

Mind the Gap: Exploring Shopping Preferences Across Fashion Retail Channels

Matthias Wölbitsch
Detego GmbH
m.woelbitsch@detego.com

Simon Walk
Detego GmbH
s.walk@detego.com

Thomas Hasler
Detego GmbH
t.hasler@detego.com

Denis Helic
ISDS, Graz University of Technology
dhelic@tugraz.at

ABSTRACT

Over the course of the last decade, online retailers have demonstrated that knowledge about customer preferences and shopping patterns is an important asset for running a successful business. For example, customer preferences and shopping histories are the foundation for recommender systems that support the search for relevant products to buy online. With the increasing adoption of modern technologies, traditional retailers are able to collect similar data about customer behavior in their stores. For example, smart fitting rooms allow to track interactions of customers with products beyond the scope of a traditional retail store. In this paper we explore how customers of a large international fashion retailer buy products online and in brick-and-mortar stores, and uncover significant differences between the two domains. In particular, we find that online customers frequently focus on buying products from one specific category, whereas customers in brick-and-mortar stores often buy a more diverse range of product types. Further, we investigate products that customers take into fitting rooms, and we find that they frequently deviate from, and complement purchases. Finally, we demonstrate how our findings impact practical applications, illustrated using recommender systems, and discuss how shopping baskets from different domains can be leveraged.

CCS CONCEPTS

• **Information systems** → *Online shopping*; *Business intelligence*; *Recommender systems*.

KEYWORDS

shopping basket; empirical analysis; fashion retail; online orders; brick-and-mortar; shopping behavior; recommender; RFID

ACM Reference Format:

Matthias Wölbitsch, Thomas Hasler, Simon Walk, and Denis Helic. 2020. Mind the Gap: Exploring Shopping Preferences Across Fashion Retail Channels. In *28th ACM Conference on User Modeling, Adaptation and Personalization (UMAP'20)*, July 14–17, 2020, Genoa, Italy. ACM, New York, NY, USA, 9 pages. <https://doi.org/10.1145/3340631.3394866>

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

UMAP '20, July 14–17, 2020, Genoa, Italy

© 2020 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-6861-2/20/07.

<https://doi.org/10.1145/3340631.3394866>



Figure 1: *Smart Fitting Room*. RFID-enabled smart fitting rooms bridge the gap between online and brick-and-mortar stores by providing features such as additional product information and product recommendations.

1 INTRODUCTION

In today's day and age the line between the physical and the virtual world is getting increasingly blurry. For example, fashion retailers nowadays regularly provide an engaging customer experience not only in their physical stores but also in their online shops. Both domains on their own exhibit strengths and weaknesses, such as the lack of personal (person-2-person) service in the online environment or the lower flexibility in terms of opening hours and limited product assortments in brick-and-mortar stores. However, customers often use both of these channels in combination (i.e., omni-channel shopping [20]) to make buying decisions by, for example, browsing, pre-selecting and reserving products online before going to the store to experience their actual look and feel.

The increasing adoption of state-of-the-art retail technologies, such as radio-frequency identification (RFID), further supports an advanced and improved customer experience in brick-and-mortar stores. For example, smart fitting rooms (see Figure 1) allow customers to experience features, previously exclusively available in online stores. Among such features are product recommendations, which provide customers with valuable suggestions either complementing products they are interested in or highlighting alternatives. **Problem.** Typically, recommendations are computed from customers' shopping histories (e.g., products that were purchased together). However, the actual customer behavior may differ across

domains as customers may have contrasting or different preferences when buying products online compared to when buying in traditional retail stores. Combining streams of data from multiple channels (i.e., online and brick-and-mortar stores) represents a challenging task that requires a sound understanding of the underlying mechanisms that drive people while shopping.

Contributions. To the best of our knowledge, we present the first extensive empirical analysis of shopping patterns in online and RFID-equipped fashion brick-and-mortar environments, highlighting and discussing some specific characteristics in the shopping baskets of customers that may reflect different intentions and motivations. Specifically, we show the distinctive differences with respect to shopping basket sizes and compositions as well as to what extent customers buy products in multiples colors or sizes. Next, we analyze how these characteristics affect product recommendations and discuss how they can potentially be leveraged to tailor recommendations towards specific purposes. Additionally, we publish our real-world datasets¹, which not only consist of sales data from both the online and offline domain, but also data that captures the interactions of customers with products in RFID-enabled smart fitting rooms.

The empirical analysis we present in this paper is an important step towards a better understanding of the shopping patterns and preferences of customers in different settings. This, in turn, builds the foundation for further research and several practical applications in RFID-based environments that require accurate user models, such as narrative-driven [8] and context-aware [1] product recommendations.

2 RELATED WORK

Blázquez [4] conducts a survey that highlights the importance of new technologies for customers that shop for fashion goods online and in brick-and-mortar stores. The author highlights that the possibility for customers to use multiple channels before and after buying goods plays an important role in their shopping experience. Wang et al. [26] investigate in their study the attitudes of customers towards buying in a multi-channel environment. They show that the choice for a channel often depends on the availability of information, the required effort to search for products, and the general convenience for customers. Similarly, Kollmann et al. [15] show that, for example, convenience is a major factor when selecting the online channel over traditional retail stores. Schramm et al. [22] analyze customer motives when they are presented with the option to buy the same product online or offline from the same retailer. Manchanda et al. [18] propose a multivariate probit model for grocery shopping baskets to capture complementary, co-occurrence, and homogeneity aspects when buying products from different categories, with the goal to provide retailers with valuable insights into purchase patterns of customers. Chu et al. [7] investigate the differences between the grocery shopping behavior of customers in online and brick-and-mortar stores, by analyzing aspects, such as brand loyalty and price sensitivity, and how they vary with the degree of internet usage of customers. In contrast, in our work we perform an empirical analysis of shopping patterns in online and RFID-equipped retail stores, and investigate the impact of the

observed differences on practical applications, illustrated using the example of recommender systems.

The *TagBooth* framework [17] can be used to collect shopping data of customers in traditional retail stores that goes beyond typical purchase patterns. Specifically, it uses RFID technology to capture interactions of customers with products in a store (e.g., picking up products from a shelf). Similarly, Zhao et al. [29] use RFID technology to identify products that are popular among customers, as well as products that are correlated with each other. Additionally, they use their collected data to find store areas that are highly frequented (i.e., hot zones), which enables retailers to optimize their store layouts. *ShopMiner* [23] and *CBID* [11] are similar approaches with the objective to gain insights into the shopping patterns of customers based on RFID technology. Al-Kassab et al. [2] discuss the benefits of RFID-enabled retail stores, based on a real-world case study in a large department store. They highlight, for example, the valuable insights into customer interaction patterns, which can be collected using smart mirrors that are placed on the sales floors and fitting rooms of the store. A similar case study was also conducted by Melià-Seguí et al. [19] in the context of the fashion retail domain. In contrast, in our work we use sales data from the online and offline domain, in addition to RFID-based fitting room interaction data, to gain insights into customer preferences across multiple retail channels.

In Hanke et al. [12] the authors discuss the benefits of recommender systems in modern brick-and-mortar fashion stores, and in particular in smart fitting rooms. Specifically, they focus on how to adapt recommender systems to take contextual information (e.g., the current season or trends) into account. Similarly, Liangxing and Aihua [16] outline the opportunity for fashion retailers to identify their customers using RFID-based membership cards that allow them to provide personalized fashion recommendations to their customers. Kang et al. [14] propose a recommender approach, which takes visual aspects of fashion products into account. Specifically, they focus on computing recommendations with the goal to “complete the look” (e.g., compile outfits that are visually compatible and fit a theme). Wan et al. [25] focus in their work on product complementarity, product compatibility, and customer loyalty patterns, and propose a new recommender method that takes these aspects into account. Moreover, structured information about product relationships (e.g., knowledge graphs or product taxonomies) is also often used to improve product recommendations [6, 28]. However, most recommender approaches specifically focus on a single shopping domain. In this work, we present an initial step towards compiling a recommendation framework that can be used across multiple retail channels, while still capturing the specific requirements of the individual domains.

3 DATASETS

For our analyses and experiments we leverage data from three different data streams of a large international fashion retailer (see Table 1 for an overview). Specifically, we focus on collections of fashion goods appearing in the same context (e.g., bought together). Such a collection consists of a number of individual physical items. Each of these items is associated with a unique product, which has certain properties (e.g., a certain design, color, and size).

¹data can be found at: https://github.com/detegoDS/mind_the_gap_dataset

Shopping Baskets. The first dataset we use consists of sales data from the online store of the fashion retailer between June 10, 2019 and September 26, 2019. Over this period we observe 1,119,570 purchases of one or more items (i.e., shopping baskets), which contain a total of 14,585 unique products. Moreover, we also leverage sales data from 112 brick-and-mortar retail stores, which we collect during the same time. In this dataset we observe a total of 2,080,072 shopping baskets, which contain a total of 17,626 unique products. Note that the purchase data from both sources has its origin in the same geographical region (i.e., the same country).

Fitting Baskets. Furthermore, we use RFID-enabled smart fitting room data to determine which products are tried on together in dressing rooms. We collect the data during a pilot project between December 2016 and March 2018 in a single store of the same retailer. Each of the 5 fitting rooms in this store is equipped with motion sensors to determine whenever a customer enters the cabin, and with RFID readers to detect the items brought into the fitting room. We denote the collection of items observed during a single session (i.e., defined by the time-span between a customer entering and afterwards leaving the fitting room) as *fitting baskets*. Hence, the fitting room dataset contains a total of 32,279 fitting baskets, which are composed of 3,790 unique products. Note that we include the dataset in our analyses only for illustration purposes, as the fitting baskets are collected within a different time period, and the pilot store is located in a different region.

Data Filtering. For online and brick-and-mortar shopping baskets as well as fitting baskets, we discard all baskets with more than 20 items. We select this threshold based on the intuition that larger basket sizes are not representative for “typical” customers, and are therefore considered outliers (e.g., due to collective orders). Using this filtering step we reduce the number of available baskets by 0.18% for the brick-and-mortar, by 0.64% for the online store, and by 24.28% for the fitting room dataset. We attribute this larger number of filtered fitting baskets to store staff occasionally using dressing rooms as temporary storage for products when restocking the sales floors while the store is closed for business.

Product Attributes. Each product in our datasets is associated with different attributes (i.e., meta data) that categorizes it with respect to different properties. For example, the product division distinguishes products on a high abstraction level. Each product belongs either to the class *apparel* (i.e., clothing), *footwear*, or *accessories*. Additionally, each product is assigned a product type, which provides additional information about the product itself. Specifically, a product is assigned to one of 94 different types (e.g.,

Table 1: Datasets Overview. We show the numbers of shopping baskets in the datasets collected from a fashion retailer, as well as the number of collected fitting baskets. We also show the number of unique products in each dataset.

Dataset	Baskets	Products
Online Shop Sales	1,119,570	14,585
Brick-and-Mortar Sales	2,080,072	17,626
Brick-and-Mortar Fittingroom	32,279	3,790

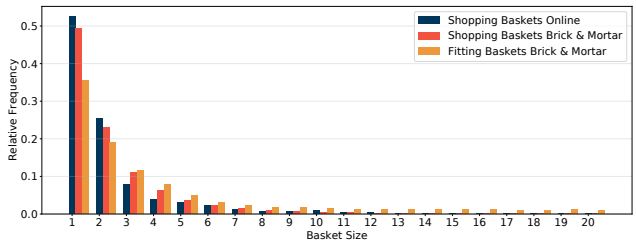


Figure 2: Empirical Basket Size Distributions. We depict the distributions of shopping basket sizes and fitting basket sizes (i.e., how many items were tried on together in the fitting room). The mean shopping basket size is 2.28 for brick-and-mortar stores, and 2.24 for the online shop, while the fitting baskets include on average 4.12 items. Nevertheless, the distribution shows that online orders often consist of few items, while interactions in physical stores tend to be more diverse. We verify the significance ($p < 0.0001$) between the difference of the distributions using the two sample chi-square test and Bonferroni correction [5].

graphic T-shirt, polo shirt, jeans, ...). To support interpretability of our analyses, we combine different product types into 8 groups (i.e., *accessories, footwear, pants, shirts, tops, socks, underwear, and others*). For example, we aggregate *graphic T-shirts* and *polo shirts* into the group *shirts*. Finally, we also assign each product to one of 48 different product categories (e.g., *casual, swimming, training, ...*), which specifies in which context a product is typically used or worn.

4 EMPIRICAL ANALYSIS

4.1 Basket Anatomy

First, we explore preferences of customers with respect to the number of items that they buy or try on in fitting rooms (i.e., the sizes of the shopping baskets and fitting baskets; see Figure 2). Additionally, we investigate the number of distinct (grouped) product types that are represented in shopping baskets and fitting baskets with more than one item, which allows us to determine the product diversity of baskets (see Figure 3a). For example, a shopping basket composed of three T-shirts, a pair of jeans, and sweatpants (i.e., a shopping basket of size 5) contains 2 product types (i.e., *shirts* and *pants*).

Particularly, for purchases made online we observe that more than half (52.6%) of all shopping baskets consist of a single item. While larger online shopping baskets are present in the dataset, their relative amount decreases quickly with increasing basket size. For example, online shopping baskets containing less than four items account for more than 89% of all baskets in this dataset (cf. Figure 2). Furthermore, we observe that 71.3% of all online shopping baskets involve only a single product type, while baskets with more than two different product types are scarcely represented in the data (cf. Figure 3a).

Shopping baskets in brick-and-mortar stores contain only one item in 49.3% of all cases. In total of 89% of all baskets contain four or less items, and larger shopping basket sizes are fairly uncommon in the brick-and-mortar setting. When comparing the sizes of fitting baskets to retail store purchases, we find that customers usually

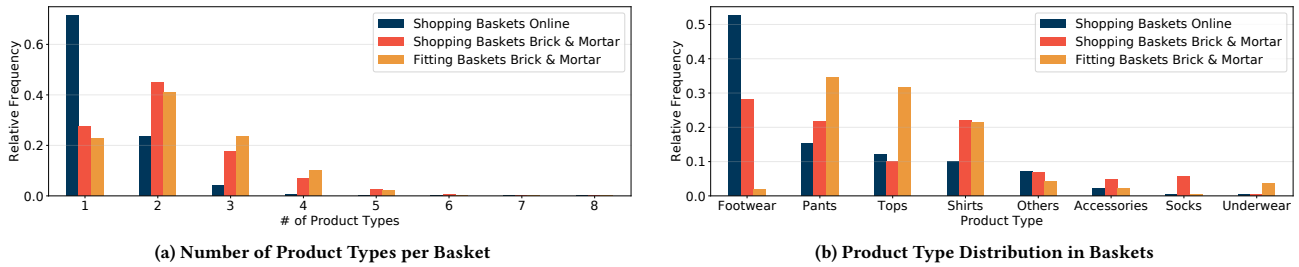


Figure 3: Product Type Properties. Each item in a shopping/fitting basket is assigned one of the 8 (grouped) product types (e.g., shirts, footwear, ...). In (a) we depict how many of these product types typically appear in baskets of size 2 or more. Customers of the online shop tend to buy items of only one type, while customers of brick-and-mortar stores tend to combine items of different types. Furthermore, in (b) we depict the relative frequency of individual products of a certain (grouped) type appearing in a fitting basket or shopping basket. Sales in the online store often include shoes or other footwear, while fittings and purchases in traditional retail stores are more homogeneous with respect to the product type, especially for apparel. Note that we verify significance ($p < 0.0001$) between distributions for (a) and (b).

take more items into the fitting room than they buy (cf. Figure 2). Fitting baskets of size 1 or 2 are less common than retail store shopping baskets, while larger collections of items that are tried on by customers occur more frequently. This is particularly evident in the tail of the distribution, where fitting baskets of size 10 or more make up more than 11% of all baskets, compared to only roughly 1% for purchases with the same number of items. Moreover, the number of distinct product types of fitting/shopping baskets in brick-and-mortar stores is similar, with most baskets containing two or more product types (cf. Figure 3a).

Online shopping is focused while offline is more diverse. Overall we find that the majority of online and brick-and-mortar shopping baskets are of size 1 or 2. However, significant numbers of larger shopping baskets still exist, and while shopping baskets of size three or more are more common in brick-and-mortar stores, smaller baskets occur more frequently online. Fitting basket sizes are on average larger, indicating that customers perform a pre-selection of products in various sizes before actually making purchase decisions in brick-and-mortar stores. Moreover, interactions with products in brick-and-mortar stores appear to be more diverse compared to the online store as most of the time two product types are involved. Although the relative number of shopping baskets of size 1 is nearly the same, the number of shopping baskets containing only a single product type is significantly higher for online than offline purchases (cf. Figure 3a). Finally, we find that fitting room sessions tend to involve a large number of different product types, which are not necessarily bought. This indicates that shopping patterns in traditional retail stores tend to span a wider range of product types, while purchases made in online stores are more focused towards a specific category.

4.2 Product Co-Purchases

Not only the quantity of items and product types that customers interact with is of great interest when exploring behavioral patterns of customers for user modeling, but also what product types they are typically interested in. This includes what types of products are bought by customers, and what kinds of products they take

with them into the fitting room (see Figure 3b). Moreover, we also investigate what types of products are bought together (i.e., co-purchases) or tried on together in fitting rooms. Hence, we map each item in a shopping/fitting basket to its corresponding product type and count the co-occurrence of product types for each possible product pair in our data (see Figure 4).

In particular, we find that shoes and other footwear are very popular in the online shop, as more than half (52.5%) of all purchases made in this setting include shoes (see Figure 3b). The remaining purchases that are made online often involve apparel, and—as can be expected for fashion retailers—the product types spanning accessories and other related product types (i.e., underwear and socks) are not sold as frequently. Further, we observe a very pronounced diagonal in the co-occurrence matrix for shopping baskets from the online store, indicating that customers usually buy products of the same type. Particularly footwear is almost exclusively bought with other footwear (i.e., 97% of the time; see Figure 4a).

In our data, shoes are also the most popular items in traditional retail stores. Their relative share in the total number of sales in our dataset is 28.2%, and they are usually not tried on in fitting rooms (i.e., 2%). On the other hand, apparel (i.e., pants, tops, and shirts) also appears very often in shopping and fitting baskets. Moreover, we can observe that socks are almost exclusively bought in brick-and-mortar stores, and that 3.6% of all fitting sessions include underwear (see Figure 3b). In general, the co-occurrence of products in brick-and-mortar stores is more diverse, while still exhibiting trends towards buying/trying on within the same product category (see Figure 4b and 4c).

Buy footwear online and matching apparel offline. In general we find that customers exhibit certain product type preferences depending on the shopping channel. Many customers buy their footwear online, despite the possibility to find the best fitting shoe size in brick-and-mortar stores. Nevertheless, shoes are also popular in brick-and-mortar stores, but are not tried on very often in the fitting rooms, as customers usually fit shoes in the shoe area of the sales floor. We hypothesize that the small percentage of shoes that we see in the smart fitting rooms can be attributed to customers

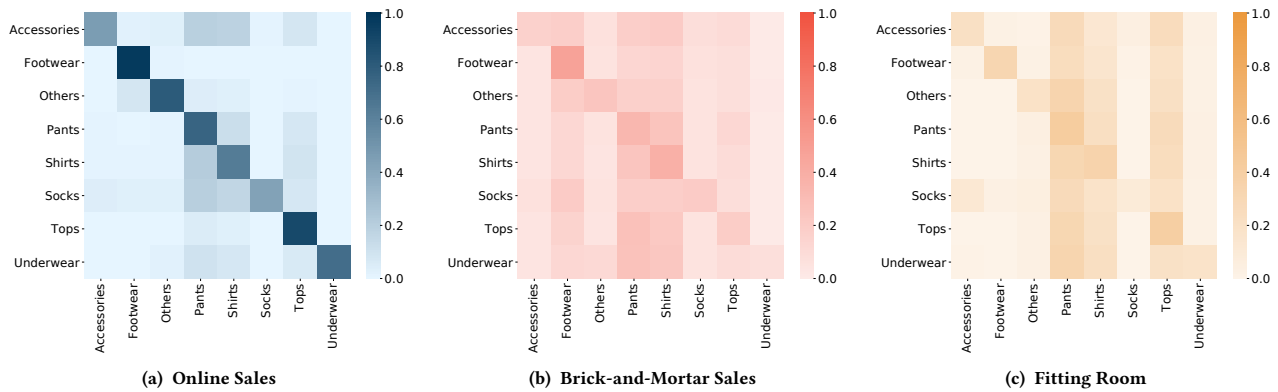


Figure 4: Product Type Co-Occurrences. We depict the relative co-occurrence frequencies between product types for (a) online shopping baskets, (b) brick-and-mortar shopping baskets, and (c) baskets observed in fitting rooms. In the online domain we observe that pairs of purchased products are often of the same product type, which is evident in the strong diagonal in the depicted heat-map. For purchases in brick-and-mortar stores we see more homogeneous shopping patterns. While this is also the case for fitting baskets, we can see that most product types are often observed with *pants*, *shirts*, and *tops*, which can be seen for the more pronounced vertical lines in the heat-map for these three product types.

looking to buy apparel while also looking for a new pair of shoes. We find that products of type *pants* and *tops* (e.g., sweatshirts, jackets, etc.) are more often tried on compared to their relative number of sales in brick-and-mortar stores. This can, however, not be observed for *shirts* (e.g., T-shirts, polo shirts, etc.) where sales and fittings are equally common.

Shop online for a single product type and offline for outfits.

We find that online customers frequently buy multiple products of the same type (i.e., the diagonal in Figure 4a), which further corroborates our observation that such customers exhibit less diverse shopping patterns. This is not the case for customers of the brick-and-mortar stores, as we are not able to observe clear patterns in the product type co-occurrence matrix. In general, while there is still a trend to buy products of the same type, the observed behavioral buying patterns across product types is more uniformly distributed. Furthermore, in the fitting rooms we find, on the one hand, a tendency of customers to try on multiple products of the same type, potentially in different sizes. On the other hand, we also see strong co-occurrences of all product types with *shirts*, *tops*, and *pants*, indicating that customers are more prone to mix-and-match products of different types. For future work we want to investigate to what extent these products visually complement each other.

4.3 Product Sizes

In Section 4.2 we discovered that customers often interact with multiple products of the same type. Therefore, we further explore to which extent this results in buying, or trying on the same product multiple times. This can either be related to buying the same product in the same size more than once, or in different sizes. Hence, we inspect all shopping baskets with more than one item, count the number of products appearing more than once in these baskets (see Figure 5a), and lastly analyze the product sizes (see Figure 5b).

We find that while the majority of purchases (63.6%) in the online store do not contain a product multiple times, 36.4% of all shopping

baskets include at least one product more than once. We can observe that buying such a product in multiple different sizes is common online (i.e., 45.9% for two sizes, and 62% in total; Figure 5b). In brick-and-mortar stores we find that most shopping baskets (87.5%) do not contain a product more than once. Whenever a product is bought multiple times, we find an equilibrium between buying the product in the same size multiple times and buying it in different sizes. The number of fitting baskets containing a product more than once is with 38.1% much more frequent compared to shopping baskets in brick-and-mortar stores (see Figure 5a). Furthermore, whenever a product is in a fitting basket more than once, more than 90% of all cases involve multiple sizes (see Figure 5b).

Buy apparel that you like more than once. We find that buying the same product in the same size multiple times is common online (38%; cf. Figure 5b). Moreover, this buying pattern is universal across the two purchase channels and potentially reflects the ambition of customers to own something they like more than once.

No fitting online: buy the product in different sizes. Ordering the same product multiple times is, in general, rather common online. This indicates that online customers counteract uncertainties about which size to buy by frequently ordering a product in multiple sizes. In general, we find that buying a product more than once is less frequent in brick-and-mortar stores. This can most likely be attributed to the fact that customers are able to try on products in traditional stores before buying them. This is further corroborated by the pattern present in the fitting baskets, where two or more sizes for the same product are very common. Nevertheless, while it is possible for customers to find their matching size of products in the fitting areas of stores, we still observe that customers occasionally buy products in different sizes. We hypothesize that these cases can be attributed to different preferences of customers such as wanting to try on the purchased products at home and returning the ill-fitting items later.

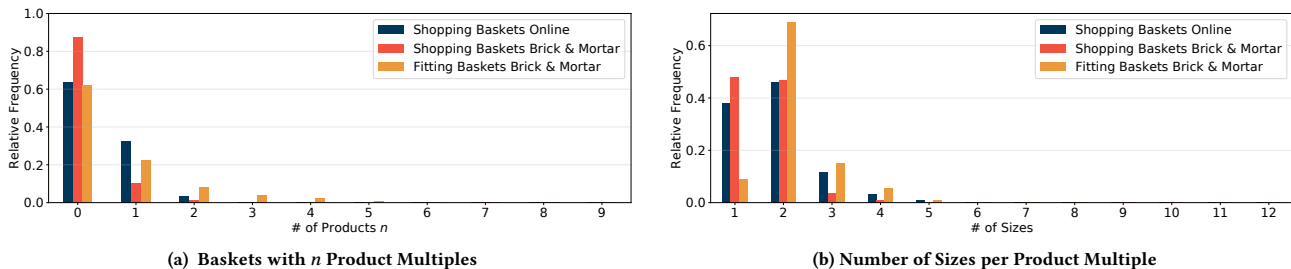


Figure 5: Product Selection. Each shopping/fitting basket can contain a product multiple times, potentially in multiple sizes. In (a) we depict the relative frequency of baskets that contain n products multiple times. In brick-and-mortar stores, baskets typically do not contain a product more than once (i.e., $n = 0$), while such products in online store shopping baskets and fitting baskets are more common. Further, in (b) we show the relative frequency of the number of sizes that are seen with a product that occurs more than once in a basket. Customers frequently buy the same product in the same size multiple times online and offline, while fitting rooms are mainly used to try different sizes. Note that we verify significance ($p < 0.0001$) between distributions for both (a) and (b).

5 IMPACT ON RECOMMENDER SYSTEMS

In this section we discuss the effects of leveraging different streams of sales data (i.e., online and brick-and-mortar shopping baskets without fitting baskets) for computing product recommendations. Based on the results of our empirical analysis we formulate the following hypotheses for computation of product recommendations:

- H1 Stronger shopping focus and more distinct co-purchase patterns should result in an easier recommendation problem for online purchases as compared to offline recommendations.
- H2 Due to different shopping basket anatomies, training on one domain while recommending on the other, as well as combining data from both domains for training, should result in diversified but less accurate recommendations.

5.1 Experiments

Recommendation Algorithms. For our experiments we use three different recommendation algorithms, each with their individual strengths and weaknesses. As a baseline approach we use the random recommendation algorithm (RAND), which selects k products from the set of products uniformly at random without replacement. Hence, this algorithm is able to recommend a wide variety of different choices. Next, we use a popularity-based recommender (POP), which recommends the k most popular products (i.e., products that are sold most often; ordered by frequency). This method often provides good results—especially in cold-start scenarios—as the recommended products may appeal to a broad audience. Finally, we use the embedding-based prod2vec [3, 10] approach (EMBED), which computes d -dimensional vector representations of products based on their co-occurrence in shopping baskets. We use a skip-gram architecture with negative sampling, where we draw 20 negative samples per example from the α -smoothed product-frequency distribution with $\alpha = 0.75$. Furthermore, we fix the number of training epochs to 100, the dimension of the resulting embedding vectors to $d = 32$, remove products from the training set which appear less than 5 times due to limited support, and randomly downsample frequently occurring products with a frequency greater than

$t = 0.0001$. We then compute recommendations based on the distance between inferred product embeddings. To that end, we use Euclidean distance to find the k closest product vectors in the d -dimensional vector space for a given product. Moreover, we also extend the shopping baskets in our training sets with information in the form of additional tokens [24, 27]. Specifically, we add the point-of-sale (POS) information (i.e., whether the purchase was made online or in a brick-and-mortar store) to the shopping baskets. We refer to this variant of prod2vec as POS-EMBED.

Evaluation Metrics. To evaluate the implications for different sources of sales information on recommendations, we compute a set of k recommendations for each product in a shopping basket. To measure the accuracy we compare the list of computed recommendations, which is ordered by relevance, against the remaining products in the shopping basket using *recall at k* ($Recall_k$), which is defined as the ratio of relevant recommended products to the actual number of relevant products (i.e., the remaining products of the shopping basket). However, recall does not meet the requirement of many real-world recommendation systems that relevant products should appear in the beginning of the list [21]. Therefore, we use as a second accuracy metric, the *normalized discounted cumulative gain at k* ($NDCG_k$) [13], which takes the rank into account as well. The metric penalizes relevant products appearing at the bottom of the list, by reducing their contribution to the score logarithmically with respect to their position.

However, the quality of recommendations can also be measured with metrics that reflect different requirements and objectives of a recommender system. To that end, we calculate *Coverage*, which measures the fraction of products that will actually get recommended. A recommender with a high coverage is able to expose customers to a wide spectrum of different products, while low coverage keeps them restricted to a limited product assortment. Related to that, we also calculate the mean *Diversity* of recommendations, which is determined by the pairwise distance between products. We calculate the distance based on weighted product attributes using Gower similarity [9]. Specifically, we use the product division with a weight of 1, the product category with a weight of 3, the

grouped product type with a weight of 5, and the specific product type with a weight of 10 (all values informed by domain knowledge). Finally, we calculate the mean *Novelty* of recommendations. The novelty of a product is related to its popularity, as popular products are less surprising, and therefore probably not of great interest. More formally, the novelty of a product is its self-information (i.e., $-\log_2(p(i))$), where $p(i)$ is the relative share of product i on the total number of sales).

Training & Test Data. We base our experiments on the same datasets already used for the empirical analysis. However, for recommender experiments, we only use shopping baskets that contain two or more distinct products. Furthermore, we do not include fitting room data due to the the limited overlap of products in the fitting room dataset compared to sales. To make results easier to interpret, we also only consider shopping baskets that contain products that appear both in the online and brick-and-mortar shopping baskets. The number of overlapping products between the datasets is 9,705, which reduces the number of available shopping baskets to 336,256 for the online shopping domain and 823,753 for the brick-and-mortar domain. Note that we perform a 85%/15% training and test split on both online and brick-and-mortar datasets.

Model Compilation. For our experiments, we fit for both datasets an individual model. For example, we calculate product embeddings using the EMBED approach solely based on the online sales training set. We then evaluate the resulting model on the test set of the online store data as well as the test set of the brick-and-mortar data. This allows us to investigate how well resulting models can be transferred to the other domain. Additionally, we compile general EMBED models, where we leverage shopping baskets from the combined training sets to compute product embeddings. Hence, we explore the possibility to simply merge the streams and investigate how this affects recommendation performance in both domains.

5.2 Results & Discussion

Our random baseline (RAND) results in both lower and upper bounds of our metrics. For example, the product catalog *Coverage* (i.e., the percentage of products getting actually recommended in the set of all products) is 100% for RAND. Also, for RAND *Novelty* and *Diversity* of the recommendations is high, which means that less popular products are frequently recommended, and the recommendation list contains products of various types and categories. On the other hand, accuracy metrics are the lowest for RAND. We now report and discuss all the results using the discussed accuracy and beyond-accuracy metrics for $k = 10$ in the context of our hypotheses.

H1: Stronger shopping focus and more distinct co-purchase patterns should result in an easier recommendation problem for online purchases as compared to offline recommendations.

When recommending products based on the most popular products (POP) in the respective domains we find that, in terms of accuracy, we achieve $NDCG_{10} = 0.0316$ in the brick-and-mortar domain, and $NDCG_{10} = 0.0197$ for the online store (cf. row (b) and (i) in Table 2), which is in contrast to our expectations. Hence, popular products are more important in brick-and-mortar stores, and therefore more significant in this domain. Moreover, *Diversity* of recommendations is also rather high for both domains (see cells marked with A in Table 2 for comparison with the random baseline). Therefore, the set of most popular products is very diverse regardless of the context. This is rather unexpected for the online shop, as the previous findings indicate a strong bias towards footwear. Hence, we expect a lower *Diversity* for the best selling products as well. However, this can be explained by the composition of the 10 most popular products, which consist for both domains of products from multiple categories (e.g., beside footwear we also find shirts and pants in the list of most popular products of the online store).

Table 2: Experimental Results. This table depicts accuracy metrics (i.e., $NDCG_k$ and $Recall_k$) as well as beyond-accuracy metrics (i.e., $Novelty_k$, $Diversity_k$, and $Coverage$) for different recommendation models (random, popularity, and prod2vec), each generating $k = 10$ recommendations. In particular, we highlight the difference in performance with respect to the different metrics when fitting them on shopping baskets from (i) brick-and-mortar stores, (ii) online store, and (iii) both of them combined, and evaluating them on a test set from the respective domain. Note that we mark cells with superscript letters (i.e., A, B, C) to highlight findings in our results.

Model	Training Set	Test Set	$NDCG_{k=10}$	$Recall_{k=10}$	$Novelty_{k=10}$	$Diversity_{k=10}$	$Coverage$
(a) RAND	-	Brick & Mortar	0.0007	0.0011	16.3763	0.7854 ^A	1.0000
(b) POP	Brick & Mortar	Brick & Mortar	0.0316	0.0461	7.7969	0.7356 ^A	0.0010
(c) POP	Online	Brick & Mortar	0.0128	0.0213	11.7956	0.7111 ^A	0.0010
(d) EMBED	Brick & Mortar	Brick & Mortar	0.0966	0.1264 ^C	10.8430	0.5297 ^B	0.5967
(e) EMBED	Online	Brick & Mortar	0.0528	0.0695	13.4118	0.4305	0.6842
(f) EMBED	Combined	Brick & Mortar	0.0969	0.1267 ^C	10.9778	0.4932	0.7861
(g) POS-EMBED	Combined	Brick & Mortar	0.0979	0.1282 ^C	11.0159	0.4884	0.7985
(h) RAND	-	Online	0.0006	0.0010	14.8616	0.7794	1.0000
(i) POP	Online	Online	0.0197	0.0319	8.3355	0.7111	0.0010
(j) POP	Brick & Mortar	Online	0.0066	0.0077	13.0900	0.7356	0.0010
(k) EMBED	Online	Online	0.1919	0.2601	11.6272	0.3735 ^B	0.6917
(l) EMBED	Brick & Mortar	Online	0.0759	0.0981	13.2781	0.5249	0.5589
(m) EMBED	Combined	Online	0.1574	0.2066	12.3362	0.4119	0.8123
(n) POS-EMBED	Combined	Online	0.1658	0.2201	12.4251	0.4006	0.8235

While the popularity model performs better for the brick-and-mortar dataset in terms of accuracy, we find the opposite is true for the EMBED models. Here, EMBED fitted on the online shop training set achieves an $NDCG_{10} = 0.1919$, while the EMBED model fitted on the shopping baskets recorded in traditional retail stores only achieves an $NDCG_{10} = 0.0966$ (cf. row (d) and (k) in Table 2). This large gap in accuracy between the domains aligns with our hypothesis, which indicates that shopping preferences can be predicted easier in an online context. This reflects, once more, that the shopping patterns in the online shop are more focused on single products compared to brick-and-mortar stores where customers buy more diverse. In contrast to the popularity-based recommender, this is also reflected in the differences in *Diversity* between the two domains (see cells marked with *B* in Table 2), where we find much higher average *Diversity* for the brick-and-mortar EMBED model. Hence, we are able to confirm our hypothesis using a more sophisticated recommendation approach, which is—in contrast to POP—able to capture the latent relationships between products.

H2: Due to different shopping basket anatomies, training on one domain while recommending on the other, as well as combining data from both domains for training, should result in diversified but less accurate recommendations.

The most popular products of one domain cannot be transferred to the other without decreasing accuracy metrics by more than 50%. Hence, popular products are very distinct within the domains, as customers express context-dependent preferences. Further, transferring improves *Novelty* by more than 50%, indicating that the most popular products of one domain exhibit far lower popularity in the other (cf. rows (b) vs. (c), and (i) vs. (j) in Table 2).

Moreover, applying the EMBED model from one domain to the other (compare rows (d) vs. (e), and (k) vs. (l) in Table 2) also affects accuracy, as evident in a decrease in accuracy by more than 60% and 45%. This further highlights the strong differences in the observed behavioral preferences. *Novelty*₁₀ increases by 14.2% and 23.7%, indicating that the popularity of products appears to be very different over the entire dataset. Average *Diversity* increases by about 40% when using the brick-and-mortar EMBED model in the online domain. However, we find a decrease by more than 18% when applying the online EMBED model to brick-and-mortar shopping baskets. Nevertheless, this is in line with heterogeneous online shopping patterns, which reflect poorly on the diversity of the generated recommendations due to purchases often happening within the same product category or type. We measure *Coverage* by comparing the set of recommended products against products appearing in the corresponding training set. *Coverage* decreases from 0.5967 to 0.5589 when testing the brick-and-mortar EMBED model on different test sets (see row (d) vs. (l) in Table 2). The online EMBED model (see row (e) vs. (k) in Table 2) follows analogously with a *Converge* decrease from 0.6917 to 0.6842.

Next, we investigate the accuracy of the EMBED model that is trained on a combined training set from the brick-and-mortar store and online store. We find no drastic difference in terms of $NDCG_{10}$ and $Recall_{10}$ for traditional retail stores. For both metrics we observe a minor improvement by 0.31% and 0.24%. Further, for the general model that leverages the point of sale information (POS-EMBED) we find a marginally better improvement by 1.35% and 1.42% for

$NDCG_{10}$ and $Recall_{10}$ (see cells marked with *C* in Table 2). Hence, mixing both data sources appears to not affect the performance for recommendations in retail stores. On the other hand, when applying the general EMBED model to the online domain we see more pronounced negative effects, with a decrease in $NDCG_{10}$ from 0.1919 to 0.1574 for EMBED, and 0.1658 for POS-EMBED. Hence, introducing more heterogeneous shopping patterns of customers of the traditional retail stores negatively influences recommendation performance in terms of accuracy in the online domain. Nevertheless, in the case of the traditional retail store domain we still improve mean *Novelty* compared to the individual model and *Coverage* to an even larger extent (i.e., more than 33%). *Diversity* still decreases, but to a smaller extent compared to using the online EMBED model in the brick-and-mortar domain. In the case of the online domain we find a consistent improvement with respect to all beyond-accuracy metrics when using our general EMBED and POS-EMBED approach. In summary, by combining training data from both domains we find mixed effects on accuracy, while improving most beyond-accuracy metrics for both domains. Hence, we can confirm our hypothesis for the online domain, while additional research is required to fully explain the effects observed for the brick-and-mortar domain.

6 CONCLUSION & FUTURE WORK

In this paper we presented a first empirical analysis that compares the behavioral shopping preferences of customers between multiple channels (i.e., online vs. brick-and-mortar). To that end, we analyzed real-world datasets² collected from a large international fashion retailer. Furthermore, we expanded our analysis by also taking customer interactions in RFID-enabled smart fitting rooms in a brick-and-mortar store into account. We showed significant differences in shopping preferences between the two domains, and further illustrated their impact on recommender systems.

For future work we plan to focus on two main research directions. First, we want to extend our empirical analysis by further investigating how customers behave in the context of traditional retail stores, especially in connection with omni-channel retailing. For example, we want to study the acceptance of features such as *Click&Collect*, where customers are able to order products online and pick them up later in the store. We also want to examine returns that customers perform offline and online, to further understand the motivation of purchases that contain products multiple times. Additionally, we are planning to investigate the generalization of our results across different domains. Second, we want to leverage our findings for practical applications. We discussed how merging different sources of sales data affects recommendation performance. In future work we plan to build on this and design recommendation approaches that strategically use this information to steer recommendations towards different directions and use cases.

In conclusion, we presented a novel field of research, relying on, and leveraging data provided from RFID technology, which is gaining more and more traction in the fashion retail industry. Our results indicate significant differences between online and brick-and-mortar retailing channels, which we can leverage in the future to build and improve on practical applications to further bridge the gap between online and offline retailing.

²data can be found at: https://github.com/detegods/mind_the_gap_dataset

REFERENCES

- [1] Gediminas Adomavicius, Ramesh Sankaranarayanan, Shahana Sen, and Alexander Tuzhilin. 2005. Incorporating Contextual Information in Recommender Systems using a Multidimensional Approach. *ACM Transactions on Information Systems* (2005), 103–145.
- [2] Jasser Al-Kassab, Philipp Blome, Gerd Wolfram, Frédéric Thiesse, and Elgar Fleisch. 2011. RFID in the Apparel Retail Industry: A Case Study from Galeria Kaufhof. In *Unique Radio Innovation for the 21st century*.
- [3] Oren Barkan and Noam Koenigstein. 2016. Item2vec: Neural Item Embedding for Collaborative Filtering. In *Proceedings of the 26th IEEE International Workshop on Machine Learning for Signal Processing*, 1–6.
- [4] Marta Blázquez. 2014. Fashion Shopping in Multichannel Retail: The Role of Technology in Enhancing the Customer Experience. *International Journal of Electronic Commerce* (2014), 97–116.
- [5] Carlo E Bonferroni. 1935. Il calcolo delle assicurazioni su gruppi di teste. *Studi in onore del professore salvatore ortu carboni* (1935), 13–60.
- [6] Yoon Ho Cho and Jae Kyeong Kim. 2004. Application of Web Usage Mining and Product Taxonomy to Collaborative Recommendations in E-commerce. *Expert systems with Applications* (2004), 233–246.
- [7] Junhong Chu, Marta Arce-Urriza, José-Javier Cebollada-Calvo, and Pradeep K Chintagunta. 2010. An Empirical Analysis of Shopping Behavior Across Online and Offline Channels for Grocery Products: the Moderating Effects of Household and Product Characteristics. *Journal of Interactive Marketing* (2010), 251–268.
- [8] Lukas Eberhard, Simon Walk, Lisa Posch, and Denis Helic. 2019. Evaluating Narrative-Driven Movie Recommendations on Reddit. In *Proceedings of the 24th International Conference on Intelligent User Interfaces*, 1–11.
- [9] J. C. Gower. 1971. A General Coefficient of Similarity and Some of Its Properties. *Biometrics* (1971), 857–871.
- [10] Mihajlo Grbovic, Vladan Radosavljevic, Nemanja Djuric, Narayan Bhamidipati, Jaikrit Savla, Varun Bhagwan, and Doug Sharp. 2015. E-commerce in Your Inbox: Product Recommendations at Scale. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1809–1818.
- [11] Jinsong Han, Han Ding, Chen Qian, Wei Xi, Zhi Wang, Zhiping Jiang, Longfei Shangguan, and Jizhong Zhao. 2014. CBID: A Customer Behavior Identification System Using Passive Tags. In *Proceedings of the 22nd IEEE International Conference on Network Protocols*.
- [12] Jannis Hanke, Matthias Hauser, Alexander Dürr, and Frédéric Thiesse. 2018. Redefining the Offline Retail Experience: Designing Product Recommendation Systems for Fashion Stores. In *Proceedings of the 26th European Conference on Information Systems*.
- [13] Kalervo Järvelin and Jaana Kekäläinen. 2002. Cumulated Gain-based Evaluation of IR Techniques. *ACM Transactions on Information Systems* (2002), 422–446.
- [14] Wang-Cheng Kang, Eric Kim, Jure Leskovec, Charles Rosenberg, and Julian McAuley. 2019. Complete the Look: Scene-based Complementary Product Recommendation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 10532–10541.
- [15] Tobias Kollmann, Andreas Kuckertz, and Ina Kayser. 2012. Cannibalization or Synergy? Consumers’ Channel Selection in Online–Offline Multichannel Systems. *Journal of Retailing and Consumer Services* (2012), 186–194.
- [16] Yu Liangxing and Dong Aihua. 2010. Hybrid Product Recommender System for Apparel Retailing Customers. In *Proceedings of the 2010 WASE International Conference on Information Engineering*, 356–360.
- [17] Tianci Liu, Lei Yang, Xiang-Yang Li, Huaiyi Huang, and Yunhao Liu. 2015. Tag-booth: Deep Shopping Data Acquisition Powered by RFID Tags. In *2015 IEEE Conference on Computer Communications*, 1670–1678.
- [18] Puneet Manchanda, Asim Ansari, and Sunil Gupta. 1999. The “Shopping Basket”: A Model for Multicategory Purchase Incidence Decisions. *Marketing Science* (1999), 95–114.
- [19] Joan Melià-Seguí, Rafael Pous, Anna Carreras, Marc Morenza-Cinos, Raúl Parada, Zeinab Liaghat, and Ramir De Porrata-Doria. 2013. Enhancing the Shopping Experience Through RFID in an Actual Retail Store. In *Proceedings of the 2013 ACM Conference on Pervasive and Ubiquitous Computing*, 1029–1036.
- [20] Wojciech Piotrowicz and Richard Cuthbertson. 2014. Introduction to the Special Issue Information Technology in Retail: Toward Omnichannel Retailing. *International Journal of Electronic Commerce* (2014), 5–16.
- [21] Stephen Robertson. 1977. The Probability Ranking Principle in IR. *Journal of Documentation* (1977), 294–304.
- [22] Hanna Schramm-Klein, Bernhard Swoboda, and Dirk Morschett. 2007. Internet vs. Brick-and-Mortar Stores-Analysing the Influence of Shopping Motives on Retail Channel Choice Among Internet Users. *Journal of Customer Behaviour* (2007), 19–36.
- [23] Longfei Shangguan, Zimu Zhou, Xiaolong Zheng, Lei Yang, Yunhao Liu, and Jinsong Han. 2015. ShopMiner: Mining Customer Shopping Behavior in Physical Clothing Stores with COTS RFID Devices. In *Proceedings of the 13th ACM conference on Embedded Networked Sensor Systems*.
- [24] Flavian Vasile, Elena Smirnova, and Alexis Conneau. 2016. Meta-prod2vec: Product Embeddings using Side-information for Recommendation. In *Proceedings of the 10th ACM Conference on Recommender Systems*, 225–232.
- [25] Mengting Wan, Di Wang, Jie Liu, Paul Bennett, and Julian McAuley. 2018. Representing and Recommending Shopping Baskets with Complementarity, Compatibility and Loyalty. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, 1133–1142.
- [26] Yu-Min Wang, Hsin-Hui Lin, Wei-Chun Tai, and Yi-Ling Fan. 2016. Understanding multi-channel research shoppers: an analysis of Internet and physical channels. *Information Systems and e-Business Management* (2016), 389–413.
- [27] Matthias Wölbitsch, Simon Walk, Michael Goller, and Denis Helic. 2019. Beggars can’t be Choosers: Augmenting Sparse Data for Embedding-Based Product Recommendations in Retail Stores. In *Proceedings of the 27th ACM Conference on User Modeling, Adaptation and Personalization*, 104–112.
- [28] Da Xu, Chuanwei Ruan, Evren Korpeoglu, Sushant Kumar, and Kannan Achan. 2020. Product Knowledge Graph Embedding for E-commerce. In *Proceedings of the 13th International Conference on Web Search and Data Mining*, 672–680.
- [29] Jumin Zhao, Like Wang, Deng-ao Li, Yanxia Li, Bin Yang, Biaokai Zhu, and Ruiqin Bai. 2018. Mining Shopping Data with Passive Tags via Velocity Analysis. *EURASIP Journal on Wireless Communications and Networking* (2018).