H. Li, "Learning to rank: From pairwise approach to listwise approach," *ACM International Conference Proceeding Series*, vol. 227, pp. 129–136, 2007.

[4] C. C. Yang, Y. C. Wong, and C. P. Wei, "Classifying web review opinions for consumer product analysis," *ACM International Conference Proceeding Series*, pp. 57–63, 2009.

[5] Hu, "Mining and summarizing customer reviews," *KDD-2004 - Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 168–177, 2004.

[6] M. Hu and B. Liu, "Mining opinion features in customer reviews," *Proceedings of the National Conference on Artificial Intelligence*, pp. 755–760, 2004.

[7] Y. Wu, Q. Zhang, X. Huang, and L. Wu, "Phrase dependency parsing for opinion mining," *EMNLP 2009 - Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: A Meeting of SIGDAT, a Special Interest Group of ACL, Held in Conjunction with ACL-IJCNLP 2009*, no. August, pp. 1533–1541, 2009.

[8] A. Trotman, "Learning to rank," *Information Retrieval*, vol. 8, no. 3, pp. 359–381, 2005.

[9] R. Tikait, "Product Aspect Identification and Ranking System," vol. 4, no. 4, pp. 1127–1131, 2015.

[10] T. Saranya, "Mining Features and Ranking Products From Online Customer Reviews," *International Journal of Engineering Research & Technology*, vol. 2, no. 10, pp. 643–648, 2013.

[11] X. Cai and S. Cao, "A keyword extraction method based on learning to rank," *Proceedings - 2017 13th International Conference on Semantics, Knowledge and Grids, SKG 2017*, vol. 2018-January, pp. 194–197, 2017.

[12] S. Kuzi, S. Labhishetty, S. K. K. Santu, P. P. Joshi, and C. X. Zhai, "Analysis of adaptive training for learning to rank in information retrieval," *International Conference on Information and Knowledge Management, Proceedings*, pp. 2325–2328, 2019.