Redefining the Offline Retail Experience: Designing Product Recommendation Systems for Fashion Stores

Jannis Hanke  
*University of Wuerzburg, jannis.hanke@uni-wuerzburg.de*

Matthias Hauser  
*University of Wuerzburg, matthias.hauser@uni-wuerzburg.de*

Alexander Dürr  
*University of Wuerzburg, alexander.duerr@uni-wuerzburg.de*

Frédéric Thiesse  
*University of Wuerzburg, frederic.thiesse@uni-wuerzburg.de*

Follow this and additional works at: [https://aisel.aisnet.org/ecis2018_rp](https://aisel.aisnet.org/ecis2018_rp)

**Recommended Citation**

[https://aisel.aisnet.org/ecis2018_rp/1](https://aisel.aisnet.org/ecis2018_rp/1)

This material is brought to you by the ECIS 2018 Proceedings at AIS Electronic Library (AISeL). It has been accepted for inclusion in Research Papers by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.
REDEFINING THE OFFLINE RETAIL EXPERIENCE: 
DESIGNING PRODUCT RECOMMENDATION SYSTEMS FOR FASHION STORES

Research in Progress

Hanke, Jannis, University of Würzburg, Germany, jannis.hanke@uni-wuerzburg.de
Hauser, Matthias, University of Würzburg, Germany, matthias.hauser@uni-wuerzburg.de
Dürr, Alexander, University of Würzburg, Germany, alexander.duerr@uni-wuerzburg.de
Thiesse, Frédéric, University of Würzburg, Germany, frederic.thiesse@uni-wuerzburg.de

Abstract

Retailers worldwide have started deploying smart service innovations in their stores to regain market share lost to online competitors. Against this backdrop, this paper focuses on the design of product recommendation systems for fashion stores. Our research particularly aims at answering the issues of whether and to what extent (i) the sensing capabilities of smart fashion retail environments and (ii) the integration of contextual information can improve the quality of such recommendations. To this end, we consider smart fitting rooms with the ability to detect products and customers as a showcase; a transaction dataset from a leading German fashion retailer; and contextual information about the time of purchase, the store type, and the weather conditions. Our preliminary analyses indicate that sensor information regarding garment and user identification, as well as further context data help to improve product recommendations in fashion stores.

Keywords: Smart Service Systems, Recommendation Systems, Context Awareness, Internet of Things, Retail Industry, Predictive Analytics, Cyberphysical Systems, Smart Fitting Rooms.

1 Introduction

Several scholars have called for more research in service-related areas over the last years (e.g., Böhmann et al., 2014; Frost and Lyons, 2017; Ostrom et al., 2010). One important aspect of service research are smart service systems (Frost and Lyons, 2017), which the National Science Foundation (2014) defines as “system[s] capable of learning, dynamic adaptation, and decision making based upon data received, transmitted, and/or processed to improve [their] response to a future situation.” One very promising “playground for service systems innovation” are cyberphysical systems, which are based on technology (e.g., sensing or communication capabilities) and allow for the integration of new sources of contextual information (e.g., location or social contexts). The design of such systems is, however, challenging because they have to bridge the boundaries between tangible and intangible resources (Böhmann et al., 2014) and need to be woven around legacy systems (Hauser et al., 2017b; Weiser, 1999). In addition, such systems should make use of the given contextual information to provide users with services that truly leverage the business value that arises from the integration of physical and virtual worlds.

The present paper is concerned with service systems in smart retail environments. Recent developments in the retail industry (e.g., the introduction of the “Amazon Go” store) point to a fundamental transformation
of traditional brick and mortar (B&M) stores into smart stores. Such stores leverage sensor technology and customer data to provide novel customer services (Gregory, 2015; Manyika et al., 2015). We consider product recommendation systems that rely on recommendation algorithms and various data sources (e.g., customer purchase histories or contextual information) to recommend products that suit individual customer preferences (Liang et al., 2006). However, in contrast to most research on recommendation systems, we do not focus on automated recommendations in e-commerce, but on systems for physical store environments. Our showcase are smart fitting rooms that offer garment recommendations on screens within individual cabins. Such systems help customers with ever-growing product ranges (Häubl and Trifts, 2000) and can lead to additional sales and increased customer loyalty (Schafer et al., 1999).

We are particularly interested in the issues of whether and to what extent (i) the sensing capabilities of smart fitting rooms (i.e., product or customer detection) and (ii) the integration of contextual information can improve the quality of such recommendations. To this end, we first review literature on recommendation systems for fashion stores focusing on (i) research that describes suitable recommendation algorithms (Section 2.1) and (ii) studies that leverage contextual information in fashion stores (Section 2.2). The review of suitable recommendation algorithms is crucial for our research, as the applicability of different algorithms depends on the sensing capabilities of smart fitting rooms. In a second step, we propose recommendation algorithms tailored to these sensing capabilities and describe a means by which to integrate contextual information (Section 3). In this context, we also investigate to what extent the choice of algorithm and the integration of selected context attributes affect recommendation quality.

2 Related Work

We conducted a comprehensive literature review on B&M recommendation systems following the suggestions put forward by Webster and Watson (2002). Google Scholar and Scopus were used as databases and searched by means of a search query consisting of two building blocks. On the one hand, “brick and mortar”, “offline”, “store”, “retail”, “stationary”, or “fitting room” had to appear in the titles of the papers. On the other hand, “recommendation system” (or commonly employed synonyms) had to appear in the papers’ full texts. We retrieved 397 papers on Google Scholar and 222 papers on Scopus, examined titles and abstracts, and discarded papers not relevant for our research. Most of the non-relevant papers either focus on the comparison of offline and online evaluations of e-commerce recommendation systems or are concerned with environments outside of the retail world (e.g., a tourist attraction recommendation system). We identified 37 relevant articles, which can roughly be categorized into

- discussions of economic potentials (Kamei et al., 2011; Al-Kassab et al., 2009; B. Keller et al., 2015; T. Keller and Raffelsieper, 2014; Kroon et al., 2007; Liaghat et al., 2013; Melià-Seguí et al., 2013; Pfeiffer et al., 2015; Pous et al., 2013; Thiesse et al., 2009),
- user acceptance studies (Daraghmi and Kadoori, 2016; Kowatsch and Maass, 2010a,b; Kowatsch et al., 2009; Y. E. Lee and Benbasat, 2010; Resatsch et al., 2008), and
- technology-focused papers (Chan and Capra, 2012; Cinicioglu and Shenoy, 2016; Giering, 2008; Hansen and Loos, 2008; Hu et al., 2016; Kamei et al., 2010; Kronberger and Affenzeller, 2011; S.-L. Lee, 2010; J. Li et al., 2014; Y.-M. Li et al., 2017; Liao et al., 2014; Luo et al., 2016; Nakahara and Yada, 2011; Pouloupolou and Kyriazis, 2017; Sano et al., 2015; Sato et al., 2015; Skaida et al., 2016; So and Yada, 2017; Walter et al., 2012; P. Wang et al., 2014; X. Zhang et al., 2016).

As per our research agenda, we are mainly interested in the technology-focused papers. To gain a more comprehensive set of such papers, we additionally conducted a forward and backward search based on the papers from all three research streams, which allowed us to identify an additional 19 relevant papers.

---

1 We used the following Google Scholar search string: (intitle:“brick & mortar” OR intitle:“B&M” OR intitle:“brick and mortar” OR intitle:“retail” OR intitle:“store” OR intitle:“offline” OR intitle:“stationary” OR intitle:“fitting room”) AND (“recommendation system” OR “recommendation agent”). For Scopus, “title” instead of “intitle” was used.
The 40 technology-focused papers investigate (i) sensing capabilities of smart fitting rooms, (ii) algorithms for B&M recommendation systems, and (iii) contextual attributes in B&M stores. We argue that for the effective deployment of smart fitting rooms, these issues must be addressed in this exact order. In order to leverage auto-ID technologies, the fitting rooms first must gather information about the environment. While some authors mention the value of identifying individual customers (Hansen and Loos, 2008), others put special emphasis on the identification of products (e.g., Hauser et al., 2017a). Secondly, recommendation algorithms that can process this information are needed as the corresponding assumption is that this helps generate better recommendations (Landmark and Sjøbakk, 2017). Finally, contextual information should also be used because users’ preferences may change due to situational circumstances.

2.1 Algorithms for B&M Recommendation Systems

Recommendation algorithms can be categorized into (i) content-based methods where similarities between item features are taken into account and (ii) collaborative-filtering approaches where product suggestions are based on the previous behavior of users with similar preferences (Adomavicius and Tuzhilin, 2005). Many papers in our article set rely on collaborative approaches (e.g., S.-L. Lee, 2010) and only a few apply content-based approaches (e.g., Wong et al., 2012). This might be because content-based filtering requires, by definition, additional information about the products (Pazzani and Billsus, 2007). This information is, however, often unavailable or does not contain valuable data with which to distinguish between items users like and dislike (Poulopoulos and Kyriazis, 2017). In the fast-changing fashion industry, the disadvantage of being dependent on static knowledge (e.g., fashion experts) is particularly severe (Landmark and Sjøbakk, 2017). In the following, we therefore focus on collaborative-filtering approaches.

In the course of the review, we identified (i) the user cold start problem and (ii) the absence of explicit product ratings as the main challenges to the design of B&M recommendation algorithms.

User cold start problem. This challenge describes the phenomenon that a system struggles to give good recommendations to users for whom only some or no information is available (Bobadilla et al., 2012). The latter case occurs if customers do not yet have a purchase history or cannot be identified by the service system. In e-commerce, users can be identified as long as they are logged in with their account while browsing. In addition, even if they are not logged in, information about their preferences can often be gained from additional data sources (e.g., click-stream data) (Bobadilla et al., 2012). In contrast, today’s B&M stores usually do not know the identity of the customers who are currently in their stores. In most cases, customer identification takes place at the checkout (e.g., through a customer card or e-payment). An earlier identification (e.g., in the fitting room) would require dedicated technical equipment. However, if such technical equipment is not available, the service systems must still be able to provide recommendations. To cope with this issue, Y.-M. Li et al. (2017) suggest using social media profiles to deduce customers’ preferences. This, however, presupposes that users are willing to provide access to their accounts. Another possibility for addressing the user cold start is the adoption of association rule mining algorithms for product recommendations (Lawrence et al., 2001; S.-L. Lee, 2010; Skiada et al., 2016; P. Wang et al., 2014). These algorithms do not require identified users, but rather attempt to derive generic rules for products purchased together (Sarwar et al., 2000). These rules enable recommendations based on the products with which customers are currently interacting (Shaw et al., 2010).

Absence of explicit product ratings. In contrast to the context of e-commerce, numerical ratings (e.g., 1-5 star scale) are particularly difficult to obtain in B&M stores (Hansen and Loos, 2008). It is thus necessary to deduce user preferences from data that only contains implicit feedback in the form of customers’ product purchases. Such data does not, however, contain negative product feedback, as not buying a product does not indicate dislike (Sahoo et al., 2012). As a result, the majority of collaborative algorithms...
that depend on explicit ratings cannot be used in cases where only purchase data is available (Rendle et al., 2009; Sahoo et al., 2012). To tackle this issue, several authors again fall back on association rule algorithms as these depend on co-purchases and not on the availability of ratings (Cinicioglu and Shenoy, 2016; Fang et al., 2012; Hsu et al., 2004; Jie et al., 2012; Kronberger and Affenzeller, 2011). A major drawback of this, however, is that the obtained recommendations usually lack personalization (P. Wang et al., 2014). For example, if a rule is decisive for recommending a specific article to a certain customer, but irrelevant for the rest of the data, it is desirable to assign this rule a high score for collaborative-filtering, but typically not for association rules (Verstrepen et al., 2017). To make association rules less generic and more customer-group-specific, Skiada et al. (2016) propose combining them with clustering approaches whereas P. Wang et al. (2014) try to capture associations using probabilistic models. Scholars that do not employ association analysis try to estimate user ratings by measuring physical distances between users and objects (Kawashima et al., 2006), considering customer movement paths (So and Yada, 2017), or using product prices and purchased quantities as additional information (Pouloupolous and Kyriazis, 2017). Another promising possibility is investigated by Sato et al. (2015), who use a ranking-based collaborative filtering approach designed for datasets that do not contain explicit product ratings (Rendle et al., 2009).

2.2 Contextual Information in B&M Stores

We identified several papers addressing the question of how contextual information can be acquired and subsequently incorporated into B&M recommendation services. Context-aware systems are able to adapt their operations to the current situation (e.g., location and time) and aim at increasing service usability and effectiveness (Baldauf et al., 2007). They have considerable potential (Villegas and Müller, 2010)—particularly in retailing (Grewal et al., 2017)—as additional environmental information influences the predictability of user preferences (Adomavicius and Tuzhilin, 2011). However, special attention must be paid to the selection of contextual attributes, as it is difficult to determine their relevance a priori, especially as the incorporation of irrelevant ones entails the risk of impairing recommendation quality (Adomavicius and Tuzhilin, 2011; Odić et al., 2013). While many papers investigate the integration of contextual data into e-commerce recommendation systems (e.g., Jiang et al., 2015; J. Lu et al., 2015), we found only a few that integrate contextual data into B&M recommendation systems. Five of them explicitly use context data for the generation of product recommendations and leverage information about trends and occasions (Wong et al., 2012), user locations (Nakahara and Yada, 2011; So and Yada, 2017), store locations (Giering, 2008), and customer interactions with products (activities) (Kawashima et al., 2006). In contrast, scholars who do not recommend products suggest points of interest (e.g., shops in a mall) or recommend paths that guide customers through stores (e.g., Y.-M. Li et al., 2017; E. H.-C. Lu et al., 2012; Z. Zheng et al., 2017) and therefore rely mainly on the locations of users.

For further investigation of relevant contextual attributes, we searched for papers examining contextual data in e-commerce recommendation systems. We again followed the approach of Webster and Watson (2002).² The search was restricted to fashion retail, as context variables are often domain-specific (Adomavicius and Tuzhilin, 2011). We retrieved 128 publications of which only 14 turned out to be relevant.

We analyzed the B&M and e-commerce papers and divided the context data used in them into categories (see Figure 1). Frequently mentioned ones are time (e.g., season), location (e.g., geolocation of IP address or in-store position), occasion, and weather. While most of the attributes can be applied in e-commerce and B&M, attributes of the categories activity (e.g., trying on garments) (Hauser et al., 2017a) and surroundings (e.g., shopping area is crowded) (Y.-M. Li et al., 2017) can only be collected in B&M stores. According to Adomavicius and Tuzhilin (2011), context can be used in two ways for generating recommendations: (i) as a search parameter to filter potentially recommendable items and (ii) for estimating user preferences in certain contextual situations by observing users while they are interacting with the

² We used the following Google Scholar search string: (intitle:fashion OR intitle:clothing) AND context AND (“recommender system” OR “recommendation system” OR “recommendation agent”). On Scopus “title” instead of “intitle” has to be used.
service system. The predominant example relying on the first paradigm is that of ubiquitous systems that use sensor information to recommend items in close proximity. All the papers from our B&M recommendation systems article set that consider collaborative-filtering approaches are based on the first paradigm. As we seek to employ context as auxiliary information to improve recommendations, we can not use the algorithms applied in these papers. In contrast, we turn to the second paradigm to gear our recommendations more towards situational user preferences. This paradigm proposes modeling the context-sensitive preferences of users and generating recommendations by adopting existing collaborative filtering methods to context-aware recommendation settings. Adomavicius and Tuzhilin (2011) identify two applicable groups of algorithms. Contextual modeling, on the one hand, aims at integrating context directly into the modeling technique and thus demands specialized algorithms. Contextual filtering, on the other hand, uses contextual data to adapt the underlying database prior to actual training (i.e., pre-filtering) or to adjust the outputted recommendations a posteriori (i.e., post-filtering) (Panniello et al., 2009).

![Figure 1. Contextual categories considered in B&M (left box) and e-commerce literature (right box)](image)

### 3 System Design and Evaluation

In this section, we first (i) propose recommendation algorithms tailored to the sensing capabilities of smart fitting rooms and (ii) describe a means by which to integrate contextual information. In this paper, we start with analyzing the contextual categories time, location, and weather. We consider this initial step crucial for assessing the importance of our project, as considering all contextual categories identified in the literature review would be pointless if context does not help to improve recommendations. In a second step, we investigate to what extent the choice of algorithm and the integration of the selected context attributes affect recommendation quality. For this preliminary evaluation, we use a pre-compiled transaction dataset from a leading German fashion retailer. Such offline assessments are common evaluation approaches in recommendation system research (Gunawardana and Shani, 2009). In our offline evaluation we follow the steps for building predictive models described by Shmueli and Koppius (2011): **Data collection.** The dataset contains 5,142,891 customer transactions, that is, information about when customers bought particular products in specific stores. The dataset contains transactions from a period of 16 months and comprises a total of 660,472 different customers and 52,902 different products. In addition, we collected historical weather data that includes the temperature and rainfall of the weather stations closest to the individual stores (Deutscher Wetterdienst, 2017). **Data preparation.** We carried out several pre-processing steps to enrich the transaction data with contextual information. We first derived the attributes season, first week (of the month) (see, e.g., Hastings and Washington, 2010), and day (weekday or Saturday) using the transaction data timestamps. We then added the attributes temperature (cold, cool, warm, or hot) and rain (yes or no). To achieve this, we first aggregated hourly measurements into daily averages and then subdivided these values into attribute categories. Finally, we derived the attribute store format (city, mall, or standalone).
Exploratory data analysis. The sales in our transaction data follow a long tail distribution, a phenomenon commonly observed in retailing (Brynjolfsson et al., 2011). 20% of all units sold are realized with only 3% of the available products, while the remaining sales are generated by the other 97% of the products. However, the recommendation of rarely-purchased articles can lead to especially interesting suggestions (Cremonesi et al., 2010). Figure 2 yields initial insights into the predictive power of the investigated contextual attributes. The individual heatmaps show the proportionate sales distributions for the ten best-selling product categories for the individual categories of the contextual attributes (e.g., yes and no for the attribute rain). The heatmaps indicate that some situational attributes could influence the customers’ buying decisions (e.g., women’s knitwear seems to be particularly seasonal and temperature-dependent).

<table>
<thead>
<tr>
<th>Season</th>
<th>Store format</th>
<th>First week</th>
<th>Day</th>
<th>Temperature</th>
<th>Rain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Men’s shirts</td>
<td>0.224</td>
<td>0.265</td>
<td>0.265</td>
<td>0.328</td>
<td>0.359</td>
</tr>
<tr>
<td>Men’s tops</td>
<td>0.236</td>
<td>0.335</td>
<td>0.197</td>
<td>0.232</td>
<td>0.368</td>
</tr>
<tr>
<td>Men’s underwear</td>
<td>0.199</td>
<td>0.244</td>
<td>0.264</td>
<td>0.303</td>
<td>0.296</td>
</tr>
<tr>
<td>Unisex underwear</td>
<td>0.256</td>
<td>0.270</td>
<td>0.230</td>
<td>0.244</td>
<td>0.366</td>
</tr>
<tr>
<td>Women’s dresses</td>
<td>0.269</td>
<td>0.292</td>
<td>0.200</td>
<td>0.293</td>
<td>0.368</td>
</tr>
<tr>
<td>Women’s knitwear</td>
<td>0.130</td>
<td>0.089</td>
<td>0.445</td>
<td>0.336</td>
<td>0.209</td>
</tr>
<tr>
<td>Women’s lingerie</td>
<td>0.225</td>
<td>0.246</td>
<td>0.248</td>
<td>0.281</td>
<td>0.326</td>
</tr>
<tr>
<td>Men’s shorts</td>
<td>0.286</td>
<td>0.265</td>
<td>0.216</td>
<td>0.235</td>
<td>0.373</td>
</tr>
<tr>
<td>Women’s tops</td>
<td>0.275</td>
<td>0.232</td>
<td>0.283</td>
<td>0.230</td>
<td>0.196</td>
</tr>
<tr>
<td>Women’s trousers</td>
<td>0.287</td>
<td>0.244</td>
<td>0.278</td>
<td>0.191</td>
<td>0.350</td>
</tr>
</tbody>
</table>

Figure 2. Purchase distribution per category subject to contextual situations

Choice of attributes. Although exploratory data analysis can indicate the predictive power of attributes, it is not suitable for excluding variables from prediction models a priori (Adomavicius and Tuzhilin, 2011; Baltrunas et al., 2012). For this reason, we follow Baltrunas et al. (2012) and first train models for each individual contextual attribute. In a second step, we consider only the contextual attributes that were used in models that actually led to improved recommendation quality (all except store format) and use them to build our final, context-aware recommendation system.

Choice of methods. Our review showed that two main challenges have to be tackled when transferring collaborative-filtering to cyberphysical contexts (i.e., user cold start and the absence of explicit product ratings). As outlined, we argue that the applicability of particular recommendation algorithms depends on the environment’s sensing capabilities and we aim to provide recommendations also in cases where such information is not available (e.g., because fitting rooms are not equipped with card readers or customers don’t have loyalty cards). Consequently, we distinguish between four different cases (see Figure 3):

- If neither customers nor products are identifiable, we use the algorithm popular item, a naive approach to recommend the most frequently-bought items in transaction datasets (Cremonesi et al., 2010).
- The identification of the garments customers bring into cabins enables us to tailor the recommendations to these items. Based on the findings from our first literature review, we suggest employing algorithms based on association rule mining (ARM) (they are used in 13 of the papers) and following particularly the approach of Sarwar et al. (2000). To this end, we first identify the products that were frequently purchased together. In a second step, we select the most interesting rules considering the evaluation metrics support and confidence. In the subsequent generation of product recommendations, these rules are used to suggest products that match products customers bring into fitting rooms.
- Targeting individual customers is only possible if they authenticate themselves (e.g., with loyalty cards). Once they are identified, algorithms can take into account individual customer purchase histories and must be able to cope with the lack of explicit product ratings. Algorithms that can handle implicit feedback can be classified as (i) prediction-based or (ii) ranking-based algorithms (Takács and Tikk, 2012). Similarly to Sato et al. (2015), we use Bayesian Probabilistic Ranking
an algorithm that falls into the second category and creates user-specific item rankings by sampling positive (i.e., items previously purchased by the customer) and negative items (i.e., items not previously purchased by the customer), as well as running pairwise comparisons (Rendle et al., 2009).

- If products and customers are detected, it is possible to use algorithms designed to incorporate both sets of information. Recently, researchers have been working on the development of (sequential) algorithms that make this possible (e.g., factorized personalized Markov chains) (Rendle, 2010).

After selecting an algorithm tailored to the hardware, the recommendations can be enriched with contextual information. We rely on contextual pre-filtering, which allows for the subsequent use of any recommendation algorithm (see Section 2.2). Specifically, we used item splitting (see, e.g., Baltrunas and Ricci, 2014) for each of the implemented algorithms. Y. Zheng et al. (2014) describe it as one of the most efficient pre-filtering techniques. Item splitting identifies relevant contextual situations for purchasing specific products. If an item is purchased more often in a particular context, the algorithm considers that item to be two different items. We used the chi-square test to split only items whose sales distributions differed to a statistically significant degree depending on the context (Baltrunas and Ricci, 2014).

Figure 3. B&M recommendation algorithms depending on environment's sensing capabilities

**Evaluation.** In this paper, we consider popular item, ARM, and BPR. We first split the transaction dataset into two sets. 20% of the customers are randomly selected to be the test customers, while the remaining 80% of the customers are used for model training. To compare the performance of the proposed algorithms, we use the frequently adopted (e.g., Herlocker et al., 1999) leave-k-out evaluation proposed by Breese et al. (1998). Following this approach, the purchases of each test customer must be divided into two different sets. The first set contains 20% of each customer’s purchases (i.e., the product purchases the algorithms attempt to predict) and is withheld. The algorithms’ objectives is to provide a ranked list of multiple (N = 1, 2, ..., 10) product recommendations for each customer. During this process, the remaining 80% of each customer’s purchases are used differently depending on the algorithm. For algorithms applied in cases with identified users, the 80% (i.e., complete customer history) is used to deduce preferences. For algorithms applied in cases with identified products, on the other hand, each product in the 80% is used separately to derive item-based recommendations. In the simplest case (i.e., no identification of customers or products) the 80% is not necessary for the evaluation. Using the withheld 20% allows for a comparison of the algorithms’ suggestions with the products that customers actually bought. We rely on the common metrics precision, recall, and F1 score (Cremonesi et al., 2010; Herlocker et al., 2004), to measure the frequency with which a recommender system makes correct or incorrect predictions about whether an item is relevant for a user or not. Precision measures the ratio of relevant items (i.e., hits = intersection of recommended and bought) to the number (N) of recommended items for a certain customer and therefore reflects the probability that a recommended product is relevant (P@N = hits / N). Recall measures the ratio of products that were actually bought by and recommended to a customer (i.e., hits) in relation to this customer’s total number (|T|) of products in the withheld 20% (R@N = hits / |T|). Therefore, recall represents the probability that a relevant product will be recommended. Precision and recall are contradictory metrics: with increasing N, the precision decreases while the recall increases (Sarwar et al., 2000). The F1 score combines both measures into a single score by using the harmonic mean. These measures are first calculated for each test customer and then averaged. Precision and recall depend heavily on the
number of items purchased per user and should therefore not be interpreted as absolute measures, but should only be used to compare algorithms (Cremonesi et al., 2008). Figure 4 presents the results for the considered algorithms with and without incorporated contextual information (i.e., season, first week, day, temperature, and rain). The results show that garment and user identification are valuable, as using ARM and BPR increases recommendation quality. In addition, the integration of context leads to improved results, but only for those algorithms that consider product or customer information. In the case of ARM, F1 improves by 24% (averaged over \( N = 1, 2, \ldots, 10 \)). In the case of BPR, context integration leads to an even stronger F1 improvement of 34% (again averaged over \( N = 1, 2, \ldots, 10 \)).

![Figure 4. Preliminary evaluation for different algorithms with and without context information](image)

### 4 Expected Contribution and Future Work

The present study is concerned with the design of product recommendation systems for fashion stores. We found that (i) the applicability of different recommendation algorithms depends on the sensing capabilities of smart retail environments and (ii) the implementation of recommendation systems in the physical world allows for the integration of additional contextual information (i.e., consideration of customer activities and information about the shopping area surroundings). Based on these findings, we propose a framework of suitable recommendation algorithms and describe a means by which to integrate context. In our preliminary evaluation, we consider smart fitting rooms (IT artifacts that offer product recommendations on screens within individual cabins); contextual information about purchase times, store types, as well as weather conditions; and a transaction dataset from a leading German fashion retailer. The evaluation shows that the ability to identify garments and users in smart fitting room cabins enables the product recommendation system to generate better recommendations. A further improvement of recommendation quality can be achieved through the integration of the considered contextual information.

Going forward, we want to conduct a more comprehensive evaluation considering all contextual categories (in particular activity and surroundings) and additional recommendation algorithms (in particular FPMC) which promises to yield further recommendation improvements. The collection of such data might require additional sensor systems (e.g., additional RFID systems or camera systems). In addition, the extraction of necessary relevant information (e.g., activities of users) from low-level sensor data streams might necessitate the application of complex data cleansing processes (see, e.g., H. Li et al., 2015). Secondly, we consider evaluations based on pre-compiled datasets only as a first step towards a comprehensive evaluation of product recommendation systems (Knijnenburg et al., 2012). To carry out a user-centric evaluation in a real-world environment, the retailer we are collaborating with has already installed several RFID-based smart fitting rooms. The data collected by this smart service infrastructure enables one to determine which products customers bring into individual fitting room cabins and which they end up buying. This allows for a detailed analysis of cause-and-effect relationships between particular product recommendations and purchase decisions. Moreover, when aiming for the adoption of recommendation services, there are a number of other important aspects that have to be considered. To this end, we are planning on conducting evaluations similar to those proposed by Pu et al. (2011) and Weinhard et al. (2017) to identify the determinants that motivate users to adopt such technologies.
References


Li et al. / Fashion Store Product Recommendation Systems


