

A Product Review Writing Recommender System based on LDA and TF-IDF

Pradnya Bhagat, Jyoti D. Pawar

Goa Business School,
Goa University,
India

{dcst.pradanya, jdp}@unigoa.ac.in

Abstract. Twitter is a micro-blogging platform where people broadcast their views and opinions to fellow users in crisp messages called Tweets. However, the platform's format of restricted character limit makes it challenging for many users to express their views exhaustively. The paper proposes a recommender system to help in writing effective product review Tweets within the restricted character limit of Twitter. The approach is divided into two phases where, the first phase uses the Latent Dirichlet Allocation (LDA) algorithm to find pivotal features from the training corpus and suggests them to the users while writing new Tweets. In the second phase, the approach suggests the most appropriate opinion words to describe the respective features by using an method based on the occurrence frequency of opinion words and TF-IDF. The evaluation results show significant improvement in the quality of product review Tweets. The percentage of good reviews corresponding to a parameter such as correct usage of feature words is found to be 17.85% higher, whereas an improvement of 23.22% is reported with regard to the correct use of opinion words using the generated recommendations.

Keywords. Recommender systems, product review tweets, feature words, topics, opinion polarity, opinion intensity, latent dirichlet allocation, term frequency, inverse document frequency.

1 Introduction

Twitter is a social media platform that allows people to broadcast their views and opinions in the form of brief messages known as Tweets. Tweets are basically short messages with a limit of 280 characters.

The platform differs from most other social media platforms in a way that the relationships formed can be asymmetric [11, 9]. For instance, if user A follows user B, user A receives all Tweets from user B, whereas user B does not receive user A's Tweets unless he/she is interested in user A's Tweets (unless user B follows user A), thereby preserving his/her interests.

Hence, Twitter acts as an interest-specific service that allows users to preserve their interests and at the same time connect with people all over the world. Furthermore, since it is a free service with a worldwide presence, people all over the globe have started using the platform to share their views and benefit from other similar-minded people.

This has driven companies as well to make their presence felt on the platform and connect with customers all over the world. Many companies have dedicated customer-relation profiles on the platform to advertise new products and to communicate directly with the public [27].

A trend has also been springing up on Twitter where people tweet about various products/services used by them and exchange their ideas and opinions with other followers. Fig. 1 shows some sample Tweets maintaining the anonymity of the authors.

As seen, people effectively use the platform to lodge complaints or express satisfaction about goods and services used by them. Twitter being a broadcasting medium, the tweeted message has the power to instantly reach the destined company

My battery has stopped charging only it's taking me over four hours to fully charge my [#iphone6s](#) and drained in one hours time !! Are you listening [@Apple](#) ????

apple ups it's game with every **new** product. Don't use earphones much but these **new #Airpods** are amazing. Thanks for adding to my collection Apple. [#AppleMusic](#) enjoying with [#AppleAirpods...](#)
[instagram.com/p/BwFeMK9Fk_M/...](https://www.instagram.com/p/BwFeMK9Fk_M/)

Fig. 1. Example Tweets

and thousands of other customers, who in turn can base their purchase decisions on the read Tweet. Also, the short message format of the platform assures the companies that the messages tweeted on Twitter are concise yet informative, unlike on other social media platforms/product reviews on ecommerce websites.

This drives the companies to give immediate attention to the suggestions and complaints raised by the customers. As a result, Twitter is emerging as an effective medium for the public to make their problems heard and addressed. A study has shown that as many as 83% of customers who used Twitter for customer service had their issues addressed and fixed [17].

However, unlike other social media platforms like Facebook or Instagram, Twitter has still not succeeded in reaching the masses. One of the major reasons is attributed to the 280 character short message format, which acts as a great challenge to express one's opinion in brief. Specifically, for people with native languages other than English, expressing one's opinion in a constrained manner using English is not always feasible.

Also, addressing of technical issues specifically, in brief, requires the use of domain specific terms; a terminology which most of the users may not be well versed with. In character intensive languages like English, where the language uses several characters just to convey one word, the users easily run out of characters while expressing themselves [20].

Moreover, although many applications exist in the market to increase the users' engagement on Twitter, we observe that none specifically focus on simplifying the process of writing the Tweets.

As a result, even after being an extremely powerful medium, its use has not been realized to its complete potential. This paper is an extension of the research work presented in [4]; an approach designed to aid people in composing better quality product review Tweets for a specific domain of products by recommending them with various product-related features and feature specific opinion words.

The recommendations to help generate review Tweets for our study are mined from the reviews on Amazon.com with the domain being restricted to cell phones and its related accessories [10, 14]. Since there is no restriction in the number of words used to write a review on Amazon.com, an average Amazon Review is more elaborate and more informative than a Twitter Tweet; hence we select it as the training corpus.

The approach works in two phases; in the first phase, we extract product features and topics from the corpus using the Latent Dirichlet Allocation (LDA) [5] algorithm. Features are the attributes that describe a particular product and a group of features together form a topic.

In the second phase, the system calculates the polarity and the intensity of opinion words and suggests the appropriate opinion words corresponding to a particular feature using Term Frequency- Inverse Document Frequency (TF-IDF) statistical measures.

Polarity of a opinion word refers to the positive/negative/neutral orientation of the word and intensity refers to how intensely the word expresses its corresponding polarity. The orientation of the sentiment of a word can largely depend on the domain in which it is used. A word that expresses a positive opinion in one domain can express negative opinion in another domain.

For example, the word "unpredictable" has shown to denote positive sentiments in the movie domain; but the same word has been shown to carry a negative sentiment in the domain of automobiles [21]. Hence, in this paper we propose a method to calculate domain dependent polarity of opinion words overcoming the limitations of a generic sentiment lexicon. The experimental study is conducted on a set of 28 users.

The users were asked to write two Tweets about the cell phones they use in 280 characters as per the specifications of Twitter on a dummy website. The first Tweet is written by the users without the aid of any recommendations. In the second Tweet, the users are provided with the recommendations generated by our system.

Both the Tweets, i.e., the one written without recommendations and the one written with the help of recommendations are evaluated by two in-house subject experts, chosen by us as judges. The generated Tweets are analyzed based on the following aspects: use of correct feature words, appropriate opinion words used to describe the features, quality of the sentences constructed, and the overall helpfulness/usefulness of the composed review Tweets.

The analysis results have shown that the quality of the reviews improved in all the aspects considered for evaluation with the introduction of the generated recommendations. The remainder of this paper is organized as follows: Section 2 presents a review on the role of social media platforms in the success of brands. The section also reviews various methods employed in the literature for feature extraction and sentiment analysis from user reviews.

Section 3 describes the proposed two-step approach for feature and opinion word recommendation. Section 4 elaborates on the implementation details explaining the tools and techniques used in the implementation of the proposed method. Section 5 presents the data sets used in the study and the experimental evaluation, section 6 asserts and discusses on the results obtained and finally section 7 states the conclusion and future work.

2 Literature Review

Social media has been a continuously evolving field with every platform rolling out new features every few days to retain their position in the market. As a result, it has been extensively studied in literature trying to cover its varied aspects. [16] explores the power of social media platforms in the incessant expansion of brands by engaging the consumers on networking platforms for regular

feedback. The study shows that a brand that collaborates with its consumers online can create, or modify, its Corporate Social Relationship (CSR) strategies to fit consumer needs in a better way. [1] does a comparative study between various social media platforms and also states that the time spent by the users on Twitter is the least compared to all social networking platforms.

[6, 7] demonstrate one of the first works in guiding users while writing content on the internet. It presents the work as a browser plugin that can be used with e-commerce websites to help users write better quality reviews. The system uses association rule mining to recommend various product features to users while writing reviews.

[8] presents an incremental work of Reviewer's Assistant and reports the use of LDA [5], for detection of keywords. [3] presents a survey and comparison of some of the major methods used by researchers for feature extraction from textual product reviews. [12] proposes a probabilistic rating framework that mines user preferences from reviews and maps them to a rating scale.

The algorithm is a step towards improving Collaborative Filtering (CF) algorithms that allow text reviews to predict user preferences. [26] works on the problem of identifying feature nouns that also imply opinions. The method works by determining the polarity of feature words by identifying the opinion words that modify the feature and analyzing the surrounding context.

[22] goes beyond the zero-one polarity and tries to compare adjectives that share a similar sentiment orientation using a semi-supervised approach. The approach tries to work on the FrameNet data [2] and derives the polarity-intensity ordering among adjectives for specific categories.

The presented approach is not entirely corpus dependent, hence the approach even attempts to find the intensity of sentiment words absent in the corpus. [21] proposes a scheme to detect domain dedicated sentiment words through an application of Chi-Square test based on the difference in the counts of the word in positive and negative documents.

3 Proposed Methodology

The work addresses the challenge of assisting users in writing better quality product review Tweets using an approach based on a recommendation of two phases:

1. **Recommending Product Feature Words:**
Recommends specific feature words to describe a product.
2. **Recommending Appropriate Opinion Words:**
Recommends correct opinion words to describe the corresponding product features.

3.1 Recommending Product Feature Words

Reviews are broken into sentences, assuming that a single sentence describes a single feature or a topic. The sentences are Parts-of-Speech (POS) tagged [23] [24] to identify the various parts of speech in the review.

Next, we identify Nouns as the parts of speech that convey the product features most of the time. The challenge lies in distinguishing feature nouns from non-feature nouns and we are interested in only the former.

We proceed our experiment on an assumption that feature nouns occur in close proximity to adjectives since the users are interested in expressing their opinions about them. On the other side, this is not the case with non-feature nouns. For example, given a sentence:

My friend advised me to buy this awesome mobile because it has this stunning look and attractive features. [4] POS tagging of the above sentence would give us;

My_PRP friend_NN advised_VBD me_PRP to_TO buy_VB this_DT awesome_JJ mobile_NN because_IN it_PRP has_VBZ this_DT stunning_JJ look_NN and_CC attractive_JJ features_NNS

The nouns occurring in the above sentence are *friend*, *mobile*, *look*, and *features*. Out of these, the nouns we would be interested in are *look*, *mobile* and *features* since they belong to the domain of cell phones.

As seen, the nouns *mobile*, *look* and *features* have some adjectives associated with them since the users want to express their opinions about the

features but the noun *friend* does not have any adjectives associated with it. As it is not a feature related to mobile phones, the user is not interested in expressing an opinion on it in a review post.

We utilize this observation to differentiate candidates to feature nouns from non-feature nouns. Next, the position of the candidate feature nouns in the sentence is retained and the pre-processed file is given to the LDA algorithm. LDA [5] algorithm helps to identify latent topics from a dataset.

Topics are basically formed by a group of words related to each other. In the case of a review dataset, it is a group of feature words related to each other forming a topic. For example, feature words like *flash*, *pixel*, *front*, *back*, *digital*, *wide* etc., can be grouped together under a single latent topic *camera*.

These related feature words can be used in making the Tweet on the topic more informative. Thus, whenever a person starts composing a Tweet on any topic, the extracted relevant features of the same topic get displayed to the user as recommendations which helps in making the Tweet more informative. Algorithm 1 summarizes the process in the form of a pseudocode.

Algorithm 1 Identify Product features and feature topics

Input: Product reviews from the corpus

Output: Product features and feature topics

```

1: for each review R in the corpus do
2:   for each sentence S in R do
3:     for each word W in S do
4:       POS_Tag(W)
5:       if POS_Tag(W) == NN then
6:         if POS_Tag(W) preceded by JJ then
7:           Retain W in S as Candidate Feature Noun
8:         else
9:           Delete W
10: Product features and topics= LDA (Review Sentences with
Candidate Feature Nouns)

```

3.2 Recommend Appropriate Opinion Words

We take the POS Tagged reviews and extract all the adjectives that occur in the dataset. The review dataset has a numeric rating associated with every review given by the reviewers in addition to the review text.

The rating is in the form of stars and it ranges from 1 star to 5 stars. We group the reviews according to the number of stars associated with the review.

To find the polarity of the sentiment words, we take the adjectives found using POS tagging and find their occurrence across all five groups of reviews. Algorithm 2 elaborates on the process of classification of sentiment words according to the star rating.

Algorithm 2 Identify the domain based polarity of sentiment words

Input: POS Tagged Review text, star rating associated with the reviews. **Output:** Sentiment words and associated polarity.

```

1: for each review R in the POS tagged Reviews do
2:   for each sentence S in R do
3:     for each word W in S do
4:       if POS_Tag(W) == JJ then
5:         Add W to the Adjectives_List
6: for each review R in the corpus do
7:   if R has 1 star rating then
8:     Classify R as 1 star review.
9:   else if R has 2 star rating then
10:    R as 2 star review
11:  else if R has 3 star rating then
12:    Classify R as 3 star review.
13:  else if R has 4 star rating then
14:    Classify R as 4 star review
15:  else if R has 5 star rating then
16:    Classify R as 5 star review
17:  else
18:    Discard R
19: for Adjective A from Adjectives_List do
20:   if A occurs majority times in 4 or 5 star reviews then
21:     Label A as positive
22:   else if A occurs majority times in 1 or 2 star reviews
23:     then
24:       Label A as negative
25:     else
26:       Label A is neutral

```

Next, we identify the occurrence frequency of all positive, negative and neutral sentiment words with respect to every feature.

To find feature specific adjectives, we calculate the Term Frequency-Inverse Document Frequency (TF-IDF) [18] of sentiment adjectives in each sentiment category (positive, negative and neutral) with respect to each feature.

The sentiment adjective having the highest TF-IDF is considered to be the most intense sentiment word in that category for that feature.

Algorithm 3 Generating feature based sentiment words recommendations

Input: 1. Product Features Extracted using Algorithm 1 denoted as F ,

2. Set of Sentiment categories: Positive, Negative, Neutral Extracted using Algorithm 2 referred as S

Output: Appropriate sentiment words with intensities for features identifies.

```

1: for Feature  $F_i$  in  $F$  do
2:   for Sentiment Category  $S_i$  in  $S$  do
3:     for Sentiment Word  $SW_i$  in  $S_i$  do
4:
5:
6:
7:
8:
9:

```

$$TF(SW_{iF_i}) = \frac{\text{Frequency}(SW_{iF_i})}{\sum_{i=1}^n (SW_{iF_i})}$$

$$IDF(SW_{iF_i}) = \log_e \frac{\text{Number of features}}{\text{Features with } SW_i}$$

$$TF - IDF(SW_{iF_i}) = F(SW_{iF_i}) * IDF(SW_{F_i})$$

```

7: for Feature  $F_i$  in  $F$  do
8:   for Sentiment Category  $S_i$  in  $S$  do
9:     Sort  $SW$  based on  $TF - IDF(SW_{iF_i})$ 

```

The top 25 sentiment adjectives with highest TF-IDF scores in positive, negative and neutral sentiment categories for every features words are retained.

These resultant opinion words are the candidates to be shown as recommendations to the users while writing their opinions about respective features in the Tweets. Algorithm 3 elaborates the process in the form of a pseudocode.

4 Implementation Details

The proposed work is implemented using Python programming language [25]. The pre-processing operations and POS tagging is carried out using the Natural Language Processing ToolKit (NLTK) [13].

Gensim [19] topic modelling library is used for LDA implementation. It is an open source library for topic modelling and Natural Language Processing tasks using machine learning techniques.

		Topic 1							Topic 2						Topic 3		
		battery	device	phone	charger	charge			case	protect	color	plastic			protector	screen	part
Positive		good	great	great	great	good	Positive		great	good	bright	hard	Positive		good	good	best
		great	good	good	portable	long			good	great	great	clear			great	clear	top
		low	portable	smart	good	great			nice	adequate	nice	cheap			easy	great	good
		new	easy	new	nice	fast			protective	decent	vibrant	soft			clear	easy	hard
		portable	compatible	easy	compact	free			hard	excellent	favourite	flexible			first	responsive	soft
Negative		low	little	little	full	little	Negative		hard	little	little	hard	Negative		little	little	top
		long	full	hard	long	cheap			little	full	dark	cheap			hard	hard	hard
		dead	high	last	last	high			cheap	hard	hard	little			last	difficult	little
		little	long	full	little	full			bad	top	wrong	flimsy			cheap	fast	back
		bad	second	difficult	less	long			bulky	less	cheap	rigid			back	full	bad
Neutral		extended	other	other	much	other	Neutral		other	much	fifferent	thin	Neutral		screen	touch	most
		external	mobile	same	quick	usb			slim	more	black	thick			other	other	bottom
		spare	electronic	much	initial	dual			much	minimal	other	glossy			bubble	anti-glare	plastic
		original	multiple	more	single	same			thin	added	red	small			same	small	rubber
		extra	same	able	same	original			more	extra	more	inner			anti-glare	big	other

		Topic 4							Topic 5				
		cable	port	car	cord	tip			quality	product	time	price	sound
Positive		great	accessible	great	retractable	soft	Positive		good	great	long	great	good
		good	easy	good	long	retractable			high	good	first	good	great
		retractable	great	long	great	replaceable			great	excellent	good	low	clear
		nice	top	new	good	great			excellent	nice	hard	reasonable	better
		long	nice	easy	nice	good			better	pleased	great	cheap	excellent
Negative		short	little	long	short	little	Negative		high	cheap	long	low	lou
		little	top	little	long	hard			poor	little	hard	cheap	little
		long	second	second	little	last			low	bad	little	little	full
		cheap	full	cheap	flat				cheap	high	last	bad	high
		flat	high	last	heavy				bad	last	short	high	low
Neutral		usb	usb	rental	usb	stylus	Neutral		sound	other	same	same	tiny
		other	dual	other	other	mesh			build	sample	few	more	other
		included	other	dual	longer	pen			audio	similar	several	retail	ear
		own	micro-usb	usb	coiled	ear			higher	more	many	other	overall
		micro-usb	standard	different	extra	micro-knit			call	same	more	current	small

Fig. 2. Features and corresponding sentiment words

5 Dataset and Experimental Evaluation

The dataset for the study presented is sourced from Amazon.com and is made available by Stanford Network Analysis Project (SNAP) [10, 14]. It is a collection of Amazon.com reviews and the associated metadata on Cell Phones and its related Accessories spanning from May 1996 to July 2014. We consider only the review text from the product and the product ratings from the dataset for our experiment. A group of 28 students with computer science background was selected for the experimental study.

The age range of the selected students for the experimental study is from 22 to 24 years,

as this age group falls in the category of users found to be most enthusiastic about mobiles and electronic accessories.

Out of the 28 students, 16 were males and remaining 12 were females. The students were asked to write reviews about the mobile phone used in the form of Tweets on a dummy website created by us expressing their satisfaction or dissatisfaction about their currently used phone.

The experiment was carried out in two phases, wherein; in the first phase, the students were told to write Tweets independently without any aid.

We imposed a restriction of 280 characters as presented by Twitter.

Without Recommendations	With Recommendations
<i>I bought my ASUS Zenfone 2 laser way back in the year 2015. It is still working flawlessly. There is no hang, no virus till date. I recommend those who want to buy budgeted mobile phone under 10000 Rs.</i>	<i>Asus zenfone 2 laser battery is enough. Device is perfect for middle range mobile. Mobile touch is still working flawlessly even after 3 years. It has corning gorilla glass 3 display which helps your device screen from damage.</i>
<i>I have Redmi Note 3, Until now I'm very satisfied with my phone. It has all the features necessary which is required to carry out all the basic functions. The Only thing I want to be better is the OS. It's full of bloatware, but the hardware suffices all the need.</i>	<i>I have Redmi Note 3, which in a lot of ways is still a good phone. The Battery is lasting and comes through the whole day usage without any problems. It charges in decent pace, and the screen is very responsive. The device using experience is also good and new apps runs smoothly.</i>
<i>xaomi redmi note 4 is quite good on battery and overall performance. I will recommend you to buy this phone if ur looking for good phone in budget category.</i>	<i>Xaomi Redmi note 4 a good device with excellent battery back up, fast charger, descent sound quality, responsive screen. I will recommend u to buy this phone.</i>
<i>Samsung grand prime has a major problem of storage plz increase it.</i>	<i>battery is removable and good ,phone is good only problem is the storage</i>
<i>Superb build quality. We easily get Android update. Stock camera is not so good. Battery life is poor.</i>	<i>Battery life is fine. Phone supports fast charging but the charger is slow charger. Case is durable and protective. Color is awesome. Plastic is durable. Screen is responsive. Cable is of high quality. Price is reasonable.</i>

Fig. 3. Some example Tweets written by subjects without and with recommendations

Table 1. Features and topics discovered using LDA algorithm

Topic 1	battery	device	phone	charger	charge
Topic 2	case	protect	color	plastic	phone
Topic 3	protector	screen	thing	one	part
Topic 4	cable	port	car	cord	tip
Topic 5	quality	product	time	price	sound

In the second phase, again the students were told to write another Tweet in 280 characters, but this time they were provided with the recommendations in the form of features and opinion words mined using the system.

Two human judges were assigned to judge the quality of the Tweets written by each student independently. The judges were given 4 parameters to judge every Tweet on a two-point scale (1 - poor or average, 2 - good).

The parameters selected for evaluation are are:

1. Use of correct feature words in the Tweet.
2. Appropriate use of adjectives/sentiment words to describe the features.
3. Overall quality of the sentences constructed.
4. Helpfulness/usefulness quotient of the review.

The agreement between both the judges was validated using Cohen's Kappa [15] inter rater reliability measure.

Table 2. Sentiment words and corresponding Normalized occurrence frequencies across review categories

1 Star	2 Star	3 Star	4 Star	5 Star
less	right	bad	last	excellent
1	1	0.57269	0.622323	1
poor	second	cheap	clear	happy
1	0.452949	0.355925	0.592842	1
second	bad	clear	good	perfect
0.547051	0.297234	0.277108	0.422779	0.800259
bad	cheap	full	little	good
0.42731	0.267953	0.235836	0.301183	0.703959
cheap	last	good	nice	great
0.376122	0.170843	0.172343	0.281524	0.695143

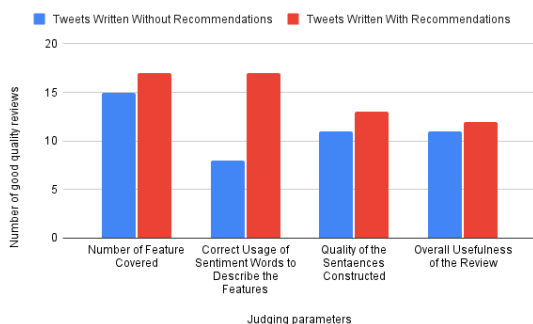


Fig. 4. Graph of ratings given by Judge 1

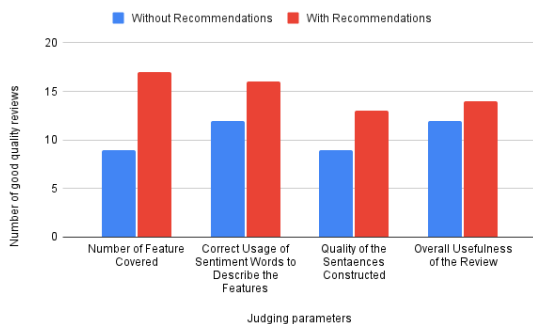


Fig. 5. Graph of ratings given by Judge 2

6 Results and Discussions

Table 1 displays the features and topics obtained using LDA algorithm. As it can be seen in the first topic, features like *battery, phone, device, charger*

and *charge* have got displayed which signifies that the features in topic 1 are related to each other i.e. they are cohesive and together form one topic. Similarly features in Topic 2 correspond to the *phone case* used for protection.

The same explanation applies to remaining topics discovered. These are the features that were displayed to the users while writing on the corresponding topics by dropping the stop words like *thing, one*, etc. using manual filtering. The second phase of the experiment deals with discovering the domain specific polarity of the sentiment words without the use of a sentiment lexicon.

The normalized occurrence frequency of first five sentiment words across review categories are shown in Table 2. As can be seen, words like *less, poor, bad* occur most of the times in 1 or 2 star reviews hence we classify them as negative sentiment words in our dataset; whereas 4 or 5 star reviews have words like *excellent, happy, perfect* and hence these words are classified as positive words.

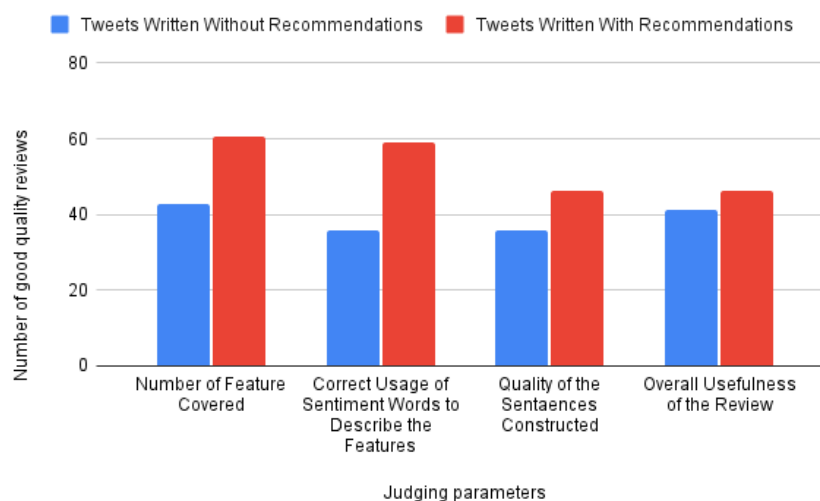
The results of the next step that deals with finding the most appropriate positive, negative and neutral opinion words for each feature using TF-IDF measure are shown in Fig. 2. For example; in Topic 2, for a feature by name *case*, the positive words discovered are *great, good, nice, protective, hard*; whereas, the negative sentiment words discovered are *hard, little, cheap, bad, bulky*.

These extracted words are recommended as sentiment words to the users to help them in explaining the respective features more effectively. The next Fig. 3 shows some example Tweets written by the subjects with and without recommendations. As can be seen, Tweets written with the help of recommendation clearly show more effective usage of feature words and sentiment words to describe them thereby making the written Tweet more effective.

The graphs in Fig. 4 and Fig. 5 show the performance of the students in writing Tweets; with and without the aid of the system as evaluated by the judges. Only good quality review Tweets were considered for plotting the graphs disregarding the poor and average quality review Tweets.

Table 3. Agreement between the judges as calculated by Cohen's Kappa

Categories	With Recommendations	Without Recommendations
Cohen's Kappa Measure	0.598361	0.515152

**Fig. 6.** Graph of average of the ratings by Judge 1 and Judge 2

This can be justified from the fact that the poor and average quality reviews have insignificant contribution in guiding other users. The parameters used for comparison of reviews were number of feature words used in a particular review, correct usage of opinion words to describe the respective features, quality of the sentences constructed and overall usefulness of the review.

It can be seen from the graph that the percentage of good quality reviews written using recommendations provided by proposed methodology are higher for all the parameters considered.

For instance, according to judge 1 in Fig. 4, only 8 people were able to make appropriate use of sentiment words while writing reviews without use of any recommendations.

Whereas, after the use of the recommender system 17 people wrote better quality reviews in terms of correct usage of sentiment words. I.e., the percentage of good reviews for a parameter such

as correct usage of sentiment words was found to be 32% higher with recommendations than without.

Similar improvement is noted in all other parameters considered for evaluation as can be seen from the graph. Also, the ratings given by Judge 2 are plotted in graph shown in Fig. 5. The product review Tweets written using the recommendations generated by the system have scored higher across all four categories.

These graphs serve as a strong evidence for the validation and usefulness of the proposed methodology. Furthermore, the validity of the ratings given by the judges is confirmed by the agreement between the two judges using Cohen's Kappa statistical measure as shown in Table 3.

Since the Cohen's Kappa score is more than 0.5 in both categories, we can clearly say that the agreement between both the judges is validated and we proceed to find the average of the ratings given by both the judges. The graph of the average ratings given by both the judges is given as Fig. 6.

As can be seen in Fig. 6 there is an overall improvement of 17.85% is observed with regard to correct usage of feature words using the recommendations. Consequently, usage of appropriate sentiment words improved a 23.22%.

Also, Both the judges found a boost of 10.72% and 5.36% with respect to the quality of the sentences constructed and the helpfulness quotient of the reviews respectively with the use of the proposed system.

7 Conclusion and Future Work

In this paper, we presented a method to help Tweeter users compose better quality product review Tweets in the restricted character limit. The method aims to generate effective and well-composed product review Tweets that are expected to help users get their Tweets desired attention and his/her problems being heard and addressed.

The approach uses LDA algorithm that helps combine related features from the training corpus and displays them as suggestions to a user while composing new Tweets. The paper also shows that the feature based polarity and intensity of the sentiment words can be calculated based on frequency of occurrence and the TF-IDF score in the dataset and we need not rely on an universal sentiment lexicon.

The Experimental results confirm that the presented method is promising to help users write better quality product review Tweets. The ratings given by both the judges and the validated inter rater agreement indicate that the Tweets written with the use of the recommendations generated by the system were of better quality than the ones written without using the system. As future work we intend to incorporate cross domain knowledge transfer in the proposed work as getting manually labelled data for every domain is not always feasible.

Acknowledgments

This publication is an outcome of the research work supported by Visvesvaraya PhD Scheme, MeitY, Govt. of India (VISPHD-MEITY-2002).

References

1. **Alhabash, S., Ma, M. (2017).** A tale of four platforms: Motivations and uses of Facebook, Twitter, Instagram, and Snapchat among college students? *Social media + society*, Vol. 3, No. 1.
2. **Baker, C. F., Fillmore, C. J., Lowe, J. B. (1998).** The Berkeley Framenet project. 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics, pp. 86–90.
3. **Bhagat, P., Pawar, J. D. (2018).** A comparative study of feature extraction methods from user reviews for recommender systems. *Proceedings of the ACM India joint international conference on data science and management of data*, pp. 325–328.
4. **Bhagat, P., Pawar, J. D. (2021).** A two-phase approach using LDA for effective domain-specific tweets conveying sentiments. In *Computational Intelligence and Machine Learning*. Springer, pp. 79–86.
5. **Blei, D. M., Ng, A. Y., Jordan, M. I. (2003).** Latent Dirichlet allocation. *Journal of Machine Learning Research*, Vol. 3, pp. 993–1022.
6. **Dong, R., McCarthy, K., O'Mahony, M., Schaal, M., Smyth, B. (2012).** First demonstration of the intelligent reviewer's assistant. *Proceedings of the 2012 ACM international conference on Intelligent User Interfaces*, pp. 337–338.
7. **Dong, R., McCarthy, K., O'Mahony, M., Schaal, M., Smyth, B. (2012).** Towards an intelligent reviewer's assistant: recommending topics to help users to write better product reviews. *Proceedings of the 2012 ACM international conference on Intelligent User Interfaces*, pp. 159–168.
8. **Dong, R., Schaal, M., O'Mahony, M. P., McCarthy, K., Smyth, B. (2012).** Unsupervised topic extraction for the reviewer's assistant. *International Conference on Innovative Techniques and Applications of Artificial Intelligence*, Springer, pp. 317–330.
9. **Gupta, P., Goel, A., Lin, J., Sharma, A., Wang, D., Zadeh, R. (2013).** Wtf: The who to follow service at twitter. *Proceedings of the 22nd international conference on World Wide Web*, pp. 505–514.
10. **He, R., McAuley, J. (2016).** Ups and downs: Modeling the visual evolution of fashion trends with one-class collaborative filtering. *Proceedings of the*

- 25th international conference on world wide web, pp. 507–517.
11. **Kwak, H., Lee, C., Park, H., Moon, S. (2010).** What is twitter, a social network or a news media? Proceedings of the 19th international conference on World wide web, pp. 591–600.
 12. **Leung, C. W.-K., Chan, S. C.-F., Chung, F.-L., Ngai, G. (2011).** A probabilistic rating inference framework for mining user preferences from reviews. *World Wide Web*, Vol. 14, No. 2, pp. 187–215.
 13. **Manning, C. D., Surdeanu, M., Bauer, J., Finkel, J. R., Bethard, S., McClosky, D. (2014).** The stanford coreNLP natural language processing toolkit. Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations, pp. 55–60.
 14. **McAuley, J., Targett, C., Shi, Q., van den Hengel, A. (2015).** Image-based recommendations on styles and substitutes. Proceedings of the 38th international ACM SIGIR conference on research and development in information retrieval, pp. 43–52.
 15. **McHugh, M. L. (2012).** Interrater reliability: the kappa statistic. *Biochemia medica*, Vol. 22, No. 3, pp. 276–282.
 16. **Okazaki, S., Plangger, K., West, D., Menéndez, H. D. (2020).** Exploring digital corporate social responsibility communications on twitter. *Journal of Business Research*, Vol. 117, pp. 675–682.
 17. **Picazo, S. (2016).** Customer service on twitter and the impact on brands.
 18. **Ramos, J. (2003).** Using TF-IDF to determine word relevance in document queries. Proceedings of the first instructional conference on machine learning, volume 242, Citeseer, pp. 29–48.
 19. **Řehřek, R., Sojka, P., others (2011).** Gensim—statistical semantics in python. Retrieved from gensim.org.
 20. **Rosen, A., Ihara, I. (2017).** Giving you more characters to express yourself.
 21. **Sharma, R., Bhattacharyya, P. (2013).** Detecting domain dedicated polar words. Proceedings of the Sixth International Joint Conference on Natural Language Processing, pp. 661–666.
 22. **Sharma, R., Gupta, M., Agarwal, A., Bhattacharyya, P. (2015).** Adjective intensity and sentiment analysis. Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, pp. 2520–2526.
 23. **Taylor, A., Marcus, M., Santorini, B. (2003).** The penn treebank: an overview. *Treebanks*, pp. 5–22.
 24. **Toutanova, K., Klein, D., Manning, C. D., Singer, Y. (2003).** Feature-rich part-of-speech tagging with a cyclic dependency network. Proceedings of the 2003 Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics, pp. 252–259.
 25. **Van Rossum, G., others (2007).** Python programming language. USENIX annual technical conference, volume 41, pp. 1–36.
 26. **Zhang, L., Liu, B. (2011).** Identifying noun product features that imply opinions. Proceedings of the 49th annual meeting of the Association for Computational Linguistics: human language technologies, pp. 575–580.
 27. **Zhang, M., Jansen, B. J., Chowdhury, A. (2011).** Business engagement on twitter: a path analysis. *Electronic Markets*, Vol. 21, No. 3, pp. 161.

Article received on 23/09/2021; accepted on 06/05/2022.

Corresponding author is Pradnya Bhagat.