

# Large Scale Cross-Category Analysis of Consumer Review Content on Sales Conversion Leveraging Deep Learning

Xiao Liu, Dokyun Lee, and Kannan Srinivasan

New York University, Carnegie Mellon University  
xliu@stern.nyu.edu, dokyun@cmu.edu, kannans@cmu.edu

## Abstract

Consumers often rely on product reviews to make purchase decisions, but how consumers use review content in their decision making has remained a black box. In the past, extracting information from product reviews has been a labor intensive process that has restricted studies on this topic to single product categories or those limited to summary statistics such as volume, valence, and ratings. This paper uses deep learning natural language processing techniques to overcome the limitations of manual information extraction and shed light into the black box of how consumers use review content. With the help of a comprehensive dataset that tracks individual-level review reading, search, as well as purchase behaviors on an e-commerce portal, we extract six quality and price content dimensions from over 500,000 reviews, covering nearly 600 product categories. The scale, scope, and precision of such a study would have been impractical using human coders or classical machine learning models. We achieve two objectives. First, we describe consumers' review content reading behaviors. We find that although consumers do not read review content all the time, they do rely on review content for products that are expensive or of uncertain quality. Second, we quantify the causal impact of content information of read reviews on sales. We use a regression discontinuity in time design and leverage the variation in the review content seen by consumers due to newly added reviews. To extract content information, we develop two deep learning models: a full deep learning model that predicts conversion directly and a partial deep learning model that identifies content dimensions. Across both models, we find that aesthetics and price content in the reviews significantly affect conversion across almost all product categories. Review content information has a higher impact on sales when the average rating is higher and the variance of ratings is lower. Consumers depend more on review content when the market is more competitive or immature. A counterfactual simulation suggests that re-ordering reviews based on content can have the same effect as a 1.6% price cut for boosting conversion.

## Introduction

Consumers increasingly read review content when making purchase decisions. As a result, e-commerce platforms need to understand how review content impacts purchase behaviors, in order to provide the most relevant content information. Merchants also need to understand the relationship between review content and conversion so as to adjust product design or pricing strategies to meet consumers' needs. And yet, despite the importance of product review content on business success, how consumers use the review content remains a black box. Prior studies on the impact of reviews were limited to easy-to-use volume, rating, and sentiment metrics that provide little information on the content of the reviews (Chevalier and Mayzlin 2006). In this paper, we shed light on this black box by diving deep into the content of the reviews and measuring its impact on sales conversion.

In contrast to the studies using aggregate review metrics such as volume and valence, review content analysis poses four key challenges that remain unresolved. 1. Data on review content reading behaviors are not available. 2. Content information in the reviews is represented by unstructured words and sentences. There is no guidance on which numerical variables to extract from reviews such that the variables can be relevant and insightful. 3. Traditional information retrieval methods are not applicable to cross-category analysis. 4. Even after knowing the relationship between review content and conversion, marketers still lack a practical strategy using reviews to boost conversion.

In this paper, we tackle each of the research challenges. 1. We leverage a unique dataset to provide a topology of consumer product purchase journeys, incorporating not only traditionally available click-stream and transactional data,

but also consumers' review content reading behaviors. 2. We open up the black box of how content affects conversion by investigating what type of content in the reviews actually shifts consumer purchase decisions. Using a marketing theoretical framework, we identify and extract six distinct quality and price content dimensions from product reviews. 3. We develop two deep learning models to tackle the problem of scalable information retrieval and quantify the impact of review content on conversion. 4. We devise a new ranking algorithm for e-commerce sites to improve conversion and provide managerial implications for when review content plays a crucial role in consumer decision making.

Several interesting insights emerge from this effort. 1. Assuming that consumers read all the reviews would result in estimates that significantly understate the impact of review content on purchase conversion. 2. Across numerous and diverse product categories, aesthetics and price content information are important dimensions. 3. Deep learning models outperform simple conventional natural language processing models without the need for time-consuming feature engineering and are scalable across different products. 4. Firms can benefit significantly by simply reordering reviews based on content.

The paper makes several contributions. Substantively, we open up the black box of how consumers use review content, by leveraging a comprehensive dataset that tracks consumers' review-reading behaviors. The findings shed light on when, where, and to what extent reviews play a role in the consumer purchase decision journey. Moreover, we conduct content analysis to extract distinct dimensions of price and quality information from the reviews. Instead of relying on surface-level measures such as volume or valence, we dissect reviews to their core information, price and quality, which differentiates product reviews from other text documents such as social media posts or news articles. We create labeled review documents that are of high precision and can be used for future studies by other researchers. Furthermore, we study the effect of review content on conversion for a wide range of product categories, whereas prior works focus on only one product category. This allows us to compare the heterogeneous effect of reviews across product categories. Methodologically, we introduce state-of-the-art deep learning natural language processing techniques to the marketing literature. We demonstrate the comparative advantages of deep learning models for automatic classification and visualization of exceedingly diverse, unstructured data. In addition, we propose a novel identification strategy to quantify the causal impact of review reading on conversion. The exogenous variation of reviews comes from the timing of when new reviews are posted, which facilitates a regression discontinuity in time design.

## Descriptions of Consumers' Review-Reading Behaviors

Our data come from a major online retailer in the United Kingdom. It is a panel of 243,000 consumers over the course of two months in February and March 2015. The data track all consumer behaviors, including page views, impressions, used interactions, and transactions. We explain them one by one. At the bottom of the product information page, a consumer can click a button to read detailed review content. The data cover two broad departments, Home and Garden as well as Technology. Each department consists of hundreds of product categories. Among all these product categories, consumers had around 2.5 million page views, 12.3 million review impressions, 500,000 used interactions, and 30,000 transactions. These actions were taken on one of the two devices, PC or mobile phone.

Looking across all consumer journeys, we find that 71.2% of the time, consumers do not read review content. But they tend to read reviews for expensive products, products without many reviews, and products with many questions and answers. Reviews might have heterogeneous effects on consumers. Approximately 43% of consumers never read reviews for any products, whereas 10% of consumers always read reviews for all products. And consumers are more likely to read reviews on PCs than on mobile. The above analysis describes that for certain products, for certain consumers, and on certain devices, consumers do not pay attention to reviews in their online purchase journeys. So in order to quantify the effect of reviews on conversion, we need to select the products for which reviews matter, select the consumers who read reviews, and differentiate the effect on different devices. In the next two sections, we build models to do so.

## Full Deep Learning Model

We first build a full deep learning model that combines consumer characteristics, product attributes, and review content in a joint framework to predict conversion.

Inspired by the works of Kalchbrenner et al. (2014) and Kim (2014), we propose the following convolutional neural network (CNN) model to examine the impact of review content information on sales conversion. Essentially, CNN models use a two-step approach: "convolution" and "pooling." "Convolution" applies a filter over each sliding window of the sentence to capture important local clues whereas "Pooling" aggregates the outputs from the filters by creating a location-insensitive summary statistic. The model architecture is illustrated in Figure 1.

Our model architecture has four layers. The first layer (the leftmost layer) is the word embedding of product reviews and the second layer is the convolutional layer. The third

layer is the max-over-time pooling layer, appended with neurons of consumer and product characteristics information. And the final layer is our outcome: conversion.

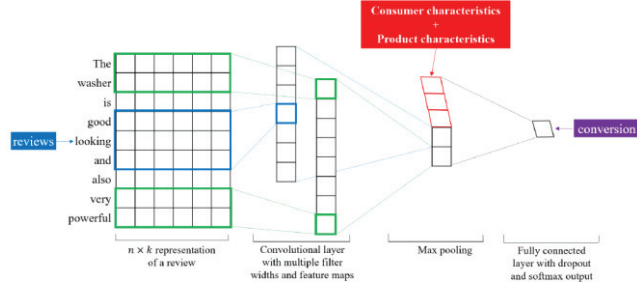


Figure 1 CNN Model Architecture

### Layer 1: Word Embedding

We use the word2vec embedding published by Google. In Figure 1, the first layer is the word embedding. Let us use one review as an example. Suppose  $n$  is the total number of words in the review and  $k$  is the dimensionality of the word embedding ( $k=300$ ). We use  $\vec{x}_i$ , a  $k$ -dimensional word vector to represent the  $i$ th word in the review. And we use  $\vec{x}_{1:n} = \vec{x}_1 \oplus \vec{x}_2 \oplus \dots \oplus \vec{x}_n$ ;  $n$  to represent the entire review. Here,  $\oplus$  is the concatenate operator which stacks all the  $n$ -word (column-) vectors to form a  $nk \times 1$  vector. This vector represents information in the entire review.

### Layer 2: Convolution Operation

The next layer is the convolutional layer, which applies the convolution operation or filter to the word embedding in the first layer. The convolution operation, or filter, is a one-dimensional vector of length  $h$ , applied to each sliding window of  $h$  words in the sentence. For example, in Figure 1, the green filter size is  $h = 2$ . The green filter can be written as  $\vec{w} \in R^{hk}$ . This is a  $hk \times 1$  vector where  $h$  is the window size and  $k$  is the dimensionality of the word embeddings in the previous layer ( $k=300$ ). The filter is first applied to the window of the first two words “the washer,” then to the next two words “washer is,” and then to the following two words “is good,” and so on. Let  $i$  be the current position and  $\vec{x}_{i:i+h-1} \in R^{hk}$  be the window of words or  $n$ -grams that the filter is applied to. The output of the convolution operation is  $c_i = f(\vec{w} \cdot \vec{x}_{i:i+h-1} + b)$  where  $\cdot$  denotes inner product,  $b \in R$  is the bias parameter, and  $f$  is the activation function. The activation function is a non-linear function, which allows neural network models to incorporate non-linear relationships between input variables. We use the rectified linear units (ReLU) as the activation function Goodfellow et al. 2016, pp. 187). The ReLU function is defined as  $f(x) = \max(x, 0)$ .

The filter is rolled over to each sliding window of  $h$  words for  $i = 1, 2, \dots$ . So the final output is a vector  $\vec{c} \in R^{n-h+1}$ ,

called the feature map. Specifically,  $\vec{c} \in [c_1, c_2, \dots, c_{n-h+1}]$ . In our setting, we try different window sizes:  $h = 2, 3, 4, 5$  to incorporate bi-grams, tri-grams, 4-grams, and 5-grams.

### Layer 3: Pooling

The third layer is the pooling layer. The pooling layer applies the max-over-time pooling operation to the feature map created in the previous layer (layer 2). The idea behind the max-over-time pooling is that we want to get the most salient information across all window positions, so

$\hat{c} = \max\{\vec{c}\}$ . The above-mentioned layer 1, 2, 3 extracts one feature from 1 filter. In reality, we can repeat the process and apply multiple filters with varying window sizes to create multiple features. Let the total number of filters be  $m$ . So the penultimate layer becomes a vector of all the features (each feature corresponds to one filter) extracted from the text data, i.e.,  $\vec{d} = [\hat{c}_1, \hat{c}_2, \dots, \hat{c}_m]$ .

In our application, we use a total of 100 filters ( $m=100$ ) for each window size  $h$  ( $h=2, 3, 4, 5$ ). This results in 400 features, each representing a distinct content dimension.

### Layer 4: Append and Output

In the last layer, the features extracted from the review text data are combined with variables of consumer characteristics ( $\vec{Z}_{it}$ , including total products searched and number of used interactions), observed product characteristics ( $\vec{X}_{it}$ , including price, average rating, cumulative number of reviews, percentage of consumers recommended, number of questions and answers), unobserved product characteristics ( $\xi_j$ ), and time fixed effects ( $Weekend_t, Daytime_t$ ) to predict the final outcome: conversion. The activation function used in the last layer is the softmax function (Bishop 2006):  $softmax(c) = e^{x_c} / (\sum e^{x_c})$ . In our case, this is equivalent to the logit choice probability function in discrete choice models (Train 2009) because we have a binary outcome of convert versus not convert.

Let  $y_{ijkt}$  denote conversion ( $y_{ijkt} = 1$  implies purchase and  $y_{ijkt} = 0$  implies no purchase) for consumer  $i$  considering product  $j$  using device  $k$  at time  $t$ . Then the specification in the last layer of the neural network is

$$y_{ijkt} = softmax(\vec{w}_k^y \cdot \vec{d}_{ijt} + b_k^y + \theta_k^y \vec{Z}_{it} + \gamma_k^y \vec{X}_{jt} + \xi_j + Weekend_t + Daytime_t)$$

We append the consumer and product characteristics to text features in the last layer of the neural network to model conversion directly. In other words, this model creates a single, joint deep neural network model to forecast conversion. We call this the “full model.” One thing to note is that the text features in this direct approach are not interpretable by face value. This is why in the next section, we introduce a two-step approach, or a “partial model.”

## Partial Deep Learning Model

In this section, we build a partial deep learning model where we use deep learning to extract interpretable content dimensions and then pass them to a classical choice model for inference. To better explain the model, we first elucidate the identification strategy.

To identify the effect of reviews read on conversion, we need exogenous variations of the reviews read by consumers. In our setting, the exogenous change of the review content comes from the timing of when new reviews are posted to the site. This allows us to use the Regression Discontinuity in Time (Lee and Lemieux 2010; Hausman and Rapson 2017) design (RDiT) to achieve identification.

The identification assumption is that before and after this review post, product characteristics stay unchanged, and other unobserved demand shocks, such as advertising, offline word of mouth, seasonality, etc., change continuously in time. However, the review content dimensions read by consumers have a discontinuous change when the new review arrives. So long as the timing of when new reviews are added are not correlated with the unobserved demand shocks, the timing of new reviews can be used to identify the effect of review content separately from unobserved demand shocks.

Now we lay out the model specification. We apply the random utility framework in classical choice models (Train 2009) and estimate the following specification

$$u_{ijkt} = \overrightarrow{ReviewContent} \cdot \overrightarrow{\tau_k} + \sum_{n=1}^3 \delta_{nk} t^n + \alpha_{ik} + \theta_k Z_{it} + \gamma_k \overrightarrow{X}_{jt} + \xi_j + Weekend_t + Daytime_t + \epsilon_{ijkt} \quad (1)$$

In equation (1),  $u_{ijkt}$  denotes the utility for consumer  $i$ , using device  $k$  considering product  $j$  at the visit time  $t$ . The term  $\overrightarrow{ReviewContent} \cdot \overrightarrow{\tau_k}$  identifies the effect of review content on conversion. As we consider multiple review content dimensions (to be introduced later), we use the vector notation for both  $\overrightarrow{ReviewContent}$  and  $\overrightarrow{\tau_k}$ . Specifically, following Lee and Lemieux (2010), we let  $\overrightarrow{ReviewContent} = \{d_{jt}\}$  be a vector of treatment variables or assignment variables. For each review dimension  $d$  (to be later),  $d_{jt} = 1$  if the consumer visit time  $t$  is after a new review for product  $j$  with content dimension  $d$  being posted to the site.

And  $d_{jt} = 0$  if the visit time  $t$  is before the review (for product  $j$  with content dimension  $d$ ) posting time. The coefficients in the vector  $\overrightarrow{\tau_k}$  are our primary parameters of interest. To account for potential time-varying factors, we include flexible time polynomials with order up to 3, represented by the term  $\sum_{n=1}^3 \delta_{nk} t^n$ . The optimal order of the polynomial is chosen using Akaike's criterion (Lee and Lemieux 2010).

As suggested by Hausman and Rapson 2017, we also add many covariates as control variables to account for potential discontinuous effects in time and consumer and product heterogeneity. These covariates include consumer  $i$ 's intrinsic

preference using device  $\alpha_{ik}$ , observed consumer activities or consumer characteristics vector  $\overrightarrow{Z}_{it}$ , product characteristics vector  $\overrightarrow{X}_{jt}$ . Product characteristics may vary over time. For example, the e-commerce site performs the dynamic pricing strategy. From our conversation with the site managers, the pricing strategy is not targeted. So the price endogeneity issue is eliminated. other unobserved product characteristics  $\xi_j$  as well as Weekend and Daytime fixed effects  $Weekend_t$  and  $Daytime_t$ . Finally we add the idiosyncratic shock  $\epsilon_{ijkt}$ . The shock term  $\epsilon_{ijkt}$  is assumed to follow a Type I Extreme Value distribution. So the conversion rate has a closed-form formula. That is,  $ConversionRate_{ijkt} = \exp(u_{ijkt}) / [1 + \exp(u_{ijkt})]$ .

We consider the "Quality" and "Price" information embedded in the review content. Price and quality of products have been the main drivers of economic transactions and consumer purchase behavior both online and offline. Thus, we look at how price and quality information within reviews influences consumer purchase decisions. While the price dimension of a product is unambiguous, the quality dimension requires a framework to define, identify, and content-code before we can measure its effect. We adopt a theory-driven approach and take a seminal work by Garvin (1984) to operationalize different dimensions of product quality found to influence purchase behavior. Garvin (1984) and Garvin (1987) introduced a set of quality dimensions that were aimed at helping organizations think about quality. The eight dimensions proposed were performance, features, reliability, conformance, durability, serviceability, aesthetics, and brand (perceived quality). We closely follow Garvin's definitions of these different dimensions to identify quality information in reviews. We combine some qualities that are conceptually close. Specifically, we include six dimensions: "aesthetics," "conformance," "durability," "feature," "brand," and "price," and each has two valences, positive and negative. We consider these attributes to be the main focus of review content analysis.

In our partial deep learning model, we utilize deep learning for two purposes. First, we use three deep learning algorithms, recurrent neural network, recursive neural network and convolutional neural network, as supervised learning classifiers to extract pre-defined price and quality content dimensions. All the algorithms are scalable across different product categories. Second, we use a salient phrase visualization technique from the deep learning literature to highlight salient sentences that are topic-relevant.

We conduct supervised learning in multiple steps. First, we collect a labeled dataset of 5,000 random reviews using Amazon Mechanical Turk. Second, the labeled data are divided into a training set comprising 70% of the observations and a test set with 30% of the observations.



## Results

### Compare Full and Partial Deep Learning Models

The results are reported in Table 1. The hit rate (1-misclassification rate) for the full model is 98.9%, much higher than the 96.2% for the partial model. This implies that combining all input variables, text features, and consumer/product characteristics in a joint deep learning framework improves prediction accuracy because this is a more efficient way of utilizing information.

| DV: Conversion              | Hit Rate | AUC   |
|-----------------------------|----------|-------|
| Full Deep Learning Model    | 98.9%    | 0.924 |
| Partial Deep Learning Model | 96.2%    | 0.873 |

Table 1 Model Comparison

### Results of the Partial Deep Learning Model

A comparative advantage of deep learning models is their capability to sift features exclusive to various domains without researchers’ domain knowledge to hand-pick features. Rather, deep learning models can accept raw data from any domain and automatically discover the representations pertinent to each domain. In our application, this comparative advantage of deep learning allows us to detect distinct review features appropriate for each product category. We use three examples to demonstrate that deep learning, or CNN in particular, can spot category-specific salient sentences which contain the six dimensions of information that reviews. In Figure 2, we exhibit examples of salient sentences in the TV, Curtain and Floorcare categories, respectively. By comparing the three categories, we stress that even for information dimensions that involve broad representations, such as aesthetics or feature, deep learning is capable of identifying vastly diverse salient sentences across different categories.

|                     | Positive   | Negative   |
|---------------------|--|--|
| Television          | Easy to setupInternal browser.Cheap price.Great connectivity and shape.Excellent picture. Definitely recommend it.   | Sounds awful. you get a vibration on speech and some music has a quarter to half volume. I'm taking mine back. The vibrations are giving me headaches. (Thought i could live with it but its really really bad |
| Curtains and blinds | Iv bought these in the past before - they are very thin - no lining at all - but they are the perfect colour for my kids rooms. Buy with black out blinds :)                             | I wouldn't say this is bubblegum pink - it is more like a lilac colour. In my daughter's pink bedroom these do not look pink. That is my only criticism - otherwise this is an ok product.                     |
| Floorcare           | I found this product dealt with animal hairs extremely efficiently - I can definitely confirm it has made a great difference to the carpets. they are much brighter and cleaner looking. | Excellent product but one day after carpet hav been washed, the rooms have negative smell for few days.  |

Figure 2 Salient Sentences for Aesthetics Information

For instance, in the TV category, aesthetics can mean “vibrations” as manifested by the review sentence “The vibrations are giving me headache”. On the other hand, in the

Curtain category, aesthetics can reflect colors as in the review “but they are the perfect colour for my kids room”. Finally, in the Floorcare category, aesthetics can regard “smell” as shown in the review “the rooms have negative smell for few days”. This distinguishing aspect of deep learning makes it particularly useful for analyzing text data from a wide realm where extracting domain-specific features is time-consuming and error-prone.

| DV: Conversion      | Read     |           | All     |         |
|---------------------|----------|-----------|---------|---------|
|                     | Mobile   | PC        | Mobile  | PC      |
| Review-Aesthetics P | 0.0157*  | 0.0862**  | 0.00308 | -0.0527 |
| Review-Aesthetics N | -0.396*  | -0.727*** | -0.0427 | 0.0708  |
| Review-Price P      | 0.1703** | 0.215**   | 0.0619  | 0.0208  |
| Product FE          | Yes      |           | Yes     |         |
| BIC                 | 42903.0  |           | 43283.8 |         |

Table 2 Effect of Review Content of Conversion

Next we present the measured causal impact of reviews on conversion. The statistically significant coefficients are presented in Table 2. Please see the supplementary material for the full table. We find that reviews containing favorable aesthetics and/or price information can significantly boost conversion on both devices, whereas other dimensions, such as conformance, durability, features, or brand, are not prominent. As mentioned in the introduction, we are concerned that without the review-reading data, the effect of review content on conversion can be considerably underestimated. To illustrate this, we repeat the regression, but assume that consumers read all reviews instead of only a limited number of reviews. The results in the right panel of Table 2 indicate that the coefficients for review content measured without the review-reading data are on average underestimated by 20%. Importantly, the positive aesthetics and price content information no longer significantly impacts conversion. This sheds light on the importance of obtaining the review-reading data to correctly quantify the impact of review content on conversion.

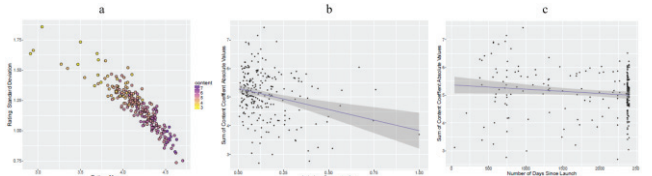


Figure 3 Content Effect by Rating, Competition and Dynamic

### Category Heterogeneity

We decompose the difference among categories in three aspects: rating, competition and dynamic.

#### Rating

Is review content more relevant when the information in ratings is tenuous? This question is of great importance in a world with rating inflation. To shed light on this issue, we plot the effect of review content on conversion by the mean

and standard deviation of ratings in Figure 3a. In this figure, each point is one product category. The color of a point denotes the sum of the absolute values of the review content coefficients. The darker (more purple, less yellow) the color, the higher the effect of content on conversion. The color pattern in the figure suggests that for categories where products have high ratings with low dispersion, review content has a high impact on conversion. This might be because the information is quite noisy, so consumers heavily rely on review content to compare products. Consistent with our expectation, the higher the variance of the rating, the less effective review content is. We also find that the higher the mean rating, the more important review content is. This might be due to the fact that consumers tend to ignore review content if the rating is lower than some minimum threshold.

### Competition

Does review content have a higher impact on conversion in a more competitive market? On the one hand, this might be true because in a more competitive market, consumer choices might be based on small differences in quality or price. Under such conditions, review content may be more likely to provide the marginal push that determines final purchases. On the other hand, in a more competitive market, firms may compete to provide information to buyers so that reviews provide no additional value. We test these competing hypotheses in Figure 3b. The x-axis is the Herfindahl–Hirschman Index (h-index), which measures market concentration of each product category. The lower the h-index, the more competitive the market. The y-axis is the review content effect on conversion, measured as the sum of the absolute values of the review content coefficients in Table 2. The downward sloping trend (slope=-1.475,  $p\_value < 0.0001$ ) indicates that review content is more useful in a more competitive market, consistent with the former view.

### Dynamic

Next we explore the dynamic effect and answer the question “Is review content more effective for early stage products?” This is intuitive because consumers have higher uncertainty with products in the early stage than with mature products. Hence, they are more likely to rely on review content to learn about the early stage products. To assess the dynamic effect, in Figure 3c we plot the effect of review content (measured by the sum of the absolute values of the review content coefficients) against the number of days since a category was launched on the e-commerce platform. The fitted regression line suggests a negative relationship (slope=-0.0002,  $p\_value=0.045$ ) between the review content effect and tenure. In other words, review content is more effective for product categories that are new to the site than for those categories that have existed for a long time.

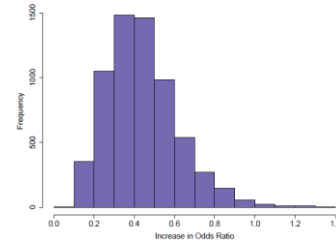


Figure 4 Increase in Odds Ratio by Re-ordering Reviews

## Counterfactual of Changing Ranking Algorithms

After discovering the relative importance of different content information in the review texts, in this section, we propose a strategy that marketers can leverage to boost conversion rate: re-ordering reviews. Our results imply that consumers pay attention not only to the summary statistics of reviews (e.g., average rating, total number of reviews), but also to the actual content of reviews. Their conversion rate is influenced by the content information embedded in the reviews. For example, aesthetics and price information have a stronger positive impact on conversion than other dimensions. As a consequence, within the set of reviews with the same rating score, marketers can display the reviews with positive aesthetics and price information before other reviews, to increase conversion.

We implement a counterfactual scenario where for each product, we randomly select an associated review that contains positive aesthetics information and move it from a lower position to the set of reviews read by each consumer. We then calculate the conversion rate odds ratio for each product and the increase in conversion rate ratio compared to what is observed in the data. Figure 4 displays the histogram of the increase in conversion rate odds ratio. The average increase in the odds ratio of the conversion rate is 44%, while the maximum is 143%. This indicates that on average, re-ordering reviews by presenting one more review with positive aesthetics information is as effective as a 1.6 percent price cut to increase the conversion rate odds ratio.

## Conclusions

The results can assist managers in multiple ways. First, managers can implement the deep learning models to automatically extract price and quality information from reviews of any product category. Second, based on our finding regarding the relative importance of review content dimensions, managers can incorporate reviews as a new marketing mix, by refining the ranking and information presentation algorithms to provide the most relevant reviews to consumers. Third, managers can collect real-time information about the consumer purchase journey, including device and reviews read, to predict final conversion more accurately.

## References

- Bishop, C.M., 2006. *Pattern Recognition and Machine Learning*. Springer.
- Chevalier, J.A. and Mayzlin, D., 2006. The Effect of Word of Mouth on Sales: Online Book Reviews. *Journal of Marketing Research*, 43(3): 345-354.
- Garvin, D.A., 1984. What Does “Product Quality” Really Mean?. *Sloan Management Review*: 25.
- Garvin, D., 1987. Competing on The Eight Dimensions of Quality. *Harvard Business Review*: 101-109.
- Goodfellow, I., Bengio, Y. and Courville, A., 2016. *Deep Learning*. MIT press.
- Hausman, C. and Rapson, D.S., 2017. *Regression Discontinuity in Time: Considerations for Empirical Applications* (No. w23602). National Bureau of Economic Research.
- Kalchbrenner, N., Grefenstette, E. and Blunsom, P., 2014. A Convolutional Neural Network for Modelling Sentences. *arXiv preprint arXiv:1404.2188*.
- Kim, Y., 2014. Convolutional Neural Networks for Sentence Classification. In *In EMNLP*.
- Lee, D.S. and Lemieux, T., 2010. Regression Discontinuity Designs in Economics. *Journal of Economic Literature*, 48(2): 281-355.
- Train, K.E., 2009. *Discrete Choice Methods with Simulation*. Cambridge university press.