University of Amsterdam
The IRLab

Understanding Multi-channel Customer Behavior in Retail

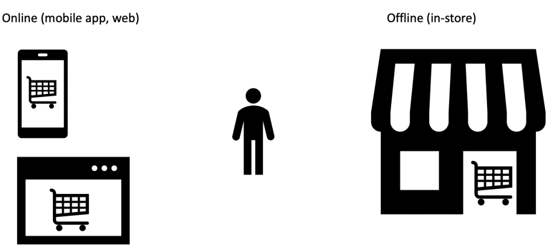
Mozhdeh Ariannezhad
Maarten de Rijke

February 25, 2022

Talk based on joint work with **Mozhdeh Ariannezhad**, Sami Jullien, Pim Nauts, Min Fang, Sebastian Schelter

Background

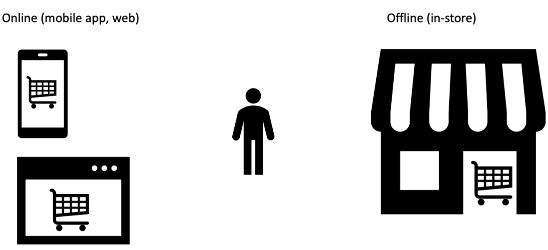
Multi-channel shopping



Ongoing development sped up by pandemic

- Complementing **offline** (in-store) shopping with **online** shopping

Multi-channel shopping



Ongoing development sped up by pandemic

- Complementing **offline** (in-store) shopping with **online** shopping
- Multiple shopping channels
  - Purchase items across all
  - Leverage one channel to prepare for purchase in another
  - Richer data for personalizing shopping experience

Understanding customer behavior

- Basis for downstream machine learning tasks
  - Product recommendation, purchase prediction, ...

Understanding customer behavior

- Basis for downstream machine learning tasks
  - Product recommendation, purchase prediction, …
- Rich body of previous work studying user behavior in online shopping
  - Click stream data, transaction data, digital receipts extracted from email, transactions logs of banks, product search engine log, …
  - Single channel only

Understanding customer behavior

- Basis for downstream machine learning tasks
  - Product recommendation, purchase prediction, …
- Rich body of previous work studying user behavior in online shopping
  - Click stream data, transaction data, digital receipts extracted from email, transactions logs of banks, product search engine log, …
  - Single channel only
- Relatively little known about multi-channel customer behavior
  - Mostly in marketing and retail research (interviews, customer surveys, …)
  - Transaction analysis based on (Ariannezhad et al., 2021)

Agenda

- Background
- Multi-channel customer behavior in retail
- Next basket recommendation
- Wrap-up

Multi-channel customer behavior in retail

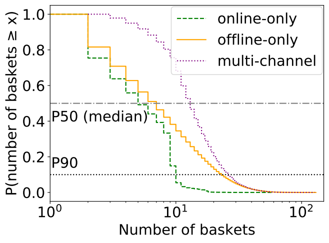
Data and terminology

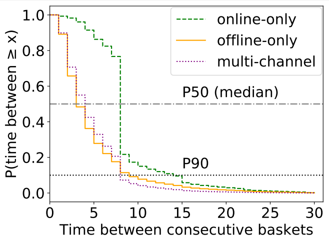
- Sample of 2.8M transactions during an 8 week period from 300,000 customers from a European food retailer, with physical stores and online platforms
- Three groups of customer
  - Online-only
  - Offline-only
  - Multi-channel
- Same product catalog for in-store and online
- Track customers across offline and online channels through their loyalty card
- Next: compare shopping behavior of these groups, then use insights gained

Extraction of transaction data

- Only customers with a loyalty card, so that we are able to track customers within and across channels
- Sample 100,000 customers at random from each group during period of interest and extract their transactions
- Each transaction marked as either online or offline: represents a basket

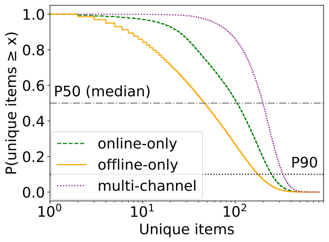
Comparison of groups: purchase frequency

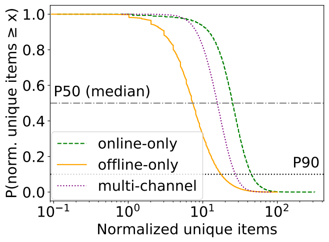




- Multi-channel customers have largest number of baskets, online only smallest
- Offline only and multi-channel customers similar behavior w.r.t. shopping times: median of 3 days between consecutive baskets, 90% percentile of 7 days
- Median for online only 7 days, partly because of possibility of selecting fixed delivery day for a series of online baskets

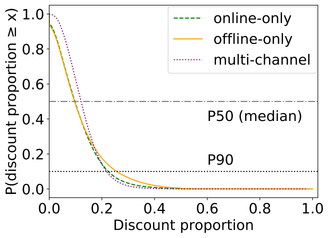
Comparison of groups: quantity and variety of items

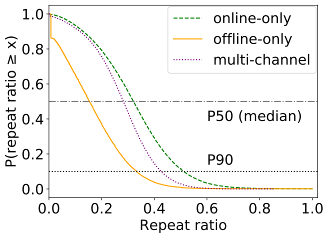




- Normalized unique items: number of unique items / number of baskets
- Offline-only customers smallest number of unique items and normalized items; multi-channel highest unique items; online-only more unique items when number of baskets is considered
- Due to (online) access to full catalog plus ease of delivery

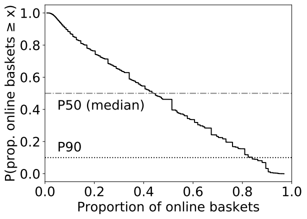
Comparison of groups: promotions & repeats

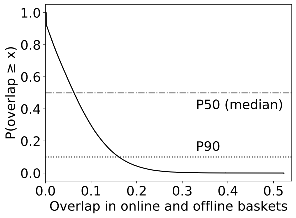




- Discount proportion (amount of discount per basket / basket value) similar for different groups
- Repeat ratio (number of unique items / total number of items purchased across baskets) highest for online-only

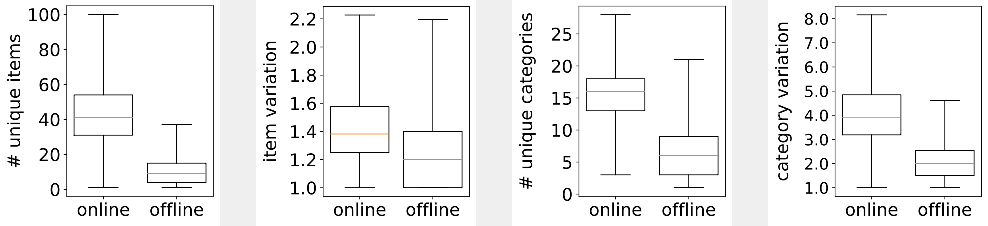
Multi-channel customers: online or offline?





- For about half of multi-channel customers, online baskets are dominant; for other half, offline baskets
- Overlap between online and offline customers is minimal for most customers: different channels serve different buying needs
- Offline shopping peaks during Fridays and Saturdays, online shopping uniformly distributed across week days
- 20% have 0.5 channel switch probability (40%: $<$ 0.5, 10%: 100%)

Online vs offline baskets



- Item variation: basket size / # unique items ; category variation: basket size / # unique item categories
- Online baskets have more unique items and categories, not much difference in variation → difference in unique items and categories mostly due to bigger basket size

Next basket recommendation

HYPOTHESIS: CHANNEL MATTERS

- Seen: Differences in behavior between different groups / different channels

Hypothesis: Channel matters

- Seen: Differences in behavior between different groups / different channels
- Hypothesis: Taking channel into account helps predictive performance of downstream tasks

Hypothesis: Channel matters

- **Seen**: Differences in behavior between different groups / different channels
- **Hypothesis**: Taking channel into account helps predictive performance of downstream tasks
- **Case study**: Next basket recommendation
  - *Predict the set of items that a customer will purchase in their next basket, given their purchase history*

- Given history of baskets for customer u, defined as $B^u = \{B_1^u, B_2^u, \ldots, B_n^u\}$, where B_i^u is a basket of items defined as $B_i^u = \{x_1, x_2, \ldots, x_t\}$, and $x_i \in X$ denotes an item from catalog X, the goal is to predict the items in the next basket of the customer, i.e., B_{n+1}^u.

Next basket recommendation

- Given history of baskets for customer u, defined as $B^u = \{B^u_1, B^u_2, \ldots, B^u_n\}$, where B^u_i is a basket of items defined as $B^u_i = \{x_1, x_2, \ldots, x_t\}$, and $x_i \in X$ denotes an item from catalog X, the goal is to predict the items in the next basket of the customer, i.e., B^u_{n+1}.
- For the basket history B^u, the recommendation model assigns a score to all items $x_i \in X$, and the top-k items are returned as the candidate items for the next basket recommendation

Next basket recommendation

- Given history of baskets for customer u, defined as $B^u = \{B^u_1, B^u_2, \ldots, B^u_n\}$, where B^u_i is a basket of items defined as $B^u_i = \{x_1, x_2, \ldots, x_t\}$, and $x_i \in X$ denotes an item from catalog X, the goal is to predict the items in the next basket of the customer, i.e., B^u_{n+1}.
- For the basket history B^u, the recommendation model assigns a score to all items $x_i \in X$, and the top-k items are returned as the candidate items for the next basket recommendation
- *How effective is next basket recommendation for different types of customer (offline, online, multi-channel)?*

What next basket recommendation method(s) to consider?

- Lessons from Li et al. (2021)
  - Comparison of three families of NBR methods …
    - Frequency-based
    - Nearest neighbor-based
    - Deep learning-based
  - … on three datasets
    - TaFeng
    - Dunnhumby
    - Instacart

What next basket recommendation method(s) to consider?

Dataset		TaFeng			Dunnhumby			Instacart		
Size	Methods	Recall	NDCG	PHR	Recall	NDCG	PHR	Recall	NDCG	PHR
10	G-TopFreq	0.0803	0.0842	0.2489	0.0987	0.1054	0.4624	0.0721	0.0820	0.4543
	P-TopFreq	0.1072	0.0959	0.3487	0.2319	0.2342	0.6569	0.3264	0.3381	0.8437
	GP-TopFreq	0.1215	0.1019	0.3706	0.2356	0.2360	0.6660	0.3273	0.3387	0.8451
	UP-CF@r	0.1257	0.1110	0.3996	0.2429	0.2471	0.6761	0.3511	0.3634	0.8642
	TIFUKNN	0.1259	0.1020	0.3871	0.2398	0.2411	0.6774	0.3608*	0.3726*	0.8640
	Dream	0.1143	0.1030	0.2991	0.0974	0.1049	0.4639	0.0722	0.0818	0.4560
	Beacon	0.1180	0.1075	0.3006	0.0991	0.1055	0.4655	0.0724	0.0820	0.4575
	CLEA	0.1184	0.1048	0.3083	0.1548	0.1726	0.5533	0.1227	0.1444	0.5633
	Sets2Sets	0.1360	0.1132	0.4104	0.1708	0.1491	0.5854	0.2125	0.1923	0.7185
	DNNTSP	0.1537*	0.1321*	0.4487*	0.2388	0.2409	0.6771	0.3337	0.3401	0.8498

What next basket recommendation method(s) to consider?

- No state-of-the-art NBR method, deep learning-based, consistently shows best performance across datasets
- Compared to a simple frequency-based baseline, improvements of SOTA methods are modest or even absent
- Clear difficulty gap and trade-off between repeat task and explore task
- Deep learning-based methods do not effectively exploit repeat behavior; they achieve relatively good explore performance

What next basket recommendation method(s) to consider?

- No state-of-the-art NBR method, deep learning-based, consistently shows best performance across datasets
- Compared to a simple frequency-based baseline, improvements of SOTA methods are modest or even absent
- Clear difficulty gap and trade-off between repeat task and explore task
- Deep learning-based methods do not effectively exploit repeat behavior; they achieve relatively good explore performance
- So? Choose P-TopFreq (personal top frequency)

Next basket recommendation with P-TopFreq

Prediction target	$k = 10$			$k = 20$			$k = 50$		
	Recall	nDCG	PHR	Recall	nDCG	PHR	Recall	nDCG	PHR
Online-only customers	0.1582	0.5873	0.9743	0.2459	0.4993	0.9882	0.3988	0.4564	0.9939
Offline-only customers	0.1773	0.2716	0.7331	0.2435	0.2664	0.7998	0.3448	0.2951	0.8664
Multi-channel customers	0.1282	0.3696	0.7688	0.1950	0.3292	0.8242	0.3085	0.3124	0.8838
Multi-channel customers, online target basket	0.1431	0.5946	0.9816	0.2265	0.5068	0.9916	0.3677	0.4372	0.9968
Multi-channel customers, offline target basket	0.1163	0.1891	0.5981	0.1697	0.1867	0.6899	0.2609	0.2123	0.7931
Multi-channel customers, target channel known	0.1373	0.3808	0.7882	0.2027	0.3387	0.8369	0.3095	0.3191	0.8846

- Strong performance of a simple method

Next basket recommendation with P-TopFreq

Prediction target	$k = 10$			$k = 20$			$k = 50$		
	Recall	nDCG	PHR	Recall	nDCG	PHR	Recall	nDCG	PHR
Online-only customers	0.1582	0.5873	0.9743	0.2459	0.4993	0.9882	0.3988	0.4564	0.9939
Offline-only customers	0.1773	0.2716	0.7331	0.2435	0.2664	0.7998	0.3448	0.2951	0.8664
Multi-channel customers	0.1282	0.3696	0.7688	0.1950	0.3292	0.8242	0.3085	0.3124	0.8838
Multi-channel customers, online target basket	0.1431	0.5946	0.9816	0.2265	0.5068	0.9916	0.3677	0.4372	0.9968
Multi-channel customers, offline target basket	0.1163	0.1891	0.5981	0.1697	0.1867	0.6899	0.2609	0.2123	0.7931
Multi-channel customers, target channel known	0.1373	0.3808	0.7882	0.2027	0.3387	0.8369	0.3095	0.3191	0.8846

- Strong performance of a simple method
- Online-only $>$ Multi-channel $>$ Offline-only – correlates with repeat ratio

Next basket recommendation with P-TopFreq

Prediction target	$k = 10$			$k = 20$			$k = 50$		
	Recall	nDCG	PHR	Recall	nDCG	PHR	Recall	nDCG	PHR
Online-only customers	0.1582	0.5873	0.9743	0.2459	0.4993	0.9882	0.3988	0.4564	0.9939
Offline-only customers	0.1773	0.2716	0.7331	0.2435	0.2664	0.7998	0.3448	0.2951	0.8664
Multi-channel customers	0.1282	0.3696	0.7688	0.1950	0.3292	0.8242	0.3085	0.3124	0.8838
Multi-channel customers, online target basket	0.1431	0.5946	0.9816	0.2265	0.5068	0.9916	0.3677	0.4372	0.9968
Multi-channel customers, offline target basket	0.1163	0.1891	0.5981	0.1697	0.1867	0.6899	0.2609	0.2123	0.7931
Multi-channel customers, target channel known	0.1373	0.3808	0.7882	0.2027	0.3387	0.8369	0.3095	0.3191	0.8846

- Strong performance of a simple method
- Online-only > Multi-channel > Offline-only – correlates with repeat ratio
- For multi-channel: online basket > offline basket; knowing about online behavior does not help offline basket prediction

Next basket recommendation with P-TopFreq

Prediction target	$k = 10$			$k = 20$			$k = 50$		
	Recall	nDCG	PHR	Recall	nDCG	PHR	Recall	nDCG	PHR
Online-only customers	0.1582	0.5873	0.9743	0.2459	0.4993	0.9882	0.3988	0.4564	0.9939
Offline-only customers	0.1773	0.2716	0.7331	0.2435	0.2664	0.7998	0.3448	0.2951	0.8664
Multi-channel customers	0.1282	0.3696	0.7688	0.1950	0.3292	0.8242	0.3085	0.3124	0.8838
Multi-channel customers, online target basket	0.1431	0.5946	0.9816	0.2265	0.5068	0.9916	0.3677	0.4372	0.9968
Multi-channel customers, offline target basket	0.1163	0.1891	0.5981	0.1697	0.1867	0.6899	0.2609	0.2123	0.7931
Multi-channel customers, target channel known	0.1373	0.3808	0.7882	0.2027	0.3387	0.8369	0.3095	0.3191	0.8846

- Strong performance of a simple method
- Online-only $>$ Multi-channel $>$ Offline-only – correlates with repeat ratio
- For multi-channel: online basket $>$ offline basket; knowing about online behavior does not help offline basket prediction
- Oracle: knowing the target channel helps improve NBR, even for P-TopFreq

Wrapping up

What have we done?

- First (transaction log based) study of customer behavior in multi-channel setting in retail
- Sample of 2.8M transactions from 300,000 customers of food retailer
- Differences in behavior across online and offline – basket size, repeat ratio
- Performance of downstream prediction task (next basket recommendation) using these insights
- Different performance levels for different types of customer (online, offline, multichannel)

What should we do next?

- Address target channel prediction
- Address explore item prediction
- Investigate treatment effect – users who repeat more, benefit more from effective NBR methods
- In all of this, take characteristics of channel / multi-channel on board for prediction tasks

Two papers

Understanding Multi-Channel Customer Behavior in Retail

Mozhdeh Ariannezhad[1] Sami Jullien[1] Pim Nauts[2] Min Fang[2]
Sebastian Schelter[3] Maarten de Rijke[3]
[1]AIRLab, University of Amsterdam [2]Albert Heijn [3]University of Amsterdam
{m.ariannezhad,s.jullien,s.schelter,m.derijke}@uva.nl,{pim.nauts,min.fang}@ah.nl

ABSTRACT

Online shopping is gaining popularity. Traditional retailers with physical stores adjust to this trend by allowing their customers to shop online as well as offline, in-store. Increasingly, customers can browse and purchase products across multiple shopping channels. Understanding how customer behavior relates to the availability of multiple shopping channels is an important prerequisite for many downstream machine learning tasks, such as recommendation and purchase prediction. However, previous work in this domain is limited to analyzing single-channel behavior only.

In this paper, we provide the first insights into multi-channel customer behavior in retail based on a large sample of 2.8 million transactions originating from 300,000 customers of a food retailer in Europe. Our analysis reveals significant differences in customer behavior across online and offline channels, for example with respect to the repeat ratio of item purchases and basket size. Based on these findings, we investigate the performance of a next basket recommendation model under multi-channel settings. We find that the recommendation performance differs significantly for customers based on their choice of shopping channel, which strongly indicates that future research on recommenders in this area should take into account the particular characteristics of multi-channel retail shopping.

CCS CONCEPTS

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

KEYWORDS

Customer behavior, next basket recommendation

ACM Reference Format:
Mozhdeh Ariannezhad, Sami Jullien, Pim Nauts, Min Fang, Sebastian Schelter, Maarten de Rijke. 2021. Understanding Multi-Channel Customer Behavior in Retail. In *Proceedings of the 30th ACM International Conference on Information and Knowledge Management (CIKM '21), November 1–5, 2021, Virtual Event, QLD, Australia.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3459637.3482208

ACM ISBN 978-1-4503-8446-9/21/11…$15.00
https://doi.org/10.1145/3459637.3482208

1 INTRODUCTION

The emergence of e-commerce in recent years has encouraged retailers with physical stores to provide the possibility of online shopping for their customers in addition to in-store (offline) shopping. Customers do not only buy goods via multiple shopping channels, but they can also leverage the online channel for exploring the product inventory, comparing products, and saving products for later purchases, before shopping offline. In addition, the data generated via online shopping provides a further opportunity for personalizing the shopping experience through recommendation [1, 11].

The ease of use of online shopping translates into a rapidly growing market share of e-commerce solutions, even in industries such as fashion. This growth is not cannibalistic – most customers who purchase goods online also purchase in-store (offline) [2]. The addition of an online channel does not isolate the offline channel. Instead, it creates a multi-channel shopping experience for customers in retail sectors like grocery, cosmetics, and apparel.

Understanding customer behavior in retail serves as a basis for many downstream machine learnings tasks, such as recommending products and predicting purchases. While there are numerous studies examining user behavior in online shopping platforms, little is known about multi-channel customer behavior. Previous work relies mainly on click stream data [3, 5, 10, 13, 17, 18, 21]. Other sources of customer behavior include transaction data [15], digital receipts of online purchases extracted from emails [9], transaction logs of a bank [16], or search logs of a commercial product search engine [12]. However, these studies utilize data from a single shopping channel only, and do not explore multi-channel customer behavior. So far, multi-channel customer behavior has mostly been studied in the marketing and retail research literature [1, 2, 4, 7, 8]. These studies rely on perceptions gathered via interviews and customer surveys, in order to model, e.g., lock-in effects or physical store surface needs. Yet, perceptions are often different from actions, and these works do not consider actual transaction data from customers.

In this paper, we provide the first study on multi-channel customer behavior in retail. Based on a sample of 2.8 million transactions from 300,000 customers, gathered from a food retailer with multiple physical stores and two online platforms, we provide a first picture of customer behavior in a multi-channel retail setting. To this end, we group the customers into three groups, namely *online-only*, *offline-only*, and *multi-channel* customers, based on their choice of shopping channels (Section 2). We first compare the shopping behavior of these customer groups in Section 3. We find that the tendency to purchase previously bought products, defined as repeat behavior ratio, is higher for online-only customers. Zooming in on multi-channel customers, our analysis reveals that there is little overlap in online and offline baskets of multi-channel customers; they use each channel for different sets of items. Our

A Next Basket Recommendation Reality Check

Ming Li
University of Amsterdam
Amsterdam, The Netherlands
m.li@uva.nl

Sami Jullien
AIRLab, University of Amsterdam
Amsterdam, The Netherlands
s.jullien@uva.nl

Mozhdeh Ariannezhad
AIRLab, University of Amsterdam
Amsterdam, The Netherlands
m.ariannezhad@uva.nl

Maarten de Rijke
University of Amsterdam
Amsterdam, The Netherlands
m.derijke@uva.nl

ABSTRACT

The goal of a next basket recommendation system is to recommend items for the next basket for a user, based on the sequence of their prior baskets. We examine whether the performance gains of deep learning-based next basket recommendation (NBR) methods reported in the literature hold up under a fair and comprehensive comparison. To clarify the mixed picture that emerges from our comparison, we provide a novel angle on the evaluation of NBR methods, centered on the distinction between repetition and exploration: the next basket is typically composed of previously consumed items (i.e., repeat items) and new items (i.e, explore items). We propose a set of metrics that measure the repeat/explore ratio and performance of NBR models. Using these new metrics, we provide a second analysis of state-of-the-art NBR models. The results help to clarify the extent of the actual progress achieved by existing NBR methods as well as the underlying reasons for any improvements that we observe. Overall, our work sheds light on the evaluation problem of NBR, provides a new evaluation protocol, and yields useful insights for the design of models for this task.

CCS CONCEPTS

• **General and reference** → Evaluation; • **Information systems** → **Recommender systems.**

KEYWORDS

Next basket recommendation; Reproducibility; Repeat behavior

ACM Reference Format:
Ming Li, Sami Jullien, Mozhdeh Ariannezhad, and Maarten de Rijke. 2022. A Next Basket Recommendation Reality Check. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '22), July 11–15, 2022, Madrid, Spain.* ACM, New York, NY, USA, 10 pages.

1 INTRODUCTION

Over the years, next basket recommendation (NBR) has received a considerable amount of interest from the research community [3, 24, 32]. Baskets, or sets of items that are purchased or consumed together, are pervasive in many real-world services, with e-commerce and grocery shopping as prominent examples [11, 23]. Given a sequence of baskets that a user has purchased or consumed in the past, the goal of a NBR system is to generate the basket of items that the user would like to purchase or consume next. Within a basket, items have no temporal order and are equally important. A key difference between NBR and session-based or sequential item recommendations is that NBR systems need to deal with multiple items in one set. Therefore, models designed for item-based recommendation are not fit for basket-based recommendation, and dedicated NBR methods have been proposed [13, 25].

Deep learning-based recommendation methods. Numerous deep learning techniques have been developed to address sequential item recommendation problems, building on the well-documented capacity of deep learning-based methods to capture hidden relations and automatically learn representations [4]. Against this background, recent years have witnessed many proposals to address different aspects of the NBR task with deep learning-based methods, e.g., item-to-item relations [18], cross-basket relations [33], and noise within a basket [22].

Recent analyses indicate that deep learning-based approaches may not be the best performing approaches for all recommendation tasks and under all conditions [15]. For the task of generating a personalized ranked lists of items, linear models and nearest neighbor-based approaches outperform deep learning-based methods [7]. For sequential recommendation problems, deep learning-based methods may be outperformed by simple nearest neighbor or graph-based baselines [8]. What about the task of next basket recommendation? Here, the unit of retrieval — a basket – is more complex than in the recommendation scenarios considered in [7, 8, 15], with complex dependencies between items and baskets, across time, thus creating a potential for sophisticated representation learning-based approaches to NBR to yield performance gains. In this paper, we take a closer look at the field to see if this is actually true.

A new analysis perspective. We find important gaps and flaws in the literature on NBR. These include weak or missing baselines, the use of different datasets in different papers, and of non-standard metrics. We evaluate the performance of three families of state-of-the-art NBR models (frequency-based, nearest neighbor-based, and deep learning-based), on three benchmark datasets, and find that no NBR method consistently outperforms all other methods across all settings.

M. Ariannezhad, S. Jullien, P. Nauts, M. Fang, S. Schelter, and M. de Rijke. Understanding multi-channel customer behavior in retail. In *CIKM 2021: 30th ACM International Conference on Information and Knowledge Management*. ACM, November 2021.

M. Li, S. Jullien, M. Ariannezhad, and M. de Rijke. A next basket recommendation reality check. *arXiv preprint arXiv:2109.14233*, September 2021. URL `https://arxiv.org/pdf/2109.14233`.