Personalized Size Recommendations with Human in the Loop

Leonidas Lefakis [1]   Evgenii Koriagin [1]   Julia Lasserre [1]   Reza Shirvany [1]

Abstract

One of the major challenges facing e-commerce fashion platform is that of recommending to customers the right size and fit for fashion apparel. In this work we study this topic in depth and demonstrate its various complexities focusing in particular on the challenging cold-start problem that arises when no order history is available for a specific customer. We demonstrate the multifaceted value of data obtained by involving the customer in the loop and show how it allows for an effective cold-start recommender system. We highlight our findings via detailed experiments performed on hundreds of thousands of customers and items in real world e-commerce scenarios. In addition, results and discussions are provided investigating the trade-off between the recommender's efficiency and the customer's experience with the goal of introducing accurate solutions with low user cognitive load.

1. Introduction

Finding fashion apparel online with the right size and fit is one of the major factors impacting customers' purchasing decisions and satisfaction, yet it is challenging for many customers. As a result, either they remain reluctant to engage in the purchasing process or they purchase articles in multiple neighboring sizes to try them out and return the ones that do not fit. Compounding the issue, customer preferences towards perceived article size and fit for their body remain highly personal and subjective, which influences the definition of the right size for each customer.

The problem is further compounded by various factors, from production processes to branding issues, that are responsible for an observed and troubling increase in its complexity. Most fashion retailers create garments for consumers they have never met, not having the opportunity to interact with

---

[1]Zalando SE, Berlin, Germany. Correspondence to: Leonidas Lefakis <leonidas.lefakis@zalando.de>.

them directly, by using body measurement data aggregated and sold by third parties intended to represent size and fit requirements of entire populations in different countries and regions. To make matters worse, each brand has their own idea and definition of their target customer base and the correlating aggregated sizes that represent that customer base. A brand might specifically target their customer base to be on the younger side, and thus, represented by say a size 8, so the products are fitted on a size 8 fit model and then "graded", i.e. sized up and down in increments to generate other sizes of the garment. This leads to distortion of the larger sizes because they have now been increased, e.g. in width, without ever being fitted on a larger model. Furthermore there is essentially no consistency with regards to sizing systems between brands (or in fact often even within brands). Adding to this the complexity of radically different sizing systems (Alpha versus Numeric), different country conventions (EU, FR, IT, UK), and *vanity sizing* (Weidner, 2010) (i.e. increase physical measurements of a nominal size to boost customer's self-esteem), the problem of making correct fashion size recommendations is greatly exasperated.

In order to highlight the scale of the size recommendation problem, we analyzed fashion articles and orders in the category of Female Upper Garments which encompasses a large variety of different fashion apparel, from dresses to denim jackets, and is strongly representative of the complexity of fashion in general and of the many obstacles that arise in personalized size recommendations in particular. In Figure 1 we present a bar plot of the number of distinct apparel sizes, in the female upper garment category, from around 2000 brands available on a large-scale e-commerce fashion platform during the 2015-2019 time period. In this plot we see that when we aggregate the list of all possible sizes for all brands (composed of all the different size systems such as numeric 38-39-40, ... standard S,M,L, ..., fractions 41 1/3, 42.5, ... confection sizes 36-38, 40-42, ... country conventions EU, FR, IT, UK ...), we reach the upper bound of 17k sizes. We also see that the scale of the size recommendation and size selection problem has grown continuously and rapidly with the number of distinct sizes more than doubling within this category over last 4 years.

As fashion e-commerce is increasingly growing, assisting customers in buying the right size presents a huge opportunity for research in intelligent size and fit recommendation
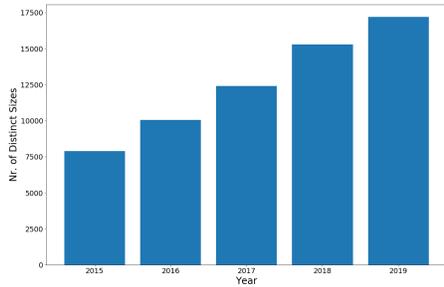
Figure 1. Bar plot of the number of distinct apparel sizes, in women upper garment category, from hundreds of brands with different size systems and country sizing conventions during 2015 to 2019 time period. The scale of the size recommendation and selection problem continues to grow rapidly in recent years.

systems which can directly contribute to increasing customer satisfaction, reducing environmental footprint and helping business profitability.

In this direction, the contributions of this work are 2-fold: 1. We demonstrate the multifaceted value of customer metadata in effectively tackling the cold-start problem by leveraging said data to build an efficient cold-start recommender comparable with those state-of-the-art solutions that have privilege access to customers order data; 2. We leverage the trade-off between the size recommender's efficiency and the required customer data and propose, for the first time to our knowledge, an *accurate large-scale cold-start size recommender with low customer cognitive load*, launched in early 2020, which has demonstrated its value for thousands of customers in six European countries with various size systems and sizing conventions.

2. Prior Work

Although customer-centric product recommendation is a well-studied field (see (Shi et al.; Zhang et al., 2017; Yi et al., 2014; Catherine & Cohen, 2016; Pazzani & Billsus, 2007)), size recommendation is still in its infancy with only a few approaches addressing parts of this problem during the past few years (Sembium et al., 2017; 2018; Abdulla & Borar, 2017; Guigourès et al., 2018; Sheikh et al., 2019; Karessli et al.; Dogani et al., 2019; Misra et al., 2018; Singh et al., 2019; Schein et al., 2002; Thalmann et al., 2011; Surville & Moncoutie, 2013; Peng & Al-Sayegh, 2014; Januszkiewicz et al., 2017; Pazzani & Billsus, 2007). A large family of these approaches depend on historical data of customer orders and returns either using statistical (Sembium et al., 2017; 2018; Guigourès et al., 2018) or deep-learning (Sheikh et al., 2019; Karessli et al.; Lasserre et al., 2020) methods, often concentrating on finding suitable em-

beddings to represent customers and orders (Dogani et al., 2019; Misra et al., 2018; Singh et al., 2019). Such approaches have the advantage of not asking customers for explicit data, thus involve a low cognitive effort from the customer. However, such solutions invariably suffer from the cold-start problem which affect thousands of existing and new customers visiting the shopping platforms everyday for which no prior order history is available in hundreds of brands and tens of fashion categories. The cold start problem has been widely studied in the context of user item recommendations (Schein et al., 2002). Prior work has typically focused either on the article side (Karessli et al.), or on entering a dialogue with the customers (Yuan & Huh, 2019; the; aso; abo; Thalmann et al., 2011; Surville & Moncoutie, 2013; Peng & Al-Sayegh, 2014). The former approaches, i.e. exploiting attributes of articles, allow to alleviate the cold-start problem by using content-based filtering (Pazzani & Billsus, 2007) and come with a low cognitive load by design. However, in the context of personalized size recommendations, article data does not bring sufficient information to tackle the problem. The latter approaches come with the advantage of allowing customers to become an direct and integral part of the recommendation system but on the other side either require customers to share a considerable amount of sensitive personal data (such as age, height, weight, tummy shape, hips form, body type, favorite brand, fit preferences, etc.) (the; aso; abo) or require multiple high resolution images, videos, and in some cases 3D scans, of customers bodies in tight clothing and canonical poses (Thalmann et al., 2011; Surville & Moncoutie, 2013; Peng & Al-Sayegh, 2014). Such imagery or 3D data is often used to create virtual-fitting solutions using recent 3D human body estimation and reconstruction approaches (Pavlakos et al., 2018; Bogo et al., 2016). As such, these approaches come with a high level of cognitive load for customers by demanding strong engagement and willingness from them to share images and scans of their bodies with fashion platforms.

A comprehensive comparison of these diverse approaches would be of high value to the community, but this remains out of the scope of this paper and constitutes a great future research direction. Here, in particular we focus on the the size recommendation problem in the so called cold-start scenario in which there is little to no order history for each customer, and thus, customers are part of the solution by providing explicit information through a questionnaire. The cold-start problem of personalized size recommendation is a new emerging field and the underlying importance and necessity of requiring a diverse set of personal and body data for providing these recommendations requires further investigation. Here we aim to investigate the size recommendation problem in this cold-start scenario, by involving customers in the loop through some sort of questionnaire

Type	Features
Overall	height, weight, age, gender
Upper body	top size, shirt collar size, shirt fit, prop. belly, top fit, prop. shoulder-waist, bust number, bust cup size, prop. shoulder-hip, blazer size
Lower body	pants size, jeans length, jeans width, prop. waist, pants waist-height , shoe size

and to create a solution that comes with a low cognitive load for the customers in a way that we burden the customers as little as possible while providing quality size advice.

3. Size recommendation Without Order History

We model the problem of cold-start size recommendation as a categorical classification task, where each size is a possible class. The idea is to directly involve customers in the process by leveraging those for which we have both customer data and purchase data to learn a mapping from customer data to ordered sizes, thus allowing us to predict appropriate sizes for any new customer.

Customer Data: We use a comprehensive set of customer data gleaned from questionnaires presented to customers as part of a specialized online fashion styling service wherein customers are paired with professional stylists who then curate personalized outfits for them. Customers are required to fill in a questionnaire covering a wide variety of fashion related areas. For this study we extract from these questionnaires the subset of information related to size and fit and use it to build a feature representation for each customer. This subset consists of 20 attributes for each customer falling into three categories as can be seen in table 1. A total of 450k questionnaires are available, each from a distinct customer which self-identified as female. The questionnaire data is projected onto an input space by calculating a vector representation for each customer. Of the 20 size-related questions on the questionnaire, 7 result in categorical variables and are one-hot encoded while the remaining 13 result in numerical variables and are normalized by mean and standard deviation.

Order data: The order data used in this work is composed of roughly 7.4 million orders placed on an e-fashion platform in the female upper body garment category. This dataset is anonymized and is not public due to various customer privacy challenges and proprietary reasons. We split these orders into training and test sets based on order timestamps. Of these 7.4 million orders, the oldest 5.6 million

comprise the training set while the most recent 1.8 million form the test set. All ordered articles are associated with a numerical size in $[34, 36, 38, 40, 42, 44, 46]$. Both in training and testing, the target variable is the size bought by the customer. There are cases where different target values correspond to the same input vector as customers will at times buy different sizes. Nonetheless allowing the classifier to handle such ambiguity was found to be the best strategy, as opposed for example to using the median or mean size in training. Using the output of the Hot-Start recommender (Guigourès et al., 2018) (presented below) as a target value also proved sub-optimal.

Classifier: We experimented with a variety of multi-class classifiers and found Gradient Boosted Trees to perform best in practice. This choice comes with the added benefit of having an easily interpretable classifier, which as we shall show in the following can prove very useful. The hyper-parameters were tuned using a grid-search and cross-validation (splitting the Training Data into Train and Validation sets). In particular we found that performance saturated at 500 trees, and did not observe any over-fitting effect when growing the ensemble beyond this point. Each tree in the ensemble has a depth of 3, while the trees themselves are built, sequentially, using a learning rate of 0.01 and a sub-sampling rate of 0.5. We note that performance on the validation set was largely robust to the latter two hyper-parameters.

Hot-Start recommender baseline: The size recommender system introduced in (Guigourès et al., 2018) has shown to be robust and efficient in a large-scale fashion e-commerce context; we thus employ it as a Hot-Start recommender baseline (built on order history data). It follows a hierarchical Bayesian approach that models jointly the probability of a size and return status (kept, too small, too big) given a customer and an article. This approach enjoys the advantages of Bayesian modeling, and in contrast with (Sembium et al., 2017; 2018) which have to predict the fit (good fit, too big, too small) for all possible sizes one by one, it can directly predict the probability of any size given a customer-article pair, conditionally on those articles being kept (good fit). It naturally fails in the case of new customers or customers with scarce order history. As one might expect however, the order data falls into the long tail problem, and as such, the shortcoming of current Hot-Start recommenders are quite pronounced in the context of online fashion where a large percentage of customers are in the cold-start category with no to scare order history.

Brand Size Offsets: Most brands suffer a high variance between their nominal and actual sizes caused by multiple design and business related factors, and as such the knowledge of customers' "usual" size alone is often insufficient information for providing both intra- and inter- brand size

recommendations. We make use of customers return trends to gain a better understanding of various brands behaviour with respect to size and fit. Customers often tend to order, for a certain brand, their "usual" size and one size up or down, either within the same order or in a later order if they returned the first-ordered size. Therefore, using data regarding kept and returned articles, one can readily define an article offset as the difference in the sizes from the (re-ordered) kept articles and the ordered (but not kept) articles. An offset for a brand is then defined as the weighted average of all the article offsets in that brand. We weigh the contributions of each article by the number of sales so that the highest selling articles contribute the most to the final offset of the brand, and calculate a weighted mean (μ) and a weighted standard deviation (σ) to fit a Gaussian. We show the range of brand offsets in Figure 2, where μ and σ are calculated for 80 popular brands using 10k distinct women upper-garment purchases per brand. We highlight the importance of exploiting brand offsets in section 4 where we present our results on cold-start recommendation.
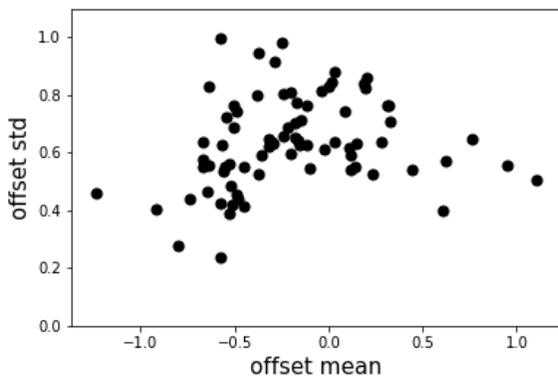

Figure 2. Brand Size Offsets for the 80 most popular brands on the e-fashion platform. The standard deviation is plotted against the mean.

Recommender predictions: The trained classifier is combined with the brand offsets presented above resulting in our Cold-Start recommender. To obtain the final prediction of the Cold-Start recommender, the predicted class c of the trained ensemble is combined with the brand offsets by adding the brand offset mean μ_{brand} to the ensemble prediction and the final size recommendation is obtained by rounding $c + \mu_{brand}$ to the closest size. We note that we use brand offsets for the Baseline recommender too, as they were found to be advantageous. We note that we use brand offsets for the baseline Hot-Start recommender too, as they were found to be advantageous.

4. Experimental Results and Discussion

In this section, we evaluate the recommender systems based on their ability to predict the sizes actually bought by the customer in the test set orders using the accuracy metric defined as the percentage of times the recommender correctly predicts the size bought by the customer on the orders in the test set.

4.1. Hot-Start and Cold-Start performances

In Figure 8 we present a comparative study of the baseline Hot-Start recommender and the proposed Cold-Start recommender. We plot the accuracy of the models against the number of prior orders of each customer in the training data (as we employ temporal split of the data into training and test set, we can also speak of training and test time). As can be seen, for low number of orders regime (< 10) the Cold-Start recommender (in green) clearly outperforms the Hot-Start one (in brown), even in cases when a customer has a substantial amount of prior orders ($> 10, < 20$). *We re-iterate that the cold-start recommender does not use any knowledge of prior orders and in theory should remain stable independent of the number of prior orders; in practice, as observed in Figure 3, the cold-start results change slightly since the number of test samples naturally reduces as the number of prior orders increases.* The Hot-Start recommender only starts outperforming the Cold-Start recommender after about 20 prior orders. This is due to the hard limitations of most current Hot-Start recommender approaches where they need a minimum set of orders per each sub-level category to perform, as has been duly noted in (Guigourès et al., 2018; Dogani et al., 2019), effectively restricting Hot-Start solutions to loyal customers with rich order history.

4.2. Impact of Brand Size Offsets

As noted, the proposed Cold-Start recommender makes great use of the brand offsets. To highlight the contribution of these offsets in the performance of the recommender we present quantitative results in Figures 4 and 5. The accuracy of two Cold-Start Recommenders are shown in Figure 4. In particular we show the accuracy of a Cold-Start Recommender based only on the ensemble method which does not exploit the brand offsets (Cold-Start Reco Without Brand Offsets) and the accuracy of the proposed Cold-Start Recommender which adds the brand offset's mean to the ensemble output. We plot these accuracies relative to a lower threshold on the standard deviation, for a given threshold θ we only include those brands which have a standard deviation above θ to calculate the accuracy. As can be seen there is a clear advantage to exploiting brand offsets when making recommendations.
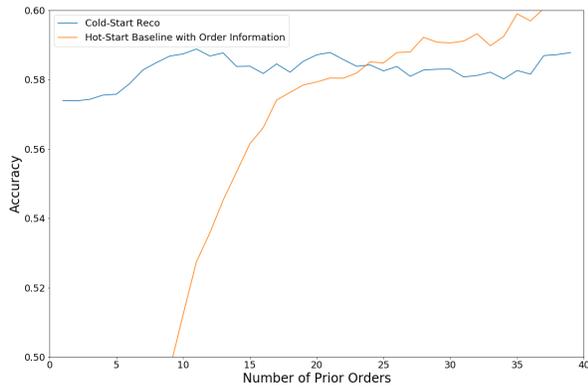
Figure 3. Comparison of the Cold-Start recommender system presented here to the Hot-Start system presented in (Guigourès et al., 2018). On the x-axis are the number of prior orders (specifically the number of prior orders of each customer in the training set). On the y-axis is the accuracy defined as the percentage of times the recommender correctly predicts the size bought by the customer on the orders in the test set.
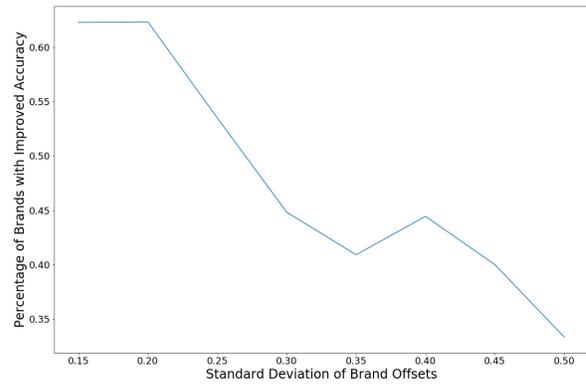
Figure 5. Percentage of brands for which adding the brand offset's mean leads to improved accuracy plotted against a lower threshold θ on the standard deviation of the brand offsets.

brand offsets diminishes as the standard deviation rises. As the standard deviation reaches 0.5 the brand offsets means lead to improved performance only in approximately 33% of cases. To address this limitation, a future work direction is to directly use the article offsets for brands suffering from such high standard deviations.

4.3. Customer coverage

In order to get a better understanding of the percentage of customers covered by various order segments [1] we plot in Figure 6 the accuracy of the recommender systems relative to these percentages. By taking all customers we achieve a customer coverage of 100% although in this case the Hot-Start Reco performs quite poorly as can be seen in the plots. On the other hand taking only those customers who have a large number of prior orders results in a Hot-Start recommender that outperforms the Cold-Start recommender but at the cost of having a low customer coverage.

4.4. Minimizing Customers' Cognitive Load

A crucial aspect of any user-in-the-loop recommendation system is the amount of cognitive load it burdens the customers with. Thus beyond the performance of the system with respect to accuracy, in practice it is of major importance to minimize the customer's cognitive load when they interact with the platform. The cold-start recommender presented in the previous section makes use of 20 explicit customer data points such as height, weight, etc. which is in reality too high. In what follows, we deep dive and
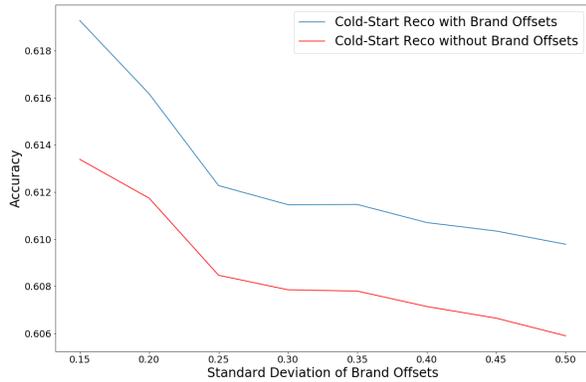

Figure 4. Accuracy of the Cold-Start recommender depending on whether brand offsets are exploited to refine size recommendations. The results show the benefits of leveraging this information in all cases, irrespective of the brand offset's standard deviation.

In Figure 5 we present the percentage of brands for which we observe improved accuracy when adding the brand offset means relative to the standard deviation of the offsets. Again we plot against a lower threshold on the brand standard deviations. As can be seen in the case of brands with an offset standard deviation of at least 0.15 adding the brand offset means results in improved accuracy in approximately 66% of cases. As expected, the positive effect of using

---

[1]The orders in the test set are segmented based on the number of prior orders in the training set of the corresponding customer.

Table 2. Production results in various countries.

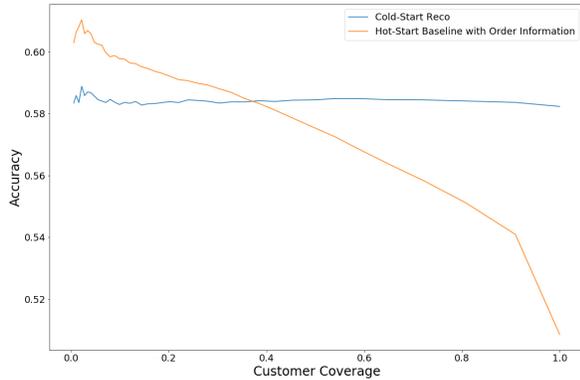| Country	DE	BE	CH	NL	SE	AT
Accuracy	58.36%	68.71%	56.56%	69.50%	66.97%	57.66%

Figure 6. Comparison of the Cold- and Hot-Start recommender systems relative to the percentage of customers covered (customer coverage).

investigate what performance can be obtained using a small subset of attributes towards providing a low cognitive load and critically reducing the intimate data requirements from the customers on their body shapes, etc.

The cold-start recommender presented in the previous section makes use of 20 explicit customer attributes such as height, weight, etc. which in reality is too high. As mentioned in section 3, one of the added benefits of using Gradient Boosted Trees is that they result in an interpretable model, which allows us to estimate the Gini importance of each individual feature in the resulting ensemble. The attributes that came out as key are Top Size, Weight and Height (with feature importance 0.849, 0.070, 0.059 respectively); the importance of the the next most important feature (pants size) was considerably lower (0.005).

Given the importance assigned to the Top Size attribute by the ensemble, the obvious question that arises is whether the customer provided top size would suffice to predict the size bought by the customer themself in any future orders. We therefore cross validated this explicit customer information with the sizes a customer actually buys on the fashion platform. We found that in fact customers only buy the size they provided in the questionnaire in roughly 57% of all orders. This observation runs counter to the intuition that customers should be good predictors of their own sizes and highlights one of the many complexities of the size recommendation problem. As customers themselves are only 57% likely to

order in their provided sizes, this leaves a remaining 43% of orders where customers are unsure of what size to order and would require accurate support in the form of a size advice.

In Figure 7 we show that, as expected, using solely the size provided by the customer leads to an under-performing recommender system (marked "Top Size" in the plot). We experimented further within this scenario by adding a minimum information to it, in particular the brand information, which in turn enables us to use the brand offsets. As can be seen this results in a significantly better recommender system (marked "Top Size + Brand" in the plot) and highlights the importance of exploiting brand offsets when making a recommendation. As expected, the Cold-Start recommender system which has access to the full questionnaire outperforms both these systems.
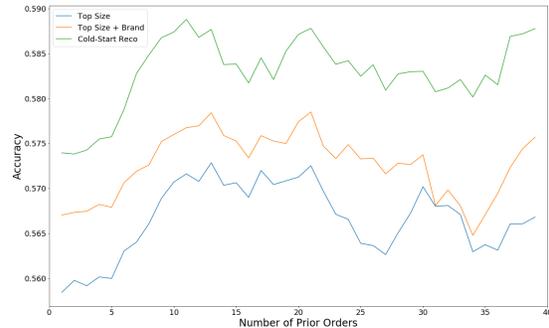

Figure 7. Accuracy of recommender systems based on the top size (with and without the brand information) compared to the Cold-Start recommender which uses all features.

In Figure 8, we show the performance of various cold-start recommender systems using these top 3 attributes: "Weight + Height" in red, "Top Size" in blue and "Weight + Height + Top Size" in purple. As expected, the Cold-Start recommender system which has access to the full questionnaire (in green) outperforms all these systems. Based on Figure 8, it is evident that asking only for Weight and Height (red curve) is not enough to reliably provide a size recommendation. However, we do note that using Weight, Height, and Top Size (purple curve) performs closely to the full Cold-Start recommender. This constitutes a good trade-off between high recommender accuracy and low cognitive effort on the side of the customer, yet it comes with a potentially underwhelming customer experience which involves having to

provide two intimate and privacy sensitive questions on an e-commerce shopping platform. In spite of its shortcomings, Top Size performs closely to all 3 attributes together. We further add brand information, which in turn enabled us to use the brand offsets. As can be seen in Figure 8 ("Top Size + Brand" in yellow), this results in a significantly better recommender system than Top Size alone (in blue). This not only highlights the importance of exploiting brand offsets when making a recommendation, but it also provides closely comparable results to that of the Weight + Height + Top Size model, and comes with the great advantage of not requiring intimate body data. Instead customers are only asked for their top size in one of their favourite brands. We consider this approach to be the best trade-off between accuracy and customer experience.

4.5. Performance in Production

Following the experiments shown in previous sections and given the encouraging results, we have rolled out our Cold-Start recommender to production in January 2020 on a large e-commerce fashion platform for the Adult Upper Garments category (both Men's and Women's categories). The model is currently live in six countries (DE : Germany, BE : Belgium, NL : Netherlands, CH : Switzerland, AT : Austria, SE : Sweden), serving approximately 3000 orders per day with an overall accuracy of 63.90%. As can be seen in Table 2 accuracy on a per country basis can vary greatly, potentially highlighting the cultural aspect of the size and fit problem which further complicates an already complex problem. This provides an exciting dimension for future work.

5. Conclusion

We addressed a major challenge for the online fashion industry, that of user-in-the-loop personalized size recommendations with low cognitive load. With the aim of building an accurate recommender system that requires only a minimum set of explicit customer data, we further investigated 20 different customer attributes such as weight, height, top size, tummy shape, etc. for the task at hand.

We experimented with different versions of the cold-start recommender system and benchmarked them against the state-of-the-art recommender systems with privileged access to rich customer order history. We presented a deep dive on the trade-off between a recommender's performance and a customer's cognitive load, and proposed a solution capable of providing accurate size advice for thousands of new and existing customers with bare minimum customer data needs.

Finally we presented our results in a production environment covering six European countries and demonstrated that our approach scales up to large-scale production requirements,

performs in practice at the level predicted by the experiments presented here, and to the level of industrial requirements.

References

ABOUT YOU. https://corporate.aboutyou.de/en/. The customer based size recommendations are accessible on product detail pages. Accessed: 2019-11.

ASOS. https://www.asos.com. The customer based size recommendations are accessible on product detail pages. Accessed: 2019-11.

THE ICONIC. https://www.theiconic.com. The customer based size recommendations are accessible on product detail pages. Accessed: 2019-11.

Abdulla, G. M. and Borar, S. Size recommendation system for fashion e-commerce. In *KDD Workshop on Machine Learning Meets Fashion*, 2017.

Bogo, F., Kanazawa, A., Lassner, C., Gehler, P. V., Romero, J., and Black, M. J. Keep it SMPL: automatic estimation of 3d human pose and shape from a single image. *CoRR*, abs/1607.08128, 2016. URL http://arxiv.org/abs/1607.08128.

Catherine, R. and Cohen, W. Personalized recommendations using knowledge graphs: A probabilistic logic programming approach. In *Proceedings of the 10th ACM Conference on Recommender Systems*, RecSys '16, pp. 325–332, New York, NY, USA, 2016. ACM. ISBN 978-1-4503-4035-9. doi: 10.1145/2959100.2959131. URL http://doi.acm.org/10.1145/2959100.2959131.

Dogani, K., Tomassetti, M., Vargas, S., Chamberlain, B. P., and Cnudde, S. D. Learning embeddings for product size recommendations. In *Proceedings of the SIGIR 2019 Workshop on eCommerce*, 2019.

Guigourès, R., Ho, Y. K., Koriagin, E., Sheikh, A.-S., Bergmann, U., and Shirvany, R. A hierarchical bayesian model for size recommendation in fashion. In *Proceedings of the 12th ACM Conference on Recommender Systems*. ACM, 2018.

Januszkiewicz, M., Parker, C., Hayes, S., and Gill, S. Online virtual fit is not yet fit for purpose: An analysis of fashion e-commerce interfaces. pp. 210–217, 10 2017. doi: 10.15221/17.210.

Karessli, N., Guigourès, R., and Shirvany, R. Sizenet: Weakly supervised learning of visual size and fit in fashion images. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops 2019*.

Lasserre, J., Sheikh, A.-S., Koriagin, E., Bergmann, U., Vollgraf, R., and Shirvany, R. Meta-learning for size and
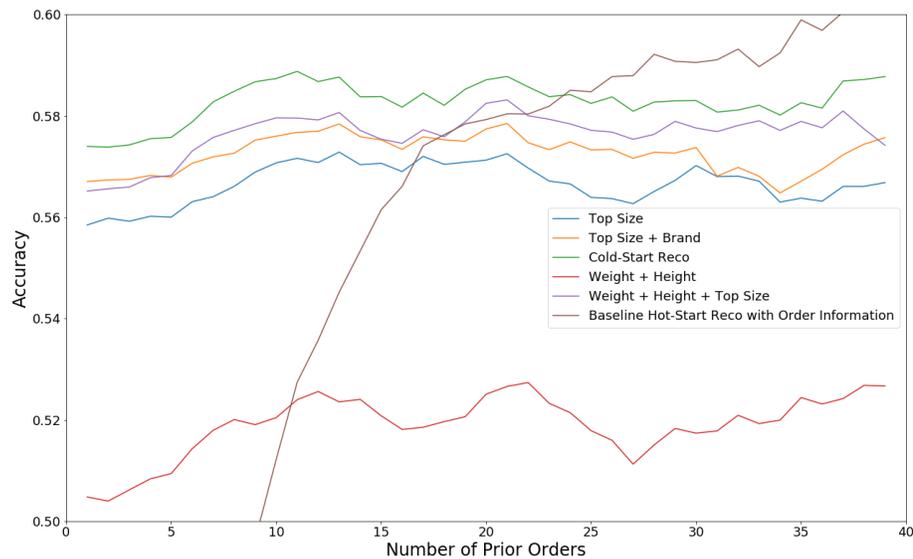
Figure 8. Comparison of the Cold-Start and the Hot-Start systems. On the x-axis are the number of prior orders (specifically the number of prior orders of each customer in the training set)

fit recommendation in fashion. In *Proceedings of the 2020 SIAM International Conference on Data Mining*, pp. 55–63, 01 2020. ISBN 978-1-61197-623-6. doi: 10.1137/1.9781611976236.7.

Misra, R., Wan, M., and McAuley, J. Decomposing fit semantics for product size recommendation in metric spaces. In *Proceedings of the 12th ACM Conference on Recommender Systems*, RecSys '18, pp. 422–426. ACM, 2018. ISBN 978-1-4503-5901-6. doi: 10.1145/3240323. 3240398.

Pavlakos, G., Zhu, L., Zhou, X., and Daniilidis, K. Learning to estimate 3d human pose and shape from a single color image. *CoRR*, abs/1805.04092, 2018. URL http://arxiv.org/abs/1805.04092.

Pazzani, M. J. and Billsus, D. Content-based recommendation systems. In *The adaptive web*, pp. 325–341. Springer, 2007.

Peng, F. and Al-Sayegh, M. Personalised size recommendation for online fashion. 2014.

Schein, A. I., Popescul, A., Ungar, L. H., and Pennock, D. M. Methods and metrics for cold-start recommendations. In *Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval*, pp. 253–260. ACM, 2002.

Sembium, V., Rastogi, R., Saroop, A., and Merugu, S. Recommending product sizes to customers. In *Proceedings of the Eleventh ACM Conference on Recommender Systems*. ACM, 2017.

Sembium, V., Rastogi, R., Tekumalla, L., and Saroop, A. Bayesian models for product size recommendations. In *Proceedings of the 2018 World Wide Web Conference*, 2018.

Sheikh, A.-S., Guigourès, R., Koriagin, E., Ho, Y. K., Shirvany, R., Vollgraf, R., and Bergmann, U. A deep learning system for predicting size and fit in fashion e-commerce. In *Proceedings of the 13th ACM Conference on Recommender Systems*. ACM, 2019. ISBN 978-1-4503-6243-6. doi: 10.1145/3298689.3347006. URL http://doi.acm.org/10.1145/3298689.3347006.

Shi, Y., Larson, M., and Hanjalic, A. Collaborative filtering beyond the user-item matrix: A survey of the state of the art and future challenges. *ACM Comput. Surv.*, 47(1): 3:1–3:45. ISSN 0360-0300. doi: 10.1145/2556270.

Singh, L., Singh, S., Arora, S., and Borar, S. One embedding to do them all. *CoRR*, abs/1906.12120, 2019. URL http://arxiv.org/abs/1906.12120.

Surville, J. and Moncoutie, T. 3d virtual try-on: The avatar at center stage. 2013.

Thalmann, N., Kevelham, B., Volino, P., Kasap, M., and Lyard, E. 3d web-based virtual try on of physically simulated clothes. *Computer-Aided Design and Applications*, 8, 01 2011. doi: 10.3722/cadaps.2011.163-174.

Weidner, N. Vanity sizing, body image, and purchase behavior: A closer look at the effects of inaccurate garment labeling. 2010.

Yi, X., Hong, L., Zhong, E., Liu, N. N., and Rajan, S. Beyond clicks: Dwell time for personalization. In *Proceedings of the 8th ACM Conference on Recommender Systems*, RecSys '14, pp. 113–120, New York, NY, USA, 2014. ACM. ISBN 978-1-4503-2668-1. doi: 10.1145/2645710.2645724. URL http://doi.acm. org/10.1145/2645710.2645724.

Yuan, Y. and Huh, J.-H. *Cloth Size Coding and Size Recommendation System Applicable for Personal Size Automatic Extraction and Cloth Shopping Mall: MUE/FutureTech 2018*, pp. 725–731. 01 2019. ISBN 978-981-13-1327-1. doi: 10.1007/978-981-13-1328-8_94.

Zhang, S., Yao, L., and Sun, A. Deep learning based recommender system: A survey and new perspectives. *CoRR*, abs/1707.07435, 2017.