Information Processing and Management 58 (2021) 102538

Contents lists available at ScienceDirect

Information Processing and Management

journal homepage: www.elsevier.com/locate/infoproman

Helpfulness of online consumer reviews: A multi-perspective approach

Satanik Mitra *, Mamata Jenamani

Department of Industrial & Systems Engineering, IIT Kharagpur, India

A R T I C L E I N F O

Keywords:
Helpfulness of review
Perspectives of helpfulness, Convolutional neural network
LSTM
Regression analysis
Deep learning

A B S T R A C T

Helpful online reviews crave the attention of many researchers as it significantly affects purchase decision. However, consumers' perception of helpfulness remains an open problem due to a lack of semantic analysis of review content and unreliable voting mechanism. In this work, we propose three qualitative perspectives considering both semantic and syntactic features of review content - *lexical, sequential* and *structural* to assess helpfulness. N-gram based semantic relation among words is explored with a D-CNN model, to predict helpfulness from *lexical* perspective. *Sequential* perspective is analysed with LSTM model, which predict helpfulness by comprehending sequence of words. *Structural* perspective is addressed with fourteen syntactic statistical features and predict helpfulness of review. These three models of qualitative perspective trained with "X of Y" ratio of helpfulness voting. Now, to decimate the unreliability of helpfulness voting mechanism and unveil the human perception of helpfulness, the manual scoring approach is implemented over a sample of reviews. With experimentation, we show that there exists a linear relationship among the perspectives with the human perceived helpfulness score. It is observed that all these perspectives have an impact on consumers' perception of helpfulness of a review. Five different product category of a benchmark dataset has been used for experimentation. A sample of 2000 reviews from five different categories has been used for human scoring of helpfulness. Finally, we estimate the weights of each of the perspectives of consumers' perception of helpfulness from online reviews and discuss the significant theoretical and practical implications.

1. Introduction

The reliance of consumers on the opinion of experienced customers in purchase decisions has led to the growth and importance of online review channels. As per studies, around 92% of consumers follow online consumer reviews before purchase (Lu, Wu & Tseng, 2018). Online reviews have become the electronic word of mouth and a tremendous resource that influences consumer decisions (Duan, Gu & Whinston, 2008; Park, Lee & Han, 2007; Zhu & Zhang, 2010). These reviews are the wellspring for gaining consumers' preferences, feeling and behavior towards the brand (Chatzipanagiotou, Veloutsou & Christodoulides, 2016; Chen, 2020; Keller, 2016; Mitra & Jenamani, 2020). However, in the vast bushel of information, it is quite time-consuming for consumers to flip through all the reviews with disproportionate qualities to make purchase decisions. A review which evaluates the product and leads to purchase decision, in other words, its usefulness or cues, is usually measured by "helpfulness" (Filieri, 2015; Wang, Li, Ye & Law, 2016). An informative and helpful review narrow down the cognitive stress; therefore consumers are more attuned by reviews which are

* Corresponding author.
E-mail addresses: satanikmitra@iitkgp.ac.in (S. Mitra), mj@iem.iitkgp.ac.in (M. Jenamani).

https://doi.org/10.1016/j.ipm.2021.102538
Received 16 October 2020; Received in revised form 7 January 2021; Accepted 31 January 2021
Available online 8 February 2021
0306-4573/

perceived as helpful (Lu et al., 2018; Malik & Hussain, 2018; Ren & Hong, 2019). It is beneficial for consumers and profitable for vendors to promote helpful reviews (Agnihotri & Bhattacharya, 2016; Fan, Feng, Guo, Sun & Li, 2019).

Despite the extensive information content, due to the unstructured nature, assessment of helpfulness of online reviews remains a challenging task, and it craves the attention of researchers across the globe (Duan et al., 2008; Singh et al., 2017; Tractinsky & Srinivasan, 2001; Weiss, Indurkhya, Zhang, & Damerau, 2005; Xu, Wang, Li & Haghighi, 2017). In literature, helpfulness is analysed predominantly with two approaches. Firstly, by involving shallow syntactic features and their statistics. Handcrafted features like structural statistics, the sentiment of the reviews, argumentative features are used to predict or classify review helpfulness (Chua & Banerjee, 2015; Diaz & Ng, 2018; Krishnamoorthy, 2015; Malik & Hussain, 2017; Singh et al., 2017). Although handcrafted features show promising results, it is criticized being costly in terms of time involvement (Olatunji, Li & Lam, 2020). Secondly, with deep learning-based automatic feature extraction process (Fan et al., 2019; Saumya, Singh & Dwivedi, 2020).

As per the uncertainty reduction theory of cognitive psychology, informational cues expressed in the review content influence the perception of helpfulness of a review (Daft & Lengel, 1986; Forman, Ghose, & Wiesenfeld, 2008). These informational cues are captured by the semantics of words and context in various sentences (Batini, 2016; Ferstl & Cramon, 2001). A subjective comprehension of review text results in helpfulness perception (Wang, Tang, & Kim, 2019). Hence, analysis of the content, intrigue connections among the words and models eliciting both syntactic and semantic perspectives, are necessary. In addition, the comparison among the existing approaches is difficult as models trained on different datasets, crawled from various e-commerce sites and which are not publicly available (Diaz & Ng, 2018; Malik & Hussain, 2017; Singh et al., 2017).

The existing works are focused on developing models and predict helpfulness of review based on user votes following "X of Y" approach. "X of Y" is the ratio of X number of votes received and "Y" number of participants in the voting (Fan et al., 2019; Malik & Hussain, 2018; Qazi et al., 2016; Saumya, Singh, Baabdullah, Rana & Dwivedi, 2018). The relevance of these votes is being questioned to assess the quality of the information delivered in the content of reviews (Cao, Duan & Gan, 2011; Diaz & Ng, 2018; G. Ren & Hong, 2019). The problem involves in this approach is, it does not say anything about how many votes are enough to consider a review to be helpful. Hence, for reliable evaluation of review helpfulness, it is necessary to label reviews with human scoring. On the other hand, labelling a massive amount of review is cumbersome for human annotators. Hence, it is essential to establish a relation between helpfulness votes and human scoring, such that the helpfulness score of a review can be inferred.

1.1. Research questions

Earlier researches mostly develop statistical features or consider automated feature extraction approaches to predict helpfulness from consumer voting mechanism (Fan et al., 2019; Krishnamoorthy, 2015; Saumya et al., 2018; Singh et al., 2017). However, due to the lack of reliability of consumer voting mechanism, these studies do not reflect consumers' perception of helpfulness. Besides, in reviews, information not only reveals through syntactic features but also significantly with the semantics of words and its sequences (Chen et al., 2019; Fan et al., 2019; Olatunji et al., 2020). As a result, syntactic statistics along with semantics of words and sequence will be useful to represent the perception of helpfulness. Hence, the relation between helpfulness votes and perceived helpfulness needs to be established. In a combination of these ideas with the existing gaps in the literature, the following research questions are formed -

RQ1. How the semantic of words, sequences of words, and syntactic statistics reflects helpfulness of a review based on helpfulness voting?
RQ2. How semantic of words, sequences of words, and syntactic statistics are related to the consumers' perception of helpfulness?
RQ3. How far helpfulness voting mechanism is enough to capture consumers' perception of helpfulness?
RQ4. How to get a reliable and content based helpfulness score of a review?

1.2. Our contributions

To address the questions discussed above, this research presents a combinational approach which includes both syntactic and semantic aspect of the text to assess the perception of helpfulness in the consumers' mind. Based on the process of comprehending a textual content, the study presents helpfulness of review from three qualitative perspectives such as lexical, sequential and structural. These perspectives are realized with deep learning and machine learning-based models involving word-based semantics, sequences and syntactic statistics. Addressing the unreliability of helpfulness voting, we apply human scoring on a sample of reviews. A regression model is implemented over the predictions of three perspectives and human scoring to analyse the contribution of each of the perspectives on human perception of helpfulness. Finally, the helpfulness score is assigned to the reviews which reflects the consumers' perception of helpfulness. The contributions of this study are–

I This research proposes an approach to encountering online review helpfulness problem from three qualitative perspectives - lexical, sequential and structural, with best of our knowledge, this is the first time in the literature where helpfulness of reviews is analysed from these perspectives.
II The study implements a Dual CNN (D-CNN) model for analysing lexical perspective. It replicates the effect of n-gram language models with the varying kernel size parameter. We predict the helpfulness of review from lexical perspective from an average of bigram, trigram, and four-gram models. LSTM captures the word to word semantic sequential connections prevailing in the review which is used for sequential perspective.

III Structural perspective covers syntactic statistics of a review apart from the semantics of words. Motivated by previous researches, fourteen structural features represent the structural perspective in this model.
IV In this work, publicly available Amazon reviews (He & McAuley, 2016; McAuley, Targett, Shi & Van Den Hengel, 2015) is used as a dataset.
V This work addresses the discrepancies involved in the helpfulness voting mechanism by scoring reviews with human volunteers.
VI Helpfulness is represented as a linear combination of lexical, sequential, and structural perspective and estimates their weights concerning human scores, which can assign context-based true helpfulness score of a review.

From the contributions of this study, several theoretical and practical implications can be drawn. Here, we present the theoretical and practical implications. The theoretical implications are -

I The perception of helpfulness of a review is conveyed through three review content-based, qualitative perspectives involving words, sequential semantics and syntactical statistics which is not attempted earlier.
II The problem involves with "X of Y" ratio has been addressed which is less discussed in the literature (Fan et al., 2019; Malik & Hussain, 2017; Singh et al., 2017). Moreover, human annotated helpfulness score gives better comprehension about the perception of helpfulness.
III We further assess the weights of each of the perspective over human annotated reviews. These weights will assign reliable helpfulness scores.
IV This research concentrates on the semantics of reviews and fairly differ from the studies where syntactic features are prevalent (Chua & Banerjee, 2015; Malik & Hussain, 2018; Saumya et al., 2018; Weathers, Swain & Grover, 2015).
V Publicly available dataset has been used in this work which is suggested in prior literature from a comparative point of view (Diaz & Ng, 2018; Olatunji et al., 2020).

The practical implications inferred from the study are –

I This research combines both semantic and syntactic aspects of reviews. For syntactic analysis, a smaller set of features has been used in the study (Malik & Hussain, 2017; Saumya et al., 2018; Singh et al., 2017).
II In contrast to earlier works (Fan et al., 2019; Olatunji et al., 2020), where deep learning-based models implemented with additional meta-information or gating mechanism, a simpler yet effective model D-CNN and LSTM model is implemented.
III In real world scenario, many reviews remain without helpfulness votes (Fan et al., 2019; G. Ren & Hong, 2019). From the weights of the helpfulness perspectives, the helpfulness score of those reviews can be estimated.

The following sections are as follows – Section 2 discussed the literature review, Section 3 is about methodology, Section 4 discussed experiment and results, discussion and conclusion is in Sections 5 and 6, respectively.

2. Literature review

Following subsections discuss the relevant literature concerning consumer reviews and deep learning-based automated helpfulness prediction mechanism.

2.1. Review helpfulness

The perceived significance of a review to the consumer is reflected through the helpfulness of the review (Mudambi & Schuff, 2010). Informative reviews are helpful for the consumers in making a purchase decision (Lu et al., 2018; Malik & Hussain, 2018; Ren & Hong, 2019).

The literature of online helpfulness is broadly channelled into two directions. Firstly, prediction based approaches to determine the helpfulness of the review (Almagrabi, Malibari & McNaught, 2015; Diaz & Ng, 2018). Second, understanding of review helpfulness based on moderating factors (Agnihotri & Bhattacharya, 2016; Diaz & Ng, 2018). Moderating factors include concepts discussed, the impact of readers' trust, review sequence, reviewers details etc. used to estimate helpfulness (Kaushik, Mishra, Rana & Dwivedi, 2018; Qazi et al., 2016; Weathers et al., 2015). For predicting helpfulness, machine learning-based classifiers, regression, and very recently deep learning-based approaches are adopted (Diaz & Ng, 2018; Fan et al., 2019; Lee, Hu & Lu, 2018).

In certain studies, based upon the quality of information, readability of the textual content, and subjectivity, reviews are categorized (Saumya et al., 2018). Features like the length of the review, timestamp, expertize of the reviewer, and writing style are being used to predict helpful reviews (Korfiatis, García-Bariocanal & Sánchez-Alonso, 2012). Korfiatis et al. (2012) conclude, highly readable reviews turn out to be most helpful. Weathers et al. (2015) address the relation between product type and helpfulness of the reviews. Experiential products like movies, books, and search-based products like mobiles, laptops are considered. It is clear from Weathers et al. (2015) that user experience-oriented reviews are more helpful for experiential products, whereas reviews that talk about features and benefits of the product are useful in the case of search-oriented products. However, both the studies lack in terms of semantic analysis and natural language processing tools is not applied. Chua and Banerjee (2016) apply the statistical approach and consider the sentiment of the review text, product type, and review quality. It is evident from their study, review sentiment impact helpfulness, and it is independent of the type of product. The researcher has been searching for a robust set of features to predict the usefulness of a

Table 1.
Few key literature of review helpfulness based on regression and classification approach.

Author	Proposed Method	Approach Statistical	NLP based	Tools Machine Learning	Deep Learning
(Weathers et al., 2015)	Explores the relation between product type and helpfulness. It concludes that experience focused reviews are helpful for experiential products like – movies and books. Feature-based reviews are helpful for search-based products like- mobiles.	✓	×	✓	×
(Chua & Banerjee, 2016)	Analysis of variance and multiple regression has been used to infer how product type and sentiment affect review helpfulness and information quality.	✓	✓	×	×
(Singh et al., 2017)	It provides a set of features for online review helpfulness prediction. These features are considered from different moderating factors.	✓	✓	✓	×
(Ma et al., 2018)	The combination of review text and the user-uploaded image shows significant improvement in helpfulness diagnosis. LSTM has been used for text and CNN to process images.	×	✓	×	✓
(Chen et al., 2019)	It captures the multi-granularity of textual features using gates governing the embedding send to the model.	×	✓	×	✓
Olatunji et al. (2020)	Use an additional context encoding mechanism to capture context between words in long sentences.	×	✓	×	✓
(Fan et al., 2019)	Implement a Bi-LSTM model on benchmark datasets. The product title is used as external metadata in this work.	×	✓	×	✓
(Saumya et al., 2020)	Proposed a two-layer CNN model including three different filters to predict review helpfulness.	×	✓	×	✓
zhou et al. (2020)	Explore the relation between review title and content. It concludes review title positively affects helpfulness of review when in high sentiment consistency.	✓	✓	×	×
Our Approach	Review helpfulness represented from three different perspectives to estimate true review helpfulness. Apart from helpfulness voting, manual scoring is applied to understand the perception of helpfulness.	✓	✓	✓	✓

review. In this context, Singh et al. (2017) rank the features as moderating factors of helpfulness and claim to discover a set of generic features for helpfulness prediction. The multilingual approach has been adopted by Zhang and Lin (2018)) by applying statistical tools. However, in these studies, the subjective aspect of review helpfulness is not considered. Chen (2020) performs an emotion-based analysis and concludes negative emotion is attracted more helpfulness votes whereas male readers' are more attracted towards reviews with positive emotions. However, word contexts and human perception of helpfulness is not assessed. Similarly, emoticons present in a review can moderate helpfulness as well (Huang, Chang, Bilgihan & Okumus, 2020). In this context, Huang et al. (2020) conduct lab-based experiments to analyse narrative-based and list-based reviews. On the other hand, Choi and Leon (2020) provides an empirical analysis of the effect of source, review and context factors over helpfulness of review. As per the study, helpfulness is affected by depth and extremity; whereas negatively affected by inconsistency in the review. It also measures the relationship among the attributes of each factors. However, semantic analysis of review contents is not considered in Choi and Leon (2020). Recently, zhou et al. (2020), includes title of the review as moderating factor and explores the relation between the helpfulness and the title of the review. Though zhou et al. (2020) incorporates both statiscal and NLP based methods, the semantic understanding of review content remain unattended.

In different dimensions of textual information quality, cohesion represents the textual characteristics that connect ideas. It is a significant process in language understanding which is perceived in the readers' mind, based upon their ability to infer knowledge from the textual content (Batini, 2016; Ehrlinger & Wöß, 2019). This attribute of textual content is captured by the semantics of words and context in different sentences (Batini, 2016). With word embeddings, deep learning modules deal with semantics and context which is inevitable in this regards.

2.2. *Deep learning-based approaches*

Deep learning models have shown substantial improvement over the handcrafted feature sets used in earlier studies. Lee and Choeh (2014) applies a neural network model to predict review helpfulness by considering features such as - review characteristics and reviewer information as features. Neural networks automatically extract features from textual reviews and, word embeddings represent the latent semantic structure of sentences (Du, Rong, Wang & Zhang, 2019; Fan et al., 2019).

In recent studies, Convolution Neural Networks (CNNs) are used to minimize prediction error and predict review helpfulness (Du et al., 2019; Wu et al., 2020). This model learns the continuous representations which encode context-aware semantics from the review text (Du et al., 2019; Fan et al., 2019). Chen et al. (2019) propose Embedding-gated CNN (EG-CNN), which implements the CNN model and capture multi-granularity text features. As each word can deviate the helpfulness of the review, additionally a word-level embedding-gates is used. Du et al. (2019) propose a deep learning architecture namely, Explicit Content-Rating Interaction to understand the relationship between review rating and helpfulness. It assigns a rating to the review content adaptively. Saumya et al. (2020) analyses the review features, which predicts the helpfulness of reviews. In this regard, a two-layer CNN model is implemented with varying filter sizes and a regression approach has been adopted to find the least mean square error. However, Saumya et al. (2020) have collected the data from two online shopping sites. Similarly, Olatunji et al. (2020) propose a context-aware encoding mechanism to analyse the context between words in long sequences. To feed context awareness into the model, they incorporate an additional context encoding mechanism.

In contrast, Long-Short Term Memory (LSTM) models can handle longer sequences with more accuracy (Ma, Xiang, Du & Fan, 2018; Ren & Ji, 2017; Sundermeyer, Schlüter & Ney, 2012). Fan et al. (2019) use a Bi-LSTM model with a benchmark dataset, including product title meta-data. Ma et al. (2018) study the review quality, considering the image as an influential factor. It uses LSTM and CNN, both the model for text and image processing, respectively. They conclude that review in association with an image is more helpful than only textual reviews. Table 1 shows some of the key literature.

2.3. Critical analysis of existing works

In prior works, the helpfulness of online reviews has been addressed either by finding complementary predictors of helpfulness involving moderating factors or by involving deep learning modules. User voting is considered in the literature as the label to build the helpfulness prediction system. However, the representation of helpfulness votes varies across websites. Some websites display the maximum helpful votes for a review and some shows the helpfulness in “X of Y” approach. Many earlier researches establish their model upon the “X of Y” approach based helpfulness score, which takes the ratio of total helpful votes and a total number of participants (Fan et al., 2019; Malik & Hussain, 2018; Singh et al., 2017). Jin and Liu (2010) discussed that human scoring differed from helpfulness voting based scores, calculated in X of Y ratio approach. They point out, as casting a helpfulness vote is voluntary, a small number of reviews get a helpfulness vote (Jin & Liu, 2010). Later on, Cao et al. (2011) commented that, as consumers read only those reviews with helpfulness score, the absence of such votes does not facilitate to identify helpful reviews effectively. Helpful voted reviews attract more votes and deprive the low voted reviews and the cycle goes on (Cao et al., 2011). The problem remains unaddressed in the literature. In recent works, Singh et al. (2017) and Olatunji et al. (2020) discussed the same issue. Moreover, in “X of Y” approach of consumer voting mechanism, there is no way to estimate the threshold vote to consider a review as helpful. However, ratio of 0.6 has been considered in earlier works as a threshold for helpfulness (Fan et al., 2019; Malik & Hussain, 2018; Qazi et al., 2016; Saumya et al., 2018). Besides that reviews with an equal number of participants and received votes result in the same ratio, and there will be no segregation line between reviews with high votes with the lower ones.

The perception of helpfulness in consumers’ mind not investigated thoroughly in the prior works. As per cognitive psychology, informational cues in the review moderates perception of helpfulness of a review (Daft & Lengel, 1986; Forman, Ghose, & Wiesenfeld, 2008). Hence, analysis of review text, and models eliciting both syntactic statistics and semantic comprehensibility is necessary. As the consumer voting mechanism is not reliable, to understand consumers’ perception about helpfulness human scoring is required. It is also necessary to find the relationship between the consumers voting and actual human perception of helpfulness. In the next section, the proposed methodology is discussed.

3. Methodology

This paper aims to understand how lexical, sequential perspective and structural statistics of review text creates the perception of helpfulness in consumers’ mind. In the next subsections, the theoretical framework and its implementation have been discussed.

3.1. The proposed framework

The quality of the content of a review satisfies the consumer with the required information, it creates the perception of helpfulness (Krishnamoorthy, 2015; Yang, Qiu, Yan & Bao, 2015). Semantic and syntactic characteristics of the review text determine the quality of information. Moreover, exploration of review content from a semantic outlook reveals the cognitive and psychological activity involved in the text, which results in the perception of helpfulness (Yang et al., 2015; Yang, Chen & Bao, 2017). We address both semantic and syntactic aspects of review text to reveal the underlying perception of helpfulness. The semantic part of review text is addressed from lexical and sequential perspective. The syntactical aspect is covered using structural perspective. Fig. 1 shows the basic framework.

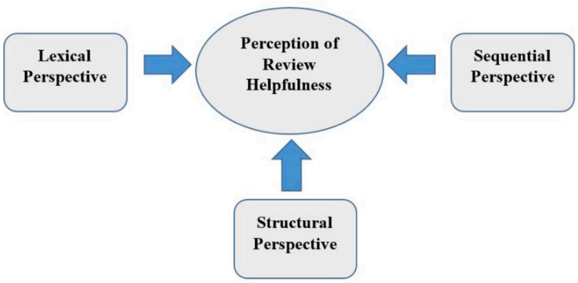

Fig. 1.. The Proposed Framework.

3.1.1. *Lexical perspective*

In textual contents, words carry significant semantics which develops the perception of information (Batini, 2016; Ehrlinger & Wöß, 2019). Context of words and semantic connection between them have an impact on the overall quality of the review and its helpfulness (Batini, 2016; Ma et al., 2018). Hence, it is essential to process words semantics along with its context. N-gram language model can capture the context of words and has various application in many natural language processing tasks (Hiemstra, 2018; Tripathy, Agrawal & Rath, 2016). However, the traditional n-gram model does not use a distributive representation of words and lacks in capturing semantic information. With non-linear neural network-based language models (Bengio, Ducharme, Vincent & Jauvin, 2003) produce features from the input sequence, based on its semantics. Hence, the word context is captured with neural network-based language models with D-CNN model.

3.1.2. *Sequential perspective*

Lexical perspective is capable of extracting local patterns corresponding to a limited length of input word sequences. However, to process a longer sequence of words, it is necessary to thrive beyond the range of lexical perspective. The sequential appearance of words with its semantics make a review comprehensible. For example, if a sentence goes like – “*The touchscreen is user friendly*”. It is evident that, till the word “*friendly*” is encountered in the sentence, the information is not clear. However, keeping semantic of each word and their interrelation in mind, clarifies the intension of the author. For example, by considering a few more words before “*friendly*”, like - “*user friendly*” or “*touchscreen is user friendly*” then it will share more information than the word “*friendly*” alone. Hence, to replicate the effect of the semantic combination of words on the readers’ perception, an LSTM model is implemented to predict the helpfulness of review. We maintain the sequence of words in which it appears in the original review.

3.1.3. *Structural perspective*

In helpfulness literature, constructs such as - count of parts of speech, readability, length of review, etc. including certain latent constructs such as overall sentiment of the review, number of polarized words, and subjectivity, are considered as an indicator of textual helpfulness (Chua & Banerjee, 2015; Krishnamoorthy, 2015; Malik & Hussain, 2017; Singh et al., 2017). These features are extracted in a handcrafted manner, which is a lengthy process than automatic feature extraction approaches. However, prior researches show, how these features represent consumers’ perception of helpfulness (P. J. Lee et al., 2018; Malik & Hussain, 2018). The works of Singh et al. (2017) and Lee et al. (2018) motivate us to construct a novel set of features for helpfulness predictions. In total, fourteen different textual features are considered.

3.2. *Implementation of the proposed framework*

Fig. 2 shows the steps involved in the implementation of the framework. Lexical perspective is explored with D-CNN, which replicates the effects of language models. LSTM model is applied to comprehend the sequential perspective. The structural perspective is explored from various language-based constraints such as noun, verb, length of sentence, etc. Later on, three different models predict the test set corresponding to three different perspectives. To unravel the contribution of each of these perspectives over the perception of helpfulness human scoring on a sample of test reviews is applied. A regression model is used to analyse the weights of each perspective. The next subsections discuss the embedding layers, which is common in both the LSTM and D-CNN models and other models.

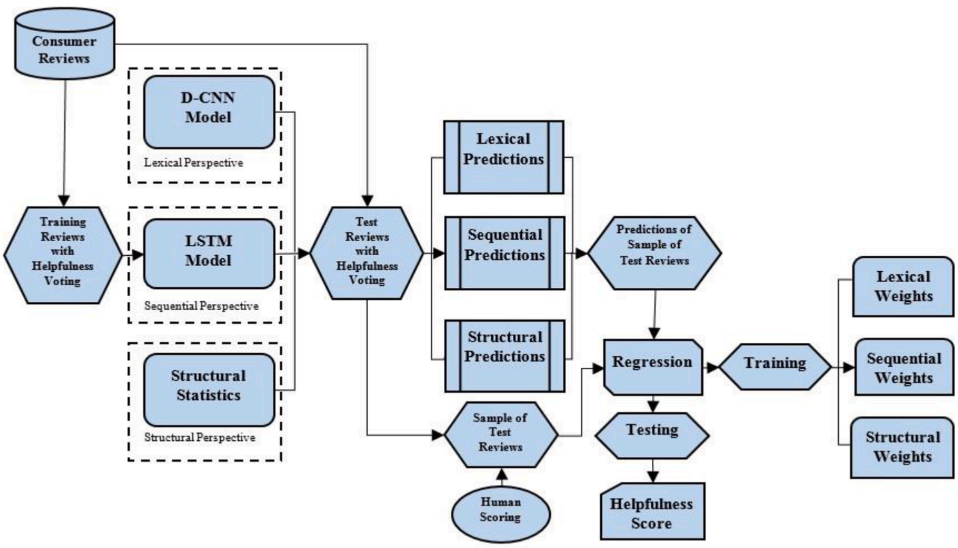

Fig. 2.. The Implementation Steps of the proposed framework.

3.2.1. Word embedding layer

Pre-trained word embedding act as a lookup table where each word is mapped to its corresponding word vector. A lookup matrix M is populated from pre-trained word embedding, where M$\in$ $R^{D\times L}$ to get embedding of each word present in the review. Here R represents the review, and D is the dimension of embedding, and L is the length of the review, which is estimated by the number of words in the review. The embedding layer of deep learning model converts the input review data into corresponding word embedding. The pre-trained embedding lookup matrix M is used as embedding weights. The embedding layer output remains the same for both D-CNN and LSTM model. GloVe pre-trained vector is the embedding weight in this model. We experiment with 50-dimensional GloVe embedding, which performs better in regression-based models.

3.2.2. Lexical perspective from dual convolutional neural network (D-CNN)

CNN is initially popularized for image processing tasks. However, it is successfully used in many natural language processing tasks such as – textual classification, text-based recommendation, sentiment analysis, and helpfulness prediction (Rezaeinia, Rahmani, Ghodsi & Veisi, 2019; Zhang, Yao, Sun & Tay, 2019). A typical CNN consists of three layers, i.e., convolutional layer, pooling layer, and a fully-connected layer. The advantage of CNN is its ability to extract significant n-grams and creates an informative latent semantic representation (Do, Prasad, Maag & Alsadoon, 2019; Goldberg, 2017). The first layer performs the convolution operations to adjusts the weights in the network, while the pooling layer does dimension reduction (Zhang, Yang, Chen, & Li, 2018).

The kernel size parameter in the CNN model defines how many words to consider in the convolution process. Variation in the kernel size, process the text data in different n-gram resolutions. Varying kernel size of 2, 3 and 4 are used for replicating the effect of bi-gram, tri-gram and four-gram language model. Fig. 3 shows a two CNN layer model. This model is motivated by Saumya et al. (2020). However, the present study is deviated from Saumya et al. (2020) by feeding embedding of review words into the first CNN layer rather than concatenating all the embeddings. The first CNN layer output document level representation of review which is feed into the second CNN layer for helpfulness prediction. Moreover, in the use of optimizer, this research gets better performance with Stochastic Gradient Decent (SGD) in respect to Adam optimizer used in Saumya et al. (2020). In general, deep neural models expects all input data to be in fixed length. However, the length of review varies across the dataset. Hence, this problem is addressed by padding review words to the same length. All the inputs to be padded to the length of 50. The semantic information corresponds to each word is generated using an additional embedding layer. The CNN model learns the relation between the words based on word embedding. In contrast to image data, text data is one dimensional. Therefore, a one dimensional (1D) Convolutional Neural Network is considered. In this work, the convolution network works as follows. Eq. (1)-(6) describes the convolutional network used in this word. Let the review x contains x_l words, so, x can be represented by –

$$x = \{x_1, x_2 \ldots x_l\} \tag{1}$$

Let n be the size of each filter and $w \in R^{nk}$ represent the filter involve in the convolution process. Hence, the convolution process is produce features defined as -

$$m_{feature} = f(w^{\circ} x_{i:i+n-1} + b) \tag{2}$$

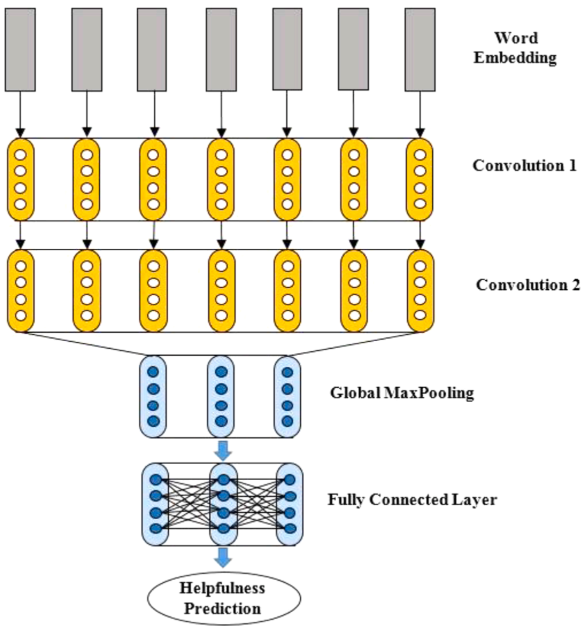

Fig. 3.. CNN model for helpfulness prediction.

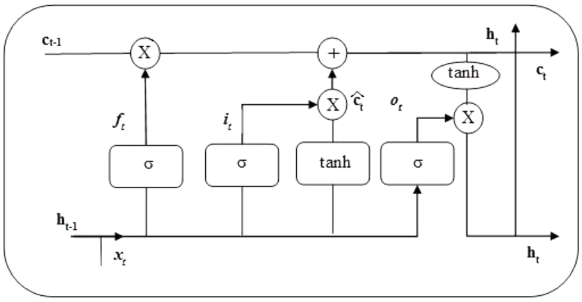

Fig. 4.. As Standard LSTM Cell.

Where $w^{\circ}x_{i:i+n-1}$ represent element-wise multiplication between w and $x_{i:i+n-1}$, f denotes the non-linearity function, and b is the bias term. Here, $x_{i:i+n-1}$ means –

$$x_{i:i+n-1} = x_i \oplus x_{i+1} .... \oplus x_{i+n-1} \tag{3}$$

$\oplus$ stands for concatenation of row vectors. After the filter runs through the review, the feature map generated as -

$$f_m = \left[m_{feature\_1}, m_{feature\_2}, ...m_{feature\_l}\right] \\ f_m \in R^{l-n+1} \tag{4}$$

Where l is the length of the review. Like this, if there are p number of filters, then it will produce features as –

$$M = \{f_{m1}, f_{m2}, ...f_{mp}\} \tag{5}$$

Now, this M represents the document representation. This feature is fed into another convolution layer and the output helpfulness prediction after passing through a Global MaxPooling layer. Feature before going to Global Max pooling layer M_{2-CNN} represented as –

$$M_{2-CNN} = \left[M_1 \oplus M_2 \oplus ....M_p\right] \tag{6}$$

To attain non-linearity in the process, we use a rectified linear unit as the activation function in the dense layer. The reviews of the test set are predicted separately with kernel sizes 2, 3, and 4, and the final prediction is given by taking the average of those individual predictions.

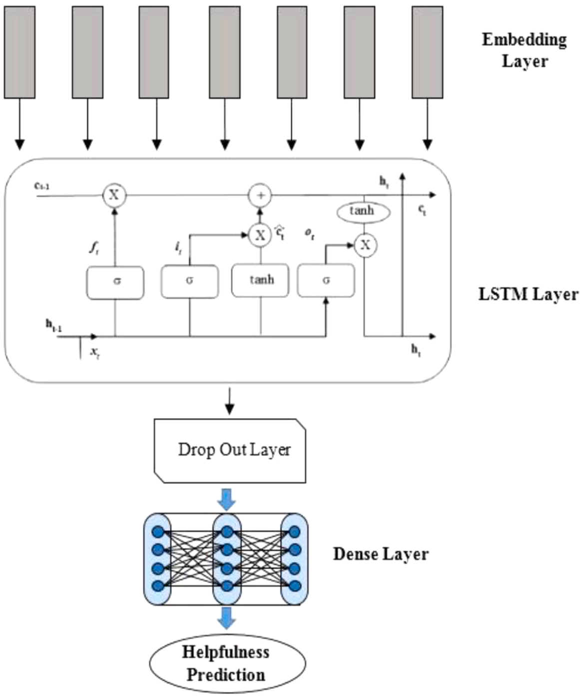

Fig. 5.. Architecture of LSTM based model.

3.2.3. Sequential perspective from LSTM

To capture the sequential effects created by words in a review, we implement a Long Short Term Memory (LSTM) model. LSTM considers text content as a sequence of words and explores the dependency among words and textual structure. LSTM sort out the vanishing gradient problem of RNN with memory cell, which in turn retain important information and forgets the unimportant ones (Otter, Medina & Kalita, 2020).

Fig. 4 shows a standard LSTM cell. A standard LSTM memory cell contains three gates, namely input, forget, and output get. Let, i, o, f denotes input, output, and forget gate at time step t. c represents memory cell and x_t is the input at time step t denoted by i_t, o_t, f_t, respectively. Eq. (7)-(12) shows the formalization of LSTM –

$$f_t = \sigma\left(W_f x_t + U_f h_{t-1} + b_f\right) \tag{7}$$

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) \tag{8}$$

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o) \tag{9}$$

$$\widehat{c_t} = \tanh(W_c x_t + U_c h_{t-1} + b_c) \tag{10}$$

$$c_t = f_t^\circ c_{t-1} + i_t^\circ \widehat{c_t} \tag{11}$$

$$h_t = o_t^\circ \tanh(c_t) \tag{12}$$

Where σ is the logistic sigmoid function, the activation vectors of input, output, and forget gate along with memory cell is represented by i_t, o_t, f_t, c_t, respectively. In this model, LSTM takes the word embeddings from the embedding layer. W_f, W_i, W_o and W_c is of the size $H \times d$ where H is the dimension of the hidden layer and d is the dimension of the memory cell. U_f, U_i, U_o, U_c and b_f, b_i, b_o, b_c is of size H and $H \times H$respectively, and the input is h. The output from LSTM is feed to a fully connected layer which generates the prediction. Fig. 5 shows the architecture of the model. We introduce a small dropout here as a regularizer.

3.2.4. Structural perspective from structural statistics

Apart from the semantics of words, the syntactic structure of reviews is used to predict review helpfulness. These features are manually extracted from reviews. In this regards, sentiments, parts of speech (POS), readability, lengths, ratings, and information entropy-based features are extracted. All these features are in numerical format. A Support Vector Machine uses these features for training. The basic idea behind this perspective is, readers will get more information from reviews with more number of POS tags, as noun and verbs establish the meaningful part of the sentence. However, many studies show, readability of review, sentiment, length of the review etc. affects helpfulness of review (P. J. Lee et al., 2018; Saumya et al., 2018). VADER sentiment analysis dictionary has been used to analyse the sentiment of words and overall reviews. Table 2 shows the details of these features.

3.2.5. Prediction of helpfulness

This paper presents helpfulness from three different perspectives lexical, sequential, and structural. To understand which perspective contribute more on creating the perception of helpfulness in consumers mind each of the models described in the previous subsection has been trained with helpfulness voting in the “X of Y” format. These values of X and Y come from consumer voting. As discussed in section1, these votes are not reliable. However, manual labelling of huge datasets requires immense effort in terms of time and labour. Hence, initially, the models have experimented with consumer voting itself which gives a clearer picture of the performance of the models. For instance, H_v is the helpfulness score of a review in X of Y format and $\mathrm{H}_{\mathrm{pred}}$ is the predicted score from one of the models discussed earlier. The focus of the experimentation is $\varepsilon = \min(\mathrm{H}_v - \mathrm{H}_{\mathrm{pred}})$. Where ε is the error in the prediction. Mean Square Error (MSE) has been used as a metric to estimate the prediction error for each model. Once training and testing of each of the

Table 2.
Description of features used to assess the structural perspective of review helpfulness.

Feature Type	Details
Sentiment	Overall sentiment score of the review.
	The subjectivity score of the review
	Review rating
Parts of Speech (POS)	Count of noun terms
	Count of Verbs
	Count of Adjectives
Readability	Dale Chall readability scores
	Flesch reading ease scores
Length	Review length
	Number of one letter words
	Number of two letter words
	Number of words longer than two letters
Information-Theoretic constructs	Type Token Ratio (TTR)
	Information Entropy

Table 3.
Description of Datasets.

Dataset	Total Review (h>0.6)	Avg Helpfulness ratio	Total Reviews
Home & Kitchen	14,750	0.65	30,000
Electronics	13,387	0.56	30,000
Sports & Outdoors	17,424	0.64	30,000
Tools and Home Improvement	14,104	0.68	30,000
Grocery & Gourmet Food	10,845	0.61	30,000

models are done using helpfulness voting, we manually labelled 400 sample reviews from each of the datasets. We keep these manually labelled reviews as the gold standard and compare the predictions from three different models. Finally, we apply regression keeping manual labels as the dependant variables and corresponding prediction from three different models as the independent variables. Formulating a regression shows the importance of each of the perspectives over review helpfulness.

4. Experiments and results

The detailed description of the experimentation and results is given in the next subsections. We use python 3.6 as the basic programming language. We incorporate deep learning models with Keras libraries. Keras is an API, which uses Tensorflow and Theano as a backend to implement deep neural network models.

4.1. Datasets and pre-processing

Amazon reviews (He & McAuley, 2016; McAuley et al., 2015) has been used as a dataset in our experiment. This is a publicly available dataset having 142.8 million reviews collected across May 1996 to July 2014. This paper experiments with five different product categories, such as – electronic, home and kitchen, automotive, grocery, and sports. Each of the reviews contains specific metadata, including product id, user id, review helpfulness, etc. The helpfulness score is mentioned herein "X of Y" format. Here "Y" is the total number of participants in voting, and "X" is the total helpful vote received by a review. Only those reviews are considered which receive at least 1 helpfulness vote. 30,000 reviews from five different product categories are randomly selected, a total of 150, 000 reviews which receive 1 or more helpfulness votes. These 30,000 reviews of each of the product category, split into training and testing groups containing 25,000 and 5000 reviews each. On the training set, we apply 80%, 20% split for validation. 10-fold cross-validation is performed for each of the models in each perspective to avoid any biases. Table 3 shows the description of the dataset used in the experiment. In literature, the ratio 0.6 is considered as the threshold for segregating reviews into helpful and non-helpful (Fan et al., 2019; Malik & Hussain, 2018; Qazi et al., 2016; Saumya et al., 2018). We use this threshold just to give an exploratory view of the dataset involved in experimentation. In this regard, the second column of Table 3 shows the count of reviews having "X of Y" ratio 0.6 or more. The texts are all pre-processed by removing unwanted characters, special characters. Stop words are essential from a grammatical point of view; however, it carries no information. Hence, all the stop words are removed from the text.

4.2. Experiment with lexical perspective

This research uses a dual CNN layer (D-CNN) based model to represent the lexical perspective of review helpfulness. To understand the effect of different n-gram models, experiment done with kernel sizes 2, 3, and 4. The models are trained with 25,000 reviews from each of the product categories. Fig. 5(a), (b), and (c) shows the training and validation curves for Home and Kitchen reviews for different kernel sizes. We test with 50-dimensional GloVe embeddings and experiment with different epochs (like- 100, 200,300 and 500) and batch sizes (like- 64,128, 512). 500 epochs and batch size 512, give the best results. We apply tenfold cross-validation for each of the models to avoid any biases. Stochastic Gradient Descent (SGD) is used as an optimizer in this process. With varying learning rates for the optimizer results differently. With a leaning rate of 0.007, the CNN performed with least MSE. Rectified Linear Unit (ReLU) is used as an activation function. The model is trained with helpfulness votes given in the dataset. Three predictions of varying kernel size are saved. Now, for each review, three distinct predictions come from three kernel sizes. With an average of these predictions for each review, the MSE for the test set is computed. Table 4 shows the prediction MSE for lexical perspective. The MSE of

Table 4.
Mean Square Errors from different models.

Datasets	D-CNN (K = 2)	D-CNN (K = 3)	D-CNN (K = 4)	Lexical	Sequential	Structural
Home & Kitchen	0.0426	0.0413	0.0422	0.0373	0.0413	0.0619
Electronics	0.0091	0.0093	0.0092	0.0089	0.0093	0.0092
Sports & Outdoors	0.0285	0.0290	0.0358	0.0249	0.0252	0.0243
Tools and Home Improvement	0.0241	0.0921	0.0320	0.0268	0.0322	0.0251
Grocery & Gourmet Food	0.0243	0.0240	0.0351	0.0222	0.0392	0.0268

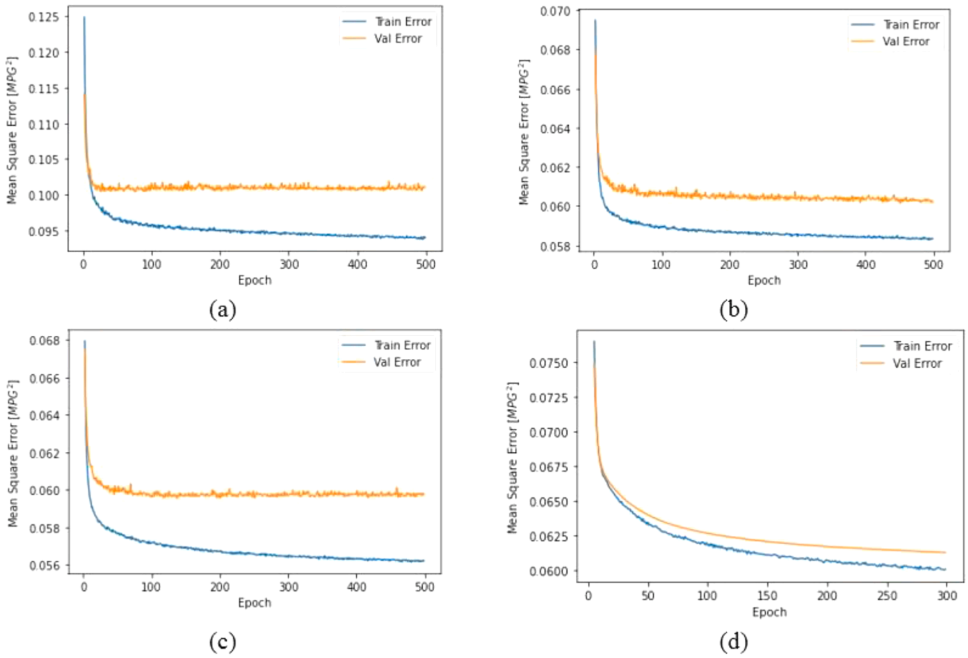

Fig. 6.. The training and validation curve for D-CNN models for Home and Kitchen reviews shown in (a) Bigram (Kernel size=2) (b) Tri-gram (Kernel size = 3) (c) Fourgram (Kernel size = 4). Fig. 5(d) shows the training validation curve of LSTM model.

average prediction is denoted in the Lexical column. It is noticeable, except for the Tools and Home Improvement reviews average of prediction gives less MSE. It has shown improvement over Kernel size =2, 3, and 4 as well. This result point towards the second perspective, which deals with the sequential processing of information. It is also evident Lexical perspective shows better results than the other two perspectives for at least three product categories.

4.3. Experiment with sequential perspective

Long short Term Memory of 100 cells is used to understand the effect of sequential perspective of review helpfulness perception. It captures long term dependencies among words which sequentially process information. Fig. 5(d) shows the LSTM training and validation error curve for Home and Kitchen. Here, SGD is used as optimizer with learning rate 0.1 and train it through 300 epochs. An additional dropout layer of 0.09 is added to control overfitting. Similar to lexical perspective, it trained with varying batch sizes and choose 512 as batch size. Like lexical perspective, the predicted values are saved for comparing with manual annotation. The MSE of sequential perspective is shown in Table 4. Sequential processing of information shows higher MSE than Lexical and Structural perspective in some cases. For electronics reviews, Sequential perspective shows the least MSE compare to other product categories.

4.4. Experiment with structural perspective

The fourteen features discussed in Table 2 has been used to analyse the structural perspective of review helpfulness. These features are extracted from pre-processed reviews. As the focus is not on finding which structural features contribute more on helpfulness, Support Vector Machine with Radial Basis kernel function is chosen. We apply tenfold cross-validation to get better predictions in terms of MSE. Finally, we predict test reviews with the trained model. The MSE of prediction has been shown in Table 4. The MSE from Structural perspective shows improvement for two product category, namely, Sports & Outdoors and Tools and Home Improvement. The results of the structural perspective suggest that the features involved in the process capable of capturing voting trends.

4.5. Comparison with human scoring

In shopping sites, helpfulness of reviews is viewed in terms of consumers voting based upon the question of whether they find the review helpful or not. As discussed earlier, the helpfulness score generated from this process is not always reliable. The ground reality is, human scoring for the helpfulness of reviews is necessary. As discussed in Section 3, manual annotators put the score for the reviews in this purpose. 400 reviews are randomly considered from every product category. A subsequent assessment of predictions, with manual annotation, determine whether the predictions complies with the consumers' perception of helpfulness. In this process, 12 annotators are involved and asked them to score each review in the range of 0–10. Where 0 represents, the review is not at all helpful, and 10 represents the review is helpful. For each review, the average scores coming from 12 different users are rescaled in the range of

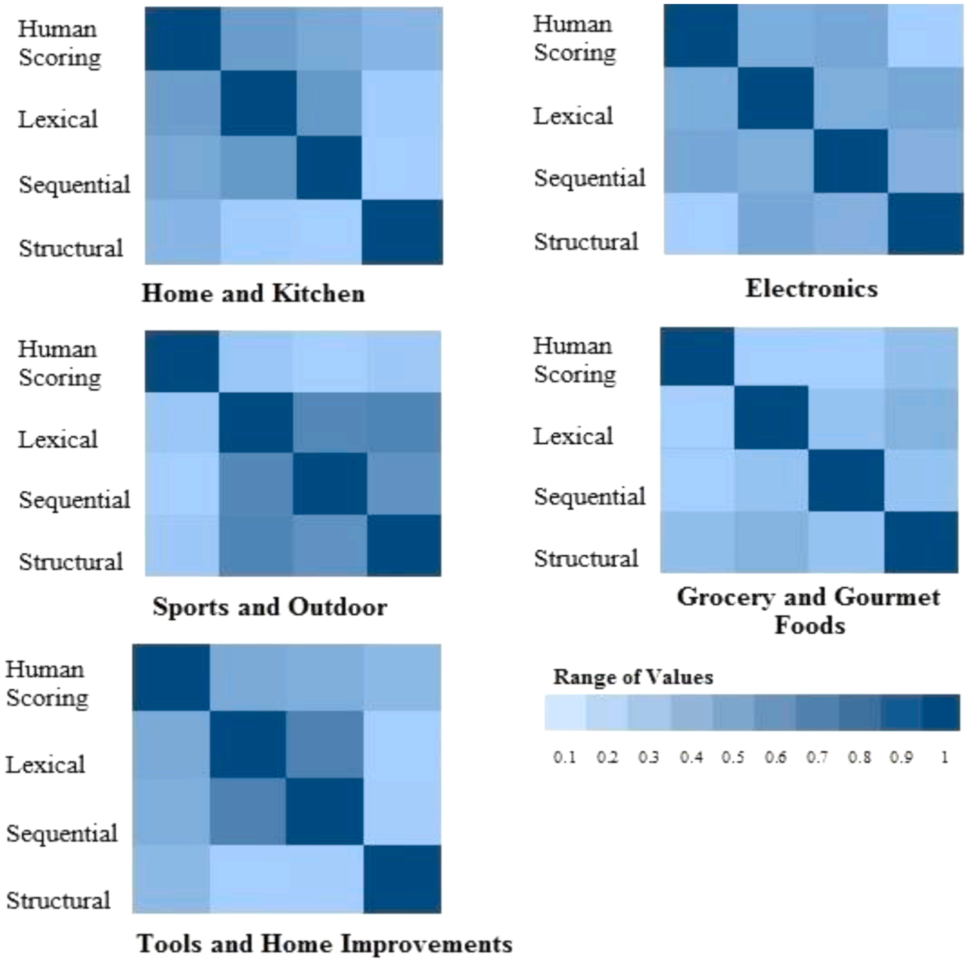

Fig. 7.. Correlation heat map between scoring and prediction from three perspectives.

0–1 to maintain generalization.

4.5.1. Comparison of helpfulness rating with human scoring

Helpfulness votes are the cumulative response of up or down votes of users who read the reviews. However, in human scoring, annotators spend time to read each review and put a helpfulness score which is much more intense in understanding and perspective point of view. To understand whether the human scoring is different from the helpfulness voting, this study performs a two-sample mean z-test. It is evident from the test the human scoring is significantly different from helpfulness voting at $p<0.001$. Now, it is necessary to know at this point whether the predicted values have significance over the human rating. To understand this, the correlation between predictions and human scoring is done. Fig. 7 shows the heat map of the correlation between human scoring and prediction of three perspectives. In Fig. 6, Home and Kitchen product category, lexical perspective shows around 40% correlation with human scoring. Whereas, the correlation between perspectives in Sports and Outdoor category, that lies between 50–60%. It is noticeable that the correlation between predictions is not very high and stays within 30–40% for different product categories. This small correlation between predictions allows exploring the possibility of how predictions from three different perspectives result in the perception of helpfulness.

4.5.2. Estimation of weights for three perspectives and helpfulness score

This study visualizes helpfulness of review from three different perspectives. Now, to understand the contributions of these perspectives on creating the perception of helpfulness in consumers' mind, a multiple regression model considering human scoring as the dependent variable and three perspectives as independent variables is implemented. However, the underlying nature of the involvement of three perspectives is not known. Hence, considering a straight forward linear model approach will shroud the possibilities of non-linearity. Experimentation with linear and non-linear functions overcome the situation. Kernel functions represent implicit non-linearity. We test the approach with linear, polynomial of degree 2 and 3 and radial basis functions. Support Vector Machine (SVM) has been used as a regressor for non-linear regression. Multiple linear regression to analyse the linear relationship is applied. The choice of SVM has been made because of two reasons. Firstly, the human scoring data is not huge to be suited for other kinds of regressors, such as deep learning modules. Secondly, SVM supports different non-linear kernel functions. 400 reviews from each product category labelled with human scoring. These reviews of each product category split into training and testing with 80% and 20% respectively. Cross-validation with ten folds has been applied to every product category. The results of the experiment shown in Table 5. The column Linear represents the multiple linear regression models. For polynomial functions, degree 2 ($D = 2$) and degree 3 ($D = 3$) are used. MSE is used as prediction metrics. The results in Table 5 clearly shows that linear regression produces lower MSE, except for Tools and Home Improvement category where polynomial function ($D = 2$) shows an improved result. This experiment clearly shows that there exists a linear relationship among lexical, sequential and structural perspectives. Therefore, proceed forward with linear models to identify the contributions of each of these perspectives over the human scoring of helpfulness. Moreover, the prediction done by the multiple linear regression model is considered as the true helpfulness score.

Multiple linear regression is obtained following Eq (13) –

$$Y_{Human\_Scoring} = \beta_0 + \beta_{lexical}x_{lexical} + \beta_{sequential}x_{sequential} + \beta_{structural}x_{structural} + \varepsilon_0 \tag{13}$$

Where,

$x_{lexical}$ = Prediction from Lexical Perspective,
$x_{sequential}$ = Predictions from Sequential Perspective,
$x_{structural}$ = Predictions from Structural Perspective,
ε_0 = Regression Error, and β_0, $\beta_{lexical}$, $\beta_{sequential}$ and $\beta_{structural}$ are regression coefficient to be determined.

Table 6 shows the respective weights of three different perspectives obtain from multiple linear regression on five product categories. Table 6 concludes that lexical, sequential and structural perspectives have a significant impact over true helpfulness of reviews. The sequential perspective shows inverse relation for Electronics and Sports categories; it can be analysed as, establishing sequence among sentences become difficult in reviews with disjoint ideas. On the other hand, human scorers read reviews, and as per their comprehension of the facts, they put their scores. Instead of that as structural perspective shows positive relation with helpfulness it is evident that statistics of language-related elements like – noun, verb, adjectives, length of review etc. enriches the review with information in a positive way, which is reflected in human scoring. An interesting observation is that lexical perspective has more influence over the perception of review helpfulness rather than other two perspectives. It is perhaps because product reviews are aspect-oriented where information is given within a few words around the product aspect rather than spending a longer sequence of inter-related words.

5. Discussion

Consumer review is an influential marketing tool for both vendors and marketers in an e-commerce scenario. It provides online businesses with a compelling, convenient, and impactful medium to reach their customers. Hence, online retailers take advantage of the opinion of experienced users to draw the attention of potential consumers. On the other hand, helpfulness rating makes the evaluation of such reviews convenient for the consumers', by addressing the problem of information overload and aid the process of decision making (Cao et al., 2011; Krishnamoorthy, 2015).

This paper analyses helpfulness of consumer reviews from three qualitative perspectives, namely lexical, sequential and structural perspective. Deep learning-based techniques have been implemented to represent lexical and sequential perspective. This research implements D-CNN, a dual convolutional layer based CNN model to capture lexical perspective of helpfulness. LSTM model is used to analyse a sequential perspective. A small set of feature address the structural perspective of helpfulness. These perspectives analyse the textual contents of reviews and capture the consumers' perception of helpfulness. Consumer voting based helpfulness score has been used for training and testing of these models. However, as consumer voting does not reflect consumers' perception of helpfulness properly, a set of sample reviews is labelled manually. The contribution of each of the perspective is then analysed using regression techniques. Finally, reviews are assigned their helpfulness score which reflects the perception of helpfulness. Several theoretical and practical implications are involved in this study. The following subsections assimilate discussion each of them.

Table 5.
Mean Square Errors (MSE) using a different Kernel function.

Datasets	Linear	Polynomial (D = 2)	Polynomial (D = 3)	Radial Basis
Home & Kitchen	**0.0255**	0.0738	0.0773	0.0285
Electronics	**0.0106**	0.0233	1.7221	0.0160
Sports & Outdoors	**0.0098**	0.1023	0.1222	0.0395
Tools and Home Improvement	0.1154	**0.0458**	0.0508	0.3483
Grocery & Gourmet Food	**0.0007**	0.0008	0.0019	0.0008

Table 6.
Regression output with human scoring.

Dataset	Perspective	Weights	Standard Error	t-Statistics
	Lexical	0.7050**	0.1296	5.4420
Home & Kitchen	*Sequential*	0.2035**	0.0746	2.7288
	Structural	0.2825*	0.0886	3.1867
	Lexical	4.3099**	0.6528	6.6020
Electronics	*Sequential*	−0.8410*	0.4346	−1.9349
	Structural	4.0182**	0.4837	8.3058
	Lexical	0.2421*	0.1067	2.2693
Sports & Outdoors	*Sequential*	0.2240**	0.0603	3.7120
	Structural	0.3629**	0.0821	4.4197
	Lexical	0.4771*	0.1466	3.2547
Tools and Home Improvement	*Sequential*	0.2083*	0.1186	1.7566
	Structural	0.1956*	0.0747	2.6180
	Lexical	1.0551**	0.2860	3.6887
Grocery & Gourmet Food	*Sequential*	−0.0627	0.1390	−0.4513
	Structural	0.3405*	0.1224	2.7815

**shows where significance at p-value < 0.001, *shows where significance at p-value < 0.05.

5.1. *Theoretical implications*

Analysis of vital predictor for online review helpfulness has drawn significant attention in the literature. The primary motive of the work is to understand the textual content of the review to analyse the helpfulness. In the earlier studies, user votes are considered as the primary source of comprehending helpfulness. The semantic of information, contextual understanding, and coherence remain unattended. Helpfulness of reviews is determined by "X of Y" voting ratio (Fan et al., 2019; Malik & Hussain, 2017; Singh et al., 2017). However, this voting ratio is not reliable due to the dependence on posting time. Older reviews are graved into the abyss of newly arrived reviews. An equal number of votes and readers create complications in separating reviews into highly read and voted category with low read and voted category. These issues are addressed in the study by human scoring. Incorporation of human labelling comes with the time constraint. Labelling a huge number of review is time-consuming. Hence, it is necessary to establish a relation between user voting and human perception of helpfulness. In this regards, models are trained and tested based on user voting, and for a small sample of reviews, the human score is given. We estimate the corresponding weights of three different perspectives by which the true helpfulness score perceived by a consumer is comprehended. With labelling of a smaller set of sample reviews reduce the cost of human scoring task in terms of time.

Earlier researches concentrate on language construct related statistics and overall sentiment of the reviews (Chua & Banerjee, 2015; Malik & Hussain, 2018; Weathers et al., 2015). Work of that sort used the syntactic structures of the reviews rather than semantics, which is undoubtedly an import criterion for a review to be helpful. Lexical perspective deals with n-gram language models involving word vectors, and a deep neural network-based Dual-CNN (D-CNN) model is implemented in this context. Deep neural models work on word vectors which represent the semantics of words (Kameswara Sarma, 2018; Yu, Wang, Lai & Zhang, 2017). Textual content is comprehended by going through the words and realising their intervening semantics. LSTM based sequential perspective replicates this idea.

The research implements the models with publicly available Amazon review dataset. In studies, the necessity of using a publicly available dataset has been discussed, considering the comparative research point of view (Diaz & Ng, 2018; Olatunji et al., 2020). Most of the researches to date have used dataset crawled by themselves from the web, making it utterly difficult for others to implement. In this context, benchmark dataset is used, which is publicly available. Furthermore, this research address an intrigue and crucial aspect that is the perception of helpfulness in consumers' mind. From the dimension of information quality, two attributes of consistency of information are explored in our approach – cohesion and coherence (Batini, 2016; Ehrlinger & Wöß, 2019). As discussed in Section 2, the combination of cohesion and coherence creates the inference of knowledge in the consumers' mind. Deep learning models address these two attributes by convolution over word embeddings features. Our approach extracts the relatedness between cohesion and coherence using human scoring and multiple linear regression. As a whole, a novel point of view to see the problem of the helpfulness of consumer reviews.

5.2. *Practical implication*

This work keep the language-based statistics as structural perspective of review helpfulness with a smaller set of features than earlier works (Malik & Hussain, 2017; Singh et al., 2017;Saumya et al., 2018). However, the semantics of word is not considered in this feature set. The approach combines both semantic and syntactic perspective of review get the complete view of the helpfulness of reviews which is not attempted earlier. In contrast to earlier works (Fan et al., 2019; Olatunji et al., 2020), additional information, in terms of review title or gating mechanism has been used to work with deep learning which increases the complications of the model. We implement the model based on review text which is visible to the consumer. Focusing only on the textual content of review makes the model simpler in terms of implementation. Moreover, the problem of reviews with no votes has been discussed in the literature (Fan et al., 2019; G. Ren & Hong, 2019). With the corresponding weights of three qualitative perspectives, reviews which have not to receive any votes can assign with their true helpfulness score.

5.3. Managerial implication

Considering the importance of review helpfulness over consumers buying behavior, estimation of content-oriented review helpfulness will be significant. From a managerial point of view, review written by any experienced consumer is essential irrespective of the polarity. Critical reviews, pointing out flaws in the product or suggesting improvement is equally significant as of helpfulness. However, prior literature suggests consumer voting of helpfulness mediated by the sentiment orientation of words (Chua & Banerjee, 2015, 2016). Human scoring of reviews alleviates this dilemma. This work demonstrates how human labelling of a small set of reviews results in identifying the intriguing nature of helpfulness. The related weights of each of the perspective of helpfulness are estimated. It is an in-hand solution for managers to implement our approach in the real world.

Moreover, as the approach deals with textual contents and its semantics, it gives a glimpse of consumer perception towards the farm. As consumers' perception is woven with the brand image, this research can be extended further in the domain of brand management. Managers can explore this area for analysis and taking the necessary strategic decision regarding the development of newer brands and products.

In the flow of consumers buying process, helpful reviews play an important role, especially in the evaluation of alternatives. An informative, well-written review can moderate the consumers' choice over other competitive alternatives. Assessment of helpfulness of consumer reviews can therefore be used as a tool for the recommendation of products as well. As experienced users give reviews, it can be seen as post-purchase behavior of consumers which will provide information regarding product performances. In such scenarios, decoding consumers' perception of helpfulness will work as a stepping stone for taking corrective measures towards product, brand and farm.

5.4. Limitations

This paper addresses the helpfulness problem from a novel dimension. However, like every other research work, this study has its limitations. Firstly, product reviews used in this work can be extended to movie or books reviews for further understanding of helpfulness perception. Secondly, deep learning-based models have many variants. CNN and LSTM models work well in this study. There are auto-encoders, attention-based deep learning models available which may open some other lane of analysis. Thirdly, we asked candidates to involve voluntarily for human scoring of reviews. We do not involve any paid services in this matter. If any professional firm provides service in terms of human volunteer, a large number of reviews can be manually scored for better analysis. Lastly, benchmark dataset has been taken for experimentation; however, real-time reviews depend upon various other factors such as – product types, time of product launch, when the product model is updated or modified etc.

6. Conclusion

This work deals with the problem of understanding consumers' perception of review helpfulness. The research defines three perspectives based on which the perception of helpfulness clarifies. Lexical perspective reveals the semantic relations among the words based on the n-gram language model. The syntactical structure of reviews is handled in structural perspective. Lastly, sequential perspective analyses the effect of a long sequence of words. To implement lexical perspective, D-CNN model, having two convolutional layers is used. Sequential perspective uses LSTM model. To estimate the structural perspective, we propose a feature set. These three models are trained and tested based on helpfulness voting in "X of Y" approach. However, "X of Y" approach is not precisely capturing review helpfulness. To address this problem, a small set of reviews is considered and follow human scoring mechanism. With the predictions from three different models, the necessary contributions of each of the perspective upon the human perception of helpfulness is estimated. It is observed that all these perspective have an impact on the perception of helpfulness and lexical perspective have a higher impact than the other two perspectives. The approach stated in this study acts as an aid to the manager to understand review helpfulness as well as predict the perceived helpfulness score of a review. On the other hand, decoding consumers' perception of helpfulness helps in framing strategies towards the growth of the product and the brand.

CRediT authorship contribution statement

Satanik Mitra: Conceptualization, Methodology, Writing - original draft. **Mamata Jenamani:** Supervision, Writing - review & editing.

References

Agnihotri, A., & Bhattacharya, S. (2016). Online review helpfulness: Role of qualitative factors. *Psychology and Marketing, 33*(11), 1006–1017. https://doi.org/10.1002/mar.20934.

Almagrabi, H., Malibari, A., & McNaught, J. (2015). A survey of quality prediction of product reviews. *International Journal of Advanced Computer Science and Applications, 6*(11). https://doi.org/10.14569/ijacsa.2015.061107.

Batini, C. (2016). Introduction to information quality. *Information Quality: The Potential of Data and Analytics to Generate Knowledge*. https://doi.org/10.1002/9781118890622.ch1.

Bengio, Y., Ducharme, R., Vincent, P., & Jauvin, C. (2003). A Neural Probabilistic Language Model. *Journal of Machine Learning Research, 3*, 1137–1155. https://doi.org/10.1162/153244303322533223.

Cao, Q., Duan, W., & Gan, Q. (2011). Exploring determinants of voting for the "helpfulness" of online user reviews: A text mining approach. *Decision Support Systems, 50*(2), 511–521. https://doi.org/10.1016/j.dss.2010.11.009.

Chatzipanagiotou, K., Veloutsou, C., & Christodoulides, G. (2016). Decoding the complexity of the consumer-based brand equity process. *Journal of Business Research, 69*(11), 5479–5486. https://doi.org/10.1016/j.jbusres.2016.04.159.

Chen, C., Zhou, J., Qiu, M., Li, X., Bao, F. S., Yang, Y., et al. (2019). Multi-domain gated CNn for review helpfulness prediction. In *The Web Conference 2019 - Proceedings of the World Wide Web Conference, WWW 2019* (pp. 2630–2636). https://doi.org/10.1145/3308558.3313587.

Chen, M. J. (2020). Examining the influence of emotional expressions in online consumer reviews on perceived helpfulness. *Information Processing and Management*, (6), 57. https://doi.org/10.1016/j.ipm.2020.102266.

Choi, H. S., & Leon, S. (2020). An empirical investigation of online review helpfulness: A big data perspective. *Decision Support Systems, 139*. https://doi.org/10.1016/j.dss.2020.113403.

Chua, A. Y. K., & Banerjee, S. (2015). Understanding review helpfulness as a function of reviewer reputation, review rating, and review depth. *Journal of the Association for Information Science and Technology*. https://doi.org/10.1002/asi.23180.

Chua, A. Y. K., & Banerjee, S. (2016). Helpfulness of user-generated reviews as a function of review sentiment, product type and information quality. *Computers in Human Behavior, 54*, 547–554. https://doi.org/10.1016/j.chb.2015.08.057.

Daft, R. L., & Lengel, R. H. (1986). Organizational information requirements, media richness and structural design. *Management Science, 32*(5), 554–571. https://doi.org/10.1287/mnsc.32.5.554.

Diaz, G. O., & Ng, V. (2018). Modeling and prediction of online product review helpfulness: A survey. In *, 1. ACL 2018 - 56th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference (Long Papers)* (pp. 698–708). https://doi.org/10.18653/v1/p18-1065.

Do, H. H., Prasad, P. W. C., Maag, A., & Alsadoon, A. (2019). Deep learning for aspect-based sentiment analysis: A comparative review. *Expert Systems with Applications*. https://doi.org/10.1016/j.eswa.2018.10.003.

Du, J., Rong, J., Wang, H., & Zhang, Y. (2019). Helpfulness prediction for online reviews with explicit content-rating interaction. In *, 11881. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (pp. 795–809). LNCS. https://doi.org/10.1007/978-3-030-34223-4_50.

Duan, W., Gu, B., & Whinston, A. B. (2008). Do online reviews matter? - An empirical investigation of panel data. *Decision Support Systems, 45*(4), 1007–1016. https://doi.org/10.1016/j.dss.2008.04.001.

Ehrlinger, L., & Wöß, W. (2019). Automated schema quality measurement in large-scale information systems. In *, 11235. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (pp. 16–31). LNCS. https://doi.org/10.1007/978-3-030-19143-6_2.

Fan, M., Feng, C., Guo, L., Sun, M., & Li, P. (2019). Product-aware helpfulness prediction of online reviews. In *The Web Conference 2019 - Proceedings of the World Wide Web Conference, WWW 2019* (pp. 2715–2721). https://doi.org/10.1145/3308558.3313523.

Ferstl, E. C., & Von Cramon, D. Y. (2001). The role of coherence and cohesion in text comprehension: An event-related fMRI study. *Cognitive Brain Research, 11*(3), 325–340. https://doi.org/10.1016/S0926-6410(01)00007-6.

Filieri, R. (2015). What makes online reviews helpful? A diagnosticity-adoption framework to explain informational and normative influences in e-WOM. *Journal of Business Research, 68*(6), 1261–1270. https://doi.org/10.1016/j.jbusres.2014.11.006.

Forman, C., Ghose, A., & Wiesenfeld, B. (2008). Examining the relationship between reviews and sales: The role of reviewer identity disclosure in electronic markets. *Information Systems Research*, 291–313. https://doi.org/10.1287/isre.1080.0193.

Goldberg, Y. (2017). Neural Network Methods for Natural Language Processing. *Synthesis Lectures on Human Language Technologies, 10*(1), 1–311. https://doi.org/10.2200/S00762ED1V01Y201703HLT037.

He, R., & McAuley, J. (2016). Ups and downs: Modeling the visual evolution of fashion trends with one-class collaborative filtering. In *25th International World Wide Web Conference, WWW 2016* (pp. 507–517). https://doi.org/10.1145/2872427.2883037.

Hiemstra, D. (2018). Language Models. *Encyclopedia of Database Systems*, 2061–2065. https://doi.org/10.1007/978-1-4614-8265-9_923.

Huang, G. H., Chang, C. T., Bilgihan, A., & Okumus, F. (2020). Helpful or harmful? A double-edged sword of emoticons in online review helpfulness. *Tourism Management, 81*. https://doi.org/10.1016/j.tourman.2020.104135.

Jin, J., & Liu, Y. (2010). How to interpret the helpfulness of online product reviews: Bridging the needs between customers and designers. In *International Conference on Information and Knowledge Management, Proceedings*. https://doi.org/10.1145/1871985.1872000.

Kameswara Sarma, P. (2018). *Learning word embeddings for data sparse and sentiment rich data sets* (pp. 46–53). https://doi.org/10.18653/v1/n18-4007.

Kaushik, K., Mishra, R., Rana, N. P., & Dwivedi, Y. K. (2018). Exploring reviews and review sequences on e-commerce platform: A study of helpful reviews on Amazon. *Journal of Retailing and Consumer Services, 45*, 21–32. https://doi.org/10.1016/j.jretconser.2018.08.002.

Keller, K. L. (2016). Reflections on customer-based brand equity: Perspectives, progress, and priorities. *AMS Review, 6*(1–2), 1–16. https://doi.org/10.1007/s13162-016-0078-z.

Korfiatis, N., García-Bariocanal, E., & Sánchez-Alonso, S. (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. https://doi.org/10.1016/j.elerap.2011.10.003.

Krishnamoorthy, S. (2015). Linguistic features for review helpfulness prediction. *Expert Systems with Applications, 42*(7), 3751–3759. https://doi.org/10.1016/j.eswa.2014.12.044.

Lee, P. J., Hu, Y. H., & Lu, K. T. (2018). Assessing the helpfulness of online hotel reviews: A classification-based approach. *Telematics and Informatics, 35*(2), 436–445. https://doi.org/10.1016/j.tele.2018.01.001.

Lee, S., & Choeh, J. Y. (2014). Predicting the helpfulness of online reviews using multilayer perceptron neural networks. *Expert Systems with Applications, 41*(6), 3041–3046. https://doi.org/10.1016/j.eswa.2013.10.034.

Lu, S., Wu, J., & Tseng, S. L. (2018). How Online Reviews Become Helpful: A Dynamic Perspective. *Journal of Interactive Marketing, 44*, 17–28. https://doi.org/10.1016/j.intmar.2018.05.005.

Ma, Y., Xiang, Z., Du, Q., & Fan, W. (2018). Effects of user-provided photos on hotel review helpfulness: An analytical approach with deep leaning. *International Journal of Hospitality Management, 71*, 120–131. https://doi.org/10.1016/j.ijhm.2017.12.008.

Malik, M. S. I., & Hussain, A. (2017). Helpfulness of product reviews as a function of discrete positive and negative emotions. *Computers in Human Behavior, 73*, 290–302. https://doi.org/10.1016/j.chb.2017.03.053.

Malik, M. S. I., & Hussain, A. (2018). An analysis of review content and reviewer variables that contribute to review helpfulness. *Information Processing and Management, 54*(1), 88–104. https://doi.org/10.1016/j.ipm.2017.09.004.

McAuley, J., Targett, C., Shi, Q., & Van Den Hengel, A. (2015). Image-based recommendations on styles and substitutes. In *SIGIR 2015 - Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 43–52). https://doi.org/10.1145/2766462.2767755.

Mitra, S., & Jenamani, M. (2020). OBIM: A computational model to estimate brand image from online consumer review. *Journal of Business Research, 114*, 213–226. https://doi.org/10.1016/j.jbusres.2020.04.003.

Mudambi, S.M., .& Schuff, D. (2010). What makes a helpful online review? A study of customer reviews on amazon.com. *MIS Quarterly: Management Information Systems*. 10.2307/20721420.

Olatunji, I. E., Li, X., & Lam, W. (2020). Context-aware helpfulness prediction for online product reviews. In *, 12004. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (pp. 56–65). LNCS. https://doi.org/10.1007/978-3-030-42835-8_6.

Otter, D. W., Medina, J. R., & Kalita, J. K. (2020). A survey of the usages of deep learning for natural language processing. *IEEE Transactions on Neural Networks and Learning Systems*, 1–21. https://doi.org/10.1109/tnnls.2020.2979670.

Park, D. H., Lee, J., & Han, I. (2007). The effect of on-line consumer reviews on consumer purchasing intention: The moderating role of involvement. *International Journal of Electronic Commerce, 11*(4), 125–148. https://doi.org/10.2753/JEC1086-4415110405.

Qazi, A., Shah Syed, K. B., Raj, R. G., Cambria, E., Tahir, M., & Alghazzawi, D. (2016). A concept-level approach to the analysis of online review helpfulness. *Computers in Human Behavior, 58*, 75–81. https://doi.org/10.1016/j.chb.2015.12.028.

Ren, G., & Hong, T. (2019). Examining the relationship between specific negative emotions and the perceived helpfulness of online reviews. *Information Processing and Management, 56*(4), 1425–1438. https://doi.org/10.1016/j.ipm.2018.04.003.

Ren, Y., & Ji, D. (2017). Neural networks for deceptive opinion spam detection: An empirical study. *Information Sciences, 385–386*, 213–224. https://doi.org/10.1016/j.ins.2017.01.015.

Rezaeinia, S. M., Rahmani, R., Ghodsi, A., & Veisi, H. (2019). Sentiment analysis based on improved pre-trained word embeddings. *Expert Systems with Applications, 117*, 139–147. https://doi.org/10.1016/j.eswa.2018.08.044.

Saumya, S., Singh, J. P., Baabdullah, A. M., Rana, N. P., & Dwivedi, Y. K. (2018). Ranking online consumer reviews. *Electronic Commerce Research and Applications, 29* (January), 78–89. https://doi.org/10.1016/j.elerap.2018.03.008.

Saumya, S., Singh, J. P., & Dwivedi, Y. K. (2020). Predicting the helpfulness score of online reviews using convolutional neural network. *Soft Computing, 24*(15), 10989–11005. https://doi.org/10.1007/s00500-019-03851-5.

Singh, J. P., Irani, S., Rana, N. P., Dwivedi, Y. K., Saumya, S., & Kumar Roy, P. (2017). Predicting the "helpfulness" of online consumer reviews. *Journal of Business Research, 70*, 346–355. https://doi.org/10.1016/j.jbusres.2016.08.008.

Sundermeyer, M., Schlüter, R., & Ney, H. (2012). LSTM neural networks for language modeling. In *, 1. 13th Annual Conference of the International Speech Communication Association 2012, INTERSPEECH 2012* (pp. 194–197).

Tractinsky, N., & Srinivasan Rao, V. (2001). Incorporating social dimensions in Web-store design. *Human Systems Management, 20*(2), 105–121.

Tripathy, A., Agrawal, A., & Rath, S. K. (2016). Classification of sentiment reviews using n-gram machine learning approach. *Expert Systems with Applications*. https://doi.org/10.1016/j.eswa.2016.03.028.

Wang, X., Tang, L. R., & Kim, E. (2019). More than words: Do emotional content and linguistic style matching matter on restaurant review helpfulness? *International Journal of Hospitality Management, 77*, 438–447. https://doi.org/10.1016/j.ijhm.2018.08.007.

Wang, Z., Li, H., Ye, Q., & Law, R. (2016). Saliency effects of online reviews embedded in the description on sales: Moderating role of reputation. *Decision Support Systems, 87*, 50–58. https://doi.org/10.1016/j.dss.2016.04.008.

Weiss, S. M., Indurkhya, N., Zhang, T., & Damerau, F. J. (2005). *Text mining: predictive methods for analyzing unstructured information*. Springer Science & Business Media.

Weathers, D., Swain, S. D., & Grover, V. (2015). Can online product reviews be more helpful? Examining characteristics of information content by product type. *Decision Support Systems, 79*, 12–23. https://doi.org/10.1016/j.dss.2015.07.009.

Wu, L., Yen, I. E. H., Xu, K., Xu, F., Balakrishnan, A., Chen, P. Y., et al. (2020). Word mover's embedding: From word2vec to document embedding. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, EMNLP 2018* (pp. 4524–4534). https://doi.org/10.18653/v1/d18-1482.

Xu, X., Wang, X., Li, Y., & Haghighi, M. (2017). Business intelligence in online customer textual reviews: Understanding consumer perceptions and influential factors. *International Journal of Information Management, 37*(6), 673–683. https://doi.org/10.1016/j.ijinfomgt.2017.06.004.

Yang, Y., Chen, C., & Bao, F. S. (2017). Aspect-based helpfulness prediction for online product reviews. In *Proceedings - 2016 IEEE 28th International Conference on Tools with Artificial Intelligence, ICTAI 2016* (pp. 836–843). https://doi.org/10.1109/ICTAI.2016.0127.

Yang, Y., Qiu, M., Yan, Y., & Bao, F. S. (2015). Semantic analysis and helpfulness prediction of text for online product reviews. In *, 2. ACL-IJCNLP 2015 - 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing, Proceedings of the Conference* (pp. 38–44). https://doi.org/10.3115/v1/p15-2007.

Yu, L. C., Wang, J., Lai, K. R., & Zhang, X. (2017). Refining word embeddings for sentiment analysis. In *EMNLP 2017 - Conference on Empirical Methods in Natural Language Processing, Proceedings* (pp. 534–539). Stroudsburg, PA, USA: Association for Computational Linguistics. https://doi.org/10.18653/v1/d17-1056.

Zhang, Q., Yang, L. T., Chen, Z., & Li, P. (2018). A survey on deep learning for big data. *Information Fusion, 42*, 146–157. https://doi.org/10.1016/j.inffus.2017.10.006.

Zhang, S., Yao, L., Sun, A., & Tay, Y. (2019). Deep learning based recommender system: A survey and new perspectives. *ACM Computing Surveys*. https://doi.org/10.1145/3285029.

Zhang, Y., & Lin, Z. (2018). Predicting the helpfulness of online product reviews: A multilingual approach. *Electronic Commerce Research and Applications, 27*, 1–10. https://doi.org/10.1016/j.elerap.2017.10.008.

zhou, Y., Yang, S., li, yixiao, chen, Y., Yao, J., & Qazi, A. (2020). Does the review deserve more helpfulness when its title resembles the content? Locating helpful reviews by text mining. *Information Processing and Management, 57*(2). https://doi.org/10.1016/j.ipm.2019.102179.

Zhu, F., & Zhang, X. (2010). Impact of online consumer reviews on Sales: The moderating role of product and consumer characteristics. *Journal of Marketing*. https://doi.org/10.1509/jmkg.74.2.133.