International Journal on Cybernetics & Informatics (IJCI) Vol. 12, No.1, February 2023

SIZE AND FIT RECOMMENDATIONS FOR COLD START CUSTOMERS IN FASHION E-COMMERCE

Jaidam Ram Tej, Jyotirmoy Banerjee and Narendra Varma Dasararaju

Flipkart Internet Private Limited, Banglore, India

ABSTRACT

Fashion e-commerce is expected to grow rapidly over the next few years. One of the main hurdles in fashion e-commerce is to recommend the right size to customers which helps customers in having a better online shopping experience. Hence Size and fit recommendation is an important problem which helps improve the confidence of a customer for making a purchase on an e-commerce platform. This also reduces the returns in fashion e-commerce. In this work we propose a novel bayesian probabilistic approach for non-personalised product size recommendation for customers. We use maximum likelihood estimation for estimating the parameters of our model. We use customer purchase and returns history to infer the true product size. Given a product we provide size recommendations to a customer, i.e. we suggest a customer to buy a size small, large or same size. In experiments with flipkart shoes datasets our model leads to an improvement of 3-4% AUC over the existing baseline. In Online AB testing for flipkart shoes categories our approach shows a performance improvement of returns by 12-24 bps.

KEYWORDS

Fashion E-Commerce, Recommendation, Size And Fit, Maximum Likelihood, A/B Testing, Catalog

1. INTRODUCTION

Fashion e-commerce has grown enormously over the last few years and is expected to grow at 11.4% per year [1]. Though Fashion e-commerce is becoming popular, customers find it challenging to buy products online and record high return rates. A key reason that leads to excessive product returns is the size and fit problem. Finding the right size and fit is among the important factors impacting customers' purchase decision making process and their satisfaction from e-commerce fashion platforms.

Customers shopping online rely on the symbolic sizes ('6', '7', '8', '9', '10') or size charts mentioned in the products for making a purchasing decision. Even though symbolic sizes are mentioned in the products they vary between brands. Sometimes there is size variation within the brand as well. Size charts provide a mapping from standard sizes to corresponding sizes in cm or inches. These size charts are usually at brand level and do not capture finer fit details of the product. Further size charts are difficult to enact requiring measuring instruments at the disposal. In addition to these brands apply vanity sizing in which they increase physical measurements of a nominal size to boost customer's self-esteem. Thus the size information mentioned in the product alone is insufficient for making a purchasing decision.

David C. Wyld et al. (Eds): NLDML, CCITT, EDUTEC, AUEN -2023
pp. 21-30, 2023. IJCI – 2023 DOI:10.5121/ijci.2023.120102

Fit in a garment is an important factor that contributes to the confidence and comfort of the customer. Customers will return if the garment does not fit. Inorder to reduce returns, online stores have to provide accurate information to the customers. Customers on the other hand should feel confident that the size they buy is the right size choice for them. Helping customers choose their garments with the right fit, is the key to build customer loyalty. To address the size issue, fashion e-commerce companies use size tips to educate and help the customers in their purchase journey [2], [3]. Size tips as shown in figure 1, are easy to consume.This mode of communication is free from customer friction. Size tips are typically the following: ‘buy a size larger’, ‘true to size’ and ‘buy a size smaller’. Implies that they recommend a new user, who usually buys UK 9, to buy UK 10, UK 9 and UK 8, respectively.

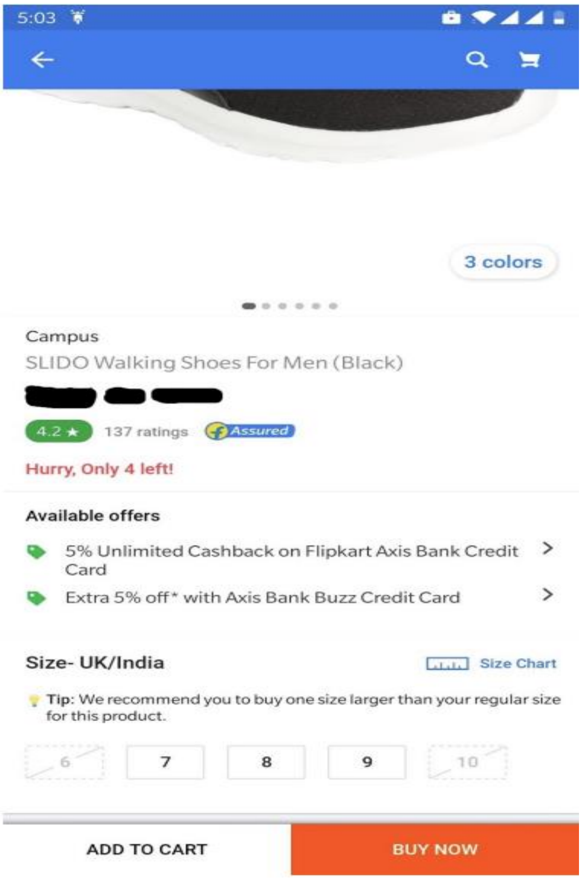

Figure 1. Size tip: We recommend you to buy one size larger than your regular size for this product.

Fashion e-Commerce companies with multiple sellers in their platform, referred to as a marketplace, are often unaware of the true sizes of the product listed in their catalog. Customers after purchasing a product may choose to return the product if they find it ‘small’ or ‘large’. If the product is not returned, then it is deemed as ‘fit’ for the customer. The fit feedback, i.e. ‘small’, ‘fit’ or ‘large’, is ordinal in nature. In our approach, instead of relying on the symbolic sizes from the sellers, we use the fit feedback from previous purchases to infer the true size of the product. The approach would provide size recommendations to a cold start customer, i.e. ‘small’, ‘fit’ or ‘large’ for a given product.

The rest of the paper is organized as follows. Section 2 describes the related works on solving size and fit recommendation problems. Our approach is detailed in Section 3. Experimental section and results are described in Section 4 and finally we provide concluding remarks in Section 5.

2. RELATED WORK

In literature, the size and fit recommendation problem has been researched very recently. Abdulla et al. [4] embed both users and products using skip-gram based Word2Vec model [5] and employ GBM classifier [6] to predict the fit. A latent factor model was proposed by Sembium et al. [2], which was later follow-up by a Bayesian formulation [3] to predict the size of a product (small, fit, large). In [3] Bayesian logistic regression with ordinal categories was used. They proposed an efficient algorithm for posterior inference based on mean-field variational inference and Polya-Gamma augmentation. Guigourès et al. employed a hierarchical Bayesian model [7] for personalized size recommendation. Misra et.al. [8] learn the fit semantics by modeling it as an ordinal regression problem. Then, they employ metric learning techniques to address the class imbalance issues.

Recently, deep learning approaches have been used to solve the size recommendation problem with encouraging results [9], [10], [11]. Deep Learning approaches unlike the traditional approaches are able to scale well with large amounts of data. SFNET [10] provides recommendations at the user cross product level using a deep learning based content collaborative approach. The approach can learn from cross-correlations that exist across fashion categories. They use both purchase and returns data as well as customer and article features for personalized size and fit prediction. Dogani et al. [11] addressed the sparsity problem by learning latent representation at a brand level using neural collaborative filtering [12]. Then, fine-tuning the product representation by transfer learning from brand representation. Lasserre et al. [9] use a deep learning based meta learning approach. Their approach is based on the premise that, given the purchase history of a customer i, products x_j and their corresponding size estimates y_{ij} share a strong linear relationship. Baier et al. [13] derive product fit feedback from customer reviews using natural language processing techniques which is then used to infer the right fit.

Most of the above approaches rely on customer features, product attributes and sales data. These approaches suffer from cold start problems for thousands of products that are on boarded on the shopping platform every day. Very few approaches explore the use of product images or 3D scan of products to address this issue. SizeNet [9] uses product images to infer whether the product will have fit issues for a customer. Their neural network employs a weakly supervised teacher-student training framework that leverages rich visual information from product images to learn size and fit related visual cues. Knowledge of body shape is important in inferring which garments will best suit a given customer. ViBE uses a computer vision approach to develop a body-aware embedding that captures garment's affinity with different body shapes [8]. The approach learns the embedding from images of models of various shapes and sizes wearing the product, displayed on catalog.

Key contributions The contributions of our work are two-fold:

1. **Novel approach** We propose a novel bayesian probabilistic approach for size and fit recommendation setting.
2. **Non-personalised recommendation** Our proposed model is effective in providing non-personalized recommendations to cold start customers.

3. APPROACH

Size and fit recommendation is vital in improving the customer experience and reducing the returns in fashion e-commerce. In this section we provide a detailed description of our novel approach for non-personalised product size recommendation for customers. Section 3.1 describes the assumptions of the data distributions for modelling the problem. Section 3.2 gives details of modelling size and fit recommendation problem. Section 3.3 describes the maximum likelihood approach to estimate the parameters of the model.

3.1. Data Distribution

We assume that for a given size, the customer population follows a normal distribution with constant variance and mean θ. θ represents the average for a given combination of product and its size. $\alpha_2 - \alpha_1$ is the tolerance band which is assumed to be constant for all product and size combinations.

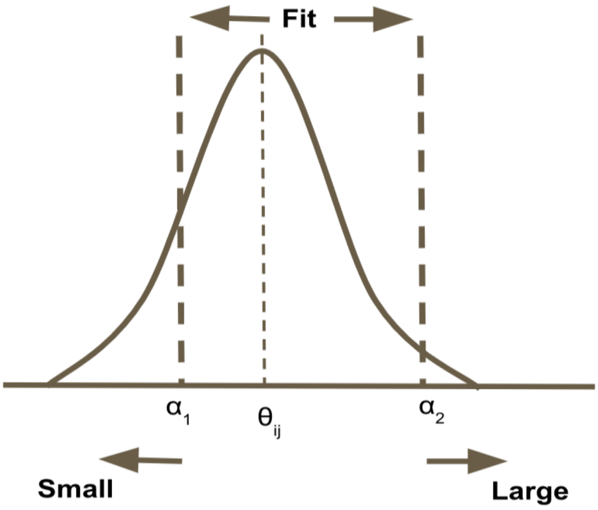

Figure 2. Population distribution for a given size

Figure 3 shows the plot of Length (in cm) on the Y-axis with UK Size on X-axis for various brands in the Shoes category. We infer that Length(in cm) increases linearly with UK size. We assume that □ follows a normal distribution with mean which is a linear function of size.

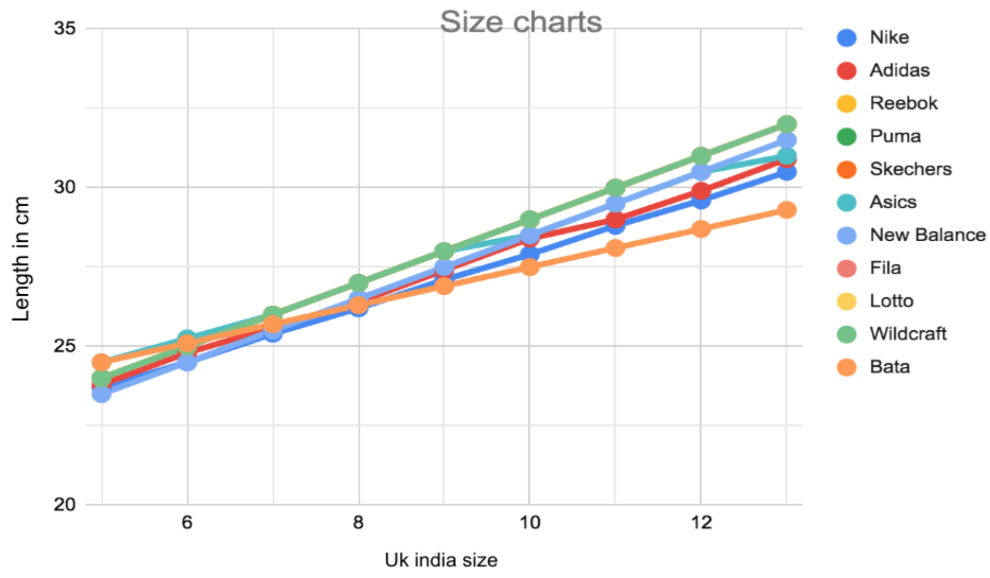

Figure 3 . Size charts for various brands in Shoes category

3.2. Modelling Size and Fit Recommendations Problem

We propose a novel bayesian probabilistic approach for non-personalised product size recommendation to customers. Customer purchase and returns history is used to infer the true product size. Given a product we provide size recommendations to a customer, i.e. we suggest a customer to buy a size small, large or same size.

Typically, a product in a catalog has multiple sizes. A combination of a product and its size is referred to as SKU. Further, a product belongs to a category and a brand. In our approach we recommend at a category level. Given a SKU, we recommend the following labels: 'size is small', 'size is fit' or 'size is large'. To this end we perform the following steps.

- We estimate a monotonically increasing function per product, where the x-axis represents standard sizes and the y-axis is the parameter θ.
- We also estimate two additional parallel lines, These lines are parallel to the x-axis and have intercepts α_1 and α_2. Shown as black parallel lines in figure 4.
- To estimate the output label for a SKU, we check if θ of the SKU is below, in between or above the parallel lines, which leads to the predictions 'size is small', 'size is fit' or 'size is large', respectively. See figure 4.

In our current modelling approach we assume monotonically increasing function with a straight line. The line equation is parameterised by its slope m_b and intercept c_i, where b is the brand and i is the product. The value of θ for a SKU is given by the expression $\{ m_b * j + c_i \}$, where j is the corresponding size. Note that, given a category, slope is the same for all the products of a brand. Examples of two such lines in purple and blue colour are shown in figure 4. As the slope of the purple and the blue lines are not the same, implies that they are from different brands.

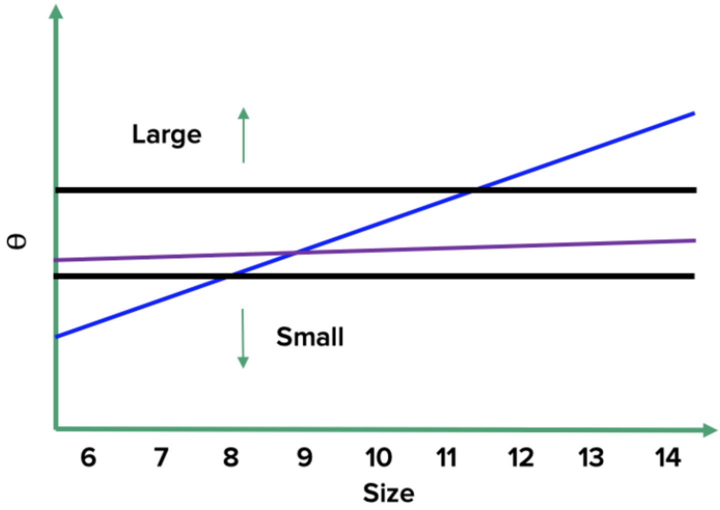

Figure 4. Approach

3.3. Maximum Likelihood Estimation

We use the maximum likelihood formulation to estimate the unknown variables. Let, B, be the number of brands, I, be the number of products, J, be number of sizes and K, be a set of size and fit feedback codes namely, 'small', 'fit' and 'large', i.e, $K \in$ { small, fit, large }. These are categorical variables fetched from the purchase and returns data. The unknown variables to be estimated are the slopes $\{m_1, \dots, m_B\}$, intercepts $\{c_1, \dots, c_I\}$ and the intercepts α_1 and α_2. The maximum likelihood formulation for estimating the unknown variables is shown in steps below :

The maximum likelihood formulation for estimating the unknown variables is shown in steps below :

$$lik\,(m_1, \dots, m_B, c_1 \dots, c_I, \alpha_1 \dots, \alpha_2) = \prod_{i=1}^{I}\prod_{j=1}^{J} p\,(y_{ij} \mid p_{ij}) \tag{1}$$

$$p\,(y_{ij} \mid p_{ij}) \propto \prod_{k=1}^{K} p_{ijk}^{\,y_{ijk}} \tag{2}$$

$$\text{with } \sum_{k=1}^{K} p_{ijk} = 1 \tag{3}$$

$$p_{ijk} = \int_{\alpha_{k-1}}^{\alpha_k} N\,(x; \theta_{ij}, 1)\, dx \tag{4}$$

$$= \Phi(\alpha_k) - \Phi(\alpha_{k-1}) \tag{5}$$

$$\theta_{ij} \sim N\,(m_b * j + c_i, \tau) \tag{6}$$

$$\tau \sim \gamma(1, 1) \tag{7}$$

N is the Normal distribution, Φ is the cumulative distribution function and γ is the Gamma distribution. For a SKU and a size and fit feedback code $k \in K$, y_{ijk} is the number of observation from the purchase and returns data.

Equation 2 gives the probability distribution of observations for a given brand and size. Equation 4 gives the probability of a size and fit feedback code k $\in$ { small, fit, large } for a given brand and size. We assume that θ_{ij} sampled from a normal distribution and its mean is a straight line parameterised by its slope m_b and intercept c_i, where b is the brand and i is the product and variance is sampled from Gamma distribution as shown in equation 6.

4. EXPERIMENTS

To measure the performance, we evaluated our proposed model on the Flipkart shoes dataset and also carried out an Online A/B Test for shoes categories. Section 4.1 provides details of the datasets used for experimentation. Our offline experimental results are described in Section 4.2. Online A/B Testing results are detailed in Section 4.3.

4.1. Datasets

We have collected purchase, returns data and product details like brand, size etc for the following categories: Mens Sport Shoes and Mens Casual Shoes. We label each purchase with small or large based on if the customer has returned the product stating small size or large size as the reason respectively. If the customer has not returned the product, then we label it as fit. The number of purchases in Mens Sport Shoes and Mens Casual Shoes are 2,773,227 and 3,045,925 respectively. This data is split into train and test with 80:20 ratio.

Figure 5&6 shows the plot of returns percentage on y-axis against uk india size on x-axis for adidas and reebok shoes. The y-axis and product names are not listed due to proprietary reasons. We observe that returns due to the feedback code 'large' increases with uk india size and viceversa for the returns due to the feedback code 'small'.

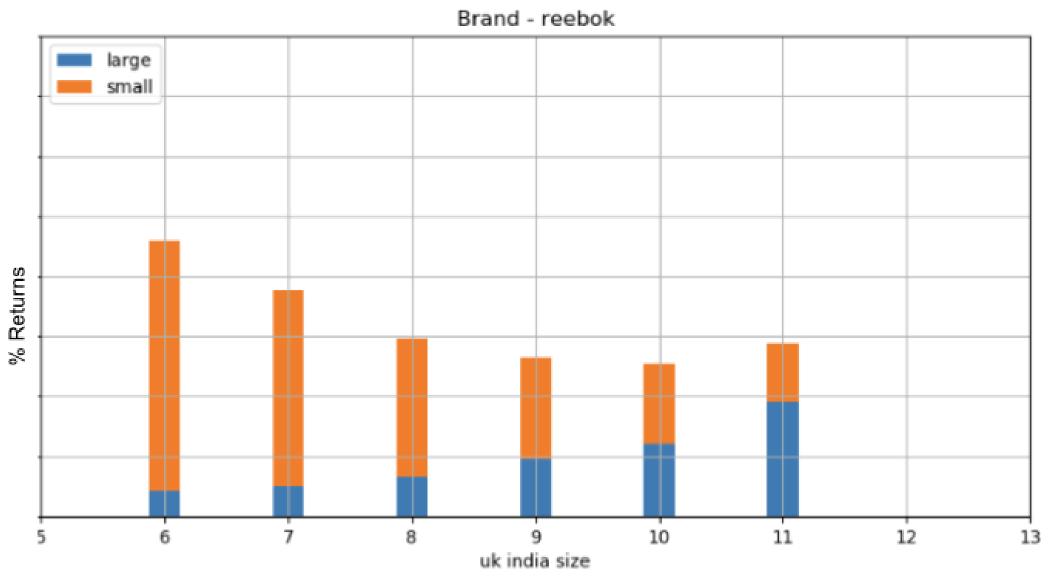

Figure 5. Return trends for Reebok shoes

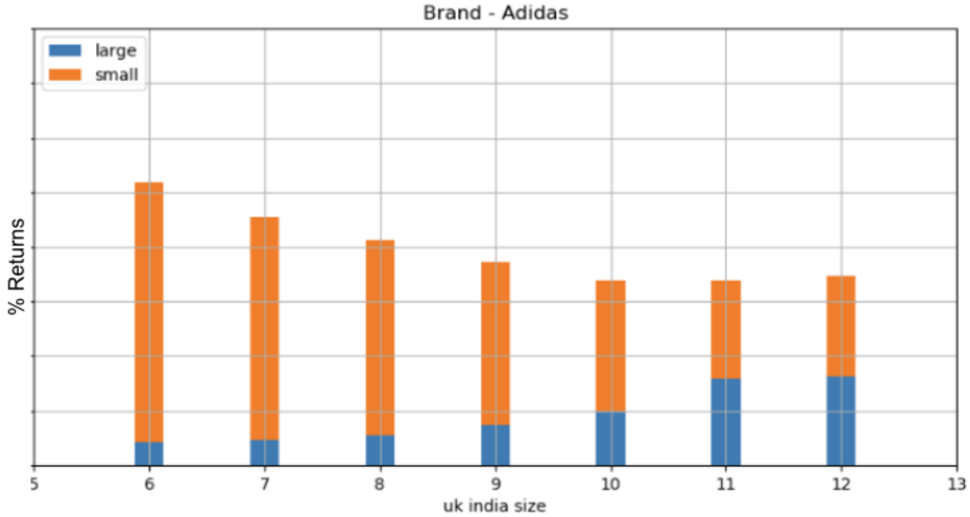

Figure 6. Return trends for Adidas shoes

4.2. Offline Experimental Results

We compare our proposed model against a baseline model which is described below.
Baseline : This approach is a rule based approach and the recommendation for a SKU is given by performing the following steps

- We check if the number of orders is more than $threshold_1$ and also the percentage of returns is more than $threshold_2$.
- We then check if the percentage of returns due to size and fit is more than threshold3 which is shown in the below equation.

$$\frac{\#returns\ size\ fit}{\#Total\ returns} > threshold_3 \qquad (8)$$

- Given a SKU, we label with 'size larger' if the percentage returns due to the reason code 'size is larger' is more than the percentage of returns due to the reason code 'size is smaller' by threshold4 which is shown in below equation.

$$\frac{\#returns\ size\ larger}{\#returns\ size\ fit} - \frac{\#returns\ size\ smaller}{\#returns\ size\ fit} > threshold_4 \qquad (9)$$

Note the $threshold_1$, $threshold_2$, $threshold_3$ and $threshold_4$ are estimated based on the purchases and return data. These values are dynamic in nature which keeps changing based on the data. The values of these are not shared due to proprietary reasons.

We split the data into 80:20 ratio. We train using the rule based approach on 80% of data.The predictions from the training data is used to infer the remaining 20% of the test data. Table 1 and Table 2 summarizes the results of our experiments performed on the flipkart shoes datasets. Performance of our proposed model (described in section 3) is compared against the baseline method described above. We report the following Metrics : AUC_SMALL, AUC_FIT, AUC_LARGE and AUC. AUC_SMALL is the AUC computed with a binary label of 1 for SMALL and 0 for LARGE and FIT. The same approach is followed for calculating AUC_FIT and AUC_LARGE. Finally the computed AUC is the unweighted mean of the individual label

AUC. From Table 1 and Table 2 we infer that our model has **3-4%** higher AUC compared to the baseline method.

Table 1. ROC AUC for Mens Sports Shoes.

Model	**AUC_SMALL**	**AUC_FIT**	**AUC_LARGE**	**AUC**
Baseline	0.580	0.552	0.521	0.551
Our Model	**0.606**	**0.571**	**0.564**	**0.580**

Table 2. ROC AUC for Mens Casual Shoes

Model	**AUC_SMALL**	**AUC_FIT**	**AUC_LARGE**	**AUC**
Baseline	0.550	0.541	0.567	0.553
Our Model	**0.575**	**0.575**	**0.609**	**0.586**

4.3. Online A/B Testing

We ran an Online A/B Testing comparing our proposed model (Test Bucket) with the baseline model (Control Bucket) for the two categories : Mens Casual Shoes and Mens Sports Shoes. The results of the online A/B test showed that we are able to reduce the returns in Mens sports shoes and Mens casual shoes by 0.12 bps and 0.24 bps respectively.

5. CONCLUSION

Fashion is a major B2C eCommerce market segment and is expected to grow rapidly over the next few years. In this paper we proposed a novel Bayesian probabilistic approach for non-personalised size and fit recommendation systems to customers based on purchases and returns data. The proposed approach uses customer purchase and return history to infer the true product size. Online A/B Testing performance showed that there is reduction of 12-24 bps in returns by using our proposed model compared to the baseline model.

REFERENCES

[1] Statista. 2020. Fashion eCommerce report 2020. Retrieved Jan 21, 2021 from. https://www.statista.com/study/38340/ecommerce-report-fashion/ (2020).

[2] Vivek Sembium, Rajeev Rastogi, Atul Saroop, and Srujana Merugu. 2017. Recommending Product Sizes to Customers. In Proceedings of the Eleventh ACM Conference on Recommender Systems, RecSys 2017, Como, Italy, August 27-31, 2017, Paolo Cremonesi, Francesco Ricci, Shlomo Berkovsky, and Alexander Tuzhilin (Eds.). ACM, 243–250.

[3] Vivek Sembium, Rajeev Rastogi, Lavanya Sita Tekumalla, and Atul Saroop. 2018. Bayesian Models for Product Size Recommendations. In Proceedings of the 2018 World Wide Web Conference on World Wide Web, WWW 2018, Lyon, France, April 23-27, 2018, Pierre-Antoine Champin, Fabien L. Gandon, Mounia Lalmas, and Panagiotis G. Ipeirotis (Eds.). ACM, 679–687.

[4] G. Mohammed Abdulla, Shreya Singh, and Sumit Borar. 2019. Shop your Right Size: A System for Recommending Sizes for Fashion products. In Companion of The 2019 World Wide Web Conference, WWW 2019, San Francisco, CA, USA, May13-17, 2019 , Sihem Amer-Yahia, Mohammad Mahdian, Ashish Goel, Geert-Jan Houben, Kristina Lerman, Julian J. McAuley, Ricardo Baeza-Yates, and Leila Zia (Eds.). ACM, 327–334.

[5] Tomás Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient Estimation of Word Representations in Vector Space. In 1st International Conference on Learning Representations, ICLR

2013, Scottsdale, Arizona, USA, May 2-4, 2013, Workshop Track Proceedings, Yoshua Bengio and Yann LeCun (Eds.).

[6] Jerome H. Friedman. 2001. Greedy function approximation: A gradient boosting machine.Ann. Statist. 29, 5 (10 2001), 1189–1232.

[7] Romain Guigourès, Yuen King Ho, Evgenii Koriagin, Abdul-Saboor Sheikh, Urs Bergmann, and Reza Shirvany. 2018. A hierarchical bayesian model for size recommendation in fashion. Proceedings of the 12th ACM Conference on Recommender Systems, RecSys 2018, Vancouver, BC, Canada, October 2-7, 2018, Sole Pera, Michael D. Ekstrand, Xavier Amatriain, and John O'Donovan (Eds.). ACM, 392–396.

[8] Rishabh Misra, Mengting Wan, and Julian J. McAuley. 2018. Decomposing fit semantics for product size recommendation in metric spaces. In Proceedings of the 12th ACM Conference on Recommender Systems, RecSys 2018, Vancouver, BC, Canada, October 2-7, 2018, Sole Pera, Michael D. Ekstrand, Xavier Amatriain, and John O'Donovan (Eds.). ACM, 422–426.

[9] Julia Lasserre, Abdul-Saboor Sheikh, Evgenii Koriagin, Urs Bergmann, Roland Vollgraf, and Reza Shirvany. 2020. Meta-Learning for Size and Fit Recommendation in Fashion. In Proceedings of the 2020 SIAM International Conference on Data Mining, SDM 2020, Cincinnati, Ohio, USA, May 7-9, 2020, Carlotta Demeniconi and Nitesh V. Chawla (Eds.). SIAM, 55–63.

[10] Sheikh, A. S., Guigourès, R., Koriagin, E., Ho, Y. K., Shirvany, R., Vollgraf, R., & Bergmann, U. (2019, September). A deep learning system for predicting size and fit in fashion e-commerce. In Proceedings of the 13th ACM conference on recommender systems (pp. 110-118).

[11] Kallirroi Dogani, Matteo Tomassetti, Saúl Vargas, Benjamin Paul Chamberlain, and Sofie De Cnudde. 2019. Learning Embeddings for Product Size Recommendations. In Proceedings of the SIGIR 2019 Workshop on eCommerce, co-located with the 42st International ACM SIGIR Conference on Research and Development in Information Retrieval, eCom@SIGIR 2019, Paris, France, July 25, 2019 (CEUR Workshop Proceedings, Vol. 2410), Jon Degenhardt, Surya Kallumadi, Utkarsh Porwal, and Andrew Trotman (Eds.). CEUR-WS.org. .

[12] Xiangnan He, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia Hu, and Tat-Seng Chua. 2017. Neural Collaborative Filtering. In Proceedings of the 26th International Conference on World Wide Web, WWW 2017, Perth, Australia, April 3-7, 2017 , Rick Barrett, Rick Cummings, Eugene Agichtein, and Evgeniy Gabrilovich (Eds.). ACM, 173–182.

[13] Stephan Baier. 2019. Analyzing Customer Feedback for Product Fit Prediction. CoRR abs/1908.10896 (2019). arXiv:1908.10896.