Helpfulness of product reviews as a function of discrete positive and negative emotions

MSI Malik[1], Ayyaz Hussain[1,2]

[1]Department of Computer Science & Software Engineering
International Islamic University, Islamabad, Pakistan
[2]College of Computing & Informatics, Saudi Electronic University,
Riyadh, Kingdom of Saudi Arabia

{msi_id@yahoo.com, ayyaz.hussain@iiu.edu.pk}

The product review plays an important role in customer's purchase decision making process on the e-commerce websites. Emotions can significantly influence the way that reviews are processed. The importance of discrete emotions embedded in online reviews and their impact on review helpfulness is not explored intensively in prior studies. This study builds a helpfulness predictive model using deep neural network and investigates the influences of emotions that contribute to review helpfulness. We present an approach that extract novel discrete positive and negative emotion features from textual content of product reviews using NRC emotion Lexicon. In addition, the type of product, reviewer, visibility, readability, linguistics and sentiment related characteristics are also used for comparison and helpfulness prediction. The experimental results on two real-life datasets demonstrate that positive emotion features are the best predictors when individual category of feature is considered. However, negative emotion features and visibility features have comparable performance. Furthermore, the hybrid set of features with positive emotion features produce the best predictive performance for helpfulness of online reviews. The empirical evaluation finds that Trust, Joy and Anticipation (positive emotions); Anxiety and Sadness (negative emotions) are most influential emotion dimensions and have greater impact on perceived helpfulness. The findings of this study highlight the importance of emotions in online reviews and have significant implications for consumer and e-commerce retailers.

Keywords

Search products, Review helpfulness, discrete emotions, linguistic, word of mouth, Trust emotion, Amazon reviews.

Helpfulness of product reviews as a function of discrete positive and negative emotions

The product review plays an important role in customer's purchase decision making process on the e-commerce websites. Emotions can significantly influence the way that reviews are processed. The importance of discrete emotions embedded in online reviews and their impact on review helpfulness is not explored intensively in prior studies. This study builds a helpfulness predictive model using deep neural network and investigates the influences of emotions that contribute to review helpfulness. We present an approach that extract novel discrete positive and negative emotion features from textual content of product reviews using NRC emotion Lexicon. In addition, the type of product, reviewer, visibility, readability, linguistics and sentiment related characteristics are also used for comparison and helpfulness prediction. The experimental results on two real-life datasets demonstrate that positive emotion features are the best predictors when individual category of feature is considered. However, negative emotion features and visibility features have comparable performance. Furthermore, the hybrid set of features with positive emotion features produce the best predictive performance for helpfulness of online reviews. The empirical evaluation finds that Trust, Joy and Anticipation (positive emotions); Anxiety and Sadness (negative emotions) are most influential emotion dimensions and have greater impact on perceived helpfulness. The findings of this study highlight the importance of emotions in online reviews and have significant implications for consumer and e-commerce retailers.

Keywords

1. Introduction

In e-commerce, Web 2.0 provides platforms for the internet users to share their knowledge, expertise and experiences on forums, review portals, blogs and other social media websites (Bertola and Patti (2016)). These platforms facilitate the users to express their personal feelings, opinions, experiences and attitudes not only for political and economic problems but also for products and services (Samha, Li et al. (2014)). Product reviews are essential part of both traditional and electronic commerce. According to a survey, Google scholar found that there are 15,600 hits for ''product reviews" and 13,200 hits for "online reviews'' respectively (Anderson and Magruder (2012)). Consumers mostly preferred reading product reviews to gather information before purchase decisions (Zhang and Piramuthu (2016)). Therefore, an intensified level of attention is to be given by the manufacturers/retailers to online product reviews because they can be an opportunity or thread for businesses (Li and Hitt (2010); Anderson and Magruder (2012); Yan, Wang et al. (2015)). The volume of product reviews on review websites is instantly increasing that result in information overload problem (Liu, Huang et al. (2008)). However, users' opinions and emotions about particular topics or services in terms of reviews aid the customers in future purchase decisions (Xu, Xia et al. (2008); Samha, Li et al. (2014)).

Several studies described that sales of products are affected by product reviews and related factors of particular product categories under some conditions (Duan, Gu et al. (2008); Forman, Ghose et al. (2008)). Recent studies highlighted that reviewer and review characteristics such as information quantity, semantic factors, reviewer location and identity opened new dimensions in the line of research (Cao, Duan et al. (2011); Ganu, Kakodkar et al. (2013)). However Mudambi and Schuff (2010) described that future research will focus on new dimensions of reviewers' status such as designation of "top reviewer" of Amazon.com. Reviews may be classified on the basis of linguistic characteristics. A comparative opinion expresses the relationship of similarities or differences between entities where as regular opinion expresses a general opinion (Jindal and Liu (2006); Jindal and Liu (2006); Liu (2012)). The suggestive opinion is based on the directives of somebody for actions in a polite way. This reveals that various sort of opinions carry various type of aspects and behaviors that are useful in desirable decision making process (Jindal and Liu (2006); Qazi, Raj et al. (2014)).

Helpful reviews facilitate consumers in terms of significant feedback and experiences of other consumers about the product (Cao, Duan et al. (2011); Li, Huang et al. (2013)). Other benefits are 1) Reviews can be effectively summarized by filtering low-quality reviews. 2) Websites that do not use voting feature could benefit from an automated helpfulness prediction system. 3) Review ranking system can be improved with better understanding of the review helpfulness factors. This implies that e-commerce websites such as Amazon and Yelp.com, containing more helpful reviews definitely maximize consumers' satisfaction by providing significant information (Kohli, Devaraj et al. (2004); Baek, Ahn et al. (2012)). Review helpfulness is an important feature among other features associated with online products. It is computed as ratio of the number of helpful votes to the total number of votes obtained by a review. This ratio is referred as the helpfulness ratio. It is a fact that reviews are not always being consistently helpful because sentiments expressed in the reviews can have varied effect on helpfulness (Forman, Ghose et al. (2008)). The variations in helpfulness of reviews also exist across product types i.e. experience and search goods (Pan and Zhang (2011). The quality of the experience products is difficult to verify before use while search products can be judged on the basis of product specifications before purchase (Park and Lee (2009); Mudambi and Schuff (2010); Willemsen, Neijens et al. (2011)). Therefore consumers looking for experience products specifically rely on others' usage experience.

2. Motivation and Research Contributions

To draw useful information about the effect of various predictors on review helpfulness, an efficient ML algorithm is required. Deep learning is a class of ML algorithms that uses multi-layer neural networks as architecture. Contrary to traditional-based algorithms, deep neural network provides architecture that is capable to counter new problems relatively easily. Deep neural network provides training stability, generalization, and scalability with big data. Since it performs quite well in a number of diverse domains, therefore it is quickly becoming the algorithm of choice for the highest predictive accuracy.

With the exponential growth of reviews on websites, reviews are not always being consistently helpful e.g. sentiments expressed in the reviews can have varied effect on helpfulness and its importance has already been

acknowledged in the previous studies (Crowley and Hoyer (1994); Chua and Banerjee (2016)). New opportunities and challenges have arisen and there is a need to measure the emotions of the large collections of product reviews. The traditional studies utilized self reports and survey based approaches to measure the emotions of small number of candidates; however e-commerce websites today have masses of customer data to gain valuable insights. Related studies have already addressed sentiment classification (Blitzer, Dredze, and Pereira 2007), review helpfulness prediction (Ghose and Ipeirotis (2011)) and impact of sentiment on review helpfulness (Chua and Banerjee (2016)). However, our research addressed the following research questions;

- How much discrete positive and negative emotions embedded in review content affect the review helpfulness ratings?

- Which emotions type (negative or positive) strongly drive the helpfulness (quality) of online reviews?

- Does the perceived helpfulness change, when product type (experience or search goods) is considered?

- Which emotion dimensions are more influential for helpfulness among positive and negative emotions?

The aim of this research is to examine the interconnection between emotional content of product reviews and their helpfulness ratio. We are the first that introduced four discrete positive and four discrete negative emotions. The influences of positive emotions and negative emotions embedded in reviews on perceived helpfulness are investigated and a binary classification model is built for helpfulness prediction based on deep neural network. Two real life review datasets are utilized to demonstrate the utility of proposed emotion features and six different models are trained to evaluate their contribution towards helpfulness prediction. Theoretically, the results of the current research has contributed to relevant literature by providing further understanding of sentimental features of review and their influence on review helpfulness. More specifically, the study takes a step further to uncover the importance of each type of positive and negative emotions that sheds light on the empirical relationship between these variables and review helpfulness. Additionally, the findings of the study extends the results found in existing literature (i.e. Yin, Bond et al. (2014); Ullah, Zeb et al. (2015)). Major contributions of the proposed study includes

1. We are the first to investigate the influences of discrete positive and negative emotions embedded in product reviews on review helpfulness prediction.

2. This article devises a method for extracting discrete positive and negative emotions features from review content and builds a review helpfulness prediction model based on deep neural network.

3. This research facilitates e-commerce managers and retailers in minimizing the cognitive processing costs for better organization of their product reviews.

4. It has been observed that proposed emotion features have vital role in predicting review helpfulness in contrast to existing state of the art baseline techniques. Results reveal that Trust, Joy, Angry and Sadness are most influential predictors for review helpfulness.

The remainder of the paper is structured as follows. Section 3 presents the related work which is followed by the Section 4, which describes problem formulation, proposed novel features, data collection and analysis method.

Subsequently, detailed results analysis is presented in Section 5, followed by the discussion (Section 6). Section 7 explains the implications of this research and Section 8 presents conclusion and directions for future research work.

3. Related Work

3.1. Review helpfulness and measurement

Review helpfulness is formulated as the observed/perceived value of a given entity to make aware about the purchase decisions (Mudambi and Schuff (2010)). It is very difficult for consumer to identify helpful reviews among large volume of product reviews posted on websites. Amazon has pioneered the idea of helpful votes (Wan (2015)). What makes a helpful review has gained growing interest in online review platforms and researchers had discovered that various product, reviewer and review features have more impacts on helpfulness ratio (Forman, Ghose et al. (2008); Lee and Choeh (2014); Krishnamoorthy (2015)). Similarly, reviews which attract more negative ratings strongly influence the helpfulness ratio (Sen and Lerman (2007); Cao, Duan et al. (2011); Willemsen, Neijens et al. (2011)) and this effect is consistent in cross disciplinary evidence showing that bad objects will create more lasting and multifaceted effects as compared to good objects (Baumeister, Bratslavsky et al. (2001); Yang and Chao (2015)).

3.2 Qualitative/Quantitative factors contributing to review helpfulness

It is important for reviewers and marketers to understand the factors that influence the online review helpfulness. Therefore, marketers are always looking for better ways to manage and improve the digital contents of e-commerce websites. The information quality, information type, reviewer expertise, semantics and sentimental characteristics of review text are the major set of factors contributing to review helpfulness. Kim, Pantel et al. (2006) build a regression model by utilizing semantic, structural, meta-data and lexical features to predict helpfulness of online reviews. The results demonstrated that review sentiments (valence), review length, its unigrams are the important predictors. Similarly, reviewer expertise, writing style and timeliness are the influential features for review helpfulness identified by Liu, Huang et al. (2008). Danescu-Niculescu-Mizil, Kossinets et al. (2009) introduced several hypotheses to explore the dependencies of review helpfulness and demonstrated that helpfulness is not only dependent on review contents but also on relationship of evaluations to other evaluations of same product. A multilayer perceptron neural network based model was built by Lee and Choeh (2014), which is trained on product, reviewer and review features. The author demonstrated that usage of neural network model outperforms the linear regression model for helpfulness prediction.

Ngo-Ye and Sinha (2014) introduced a novel idea of reviewer engagement (RFM) features to improve the helpfulness prediction performance. The authors proposed that hybrid model consists of RFM and textual features of review produce best predictive results. However, they do not incorporate other significant features of review that are experimentally proved to be better predictors such as subjectivity, readability and meta-data features (Kim, Pantel et al. (2006); Liu, Cao et al. (2007); Ghose and Ipeirotis (2011)). Recently, Quaschning, Pandelaere et al. (2015) investigated the influence of nearby reviews by exploiting the relationship of valence consistency and review

helpfulness. Authors concluded that consistent reviews are more helpful than inconsistent reviews. By dividing product reviews into low, medium, high, spam and duplicate categories, Chen and Tseng (2011) proposed a model to evaluate the information quality of reviews for the review helpfulness. Similarly, Chen and Tseng (2011) developed a model by using readability, subjectivity and information related predictors to identify low/high quality reviews. A probability distribution model for review helpfulness prediction is proposed by Zhang and Tran (2010) and author utilized the expectation maximization algorithm. The goal of the study was to investigate the maximized distribution probability for review helpfulness.

Recently, Zhang, Wei et al. (2014) conducted various experiments to investigate the relationship between ratio of helpfulness, helpfulness distribution and confidence interval by using synthetic and real datasets. Krishnamoorthy (2015) presented the idea of extracting novel linguistic features from textual content of reviews. The author demonstrated that linguistic features present better predictive accuracy for review helpfulness and hybrid set of features produce best prediction results. A study was conducted by Huang, Chen et al. (2015) to explore the impacts of quantitative and qualitative factors of reviews and reviewers respectively such as reviewer impact, experience and cumulative helpfulness. The authors demonstrated that word count with a certain threshold is effective and reviewer experience has a varying effect on helpfulness prediction. Similarly Liu and Park (2015) build a text regression model using combination of review and reviewer features to predict the helpfulness of reviews. The feature set includes readability and valence of reviews, reviewer's identity, expertise and reputation. Other studies that utilizes regression models explore significant textual and non-textual features include (Cao, Duan et al. (2011); Ghose and Ipeirotis (2011); Pan and Zhang (2011); Korfiatis, García-Bariocanal et al. (2012); Chua and Banerjee (2015); Chua and Banerjee (2016)).

3.3 Consumer sentiments, emotion features and review helpfulness

Related to sentiments, a model was proposed by Hong, Lu et al. (2012), that uses information reliability, sentiment measure and need fulfillment based features for review classification problem. Authors used WordNet library to determine the polarity of each review and SVM based method is used for classification. The results produced are better as compared to previous studies (Kim, Pantel et al. (2006); Liu, Cao et al. (2007)). Zheng, Jin et al. (2014) presented a multi-dimensional sentiment analysis approach and propose an algorithm for sentiment analysis based on rules and a dimensional sentiment lexicon. The results demonstrated that proposed method outperforms the traditional approaches. Bai (2011) proposed a model for the purpose of extracting sentiments from reviews by capturing the dependencies among words. The results reveal that sentiments can be detected by conditional dependencies among words as well as by keywords or high-frequency words.

Tripathy, Agrawal et al. (2016) utilized ML algorithms for classification of human sentiments using n-gram techniques. The proposed method classifies the reviews and blogs into positive, negative or neutral polarity. Salehan and Kim (2016) presented a model that predicts the performance of OCR using a sentimental mining approach. The results reveal that reviews with higher levels of positive sentiment in the title receive more readerships while sentimental reviews with neutral polarity in the text are also perceived to be more helpful. An unsupervised

approach for domain independent feature specific sentiment analysis has been proposed by Shelke, Deshpande et al. (2017). Author used SentiWordNet to determine the polarity of identified features. The results reveal that proposed method outperformed the previous techniques.

Yin, Bond et al. (2014) examines the impact of anxiety and anger negative emotions for review helpfulness using seller reviews. The empirical results demonstrated that their impacts have varying effects on perceived helpfulness. The effects of emotional contents of review texts on the number of helpfulness votes are examined by Ullah, Zeb et al. (2015). The findings suggest that positive emotional content have a positive effect and negative emotional content have no effect on perceived helpfulness. More recently, emotional content of product reviews are examined by Ullah, Amblee et al. (2016) using NLP techniques. The results reveal that more extreme reviews have a greater proportion of emotional content as compared to less extreme reviews. We believe that these studies are relevant in the context of present work. The use of proposed discrete positive and negative emotions characteristics provide better prediction accuracy for review helpfulness prediction problem.

4. Proposed Method

In the first step, we formally define the problem. The examined case is a classification problem. Then, we describe the features used in our prediction model.

4.1 Problem Formulation

Suppose $\mathbf{F}$ is a T × K feature matrix where T is the total number of reviews in $\mathbf{U}$ and K is the total number of features. Where $\mathbf{U}$ is the set of reviews ($\mathbf{U} = u_1, u_2, u_3, u_T$) in the dataset and $\mathbf{F}_i$ is the feature vector for review u_i, $\mathbf{F}_i \in R^K$ for all i. Each review u_i may be illustrated as a record that comprised of following elements [P, R, C, M, H]. P represents the product features, R describes reviewer features, C is textual features of review and M describes metadata features of review. $\mathbf{H}$ denotes the helpfulness score which is computed as the ratio of helpful votes to the total votes, where H $\in$ [0, 1].

Let $\mathbf{Z}$ be the vector comprised of predicted values for all reviews and Z_i represents that review u_i is either helpful (1) or not helpful (0). Finally, Z_i for review u_i is computed as follows:

$$Z_i = \begin{bmatrix} 1, & \text{if } H_i > \beta \\ 0, & \text{otherwise} \end{bmatrix} \qquad (1)$$

β is the helpfulness threshold. Our goal here is to build a model that minimizes the prediction error of $\mathbf{Z}$ given $\mathbf{F}$. The learned model is then utilized to predict the test samples.

4.2 Model features

In this section, we introduce the novel positive and negative emotion characteristics in addition to the state of the art features. The utilized features are 1) discrete positive and negative emotions extracted from textual content of

reviews 2) product features 3) reviewer related features 4) review visibility features 5) linguistic features 6) readability features and 7) sentiment features. The final feature matrix (**F**) is constructed by combining all types of features. Following subsections will describe each type of features used in this article.

Table 1: NRC Emotion Lexicon

Emotion type	Number of words
Trust	1231
Anticipation	839
Joy	689
Surprise	534
Anxiety	1476
Anger	1247
Sadness	1191
Disgust	1058
Total	8265

4.2.1 Extraction of positive and negative emotion features

In this study, four positive and four negative discrete emotions are introduced to examine the influences of positive and negative emotions embedded in review contents on the review helpfulness. Previously, Yin, Bond et al. (2014) utilized two discrete negative emotions and investigated their influences for quality of reviews. We are the first one that considers four positive and four negative discrete emotions simultaneously and investigate their influences on review helpfulness prediction. The idea of eight basic emotions is introduced by Plutchik (1994). The positive emotions are Joy, Surprise, Anticipation and Trust where as Angry, Anxiety, Sadness and Disgust are negative discrete emotions. The National Research Council Canada (NRC) presented a large word-emotion association lexicon that is comprised of more than 8265 word types. This lexicon is referred as the NRC Emotion Lexicon (Mohammad and Turney (2013)) which categorizes the word into eight emotion dimensions proposed by Plutchik (1994). The details of NRC Emotion Lexicon, list of emotion dimensions and number of words related to each emotion dimension is described in **Table 1**.

Algorithm 1: Discrete emotions features scoring mechanism
1: for each review, u_i in U do
2:    DE_i = [0,0,0,0,0,0,0,0]   // DE_i = [trust, anticipation, joy, surprise, anxiety, anger, sadness, disgust]
3:    word-count = 0   // initialize the word count
4:    for each sentence s_j in u_i do
5:       for each word w_k in s_j do
6:             Increment word-count
7:          for each emotion dimension, e_m in NRC do
8:                if(w_k belongs to NRC[e_m])
9:                   Increment the score in DE_i[m]
10:                end if
11:          end for
12:       end for
13:    end for
14:    for each element, n in DE_i do
15:       compute DE_i[n] = (DE_i[n] * 100) / word-count
16:    end for
17:    end for
18: Compute mean and standard deviation of DE scores for product category
19: for each review u_i in U do
20: Compute z-score to obtain the final DE_i
21: end for

The measurement of scores for all discrete positive and negative emotions embedded in reviews' content is computed similar to LIWC text analysis software tool developed by the Pennebaker, Booth et al. (2007). LIWC only facilitates to compute measurement of two negative emotions (angry and anxiety), therefore a custom made software is developed to compute the measurement of each discrete emotion in reviews contents using NRC emotion lexicon. The example words related to each emotion dimension in NRC Emotion Lexicon is presented in **Table 2**. By using a generic emotion lexicon like NRC, Our scoring procedure for measurement of discrete emotions may be applicable to any digitally available lexicon.

After receiving each review text, all words are extracted from the body of review. Each word is processed at a time and emotion features are computed by adapting the scoring procedure presented in Algorithm 1. For each word, the scoring procedure (lines 7-9) searches a match to number of words associated with each emotion dimension using NRC Emotion Lexicon. If a match occurs, then value of the corresponding emotion dimension in DE (discrete emotions) features is incremented. This process is repeated for all words extracted from all sentences in the review text. The algorithm computes the eight emotions' scores/weights for each review in the dataset (lines 1-17). Then final scores of all emotion dimensions are computed for each review (lines 14-16). The structure of emotion dimensions in DE features space is described as: DE <trust, anticipation, joy, surprise, anxiety, anger, sadness, disgust >. We can say, DE_1 is a tuple which contains eight features (emotions scores) of first review. The final score of a particular emotion dimension related to first review is mathematically computed as

$$DE_1<trust> = \quad (\text{\# trust-related words} / \text{\# words in a review}) * 100 \quad (2)$$

The algorithm then computes mean and standard deviation of DE scores at product category level (line 18). At the end, z-score normalization (lines 19-21) is applied to each review to get the final feature scores (DE) of all reviews in the dataset. The z-score for review u_i and feature j is mathematically computed as

$$\text{z-score}_i \quad = \quad \frac{DE_{ij} - mean}{st.deviation} \quad (3)$$

Example:

Let us demonstrate the Discrete Emotions (DE) features score computation by an example. Consider the following three Amazon product reviews:

1. I want to start by saying Fred Flare- shipped this product very fast!! And the transaction itself was very smooth. I do however; have extreme problems with the product itself. The product is not leather, its nylon, and it sort of looks cheap? The inside material is sued, but that's only the lining for the base of the wallet. Also, the wallet part is very hard (Trust) to use. You can't really put too much in the wallet- The credit card slots are a little too snug, and there is no place for my I.D. The wallet included a small "note book" but it also doesn't fit in the wallet? I was very excited about this product, but now I feel duped. The picture made the wallet seem like it was of higher quality, and that it was user friendly, but it's not. I do not recommend this product.

The Initial DE feature values for the first review are: DE1 = <5, 3, 1, 1, 2, 1, 1, 0>. There are 148 words in the review. Final DE feature values after applying procedure (lines 14-16) is computed as DE1 = <3.38, 2.03, 0.68, 0.68, 1.35, 0.68, 0.68, 0>

2. I have to say that I was disappointed when I opened up the package containing my iPod wallet. It's cute, but not $60ish cute. First of all, it's not leather, it's nylon. The lining is indeed suede, but the photos in the product listing are misleading. I'm keeping it because the hassle of shipping it back, etc. isn't worth it. It does the job, but it wasn't what I was expecting. I feel ripped off

   The Initial DE for the second review is: DE2 = <0, 1, 0, 0, 0, 1, 1, 1>. There are 75 words in the review. Final DE feature values after applying procedure (lines 14-16) is computed as DE2 = <0, 1.33, 0, 0, 0, 1.33, 1.33, 1.33>

3. Perhaps it is my own fault for not reading more closely, or failing to question the somewhat ambiguous product description. The fact is, I thought that I would be receiving 84 plastic stays (3x28), 28 each in 3 different sizes. At $8.00, that seemed like a reasonable deal for plastic collar stays. What I got, and I confess that the product description could be interpreted this way, was a total of 28 stays in assorted sizes. When you add in the $4.00 I was charged for shipping (actual US Postal Service postage on the package was $1.84), this really is a poor deal compared to other available options. About 43 cents for each plastic collar stay? A bad choice.

   The Initial DE for the third review is: DE3 = <3, 5, 3, 3, 2, 2, 3, 1>. There are 119 words in the review. Final DE feature values after applying procedure (lines 14-16) is computed as DE3 = <2.52, 4.20, 2.52, 2.52, 1.68, 1.68, 2.52, 0.84>

Then mean and standard deviation across three reviews are computed. Finally, z-scores are computed for each review to obtain the final feature score.

Table 2: Example words related to each emotion dimension

Emotion type	Example words
Trust	Admirable, authoritative, brotherhood, comfort, pleased, commandant
Anticipation	Arouse, charitable, confession, denying, eventuality, importance, opportunity
Joy	Blessed, charity, celebrated, cheerful, comfort, encourage, festival, hope, jump
Surprise	Bang, differently, blast, bomb, chance, bonus, curiosity, deal, death, detonate
Anxiety	Warning, tramp, tearful, teasing, suppression, supremacy, socialism, snake, slam
Anger	Bear, collusion, brute, casualty, clash, denounce, defense, disobey, disaster, detest
Sadness	Alienated, anthrax, bad, bankruptcy, deterioration, cutting, dark, decay, depressed
Disgust	Cancer, dabbling, coldness, creature, crude, dislike, dirty, disallowed, disappoint

4.2.2 Product features

Previous studies show that the product features play a decisive role for the prediction of helpfulness of online reviews (Ghose and Ipeirotis (2011); Lee and Choeh (2014)). This article also considers four product features and examines their influence on the perceived helpfulness in comparison to other type of features. Following product

features are included namely, 1) Immediacy index of product 2) Amazon sale rank 3) Amazon list price of product and 4) Elapsed time from the product release date. The list of all features used in this study is presented in the **Table 3.**

The last three features are taken from the literature (Lee and Choeh (2014)) where as Immediacy index of the product is proposed in this article. Mathematical, Immediacy index is computed as:

$$Immediacy\_index \quad = \quad \text{Average review rating of the product} \quad \diagup \text{Total reviews}$$

(4)

***Immediacy Index*:** is the potential score of the product and is computed using equation (4). A product with large potential score shows that there is less number of reviews posted on the product but all reviews got high rating. Furthermore, the product with high review rating and small number of total reviews will receive high potential score because small value of denominator (number of reviews) will generate higher potential score for the product.

Table 3: List of features

Type	Feature Name	Operationalization	Notes
Positive Emotions	Joy	(# joy-related words / # words in a review) * 100	
	Surprise	(# surprise-related words / # words in a review) * 100	
	Anticipation	(# anticipation-related words / # words in a review) * 100	
	Trust	(# trust-related words / # words in a review) * 100	Range: [0, 100] Similar to LIWC
Negative Emotions	Angry	(# angry-related words / # words in a review) * 100	
	Anxiety	(# anxiety-related words / # words in a review) * 100	
	Sadness	(# sad-related words / # words in a review) * 100	
	Disgust	(# disgust-related words / # words in a review) * 100	
Product	Imm_indx	Average review rating / total number of reviews	
	Sale_rank	Amazon sale rank of product	
	Price	Amazon price	
	Elps_time	Elapsed time from the date of product release	
Reviewer	RV_Activity_Len	Time between first review and last review	
	RV_tt_reviews	Reviewer total reviews	
	RV_tt_votes	Reviewer total votes	
Visibility	Rating	Review Rating	
	Day	Which week day the review was written	
Linguistic	Noun	The number of nouns in the review text	
	Verb	The number of verbs in the review text	
	Adjective	The number of adjectives in the review text	
	Adverb	The number of adverbs in the review text	
Readability	SMOG	Simple Measure Of Gobbledygook	
	FKGL	Flesch–Kincaid Grade Level	
	ARI	Automated Readability Index	
	GFI	Gunning Fog Index	
	CLI	Coleman–Liau Index	
	FKRE	Flesch-Kincaid Reading Ease	
Sentimental	Polarity	Polarity of review text	Range: [-4, 4]
	Sentiment	Sentiment of review text	Range: [0, 8]

4.2.4 Reviewer features

Previous studies demonstrated that reviewer characteristics are the influential features to predict the helpfulness of online reviews. Total votes and Total number of reviews of the reviewer are proved to be important predictors in the literature (Lee and Choeh (2014)). In this research, we have introduced a novel reviewer characteristic namely

'Reviewer Activity Length' along with 'total votes' and 'total number of reviews'. Mathematically reviewer activity length is computed as:

$$RV\_Activity\_Length = \text{Time between first review and last review (number of days)} \qquad (5)$$

Reviewer Activity Length: It has been observed in literature that temporal dimension plays a significant role in prediction of helpfulness. Therefore, Reviewer Activity length feature is being proposed in this study which is calculated as the number of days between first review and the last review written by the reviewer. This represents that how much reviewer is actively participating in writing the recent reviews on any products. Longer activity time shows that reviewer is actively involved in writing recent reviews. Higher activity length of a reviewer corresponds to higher number of potential reviews.

4.2.5 Visibility features

Review characteristics are the most important features for the prediction of review helpfulness. In this article, we have introduced a novel feature 'Review Day' along with 'Rating of the review'. Review day is computed as the week day on which review is written.

Review day: It has been observed from review data statistics that reviews posted at weekends receive more helpful votes than reviews posted in week days. We utilize this feature to examine the effect of posted day of review on helpfulness and try to find which posted week days are influential.

4.2.6 Linguistic Features

Linguistic characteristics of textual content of review are another important predictors that can influence the helpfulness of online reviews. Utilizing Nouns, Verbs and Adjectives features are demonstrated to be influential predictors in the literature (Singh, Irani et al. (2017)). We have proposed 'Adverbs' feature along with 'Nouns', 'Verbs' and 'Adjectives'. These features are calculated directly from the review by counting their percentages in the review text.

4.2.7 Readability Features

Existing studies (Hu and Chen (2016)) demonstrated that Readability of a review text is an influential feature for review helpfulness prediction. A review with high readability is probably to be read and perceives more votes from users. We utilize following grade-level readability indexes in this study: (1) Automated Readability Index (ARI), (2) Simple Measure of Gobbledygook (SMOG), (3) Flesch–Kincaid Grade Level (FKGL), (4) Gunning Fog Index (GFI), and (5) Coleman–Liau Index (CLI), (6) Flesch-Kincaid Reading Ease (FKRE).

4.2.8 Sentiment Features

Sentiment characteristics are also proved to be important determinants for review helpfulness prediction. In this article, SentiStrength software (Thelwall, Buckley et al. (2010)) is used for the sentimental analysis of the review

text. The scores of the two features (Polarity and Sentiment) are calculated by adapting the technique used by (Stieglitz and Dang-Xuan (2013)). Mathematically, polarity and sentiment are computed as:

$$\text{Sentiment} = (\text{Positive Sentiment} - \text{Negative Sentiment}) - 2 \qquad (6)$$

$$\text{Polarity} = \text{Positive Sentiment} + \text{Negative Sentiment} \qquad (7)$$

4.3. Review data collection

In this study, we used two real-life review datasets for the demonstration of experiments. First dataset is a publicly available multi-domain sentiment analysis dataset (Blitzer, Dredze et al. (2007)), comprised of customer reviews across twenty one different product categories. The second dataset is a more recent review dataset which is obtained by crawling reviews from amazon.com. This dataset is comprised of customer reviews across thirty four different product categories. The descriptions of both the datasets are presented in **Table 4**. The distribution of helpfulness scores versus frequency of the both the datasets is presented in **Fig. 1**. Helpfulness ratio (score) varies from 0 to 1 and is represented by the x-axis whereas frequency is described by the y-axis. It shows the relative frequency of the number of reviews at different helpfulness scores. For instance, **Fig. 1** shows that both the data sets contain large collection of reviews having helpfulness range from 0.9 to 1, which indicates density of helpfulness score skewed towards the right.
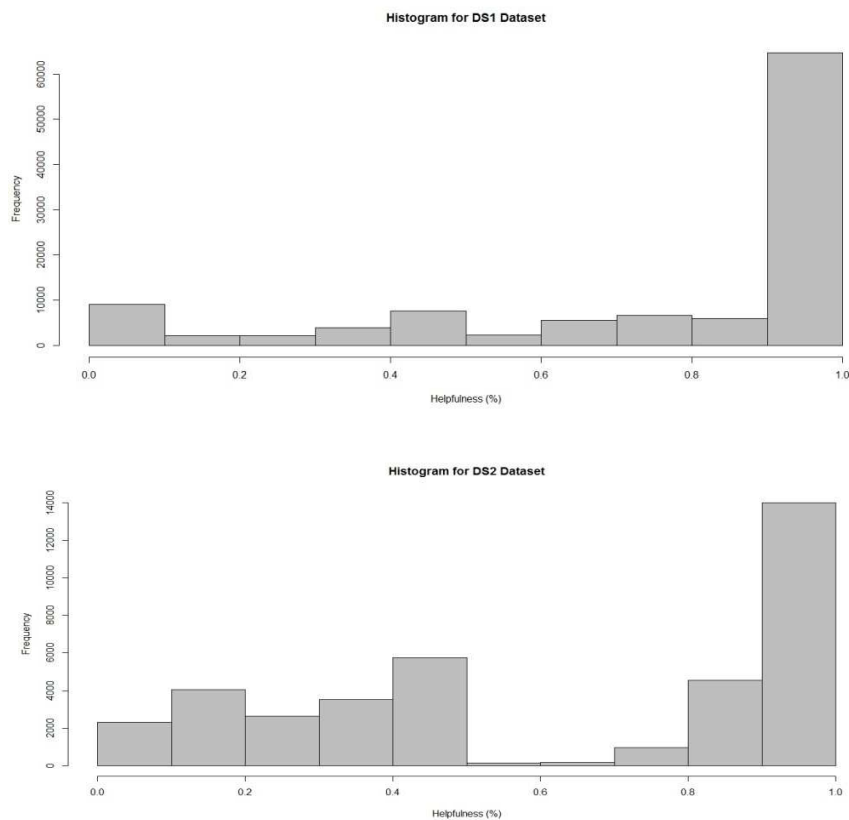
Fig. 1: Distribution of helpfulness scores.

There are total 40 different categories of products on the amazon.com and 34 product categories are considered for second dataset from which review data is crawled. We only considered those products which comprises of top-10 best sellers from each category by the amazon.com. Out of 40 categories, only 34 categories of amazon.com are selected because products of these 34 categories contain reviews with at least two helpful votes. The products from remaining 6 categories did not have required number of helpful votes that's why they are not considered. Almost 1200 product reviews are collected from each product category of amazon.com and initial data set consists of 40800 reviews. The number of different products in the sample is 3360. Then data cleaning process (Liu, Huang et al. (2008)) is applied on both datasets using following three steps, 1) Duplicate reviews are identified and removed. 2) Reviews having high helpful votes but very low total votes are less useful. Therefore only those reviews are selected which have at least ten total votes. 3) Reviews with blank text are also removed from the dataset. After data cleaning, first dataset is comprised of 32,434 reviews and second dataset is comprised of 109357 reviews. The names of various product categories for both datasets are listed in **Table 4**.

Table 4: Dataset

Dataset	No of reviews	Product categories
DS1	109357	Apparel, automotive, baby, beauty, camera and photo, cell phones and service, computer and video games, electronics, gourmet food, grocery, health and personal care, jewelry and watches, kitchen and house wares, magazines, musical instruments, office products, outdoor living, software, tools and hardware, toys and games, video
DS2	32434	Camera & Photo, Cell Phones & Accessories, Clothing, Appliances, Books, Appstore for Android, Arts, Crafts & Sewing, Computers & Accessories, Electronics, Gift Cards, Beauty, Industrial & Scientific, Automotive, Baby, Grocery & Gourmet Food, Health & Personal Care, Home & Kitchen, Home Improvement, Jewelry, Kindle Store, Kitchen & Dining, MP3 Downloads, Magazines, Movies & TV, Music, Musical Instruments, Office Products, Patio, Lawn & Garden, Pet Supplies, Shoes, Software, Sports & Outdoors, Toys & Games, Video Games, Watches

4.4 Classifiers and Evaluation Metrics

We used the popular machine learning algorithms to classify product reviews as helpful or not helpful on the basis of emotion and state of the art features. Naïve Bayes (NB), Regularized discriminant analysis (RDA), Support Vector Machine (SVM), Random Forest (RF), pruned C4.5 decision trees (PART) and Deep Neural Network (DNN) are used as learning methods to build the helpfulness predictive model. It is the first study that utilizes the Deep Neural Network method to build the predictive model for helpfulness of reviews. R statistical programming language is used for all model development and testing. It has built-in packages which provide interfaces for these algorithms. There is at most twenty nine attributes and maximum size of feature matrix will be N *29, where N is the number of tuples. All sort of experiments are conducted using standard 10-fold cross validation method. The evaluation of our predictive models is conducted using standard classification error based metrics such as f-measure and accuracy metrics.

Helpfulness of online reviews ranges from 0 to 1. In this study, we build a binary classification model for helpfulness prediction and try to identify helpful reviews from collection of product reviews. One of the important

parameter in review helpfulness prediction model is the helpfulness threshold value; β. For our experiments, β is set as 0.60 (default value). This value is selected based on its proven efficiency in previous studies (Ghose and Ipeirotis (2011); Krishnamoorthy (2015)). This indicates that reviews which have more than 60% helpful votes are labeled as helpful reviews and vice versa. The Helpfulness score **H** described in problem formulation is the ground truth and is provided with data samples. Z_i is the predicted value of a review that classifies the review as helpful or not helpful as described in **Eq. 1**.

In this study, two evaluation metrics (F measure and Accuracy) are utilized to evaluate the performance of applied learning classifiers. The mathematical definition of these metrics is described as:

$$\text{Precision} \quad = \quad \frac{TP}{TP + FP} \tag{8}$$

$$\text{Recall} \quad = \quad \text{sensitivity} \quad = \quad \frac{TP}{(TP + FN)} \tag{9}$$

$$\text{F1} \quad = \quad 2 . \quad \frac{\text{precision. recall}}{\text{precision} + \text{recall}} \tag{10}$$

$$\text{Accuracy} \quad = \quad \frac{TP + TN}{TP + TN + FP + FN} \tag{11}$$

4.4.1 Deep Neural Network (DNN)

Deep neural networks are able to learn necessary information using past instances and recommend/predict the degree of helpfulness for future instances. The helpful or not helpful classification of online reviews is estimated by the learned relationships between the predictors and helpfulness using past instances. Since deep neural networks perform quite effective in diverse problems, therefore it is gaining more popularity due to highest predictive accuracy.

In this study, we adapted five-layer feed-forward ANN that will be trained with stochastic gradient descent using back-propagation algorithm for the prediction of review helpfulness. The five layers are input, hidden1, hidden2, hidden3, and output layer. The maximum neurons that are configured in the input layers are twenty nine. Therefore same number of neurons is configured in the three hidden layers to moderately perform the task of both training and testing. The output layer contains only one neuron and each neuron in one layer is connected in the forward direction to every neuron in the following layer. The other parameters of the deep neural network are described in **Table 5**. A tanh function is used as an activation function, where Net is the sum of the weighted input and output of the f (NET) is bounded by (0, 1).

Table 5: DNN parameters

S #	Parameters	Value
01	Activation function	$\text{Tanh} = \text{f}(NET) \quad = \quad \frac{e^{NET} - e^{-NET}}{e^{NET} + e^{-NET}}$
02	Epochs	100
03	Learning rate	0.05
04	Momentum rate annealing	1.0E-6
05	Learning rate decay	0.5
06	Momentum start	.0005
07	L1(constrains the weights)	

08	Loss function	Cross entropy
09	Distribution function	Gaussian
10	Initial Weights	0.3

5. Experimental results

Series of experiments are conducted to examine the influence of proposed discrete positive and negative emotion features on the review helpfulness prediction. The experiments comprised of helpfulness prediction analysis, feature-wise analysis, relative importance of positive and negative emotions, and effect of product type (search and experience goods) on review helpfulness.

5.1 Helpfulness prediction

In the first set of experiments, the predictive models are constructed and their accuracies are compared using most commonly used 10-fold cross validation technique. Six learning algorithms are trained using the hybrid set of features (combination of product, reviewer, visibility, linguistic, readability and sentiment features) with positive emotions and negative emotions on both datasets. The results of the experiments using hybrid set of features with positive emotions are presented in **Fig. 2**. Deep neural network learning classifier delivers the best performance as compared to other five learning methods. The experiments reveal the outperformance of DNN by using positive emotions with hybrid set of features and negative emotions with hybrid set of features for both datasets as depicted in **Fig.2 & 3**. We find that positive emotions are better predictors than negative emotions and the DS1 dataset delivers relatively better performance than DS2 dataset.
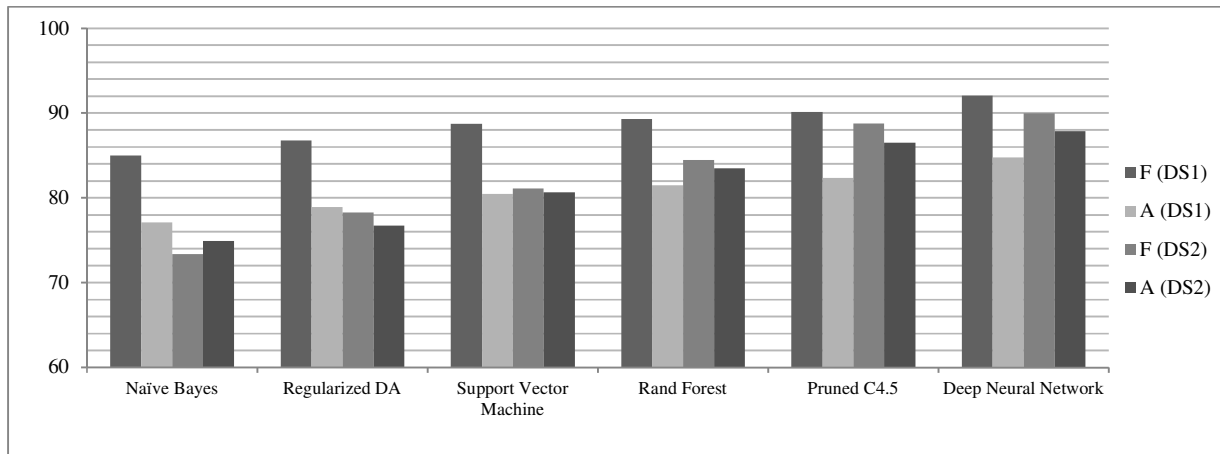

Fig 2: Prediction performance using hybrid features with positive emotions
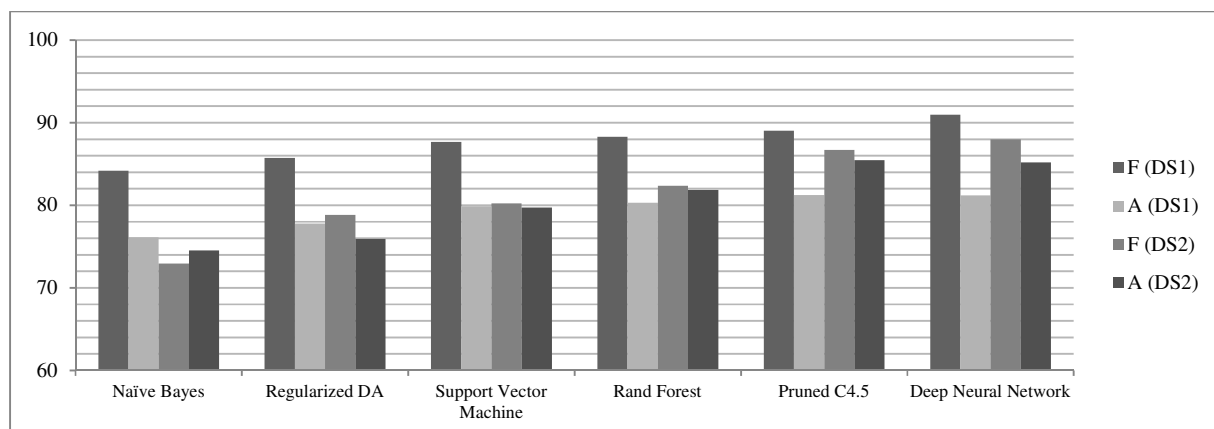
Fig 3: Prediction performance using hybrid features with negative emotions

By using positive emotions with hybrid set of features, DS1 dataset delivers 92.05 % f-measure and 84.78% accuracy and DS2 delivers 89.95 % f-measure and 87.85% accuracy with DNN classifier. Similarly, by using hybrid set of features with negative emotions, DS1 dataset outputs 90.98 % f-measure and 81.20 % accuracy and DS2 delivers 87.98 % f-measure and 85.20 % accuracy. The results of experiments clearly show that the overall performance is quite promising and proves the utility of proposed positive and negative emotions characteristics in terms of f-measure and accuracy metrics for the review helpfulness prediction. The other series of experiments are conducted using the DNN classifier as it presented the best performance.

5.2 Feature-wise analysis

Another set of experiments are conducted to analyze the impact of various types of features on the review helpfulness prediction using DNN learning algorithm using both datasets. The types of features are proposed positive and negative emotions; visibility, product, reviewer, readability, linguistic and sentiment features; hybrid set (visibility, product, reviewer, readability, linguistic and sentiment features), hybrid with negative emotions and hybrid with positive emotions respectively (**Fig. 4**). Considering individual type of feature, results demonstrate that positive emotion features offer best predictive accuracy as compared to other types of features using both datasets. Similarly, negative emotion characteristics offer slightly lower predictive performance as compared to positive emotions but better than all other types of features using both datasets. It is evident from the results (**Fig. 4**) that positive emotion and negative emotion features are most effective and influential than others (visibility, product, reviewer, readability, linguistic and sentiment features) to predict the helpfulness of online reviews.
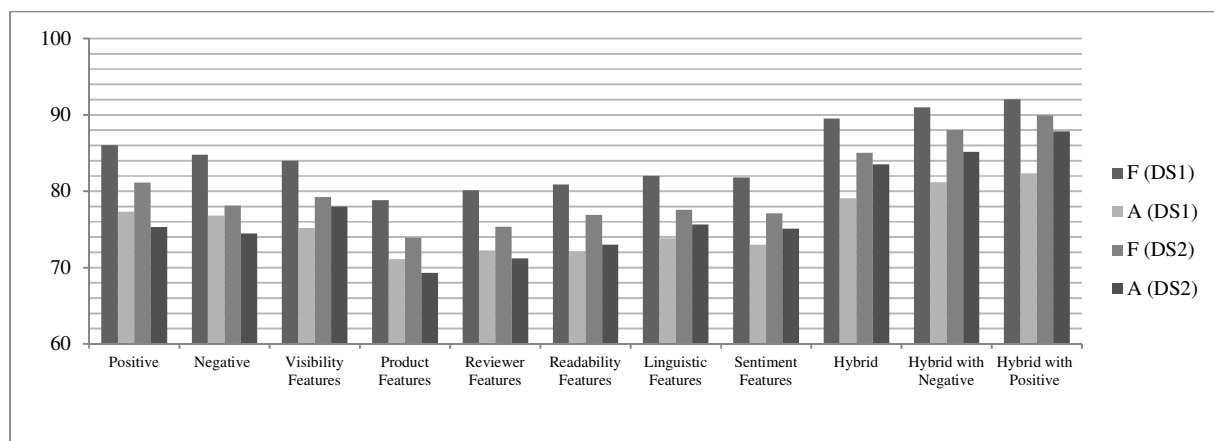
Fig 4: Feature-wise performance analysis

The hybrid set with positive emotions deliver the best performance and hybrid set with negative emotions deliver slightly lower performance as compared to former but better than hybrid set of features when combined set of features are considered using both datasets. However, DS1 dataset presents higher predictive accuracy than DS2 dataset for all sorts of experiments. The prior studies demonstrated that the type of visibility and reviewer features are most effective for the review helpfulness prediction (Korfiatis, García-Bariocanal et al. (2012); Lee and Choeh (2014)). But our experimental results show the effectiveness and outperformance of proposed positive and negative emotions with promising results using both datasets. In addition, product and reviewer features deliver the lower predictive performance as shown in **Fig. 4**.

The next set of experiments is conducted to analyze the relative importance of each positive and negative emotion characteristics on review helpfulness using both datasets. The aim of this activity is to investigate that how much each type of emotion embedded in review text influences the review helpfulness and which emotion influences most.

Using dataset DS1, we find that reviews that contain Trust emotions in the body perceive more helpful votes than reviews that contain other type of positive or negative emotions. Because online readers always seek trust-worthy reviewers and more Trust-embedded reviews represent the more genuine opinions of product reviewers, therefore Trust positive emotion is more influential. The next most influencing emotions for review helpfulness are Anticipation and Joy positive emotions as shown in **Fig. 5**. This shows that consumers expect some adventure and opportune; satisfied and pleased opinions about the products to be highlighted in the reviews. Among negative emotions, Anxiety-embedded reviews perceive more helpful votes than reviews embedded by other negative emotions. The next most influencing negative emotions for helpfulness are Sadness and Angry. The high importance of Anxiety and Sadness among negative emotions indicate that negative sentimental aspects emphasized in the reviews content derive the customers whether to purchase the products or not.
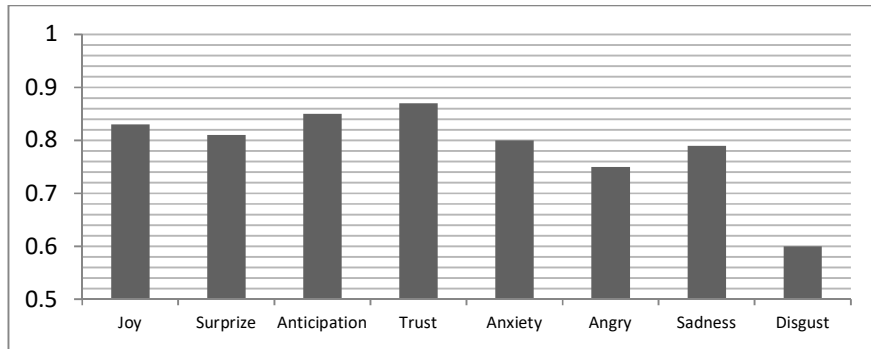
Fig. 5: Relative importance using DS1 dataset

The results of experiments to investigate the importance of positive and negative emotions using dataset DS2 for review helpfulness is slightly different. Trust positive emotion influences the review helpfulness most as shown in **Fig. 6**. However, next most influencing emotions are Joy and Anticipation positive emotions where as Surprise emotion attracts the least attention of potential buyers in review contents. Among negative emotions, Sadness emotion has a high impact on review helpfulness than other type of emotions. The next most impactful emotions are Anxiety and Angry where as Disgust emotion has some marginal effects on review helpfulness.
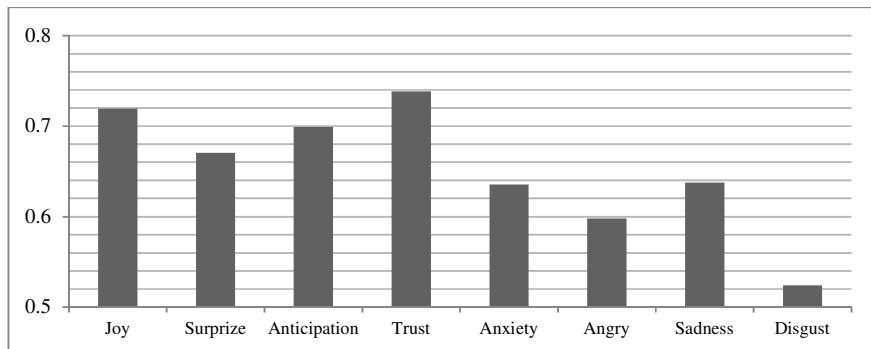

Fig. 6: Relative importance using DS2 dataset

5.3 Effect of product type on review helpfulness

Previous research reveals that experience goods and search goods are the two types of products and helpfulness of product reviews exhibits different response for both types of products (Nelson (1970); Huang and Yen (2013)). The next set of experiments are conducted to investigate the effect of product type on review helpfulness and to examine the effectiveness of proposed positive and negative emotions in case of search and experience goods for review helpfulness prediction using DS1 and DS2 datasets. The product reviews related to experience goods and search goods are separated from both datasets and deep neural network is applied to analyze the influence of various types of features on review helpfulness.

Table 6: Prediction performance for search and experience goods

Learning Method	DS1 Search		DS2 Search		DS1 Exp.		DS2 Exp.	
	F	A	F	A	F	A	F	A
Positive	90.15	82.38	82.09	77.02	83.02	74.22	78.23	74.67
Negative	86.96	79.4	79.85	74.97	81.9	73.32	75.19	71.04
Visibility Features	86.02	78.93	79.34	74.59	81.2	73.06	75.02	70.91
Product Features	80.27	72.5	74.74	70.56	75.39	68.7	70.87	67.1
Reviewer Features	82.45	74.87	76.73	72.23	78.23	71.01	72.65	68.54
Readability Features	82.9	75.23	77.34	73.79	78.9	71.89	73.85	69.23
Linguistic Features	84.56	76.33	78.89	74.14	80.22	72.96	74.79	70.42
Sentiment Features	83.24	76.24	78.34	74.03	79.32	72.23	74.08	69.70
Hybrid	92.1	84.54	86.31	84.02	85.89	77.1	82.94	77.74
Hybrid with Negative	94.34	86.3	88.37	86.14	87.1	78.84	84.55	81.23
Hybrid with Positive	95.13	88.12	90.45	87.97	89.15	79.02	86.87	83.76

It is evident from literature that a helpful review of search good is likely to have useful information about product attributes and aspects where as a helpful review of experience good is likely to contain more information about customer experiences. The results of the experiments demonstrate that proposed positive and negative emotion characteristics are more effective in case of search goods as compared to experience goods. In addition, for search good, the hybrid set of features with positive emotions and with negative emotions present best predictive performance compared to experience goods as shown in **Table 6**. However, visibility features present comparable performance to negative emotions considering both product types. Product features influence the review helpfulness least for search and experience goods. In addition, linguistic features present better performance as compared to reviewer, readability and sentiment features for both types of goods.

6. Discussions

Series of experiments and findings of the study produce strong empirical support to our research contribution. The results are robust and consistent across two real world Amazon data sets, discrete positive and negative emotions and two performance metrics. This research proposed a new helpfulness prediction model that utilized eight types of features namely, discrete positive and negative emotions; product, reviewer and visibility, readability, linguistic and sentiment type characteristics. It also presents the idea of extraction of novel discrete positive and negative emotions from review contents for review helpfulness prediction. The most interesting contribution is the embodiment of discrete positive and negative emotions with state of the art features that leads toward the best predictive performance. Our proposed helpfulness model is very effective to identify the list of helpful reviews that aid consumers in their product purchasing decisions. The proposed prediction model obtained predictive f-measure of over 89 % on two real-life review datasets using hybrid set of features with discrete positive emotions features (**Fig**

2). It also obtained predictive f-measure of over 87 % on two real-life review datasets using hybrid set of features with discrete negative emotions features (**Fig 3**).

The type of discrete positive and negative emotion features are established to be strongly influential characteristics for improving accuracy of prediction model. As a standalone model, discrete positive emotion features deliver the best performance for both datasets while discrete negative emotion features or visibility features produce comparable performance (**Fig. 4**). In addition, both (positive and negative) emotion features deliver superior performance as compared to reviewer, product, readability, linguistic and sentiment features when individual feature type is considered. The hybrid set of features with positive emotions deliver the best performance. Thus series of experiments and their results (**Fig. 2, 3 & 4**) provide the evidence of our judgments.

The deep neural network (DNN) ML algorithm is found to be the best predictive classifier and outperform the Pruned C4.5, Random Forest, SVM, RDA and NB learning algorithms as demonstrated by the results of experiments (**Fig. 2 & 3**). We are the first that applied deep neural network method to build prediction model for review helpfulness. The importance of each positive and negative emotion for review helpfulness is also investigated. The Trust positive emotion show high importance to review helpfulness, Joy and Anticipation positive emotions stand next where as Surprise emotion show least effects on review helpfulness (**Fig. 5 & 6**). Similarly, Anxiety and Sadness negative emotions show high importance to review helpfulness where as Angry stands next and Disgust negative emotion show marginal effect on the review helpfulness.

The effect of product type on review helpfulness and effectiveness of proposed positive and negative emotions features in case of search and experience goods for review helpfulness prediction are critically analyzed (**Table 6**). It is established from results that proposed positive and negative emotion characteristics are more effective for search goods as compared to experience goods (**Table 4**). In addition, for search good, the hybrid set of features with positive emotions and with negative emotions present best predictive performance compared to experience goods. However, product features influence the review helpfulness least for search and experience goods. In addition, linguistic features present better performance as compared to reviewer, readability and sentiment features for both types of goods.

7. Implications

The outcomes of this work have some valuable implications for developments in this domain. Most prior studies have adapted distinctive approaches, although drew important conclusions but inconsistent and diverse in nature. The main aim of this research is to present an effective predictive model for helpfulness of online product reviews. We are the first ones that utilize consumers' cognitive characteristics (emotional expressions) such as discrete positive and negative emotions embedded in online product reviews and investigate their influences on the perceived review helpfulness.

To predict review helpfulness is generally a complex concept. It is the first study that used deep learning based neural networks (DNN) to predict the helpfulness of online reviews. Deep neural network algorithms mimic the human brain and they require less data to train. They provide an architecture that can be adapted to new problems relatively easily. In this article, DNN based method delivers best performance as compared to PART, RF, SVM, RDA and NB learning methods. The empirical results indicate that reviews which contain more positive emotional product attributes and aspects may receive more helpful votes. In addition, reviews comprised of Trust, Joy and Anticipation (positive emotions) related opinions will receive usually large helpful votes. Similarly, reviews which contain negative emotions such as Anxiety and Sadness related words may receive more helpful votes. The negative or positive emotional contents in reviews show the opinions of users about product aspects and attributes and sentiments of users' about the experiences of product usage. However, reviews which contain only Surprise or disgust related emotional contents receive less helpful votes. This study extends the idea of the prior study (Yin, Bond et al. (2014)) and proposed novel positive and negative discrete emotions based features to investigate the effects of emotional contents embedded in online reviews on review helpfulness.

One practical implication is that retailers, managers and customers are able to view the valuable evaluations of their products and services on websites and forums. Everyone wants to find more helpful and useful reviews for evaluations because these reviews support potential values to e-commerce business (Chen, Dhanasobhon et al. (2008); Zehrer, Crotts et al. (2011)). In addition, consumers often face high cognitive processing costs while reading large collection of reviews. The contribution of this research offers facility to retailers and managers to minimize the cognitive processing costs for better organization of online product reviews.

8. Conclusions

This study analyzes the consumers' cognitive characteristics such as discrete emotions embedded in review contents and examines their influences on review helpfulness. The problem of helpfulness prediction of online reviews is addressed and a predictive model is built using deep neural network based classifier. We are the first one that utilizes distinct positive and negative emotions based feature for review helpfulness prediction. Four discrete positive emotions namely Surprise, Trust, Anticipation and Joy; and four negative emotions such as Anxiety, Disgust, Angry and Sadness are considered and their features values are extracted from review contents. In addition, the state of the art product, reviewer, visibility, readability, linguistics and sentiment features are also considered for comparison and helpfulness prediction analysis. The hybrid set of features with positive emotions produce the best predictive performance. The performance of DNN learning method is compared with PART, RF, SVM, RDA and NB learning methods and deep neural network outperforms. It is evaluated through experimental results that set of positive emotions present the best performance when individual category of feature is considered. However, negative emotion features and visibility features have comparable performance. The results also demonstrate that Trust and Joy; Anxiety and Sadness are relatively most influential positive and negative emotions respectively.

Regarding future research, there are several significant extensions of the current study. One can apply hybrid evolutionary algorithms to improve the predictive accuracy of helpfulness of reviews using same set of features. One

of the future extensions is to explore the use of influential features such as semantic and sentimental characteristics, reviewer identity and social features and examine their impacts on review helpfulness. One can explore the impacts of more emotion types for review helpfulness prediction. Another interesting research direction is the application of novel distinct positive and negative emotion features in similar domains such as to rank the products on the basis of helpful reviews and to identify the influential reviewers for their effective ranking on the basis of hybrid set of features.

References

Anderson, M. and J. Magruder. (2012). Learning from the crowd: Regression discontinuity estimates of the effects of an online review database. *The Economic Journal*, **122** (563), 957-989.

Baek, H., J. Ahn, et al. (2012). Helpfulness of online consumer reviews: Readers' objectives and review cues. *International Journal of Electronic Commerce*, **17** (2), 99-126.

Bai, X. (2011). Predicting consumer sentiments from online text. *Decision Support Systems*, **50** (4), 732-742.

Baumeister, R. F., E. Bratslavsky, et al. (2001). Bad is stronger than good. *Review of general psychology*, **5** (4), 323.

Bertola, F. and V. Patti. (2016). Ontology-based affective models to organize artworks in the social semantic web. *Information Processing & Management*, **52** (1), 139-162.

Blitzer, J., M. Dredze, et al. (2007). Biographies, bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification. *ACL*,

Cao, Q., W. Duan, et al. (2011). Exploring determinants of voting for the "helpfulness" of online user reviews: A text mining approach. *Decision Support Systems*, **50** (2), 511-521.

Crowley, A. E. and W. D. Hoyer. (1994). An integrative framework for understanding two-sided persuasion. *Journal of Consumer research*, **20** (4), 561-574.

Chen, C. C. and Y.-D. Tseng. (2011). Quality evaluation of product reviews using an information quality framework. *Decision Support Systems*, **50** (4), 755-768.

Chen, P.-Y., S. Dhanasobhon, et al. (2008). All reviews are not created equal: The disaggregate impact of reviews and reviewers at amazon. com. *Com (May 2008)*,

Chua, A. Y. and S. Banerjee. (2015). Understanding review helpfulness as a function of reviewer reputation, review rating, and review depth. *Journal of the Association for Information Science and Technology*, **66** (2), 354-362.

Chua, A. Y. and S. Banerjee. (2016). Helpfulness of user-generated reviews as a function of review sentiment, product type and information quality. *Computers in Human Behavior*, **54** 547-554.

Danescu-Niculescu-Mizil, C., G. Kossinets, et al. (2009). How opinions are received by online communities: a case study on amazon. com helpfulness votes. *Proceedings of the 18th international conference on World wide web*, ACM.

Duan, W., B. Gu, et al. (2008). The dynamics of online word-of-mouth and product sales—An empirical investigation of the movie industry. *Journal of retailing*, **84** (2), 233-242.

Ganu, G., Y. Kakodkar, et al. (2013). Improving the quality of predictions using textual information in online user reviews. *Information Systems*, **38** (1), 1-15.

Ghose, A. and P. G. Ipeirotis. (2011). Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics. *IEEE Transactions on Knowledge and Data Engineering*, **23** (10), 1498-1512.

Forman, C., A. Ghose, et al. (2008). Examining the relationship between reviews and sales: The role of reviewer identity disclosure in electronic markets. *Information Systems Research*, **19** (3), 291-313.

Ngo-Ye, T. L. and A. P. Sinha. (2014). The influence of reviewer engagement characteristics on online review helpfulness: A text regression model. *Decision Support Systems*, **61** 47-58.

Hong, Y., J. Lu, et al. (2012). What reviews are satisfactory: novel features for automatic helpfulness voting. *Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval*, ACM.

Hu, Y.-H. and K. Chen. (2016). Predicting hotel review helpfulness: The impact of review visibility, and interaction between hotel stars and review ratings. *International Journal of Information Management*, **36** (6), 929-944.

Huang, A. H., K. Chen, et al. (2015). A study of factors that contribute to online review helpfulness. *Computers in Human Behavior*, **48** 17-27.

Huang, A. H. and D. C. Yen. (2013). Predicting the helpfulness of online reviews—a replication. *International Journal of Human-Computer Interaction*, **29** (2), 129-138.

Jindal, N. and B. Liu. (2006). Identifying comparative sentences in text documents. *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval*, ACM.

Jindal, N. and B. Liu. (2006). Mining comparative sentences and relations. *AAAI*,

Kim, S.-M., P. Pantel, et al. (2006). Automatically assessing review helpfulness. *Proceedings of the 2006 Conference on empirical methods in natural language processing*, Association for Computational Linguistics.

Kohli, R., S. Devaraj, et al. (2004). Understanding determinants of online consumer satisfaction: A decision process perspective. *Journal of Management Information Systems*, **21** (1), 115-136.

Korfiatis, N., E. García-Bariocanal, et al. (2012). Evaluating content quality and helpfulness of online product reviews: The interplay of review helpfulness vs. review content. *Electronic Commerce Research and Applications*, **11** (3), 205-217.

Krishnamoorthy, S. (2015). Linguistic features for review helpfulness prediction. *Expert Systems with Applications*, **42** (7), 3751-3759.

Lee, S. and J. Y. Choeh. (2014). Predicting the helpfulness of online reviews using multilayer perceptron neural networks. *Expert Systems with Applications*, **41** (6), 3041-3046.

Li, M., L. Huang, et al. (2013). Helpfulness of online product reviews as seen by consumers: Source and content features. *International Journal of Electronic Commerce*, **17** (4), 101-136.

Li, X. and L. M. Hitt. (2010). Price effects in online product reviews: an analytical model and empirical analysis. *MIS quarterly*, 809-831.

Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*, **5** (1), 1-167.

Liu, J., Y. Cao, et al. (2007). Low-Quality Product Review Detection in Opinion Summarization. *EMNLP-CoNLL*,

Liu, Y., X. Huang, et al. (2008). Modeling and predicting the helpfulness of online reviews. *2008 Eighth IEEE International Conference on Data Mining*, IEEE.

Liu, Z. and S. Park. (2015). What makes a useful online review? Implication for travel product websites. *Tourism Management*, **47** 140-151.

Mohammad, S. M. and P. D. Turney. (2013). Nrc emotion lexicon.NRC Technical Report.

Mudambi, S. M. and D. Schuff. (2010). What makes a helpful review? A study of customer reviews on Amazon. com. *MIS quarterly*, **34** (1), 185-200.

Nelson, P. (1970). Information and consumer behavior. *Journal of political economy*, **78** (2), 311-329.

Pan, Y. and J. Q. Zhang. (2011). Born unequal: a study of the helpfulness of user-generated product reviews. *Journal of Retailing*, **87** (4), 598-612.

Pennebaker, J. W., R. J. Booth, et al. (2007). LIWC2007: Linguistic inquiry and word count. *Austin, Texas: liwc. net*,

Plutchik, R. (1994). The psychology and biology of emotion, HarperCollins College Publishers.

Qazi, A., R. G. Raj, et al. (2014). Enhancing business intelligence by means of suggestive reviews. *The Scientific World Journal*, **2014**

Quaschning, S., M. Pandelaere, et al. (2015). When consistency matters: The effect of valence consistency on review helpfulness. *Journal of Computer-Mediated Communication*, **20** (2), 136-152.

Salehan, M. and D. J. Kim. (2016). Predicting the performance of online consumer reviews: A sentiment mining approach to big data analytics. *Decision Support Systems*, **81** 30-40.

Samha, A. K., Y. Li, et al. (2014). Aspect-based opinion extraction from customer reviews. *arXiv preprint arXiv:1404.1982*,

Sen, S. and D. Lerman. (2007). Why are you telling me this? An examination into negative consumer reviews on the web. *Journal of interactive marketing*, **21** (4), 76-94.

Shelke, N., S. Deshpande, et al. (2017). Domain Independent Approach for Aspect Oriented Sentiment Analysis for Product Reviews. *Proceedings of the 5th International Conference on Frontiers in Intelligent Computing: Theory and Applications*, Springer.

Singh, J. P., S. Irani, et al. (2017). Predicting the "helpfulness" of online consumer reviews. *Journal of Business Research*, **70** 346-355.

Stieglitz, S. and L. Dang-Xuan. (2013). Emotions and information diffusion in social media—sentiment of microblogs and sharing behavior. *Journal of Management Information Systems*, **29** (4), 217-248.

Tripathy, A., A. Agrawal, et al. (2016). Classification of sentiment reviews using n-gram machine learning approach. *Expert Systems with Applications*, **57** 117-126.

Thelwall, M., K. Buckley, et al. (2010). Sentiment strength detection in short informal text. *Journal of the American Society for Information Science and Technology*, **61** (12), 2544-2558.

Ullah, R., N. Amblee, et al. (2016). From valence to emotions: Exploring the distribution of emotions in online product reviews. *Decision Support Systems*, **81** 41-53.

Ullah, R., A. Zeb, et al. (2015). The impact of emotions on the helpfulness of movie reviews. *Journal of applied research and technology*, **13** (3), 359-363.

Wan, Y. (2015). The Matthew Effect in social commerce. *Electronic Markets*, **25** (4), 313-324.

Willemsen, L. M., P. C. Neijens, et al. (2011). "Highly Recommended!" The content characteristics and perceived usefulness of online consumer reviews. *Journal of Computer-Mediated Communication*, **17** (1), 19-38.

Xu, R., Y. Xia, et al. (2008). Opinion Annotation in On-line Chinese Product Reviews. *LREC*,

Yang, H.-L. and A. F. Chao. (2015). Sentiment analysis for Chinese reviews of movies in multi-genre based on morpheme-based features and collocations. *Information Systems Frontiers*, **17** (6), 1335-1352.

Yan, X., J. Wang, et al. (2015). Customer revisit intention to restaurants: Evidence from online reviews. *Information Systems Frontiers*, **17** (3), 645-657.

Yin, D., S. Bond, et al. (2014). Anxious or angry? Effects of discrete emotions on the perceived helpfulness of online reviews. *Mis Quarterly*, **38** (2), 539-560.

Zehrer, A., J. C. Crotts, et al. (2011). The perceived usefulness of blog postings: An extension of the expectancy-disconfirmation paradigm. *Tourism Management*, **32** (1), 106-113.

Zhang, J. and S. Piramuthu. (2016). Product recommendation with latent review topics. *Information Systems Frontiers*, 1-9.

Zhang, R. and T. T. Tran. (2010). A novel approach for recommending ranked user-generated reviews. *Canadian Conference on Artificial Intelligence*, Springer.

Zhang, Z., Q. Wei, et al. (2014). Estimating online review helpfulness with probabilistic distribution and confidence. Foundations and Applications of Intelligent Systems, Springer**:** 411-420.

Zheng, L., P. Jin, et al. (2014). Multi-dimensional sentiment analysis for large-scale E-commerce reviews. *International Conference on Database and Expert Systems Applications*, Springer.