Study on Price Consistency regarding Pack Size via Product Variant Retrieval and Pack Size Extraction

Yang Liu
Amazon.com
Seattle, US
yliuu@amazon.com

Chu Wang
Amazon.com
Seattle, US
chuwang@amazon.com

Zuohua Zhang
Amazon.com
Seattle, US
zhzhang@amazon.com

Yongning Wu
Amazon.com
Seattle, US
yongning@amazon.com

ABSTRACT

Price perception is extremely important for retailers. Customers assess the price of a product not only from the product's own price history, but also from the prices of the product's close variants. One particular kind of variant considered is the same product sold in different sizes, where a reduced unit price is generally expected for the ones sold in large quantities. Such price consistency between product variants could be important for customer experience, yet very challenging for retailers which carry millions of products with possibly missing and noisy catalog information. We propose a framework to measure pricing consistency between product size variants by retrieving product variants via search and extracting product size information with natural language processing methods. We evaluate three monotonic regression models that regularize the unit price instead of simple heuristics. To quantify the extent of price inconsistency, we define new metrics and demonstrate that one method can lower the inconsistency measure by up to 45% on the experiment sample set.

KEYWORDS

Price, Information Retrieval, Information Extraction, Regression, Named Entity Recognition

ACM Reference Format:
Yang Liu, Chu Wang, Zuohua Zhang, and Yongning Wu. 2018. Study on Price Consistency regarding Pack Size via Product Variant Retrieval and Pack Size Extraction. In *Woodstock '18: ACM Symposium on Neural Gaze Detection, June 03–05, 2018, Woodstock, NY*. ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/1122445.1122456

1 INTRODUCTION

Price perception has direct impact on customers' purchase decisions, satisfaction, and their intention to return. It is especially true for retailers since customers have many choices where to buy products [4]. Customers perceive the attractiveness of a price by comparing it with their own reference price, which is a price point that customers think is fair to charge for the product [3, 5]. Apart from the past price history of the same product, customers can also develop their reference price from the current prices of other similar products, including those from other retailers. For example, the price of a pair of shoes is expected to be close to the same shoes of different sizes. A big bottle of soda is expected to have a lower unit price than the same kind sold in a small can. The inconsistency between the offered price and customers' reference price tends to make customers puzzled, increases customers' cognitive load, and leads them not to purchase a product that they otherwise might want to. Figure 1 gives an example of such an inconsistent pricing scheme.

Customers may particularly expect consistent pricing schemes when they are shopping in one retail store. However, retailers often price products individually and might not necessarily consider product variants in their general strategies. The unconventional pricing relationship among the variants of the same product in a store could lead to bad customer experiences and damage the customer trust in pricing in the long term. This creates a unique challenge for retailers with millions of products to be priced every day. The measurement of price differences between the same product variants could be one of many signals for retailers to consider when they make decisions.

In this paper, we focus on the price consistency of the product variants with different pack sizes. Pack size is one of the most common types of product variation and directly relevant to pricing strategies. Customers are used to comparing unit price, or price per unit (PPU), when they shop products sold by different weights or volumes. Lower unit price is generally expected for the same product with a larger pack size. Inconsistent pricing schemes may negatively impact customers' shopping experience, and could raise doubts about price reasonableness. So the pricing consistency of the size variants of the same product could be important for sales and customer satisfaction.

There are several challenges that make the measurement hard to do. First, product relationships such as size variations are often incomplete and hard to maintain in the catalog, especially when the product selection is large and diverse, and new products are added into the catalog every day. Next, as a previous study suggests, product attribute information in the catalog can be noisy and

Woodstock '18, June 03–05, 2018, Woodstock, NY

ACM ISBN 978-1-4503-9999-9/18/06…$15.00
https://doi.org/10.1145/1122445.1122456

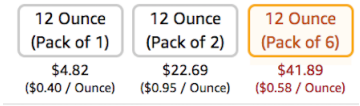

Figure 1: An example of the price information of a product with three pack sizes, Pack of 1, Pack of 2 and Pack of 6. The unit price of Pack of 1 is the cheapest among the three size variants.

missing [14]. A retailer needs to accurately extract products' pack size to correctly calculate and compare the unit prices. Then, even with all the information available, it is not clear what pattern the unit prices among the size variants would have. Finally, as a novel problem, how to measure and quantify the price consistency itself needs to be understood.

To address these challenges, we design a three-step approach using information retrieval and natural language processing techniques. Given a product to be considered, we first determine whether the product is sold by weight or volume by applying a binary text classifier using the recently developed BERT embedding [1]. If the product is sold by weight or volume, we feed the product as a query to retrieve its reference products. We further detect the size variants of the query product from the top results using a size-variant classifier. After extracting the pack size from the product title using a sequence labeling method [6], common monotonic regression models are used to estimate the average unit prices. Finally, the price of the query product for the later evaluation is calculated based on the unit price estimate. To measure the price consistency, we define new metrics to quantify the unit price margin to reach price consistency. The results show that one model can lower the inconsistency measure by 25% to 45%, based on a selected sample set of publicly posted prices.

2 BACKGROUND

Price perception has been studied extensively in marketing and psychology literature. Customers comparing a price to a reference price which depends on their past and present contexts of experiences [3, 5]. Customers' price perception has a direct effect on their satisfaction and intention to return [4, 11].

For products sold by volume, unit price information can help customers save money [10]. It has also been shown that comparing imputed price to explicitly displayed unit price, customers may delay purchase decisions when unit price is not displayed explicitly or customers are too distracted to estimate the unit price [2, 13].

For retailers with millions of products, even showing customers the correct unit price is not trivial, since it relies on the accuracy of pack size information in the catalog [9]. Given the size of the product catalog and various sources of input, the structured form of product attribute information including pack size can be missing and inaccurate.

Consistent prices with respect to unit price between product size variants further requires establishing the variation relationship between products. Maintaining such product relationships in the catalog is not easily scalable as the size of catalog increases and updates become frequent. Instead of curating various relationships between each pair of products in the catalog, we propose to leverage product search to retrieve reference products ad hoc and classify the retrieved results with respect to different types of relationship at query time.

Our work has three major contributions. First we define new metrics to quantify price inconsistency with respect to pack size. Next we design a system to detect product variants with reference product retrieval and extract pack size, which can address the potential catalog issue. Finally, we compare three data-driven methods to understand the price per unit pattern among those pack size variants.

3 PRICE CONSISTENCY DEFINITION

In this section, we define our measure of price inconsistency in the context of pack size. Given a pair of the same products with different pack sizes, we consider the consistent pricing scheme as: 1) the price of the bigger size variant should be higher than the price of the smaller size variant; 2) the PPU of the bigger size variant should be less than or equal to the PPU of the smaller size variant. Therefore, we define two types of pairwise price inconsistencies: *type 1 inconsistency* and *type 2 inconsistency*, to measure the violations of the two assumptions respectively.

Formally, we define type 1 inconsistency as:

$$Inconsistency_{type1} = \max\left(0, \frac{price_{smallsize} - price_{bigsize}}{price_{bigsize} + price_{smallsize}}\right), \quad (1)$$

where $price_{smallsize}$ and $price_{bigsize}$ are the prices of small and big size variants respectively.

Similarly, we define type 2 inconsistency as:

$$Inconsistency_{type2} = \max\left(0, \frac{PPU_{bigsize} - PPU_{smallsize}}{PPU_{bigsize} + PPU_{smallsize}}\right), \quad (2)$$

where $PPU_{smallsize}$ and $PPU_{bigsize}$ are the price per unit of small and big size variants respectively.

Under this definition, we can now quantify the price inconsistency shown in Figure 1. The big size variant (pack of 2) has a unit price of \$0.95 per ounce. The small size variant (pack of 1) has a unit price of \$0.40 per ounce. Thus the prices of this pair of products suffers from type 2 inconsistency of 0.41. We further sampled around 110 thousand product variant pairs in one category. The sampled prices of these pairs on average have type 1 inconsistency of 0.05 and type 2 inconsistency of 0.04. We will describe the details of price consistency evaluation in Section 5.

4 METHODS

The analysis of price consistency requires a query product and a set of its variants, together with their size information. Given low coverage and accuracy of all such information in the catalog, we proceed via the following steps. We first retrieve a list of reference products given a query product with size information, then we detect a set of its pack size variants and extract their size information. After that, the set of variants and their size and price information is readily available to be used for price consistency analysis. In figure 2, we demonstrate the logic flow of our work, and we discuss each component in the following subsections.

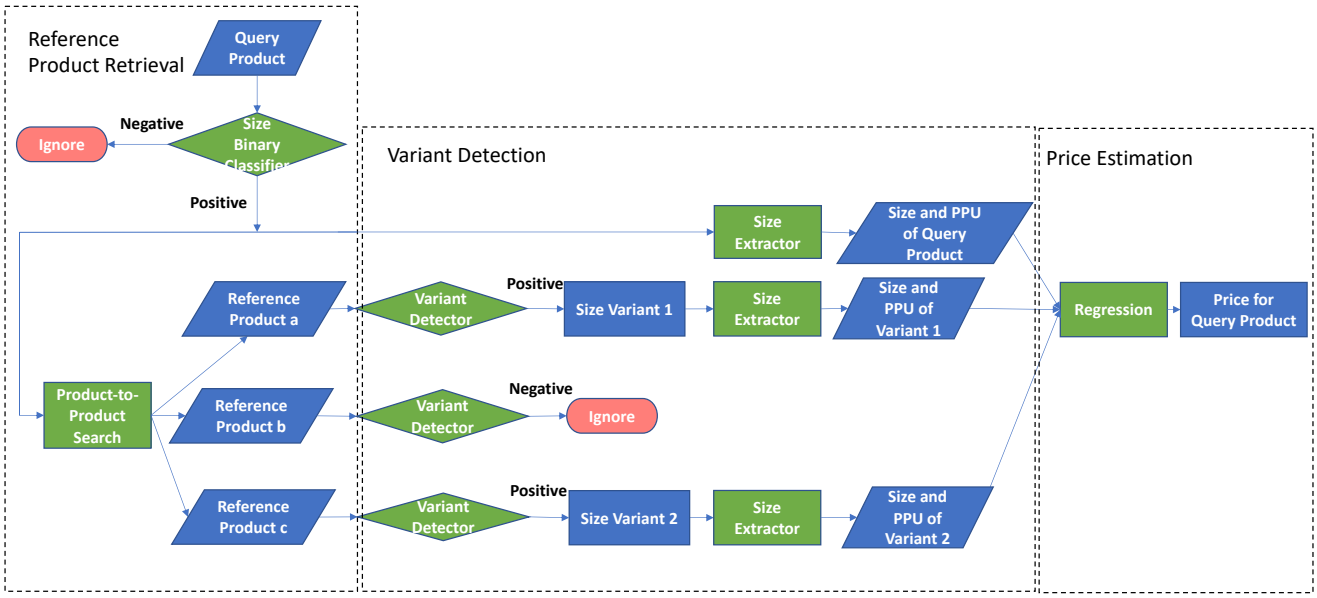

Figure 2: The flow chart of the price consistency analysis.

4.1 Reference Product Retrieval

We train a binary text classifier on the product titles to determine whether a product is sold by weight or volume. Products without such information will be filtered out as they are not proper candidates for price consistency analysis. During training, the positive instances are product size variants extracted from the product catalog, while the negative instances are sampled from products with empty size attributes and without size-related keywords. We apply the state-of-the-art language representation model BERT [1] to generate title embeddings and feed them into a binary classification layer. Specifically, for each product title, we obtain the final hidden state for the special *[CLS]* word embedding denoted as vector $\vec{u} \in \mathbb{R}^d$ using BERT. $\vec{u}$ is then fed into a classification layer $W \in \mathbb{R}^{2\times d}$. The label probability P is computed as $P = \text{softmax}(\vec{u}W^T)$. We fix $\vec{u}$ and only fine-tune W to maximize the log probability of the correct label.

We then use the products passed by the binary classifier as queries to retrieve a list of similar products using reference product search [12]. These reference products will serve as input for the variant detector and product size extractor.

4.2 Variant Detection and Size Extraction

For a query product and a list of reference products similar to the query, we apply a size variation ranking model to obtain the same products with different pack sizes. The size variation ranking model is a Siamese type model consisting of two identical attention-LSTM structures. The model structure is illustrated in Figure 3. For the query product and one of its reference products, the model applies the same attention-LSTM module to both titles for raw feature extraction, and then applies the same attention module to optimally average the signals extracted from each token in the title. The absolute difference between the two resulting hidden layers from the attention-LSTM network is then fed to a fully connected network for the final binary classification to determine whether the reference product is a size variant of the query product.

For pack size extraction from the product title for the variant detected, we adopt a similar approach to [8] and [14] by treating the problem as a sequence labeling task. Specifically, for each token in a product title or description, we classify it as one of the three labels among: single pack size, multi-pack count or other. Figure 4 illustrates an example of the labeling scheme. We generate the training data using size attributes in the catalog and train a sequence labeling model using the BiLSTM-CNNs-CRF framework introduced by Ma and Hovy [6].

After the position of single pack size mention is predicted, we further extract the unit information around it using regular expression. Finally, we multiply the single pack size by pack count to get the total quantity of the product, and write the value in the standardized unit for downstream calculations.

4.3 Price Estimation with Regressed Unit Price

Retailers often offer a quantity discount to incentivize customers to purchase products in large quantities [7]. Instead of a global quantity discount factor, for each query product with its size variants detected, the price of the query product for the later evaluation is calculated based on the unit price by using the total quantity as the independent variable. We experiment with three regression models fitting monotonic functions, namely, linear, exponential, and isotonic regression, to ensure that a lower unit price is associated with a higher pack size. The parameters are estimated by minimizing the mean squared error with the monotonically decreasing constraint in all cases.

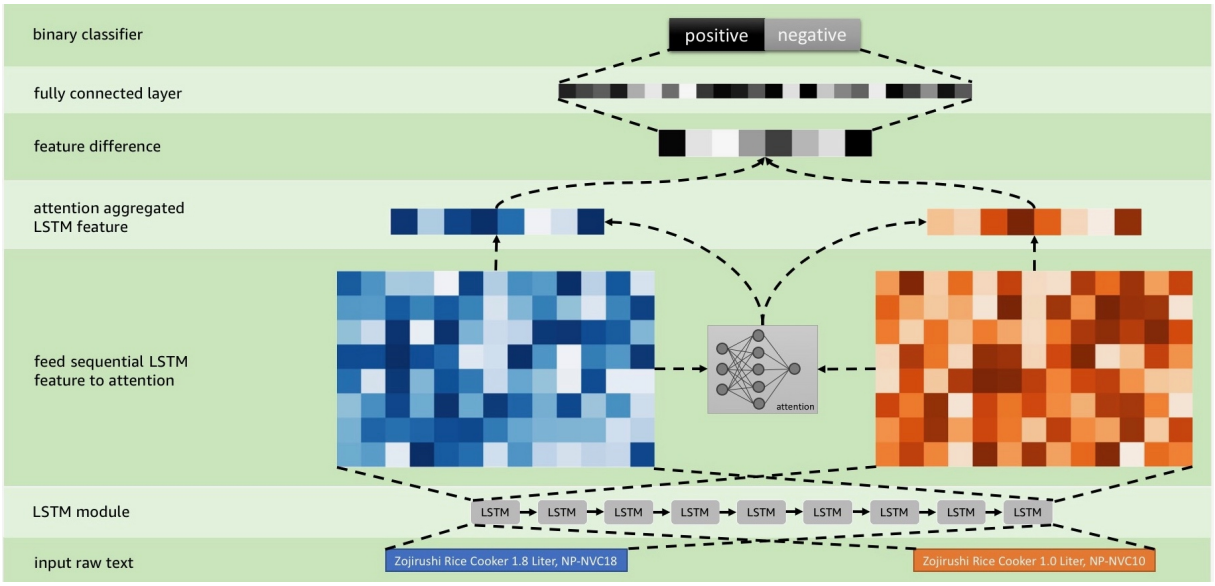

Figure 3: The structure of the size variation ranking model.

Perrier	Carbonated	Mineral	Water,	16.9	fl oz.	Plastic	Bottles	(	24	Count)
O	O	O	O	Single_Size	O	O	O	O	Pack_Count	O

Figure 4: An example of sequence labeling results on product titles. In this case, the single pack volume is 16.9 fl oz., and there are 24 counts. So the total product quantity is 405.6 fl oz.

We assume that most of the publicly posted prices are reasonable. We use the variants' prices to train the regression model, and only calibrate the query product's price for the evaluation purpose. To avoid potential price mistakes or size errors introduced from our upstream models, we skip the query products when the regression models have extremely bad fits. Specifically, we compute the distribution of the mean squared errors of linear regression models and skip the ones with the mean squared error greater than the upper inner fence ($Q3 + 1.5IQR$). [1]

5 EXPERIMENTS

5.1 Data Set

We run our experiment on one category where the size variants are most common. We generated our query product set in the following process. We first collected product information from the catalog. Among those products, about 55% have non-empty catalog size attributes. We consider these products as products with pack size which can skip our classifier. The remaining products which have empty values of size attributes were fed into our binary classifier to further identify ones with pack size information.

[1] 18% of query products were skipped in our experiment.

5.2 Binary Classification of Products with Pack Size

We leveraged the catalog information to build our training data for the classifier. We first extracted a list of size variants from catalog as positive examples. For the negative examples, we extracted products with empty catalog size attributes and filtered them by size-related keywords such as 'oz', 'ct', and 'lbs'. We sampled 60,000 products from both positive and negative products and used their product titles to train our classifier. The product titles were transformed into 768-dimensional embeddings using BERT-Base pretrained model [1] and fed into the final classification layer. The accuracy is 99.2% on the 20% holdout dataset.

With our trained classifier, we further identified 67% as sold by size from those without size attributes. We combined them with the products which have catalog pack size attributes, filtered products without page views for a certain period, and randomly sampled 100,329 products as our final query product set.

5.3 Pack Size Variant Detection

We gathered more than 10 million non-media products for training the model, after filtering out certain products that were more likely to have low quality data. We used the variation data from catalog as the positive training data. For the negative samples, we first conducted reference product retrieval to get a list of similar products, then filtered out the size variants indicated by the catalog, and used random sampling from the rest of the products. Based on our experiment in one category, the accuracy of the binary classifier is above 97%, which is high enough to generate reliable candidates for the following pack-size experiments.

For each of the query products, we applied reference product retrieval to get 50 reference products serving as the size-variant candidates. Then each reference product was coupled with the query product and the pair was fed to the size variation ranking model to

get a probability of both being size variants. Note that it is possible that the model predicts negative for all the reference products, which means that the query does not have any size variants though it may contain size information itself. During our experiment for the 100,329 query products, our variant detector was able to catch at least one variant for 80.9% of them. On average, 10 variants were found for each query. In contrast, the catalog only has variant records for 34.3% of the query products. Therefore, our variant retrieval and detection process is able to extend the existing variation relations by over 200% to make our measurement applicable to more query products.

5.4 Pack Size Extraction

To prepare the training set for our sequence labeling model, we matched the numerical tokens in the product title with the catalog size attributes. One issue with this approach is that it lacks examples of numerical tokens which are not related to pack size. Because of the potential data quality issue of the catalog data, we cannot simply tag all the nonmatched numerical tokens as negative sample. To overcome this problem, we manually annotated 500 products with non-matching numerical tokens. Some common examples of size-irrelevant numerical tokens include Omega-*3*, stage *2* baby food, *6*g of fiber per serving. The final training set contains 100,000 product titles. Table 1 shows the token level scores in the holdout set.

Table 1: Token level scores of the pack size extraction.

Label	Accuracy	Precision	Recall	F1
Single_Size	1.00	0.99	0.99	0.99
Pack_Count	1.00	0.97	0.97	0.97
Other	1.00	1.00	1.00	1.00

We applied our trained model to the 65,325 out of 104,834 query and variant products that are active. Table 2 compares the numbers of missing size attribute between catalog and our model output. It shows that our pack size extraction model was able to reduce the missing rate of the size attribute from catalog by one order of magnitude.

Table 2: Number of missing size attributes from catalog and our model output on 65,325 products.

Attribute Type	Source	Attribute	# NA
Number of Packages	Catalog	case_pack_quantity	35,347
		item_pack_quantity	1,800
		number_of_items	3,871
	Our Model	pack_count	**275**
Pack Size	Catalog	unit_count	1,712
	Our Model	single_size	**276**

5.5 Unit Price Regression

To fit the regression models, we calculated the PPUs for both the query and its variant products with their prices and total quantities extracted by our pack size extraction model.

As we trust most of the publicly posted prices as of the date of the data collection to be reasonable, we can evaluate how the regression model fits the posted prices of the variant products. Table 3 shows the goodness of fit of the three regression models. Isotonic regression has both the lowest mean squared error and the highest R^2 value. Among the two parametric models, the exponential regression model has a better fit. It suggests that at least for this category, the quantity discounts are usually nonlinear.

Table 3: Goodness of fit of regression models

	MSE	R^2
Linear	0.11	0.45
Exponential	0.06	0.62
Isotonic	0.05	0.67

Figure 5 illustrates an example of regression results on the PPUs of a group of product size variants. It shows that isotonic regression has the best fit, but tends to overfit the posted PPU. In addition, its estimate on a product with size outside of the variant products' range will heavily rely on the PPU associated with the smallest or largest one. Linear regression has the worst fit when the actual quantity discount factor is not linear. The exponential regression has a good balance between good fit and generalizability.

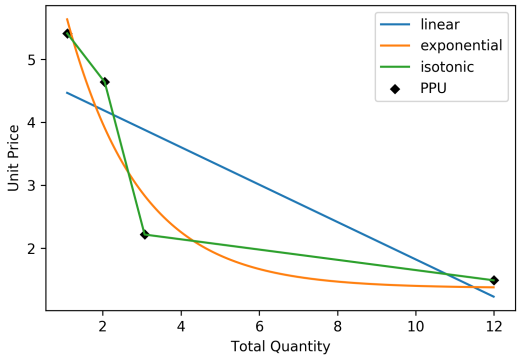

Figure 5: An example of regression results fitting linear, exponential, and isotonic functions.

We show two examples of unit price estimates by exponential regression in Table 4. The first product in each group is the query product. The rest are the variant products retrieved from the product variant detection. In the first case, the PPU of the query product was too low and resulted in type 1 inconsistency. In the second case, the query product had the largest size but also the highest PPU, which led to type 2 inconsistency. In both cases, the unit price regression can help identify and mitigate the inconsistencies. We also compared the price estimates with the MSRPs. The results are close to the MSRPs, which are often manually curated by the vendors and reflect the quantity discount factor that the vendors desire.

Table 4: Examples of PPU calibration. The query products are highlighted.

Item Name	Total Quantity	MSRP	Posted Price	Posted PPU	PPU (exp)	Price (exp)
Product A, 3.4 lb	54.40 oz	35.56	6.95	0.13	0.65	35.38
Product A, 3.5 oz (Pack of 3)	10.50 oz	13.26	12.66	1.21	n/a	n/a
Product A, 3.5 Ounce	3.50 oz	5.45	5.45	1.56	n/a	n/a
Product B, Flavor X, 7 Ounce (Pack of 12)	84 oz	17.5	26.99	0.32	0.21	17.62
Product B, Flavor X, 5 Ounce (Pack of 15)	75 oz	19.99	16.07	0.21	n/a	n/a
Product B, Flavor Y, 16 Ounce (Pack of 12)	192 oz	32.8	36.41	0.19	n/a	n/a

5.6 Price Consistency Evaluation

We report the average inconsistency measure we defined in Section 3 between the query products and their size variants. In addition to the total quantity of the product, the single pack size itself may have some effect on the unit price. For example, a pack of twelve 12 fl oz. Coca-Cola cans does not necessarily have a lower unit price than that of a 2-liter bottle one. Therefore, we further report the strict version of price inconsistencies, in which we only consider the variant pairs with the same single pack size.

We compare three regression models with the publicly posted price as the baseline. Table 5 shows that isotonic regression has the lowest type 2 inconsistency. It lowers type 2 strict measure by 45% comparing to the posted price. Exponential regression has the lowest type 1 inconsistency, where it lowers the type 1 strict measure by 25%.

Table 5: Price inconsistency measure of the query-variant pairs

	Type 1	Type 1 Strict	Type 2	Type 2 Strict
Baseline	4.68%	2.13%	3.70%	2.93%
Linear	4.07%	2.20%	2.53%	2.10%
Exponential	**3.23%**	**1.59%**	2.32%	1.97%
Isotonic	3.61%	1.60%	**2.09%**	**1.62%**

6 CONCLUSION

To understand price consistency regarding pack size is challenging for retailers offering large product selections, especially when the catalog information can be missing and inaccurate. In this paper, we introduced a three-step method that can discover price inconsistencies between product variants and provide price input signals which may mitigate price inconsistency given the challenge of catalog data quality. We proposed to apply product search to retrieve product variants instead of maintaining product relationship explicitly. As side products, we also built high accuracy machine learning models to enrich the product size information and product variants in the catalog, which can be beneficial to many other applications beyond this study. The modular design of our approach allows each component to be tuned and improved independently in future. One direction of our future work is to extend our current framework to other variation types.

REFERENCES

[1] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv preprint arXiv:1810.04805* (2018).

[2] Clive WJ Granger and Andrew Billson. 1972. Consumers' attitudes toward package size and price. *Journal of Marketing Research* (1972), 239–248.

[3] Chris Janiszewski and Donald R Lichtenstein. 1999. A range theory account of price perception. *Journal of Consumer Research* 25, 4 (1999), 353–368.

[4] Pingjun Jiang and Bert Rosenbloom. 2005. Customer intention to return online: price perception, attribute-level performance, and satisfaction unfolding over time. *European Journal of Marketing* 39, 1/2 (2005), 150–174.

[5] Gurumurthy Kalyanaram and Russell S Winer. 1995. Empirical generalizations from reference price research. *Marketing science* 14, 3_supplement (1995), G161–G169.

[6] Xuezhe Ma and Eduard Hovy. 2016. End-to-end Sequence Labeling via Bi-directional LSTM-CNNs-CRF. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, Vol. 1. 1064–1074.

[7] James P Monahan. 1984. A quantity discount pricing model to increase vendor profits. *Management science* 30, 6 (1984), 720–726.

[8] Ajinkya More. 2016. Attribute Extraction from Product Titles in eCommerce. In *Workshop on Enterprise Intelligence, Aug 14, KDD 2016.*

[9] Jagdish Ramakrishnan, Elham Shaabani, Chao Li, and Mátyás A Sustik. 2019. Anomaly Detection for an E-commerce Pricing System. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM.

[10] J Edward Russo. 1977. The value of unit price information. *Journal of Marketing Research* (1977), 193–201.

[11] Sajeev Varki and Mark Colgate. 2001. The role of price perceptions in an integrated model of behavioral intentions. *Journal of Service Research* 3, 3 (2001), 232–240.

[12] Chu Wang, Lei Tang, Shujun Bian, Da Zhang, Zuohua Zhang, and Yongning Wu. 2019. Reference Product Search. In *Companion Proceedings of the 2019 World Wide Web Conference.*

[13] Dengfeng Yan, Jaideep Sengupta, and Robert S Wyer Jr. 2014. Package size and perceived quality: The intervening role of unit price perceptions. *Journal of Consumer Psychology* 24, 1 (2014), 4–17.

[14] Guineng Zheng, Subhabrata Mukherjee, Xin Luna Dong, and Feifei Li. 2018. OpenTag: Open Attribute Value Extraction from Product Profiles. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD '18)*. ACM, New York, NY, USA, 1049–1058.