Characterizing and Predicting Repeat Food Consumption Behavior for Just-in-Time Interventions

Yue Liu
Singapore Management University
Singapore, Singapore
yueliu@smu.edu.sg

Helena Lee
Singapore Management University
Singapore, Singapore
helenalee@smu.edu.sg

Palakorn Achananuparp
Singapore Management University
Singapore, Singapore
palakorna@smu.edu.sg

Ee-Peng Lim
Singapore Management University
Singapore, Singapore
eplim@smu.edu.sg

Tzu-Ling Cheng
National Taiwan University
Taipei, Taiwan
nancy.cheng.tl@gmail.com

Shou-De Lin
National Taiwan University
Taipei, Taiwan
sdlin@csie.ntu.edu.tw

ABSTRACT

Human beings are creatures of habit. In their daily life, people tend to repeatedly consume similar types of food items over several days and occasionally switch to consuming different types of items when the consumptions become overly monotonous. However, the novel and repeat consumption behaviors have not been studied in food recommendation research. More importantly, the ability to predict daily eating habits of individuals is crucial to improve the effectiveness of food recommender systems in facilitating healthy lifestyle change. In this study, we analyze the patterns of repeat food consumptions using large-scale consumption data from a popular online fitness community called MyFitnessPal (MFP), conduct an offline evaluation of various state-of-the-art algorithms in predicting the next-day food consumption, and analyze their performance across different demographic groups and contexts. The experiment results show that algorithms incorporating the exploration-and-exploitation and temporal dynamics are more effective in the next-day recommendation task than most state-of-the-art algorithms.

CCS CONCEPTS

• **Information systems** → **Data mining**; • **Applied computing** → Consumer health.

KEYWORDS

Food Recommendation; Implicit Feedback; Repeat Consumption

1 INTRODUCTION

Habitual consumption of balanced and healthful diets is known to positively correlate with long-term physical well-being of individuals [2]. Yet, a vast number of people, including members of online weight-loss community [2], did not tend to eat as healthy as they should on a daily basis. Personalized digital health interventions could play a crucial role in helping individuals develop and maintain healthy eating habits. The emerging paradigm of just-in-time health interventions [29] requires computational models capable of adapting to varying needs of individuals and changing context. Due to their ability to computationally model user preference from past data, food recommender systems can serve as an effective facilitator of the just-in-time healthy eating interventions through the personalized recommendation of healthy food items which are tailored to the individuals' tastes and dietary preferences [12, 32].

In this study, we identify three main research gaps in current food recommendation research pertaining to its ability to capture the individuals' day-to-day food consumption patterns, which is crucial to its effectiveness in the just-in-time interventions.

Firstly, existing food recommender systems have been largely focused on the *utility* or the *novelty* aspect of the recommendations, i.e., the effectiveness of the system is often measured by the user satisfaction of new recommended food items, whereas the *repetitiveness* or the *habitual* aspect of food consumption behavior has so far been under-explored. As a creature of habit, many of our consumption behaviors, including food consumption, exhibit both the *novelty-seeking* [13] and the *repetitive* [35] characteristics. The dynamics of novel and repeat consumption behaviors has also been modeled as the exploration (novel) and exploitation (repeat) phenomenon [4, 19, 22]. By taking into account this nature of food consumption behavior, the systems can gain a better understanding of the users' eating habits in various contexts and improve the effectiveness of the recommendations.

Secondly, the self-report food consumption data as a form of implicit feedback have not been extensively investigated in food recommendation research. Several food recommender models learn the general user preference from the past user-item rating data on cooking recipes [32]. However, in the just-in-time health intervention scenario, this form of explicit feedback may be difficult to obtain since they are likely to impose significant burden on the users having to continually evaluate their own satisfaction of a

large amount of food items on a daily basis, many of which have been previously consumed. Given the limitations of the current food logging methods, it is important to understand the effectiveness of implicit-feedback food recommender systems where only the past user-item consumption data are available.

Lastly, research in general recommender systems has just begun to study the fairness in recommendation issue [10]. Similarly, algorithmic bias is a potential problem for food recommender systems and just-in-time health interventions. Users from diverse demographic backgrounds and behaviors may not receive the same benefits from the food recommender algorithms due to the potential uneven distribution of effectiveness across different groups. In the worst case scenario, inaccurate recommendations may negatively affect the users' long-term health. To our knowledge, this issue has remained unexplored in the evaluation of food recommender systems.

Therefore, we aim to address these gaps in prior work in this study. Specifically, we formulate the following research questions (RQs):

RQ1: *What are the overall characteristics of repeat food consumption behavior? Additionally, how do repeat consumption patterns differ across diverse contexts, such as meal occasions, temporal lifestyle factors, and demographic groups?*

RQ2: *What is the effectiveness of different state-of-the-art implicit recommender systems in predicting daily food consumption patterns of individuals in a just-in-time recommendation setting?*

RQ3: *To what extent does the state-of-the-art food recommender system exhibit an algorithmic bias when generating recommendations for diverse groups of users and eating contexts?*

Research Contributions: In pursuing those questions, we made two major research contributions which are summarized as follows. Firstly, we conducted a quantitative study to thoroughly examine the repeat food consumption behavior of individuals across meal occasions, temporal factors, and demographic factors in a large population of nearly 8K MFP users consisting of 2.7M daily food consumption data over 6 months. To our knowledge, no prior studies have examined these types of food consumption behaviors using self-report food consumption data. Findings from the analysis help establish a better understanding of the food consumption behaviors of the heterogeneous groups of users and the impact of their behaviors on the performance of food recommender systems.

Secondly, we conducted an offline evaluation of many state-of-the-art recommender system algorithms in the just-in-time food recommendation with implicit feedback data. Specifically, we showed that the simple multinomial mixture model with time weighting, proposed in this paper, significantly outperformed most state-of-the-art algorithms. In addition to the aggregate performance, we also performed a context-aware evaluation to examine the algorithmic bias across different demographic groups and eating contexts.

In what follows, we briefly review related work and describe the dataset and the data preprocessing steps used in the study. Next, we present the analysis of repeat food consumption (RQ1) and the just-in-time food recommendation experiment and the results (RQ2). Then, we present the results of the context-aware evaluation (RQ3), summarize the findings, and discuss their implications. Lastly, the limitations of the study and future work are described.

2 RELATED WORK

First, we begin by reviewing past research related to computational studies of food consumption behavior, particularly its exploration (novel) versus exploitation (repeat) nature, and their applications in food recommender systems.

Studying food consumption using online data: Past computational studies have demonstrated various public health monitoring applications, especially pertaining to healthy food consumption [1, 23], through the use of large-scale data from popular online platforms, such as Twitter, Instagram, Allrecipes, and MyFitnessPal. Next, recent studies have investigated the healthiness cues uncovered from food images posted by Instagram users [24] and online cooking recipes from Allrecipes [27]. Lastly, online food diaries data from MyFitnessPal users have been used to study individuals' dieting [8, 16, 34], food substitutes extraction [3], and healthy eating behaviors [2]. In contrast to the public health monitoring aspect of previous work, our work focuses on predicting food items likely to be consumed in the next consumption session which has a direct application to the just-in-time health interventions.

Modeling novel and repeat consumptions: Novel and repeat consumptions have been studied in psychology and consumer behavior research from either the hedonic aspect [13, 18] or the habitual aspect [20, 35] using conventional methodology, such as questionnaires and interviews. Compared to past consumer behavior studies, our work is one of a few computational studies which quantitatively characterized novel and repeat food consumption behaviors using publicly-available online data. Recently, the novel and/or repeat consumption phenomenon has also been studied in data mining and machine learning research. These studies can be divided into two broad categories. The first category focuses on predicting novel items for future consumptions from historical observations – a common task of general recommender systems. Next, the second category of work examines the recurring nature of past events and consumptions from a variety of online and offline domains, e.g., web search logs [28], online media consumptions and geolocation check-ins [4, 5, 19, 22], and online product purchases [6]. Within this category, several methods have been proposed to capture the exploration-and-exploitation dynamics underlying the consumption behaviors [4, 5, 19, 22]. Compared to the existing work which mainly investigated online consumptions, our study specifically focuses on characterizing and predicting repeat consumptions from self-report *offline* food consumption data. Our experimental setup is similar to that of Kotzias et al. [22]. In particular, we adopted the recommender algorithms used in their experiments as our baselines. In contrast, we further extended their proposed mixture model by decaying count data over time and analyzed the algorithmic performance and biases in a context-aware evaluation.

Food recommender systems: Past food recommendation research primarily aimed to predict user ratings of online cooking recipes from historical user-item rating data [31]. Several general recommendation methods have been applied to this domain, including content-based filtering [12, 30], collaborative filtering via k-nearest neighbors algorithms [17, 32] and matrix factorization [11, 14, 32], etc. Unlike the well-explored problem of rating prediction in food recommender systems, our work focuses on a problem of just-in-time implicit-feedback food recommendation. Specifically,

we aim to generate the lists of recommended food items containing both repeat and novel items for the next consumption session (i.e., the next day in our study) by learning from the users' consumption history. Our task setup is comparable to next-basket and sequential recommendation [25, 33] which explicitly incorporate sequential information and temporal dynamics [9, 21] of past behavior in the models. To our knowledge, this particular task has not yet been extensively studied in the food recommendation research [31].

3 DATASET

Popular food diary and health tracking applications, such as MyFitnessPal (MFP), Fitbit, etc., provide a useful and publicly available source of granular data suitable for the study of food consumption behaviors of individuals. In this study, we used a MyFitnessPal food diary dataset[1] created by Weber and Achananuparp [34]. The original dataset contains: (a) 6.5M food diary entries collected from 9.9K users covering a 6-month period from September 2014 to April 2015; and (b) user profile information (e.g., gender, age, location, etc.) of 8.8K users. Each food diary entry (a data row) consists of textual description of a food item and its portion size, meal occasion (e.g., breakfast, lunch, etc.), nutrition (e.g., calories, protein, fat, etc.), and a set of high-level food categories (e.g., meats, vegetables, etc.) and sub-categories (e.g., beef, wheat, etc.) annotated by the dataset creators. To reduce sparsity in the user-item consumption data, we removed keywords mentioning: (a) specific commercial entities, such as brand and restaurant names, and (b) quantities, from the textual description of food diary entries. We further selected a subset of the original data from October 12, 2014 to March 14, 2015 (22 weeks) for their frequent activity level to be used in this study.

We performed the following data cleaning steps. First, we removed outlier entries such as those with negative portion sizes, food with calories higher than 3,000 kcal, and non-food entries such as "quick add calories" (33.9K, 0.52% of all records). Next, we removed entries containing auxiliary items, such as dietary supplements and condiments (1.1M, 17.50%). For the prediction task, we only considered records from meals with breakfast, lunch, dinner, or snack labels. Meals with other labels (1.4M, 22.17%) were excluded. Lastly, we performed a p-core filtering by recursively removing: (a) food items that were not consumed by more than 5 users; and (b) users who consumed less than 20 remaining food items (1.7M, 25.71%). After the data cleaning and preprocessing steps, the dataset contains 2.7M food diary entries involving 55K unique food items and 7.7K unique users as shown in Table 1. Demographically, a vast majority of users were female (82.62%), young adults 18-44 years of age (79.09%), and lived in the United States (71.72%).

Figure 1 shows the number of active users – those who recorded their food diary on any given day, over time (mean daily active users = 3,034). The data exhibit a clear cyclic pattern where the users tended to be more active on weekdays than weekends. Furthermore, fewer users continued to record food diaries during the holidays. For example, the numbers of daily active users decreased by 23.3% and 24.3% on Thanksgiving day and Christmas day, respectively. At the start of 2015, a large number of active users emerged, possibly due to the effect of the new year's resolution. Next, the food consumption pattern follows a near power-law distribution. Figure 2 shows an

[1] http://bit.ly/2XMywQO

Table 1: Data statistics

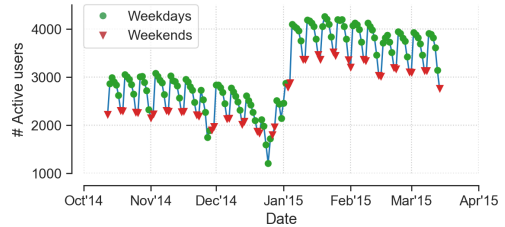|	
# users	7,721
# items	55,584
# meals	1,149,692
# diary entries	2,737,885
# items per user (mean ± S.D.)	115 ± 85.57
# items per user per day (mean ± S.D.)	5.86 ± 3.57
% female	82.62%
% 18-44 years old	79.09%
% United States	71.72%

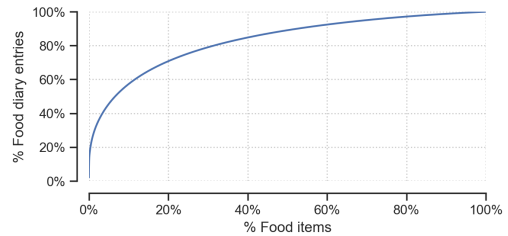**Figure 1: Number of active users over time**

Figure 2: CDF of food items

empirical cumulative distribution function (CDF) plot of the food items recorded across all food diary entries (i.e., food consumptions). As we can see, up to 30% of food items accounted for 80% of food diary entries, suggesting that a small fraction of items tended to be reconsumed most of the time.

4 RQ1: REPEAT CONSUMPTION ANALYSIS

In this section, we present a quantitative analysis of repeat food consumption behavior. Specifically, we investigate repeat consumption patterns across meal occasions, temporal lifestyle factors, and demographic groups. Firstly, let us define the notion of repeat food consumption used in the analysis. A food item by a user is considered repeated if the user has consumed the same item within the last k time steps.

Formally, let U be the set of all users, F be the set of all food items, and S be the set of food consumption sequences of all users, $S = \{S_1, \cdots, S_{|U|}\}$. Each user $u_i \in U$ has a consumption sequence $S_i = \{X_i^1, \cdots, X_i^T\}$, where X_i^t denotes the set of food items consumed by user u_i at time step t; $X_i^t \subset F$. Next, we define the food consumption sequence of u_i during the interval $[t_l, t_r)$ as $S_i^{[t_l, t_r)} = \{X_i^{t_l}, \cdots, X_i^{t_r-1}\}$ and use $n_{i,j}^{t,k} = |\{X \in S_i^{[t, t+k)}, f_j \in X\}|$ to denote the number of days the food item f_j is consumed by

user u_i during the interval $[t, t + k]$. If $n_{i,j}^{t,k} \geq 2$, f_j is said to be *reconsumed* by u_i during the holding time window $[t, t + k]$.

Next, we measure the propensity to reconsume of user u_i as a fraction of repeat consumptions over all consumptions of u_i. That is, let $\mathbb{S}_i^{t,k} = \{X \in S_i^t | \exists f_j \in F, n_{i,j}^{t,k} \geq 2\}$ be the repeat consumptions of user u_i at time step t, the fraction of repeat consumptions of u_i at time step t and the average fraction of repeat consumptions of u_i given a k-day window size are defined in equations 1 and 2, respectively.

$$R_i^{t,k} = \frac{|\mathbb{S}_i^{t,k}|}{|S_i^t|} \quad (1) \qquad R_i^k = Avg_{t \in [1, T-k+1]} R_i^{t,k} \quad (2)$$

The CDF plot of the fraction of repeat consumptions (R_i^k) at different k-day window sizes, i.e., 2 days, 7 days, 30 days, and lifetime (∞) are shown in Figure 3. With smaller k values, the bounded consumption sequence is shorter and the set of food items consumed by the user is generally smaller. Therefore, fewer food items can be considered and the fraction of repeat consumptions is expected to be lowered. When $k = 7$, we observe that about 50% of the users reconsumed the same food items up to 40% of the time, whereas when considering past consumptions over the entire users' lifetime ($k = \infty$), about 50% of the users reconsumed up to almost 60% of past items. For the rest of this section, we set $k = 7$ as the default time window given a strong recency bias towards food consumptions of the previous week.
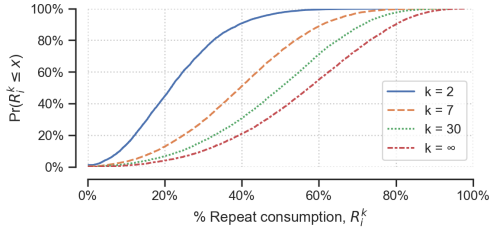

Figure 3: CDF of the fraction of repeat consumptions (R_i^k) at different k-day window sizes

4.1 Meal Occasions and Carryover Effects

Next, we investigate the presence of *carryover* effects – the influence of past food consumption behavior on current food consumption behavior [20], and the repeat consumption patterns in different meal occasions, i.e., breakfast, lunch, dinner, and snack, in our dataset. Specifically, we aim to characterize the repeat consumption behavior within the same meal occasions (e.g., breakfast → breakfast) and across different meal occasions (e.g., breakfast → lunch). First, we define the fraction of *within-meal* repeat consumptions $R_i^{m,k}$ of user u_i in a similar manner as the fraction of repeat consumptions R_i^k (equation 2). Let $X_i^{b,t}$, $X_i^{l,t}$, $X_i^{d,t}$ and $X_i^{s,t}$ denote the subsets of items consumed by user u_i at time step t in four different meal occasions: breakfast, lunch, dinner, and snacks, respectively, we computed $R_i^{m,k}$ from $X_i^{b,t}$, $X_i^{l,t}$, $X_i^{d,t}$ and $X_i^{s,t}$ for the corresponding meal occasions. According to the CDF plot of fraction of within-meal repeat consumptions in Figure 4, breakfast has the highest within-meal fraction of repeat consumptions ($R_i^{m,k}$)

amongst all meal occasions. That is, about 50% of users reconsumed up to 50% of the past breakfast items in the last 7 days. On the other hand, dinner has the lowest $R_i^{m,k}$, where 50% of users only reconsumed up to 17% of the past dinner items.
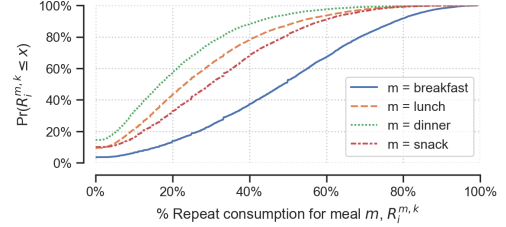

Figure 4: CDF of fraction of within-meal repeat consumptions ($R^{m,k}$) of 4 different meal occasions
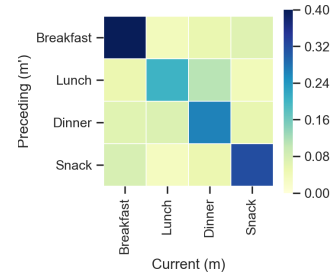

Figure 5: Fractions of within-meal ($R^{m,k}$) and across-meal ($R^{m,m',k}$) repeat consumptions, averaged across all users

Next, we define the fraction of *across-meal* repeat consumptions $R_i^{m,m',k}$ of user u_i for meal m with respect to the past consumptions in the holding time window of meal m'; $m \neq m'$. Then, we computed $R_i^{m,m',k}$ for the twelve corresponding meal occasion pairs for all users. Figure 5 displays the fractions of within-meal ($R^{m,k}$ in the diagonal cells) and across-meal ($R^{m,m',k}$ in the non-diagonal cells) repeat consumptions, averaged across all users. As we can see, the across-meal carryover effects are much weaker than the within-meal carryover effects. The strongest across-meal carryover effect is found between the preceding lunch and the current dinner (lunch → dinner); $R^{m,m',k} = 0.111$ (S.D. = 0.118). This is within our expectation as the food items consumed at lunch and dinner are generally more similar and interchangeable than other meals. It is rather common to eat lunch leftover at dinner. These findings are also in line with prior consumer research [20].

4.2 Temporal Dynamics of Consumption

Furthermore, we explore the impact of temporal lifestyle factors, such as the weekday-weekend cycle, on the repeat consumption behavior over the 6-month period. At each time step t, we computed the fraction of *daily* repeat consumptions across all users as $R^{t,k} = Avg_{u_i \in U} R_i^{t,k}$ where t represents day of the year. As shown in Figure 6, there is a clear cyclical and habitual pattern in the repeat consumption behavior where the fractions of daily repeat consumptions fluctuate in a weekly cycle yet the trend remains more or less
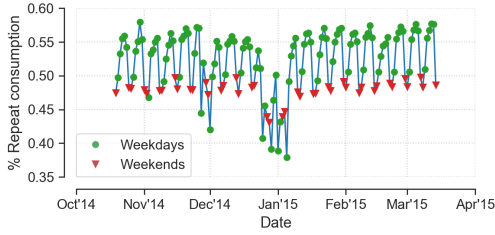
Figure 6: Fraction of daily repeat consumptions ($R^{t,k}$) over time
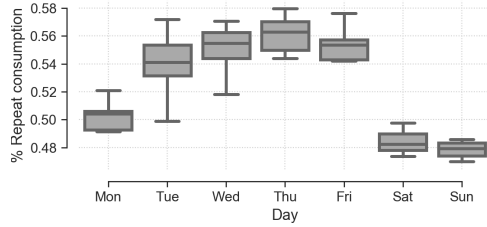

Figure 7: Distribution of fraction of daily repeat consumptions ($R^{t,k}$) for days of the week

constant over a long period of time. The fractions of daily repeat consumptions are greatly lower during the Thanksgiving and the Christmas holidays in the US, possibly due to temporary changes to seasonal food choices. The day with the lowest fraction of daily repeat consumption is the first Monday of 2015 ($R^{t,k}$ = 0.379), largely due to the surge in newly active users, whose past consumption data were far fewer than the users in the preceding periods. Next, Figure 7 displays the distribution of $R^{t,k}$ for different days of the week. As shown here, the medians and the variability of $R^{t,k}$ for all weekdays are higher than those of weekends, suggesting that the users were more likely to engage in variety-seeking behavior during the weekends than the weekdays. Within the weekdays, the fractions of repeat consumptions are the lowest on Monday and continue to rise as the week progresses until reaching the peak on Thursday. This may also indicate the presence of carryover effects as the users' daily eating habits were picked up from one weekday to the next.

4.3 Demographic Differences

Does the propensity to reconsume differ significantly between demographic groups? In this section, we compare the repeat consumption behaviors of users in the following subgroups based on their demographic attributes: genders (female and male users), age groups (younger adults aged 18-44 years old and older adults aged 45 years and older), and regions of residence in the US including northeast (NE), midwest (MW), south (S), and west (W). For each user u_i, we computed the fraction of repeat consumptions R_i^k and averaged all values of R_i^k across all users belonging to the same demographic subgroups to get the fraction of aggregated repeat consumptions $R^{d,k}$ for the subgroups. To measure the differences of $R^{d,k}$ between the demographic subgroups, we performed Kruskal-Wallis H test with Dunn's multiple comparison test at the significance levels of

0.01 and 0.05. According to Figure 8, male users, older adults, and those in the northeast generally had a higher propensity to reconsume than their counterparts. Specifically, males had a significantly higher $R^{d,k}$ than females (p<0.01), whereas older adults had a significantly higher $R^{d,k}$ than younger adults (p<0.01). Next, there was a significant difference between $R^{d,k}$ of users in different regions (p<0.05), specifically, users in the northeast had a significantly higher $R^{d,k}$ than those in the south (p<0.05).
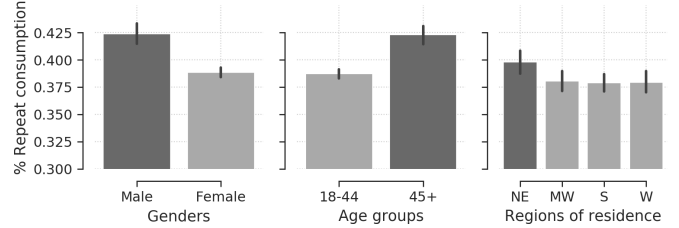

Figure 8: Fraction of aggregated repeat consumptions of different demographic groups. The groups with the highest mean are highlighted in darker color.

Compared to previous research, our finding about the age differences is in line with prior knowledge about differences in sensation seeking traits [26] in which younger people are more likely to seek out novel and varied sensations and experiences than older people. However, our finding about the gender differences is in contrary to previous findings. Specifically, male users in our study had a significantly lower tendency for novel consumptions than female users, whereas male adults generally score higher in sensation seeking traits than females [26]. One possible reason could be due to the potential interaction effect between gender and age since the average age of the male users in our study (39.5 years; S.D. = 10.7) is much higher than that of the female users (35.4 years; S.D. = 10). Lastly, to our knowledge, this study was the first to report the inter-regional differences in repeat consumption tendencies.

Summary of Findings: According to the quantitative analysis, repeat food consumption patterns displayed a recency bias. The majority of users repeatedly consumed at least 50% of food items recently consumed within the last 7 days, whereas the repeat consumption rate went up to 60% or higher once the entire consumption history (up to six months) was considered. Furthermore, the individuals' repeat consumption tendencies significantly differed across meal occasions, temporal lifestyle factors, and demographic groups. First, there were greater patterns of the within-meal carryover effect than the between-meal carryover effect. In particular, users tended to reconsume more during breakfast and snack. However, they tended to explore novel food choices more frequently during lunch and dinner. Next, the repeat consumption behavior clearly exhibited a weekday-weekend. That is, users were significantly more likely to engage in variety-seeking behavior in food consumption during the weekends (and holidays) than the weekdays. Furthermore, we observed a significantly higher repeat consumption tendency amongst male users, older adults, and users residing in the northeast of the United States, compared to their respective counterparts.

5 RQ2: JUST-IN-TIME RECOMMENDATION

Due to the novel and repeat consumption dynamics and the context-sensitive nature of food consumption, we argue that predicting a complete set of repeat and novel food items for the next consumption sessions is a more pertinent task for food recommender systems than the general rating prediction task, especially in the just-in-time health intervention scenario. Once the users' daily eating habits are learned, different behavioral interventions can be taken, for example, recommending healthier and similar substitutes to replace the less healthy but highly reconsumed items, recommending healthy novel items complementary to the basket of highly reconsumed items, etc. In addition, many state-of-the-art algorithms in food recommender systems, such as matrix factorization, are effective in recommending new items to the users, their performance in predicting both the repeat and novel food items for future consumptions is currently not known. Lastly, traditional food recommender systems often rely on the user-item rating data which may be scarce or difficult to obtain given the burden of data collection.

Therefore, we present an offline experiment of *just-in-time* food recommendation to investigate the effectiveness of many recommender system algorithms using *implicit feedback* data in which only the historical consumptions and no rating data are provided. Specifically, we define the food recommendation task as generating the top-N recommendation list for the *next-day* consumption. Moreover, as food consumption behavior is highly context-sensitive, we further conduct a context-aware evaluation to examine the performance of different algorithms across subgroups of users based on their demographic attributes and contexts. The results of the context-aware evaluation will be discussed later in section 6.

5.1 Algorithms

We compare the performance of eight algorithms in the just-in-time food recommendation evaluation. These algorithms can be categorized into 4 following groups: multinomial mixture models, sequential recommender models, latent-factor models, and rule-based methods.

Multinomial Mixture Models: Motivated by the recent success of the multinomial mixture model (**Mixture**) [22], which has been shown to outperform most state-of-the-art algorithms in several implicit-feedback recommendation tasks, we propose a time-weighted mixture model (**MixtureTW**) as a simple extension of **Mixture**. The original **Mixture** consists of two multinomial components that capture the balance between the *individual* exploitation component (θ^I) and the *population* exploration component (θ^P), i.e., the repeat and the novel consumptions, respectively. This exploration-exploitation framework seems naturally suitable for modeling the food consumption behaviors of individuals due to the inherently recurring nature of the data. In **Mixture**, the probability P_{ij} of user u_i consuming item f_j is formally defined as:

$$P_{ij} = \pi_i \theta^I_{ij} + (1 - \pi_i)\theta^P_j \qquad (3)$$

where the personalized mixture weight π_i represents the trade-off between the two components $\theta^I_{ij} = \frac{C_{ij}}{C_i}$ and $\theta^P_j = \frac{C_j+1}{C+|F|}$; C_{ij} denotes the user-item consumption count of user u_i and item f_j; F denotes the set of all food items. The Expectation-Maximization

(EM) algorithm is used to learn π_i from the training and validation data [22].

In **MixtureTW**, we incorporate the idea of time-weighted recommender systems [9] into **Mixture** by decaying user-item consumption counts over time such that recent consumption counts are weighted higher than old consumption counts. Specifically, let T be the most recent time step, the user-item consumption count C_{ij} for the time steps $[1, T]$ is discounted by a decay rate $\lambda \in (0, 1)$ as follow:

$$C_{ij} = \sum_{t=1}^{T} \lambda^{T-t} c^t_{ij} \qquad (4)$$

.

where c^t_{ij} is the number of times user u_i consumed item f_j in the time step t. Similar time weighting is applied to other consumption count data, e.g., C_i, C_j, etc., to derive θ^I_{ij} and θ^P_j.

In the experiment, we consider both **MixtureTW** and **Mixture** as competitive algorithms. Our implementations of **MixtureTW** and **Mixture** are based on the original authors' code[2].

Sequential Recommender Models: Additionally, we employ a well-known Factorizing Personalized Markov Chains algorithm (**FPMC**) [25] as a representative baseline for sequential recommender systems. **FPMC** takes into account both sequential information of items in different time steps and general user preferences when generating recommendations. In the experiment, we modified a python implementation of **FPMC**[3] (v.0.1) to allow for variable-sized baskets.

Latent-Factor Models: Next, we include three latent factor models widely-used in the implicit-feedback recommender systems, i.e., non-negative matrix factorization (**NMF**), hierarchical Poisson factorization (**HPF**) [15], and latent Dirichlet allocation (**LDA**) [7]. These algorithms have been empirically shown to perform effectively in both the user-item rating prediction and the implicit-feedback general recommendation tasks [15, 22, 32]. In the experiment, we used the scikit-learn[4] (v.0.20) implementations of **NMF** and **LDA** and the hpfrec[5] (v.0.2.2) implementation of **HPF**.

Rule-Based Methods: Lastly, we define two rule-based algorithms as simple baselines: global popularity (**Global**) and personal favourite (**Personal**). **Global** is a naive baseline in which each item f_j is assigned a score proportional to its global consumption frequency n_j in the training set. Thus, every user was recommended the same set of globally popular items in each session. Next, **Personal** is another naive baseline where a score of user u_i consuming item f_j is proportional to the consumption frequency n_{ij} from u_i's past consumptions of f_j in the training set. That is, the method simply assumes that the users always reconsumed their personally favourite items and never tried new items, i.e., exploitation-only behavior.

5.2 Experimental Protocols

We used the MFP dataset previously described in Table 1 in the experiment. The dataset was split into 146 sliding-window sessions. Each experiment session contains 9 days (or 9 time steps) of food consumption sequence data and the next session was incremented

---

by one time step from the previous session. In each session, data at the most recent time step t were used as a held-out test set for evaluation, those at $t-1$ were used as a validation set for optimizing hyperparameters, and those at $[t-8, t-1)$ were used as a training set. Given a set of all food items in the training set F, for each session, the goal is to estimate for each user u_i, $\theta_i = [\theta_{i1}, \ldots, \theta_{i|F|}]$, $\sum_{f_j \in F} \theta_{ij} = 1$ where θ_{ij} is the probability that user u_i consumed item f_j at time step t. Moreover, we removed from the test set (i) unseen items (those not existing in the training set) and (ii) unseen users (those not existing in the training and the validation sets) to ensure that the mixture models were able to estimate a personalized mixture weight π_i for all the users in the test set. On average, 3.6K unseen items (S.D. = 928) and 595 unseen users (S.D. = 252) were removed per session. The statistics of the dataset used in the experiment are summarized in Table 2.

For the training data used in all algorithms (except **HPF**), the user-item consumption frequency matrix is L1 normalized, such that the item consumption frequencies add up to 1 for each user. This allows the algorithms to be more robust to outliers. Since **HPF** inherently models the skew in item popularity, we used the original user-item consumption frequency matrix as input for **HPF**. Next, we used the default hyperparameters in the respective packages for **NMF**, **HPF**, **LDA**, and **FPMC** and evaluated different numbers of latent factors using a subset of the experiment data spanning 3 days (January 20 - 22 of 2015). Lastly, we optimally set the number of latent factors for **NMF**, **HPF**, **LDA**, and **FPMC** to 100, 500, 50, and 500, respectively, for all sessions. For **MixtureTW**, the decay weight λ was optimally tuned for each test session.

Table 2: Experiment data statistics

# sessions	146
# users/session (mean ± S.D.)	2,461 ± 683
# items/session (mean ± S.D.)	23,651 ± 4,385
# items/user in training (mean ± S.D.)	21.81 ± 10.64
# items/user in validation (mean ± S.D.)	5.27 ± 3.16
# items/user in testing (mean ± S.D.)	5.13 ± 3.12
# novel items/user in testing (mean ± S.D.)	3.55 ± 2.51
# repeat items/user in testing (mean ± S.D.)	1.58 ± 1.89

In addition, to obtain the meal-specific food recommendation results for the context-aware evaluation, we first split the experiment data into 4 disjoint subsets for breakfast, lunch, dinner, and snack. Then, the same protocols described earlier were performed on each meal-specific subset. Due to space constraints, we only reported their results in section 6.

5.3 Evaluation Metrics

We used three standard metrics commonly used in the implicit-feedback recommendation evaluation: recall, precision, and normalized discounted cumulative gain (nDCG), to measure the effectiveness of different algorithms in generating the top-N recommendation lists. Firstly, **recall@N** is defined as the proportion of actual consumption in the test set was identified correctly in the top-N recommendation list, over all items actually consumed by u_i, i.e., the size of the test set. Particularly, we adopted the definition of weighted recall used in Kotzias et al. [22] shown in equation 5.

$$\text{Recall@N} = \frac{1}{|U|} \sum_{i \in U} \sum_j \frac{n_{i,j} \cdot \mathbf{I}\big(rank(i,j) \le N\big)}{\sum_j n_{i,j}} \quad (5)$$

Next, **precision@N** is defined as the fraction of correctly recommended items in the top-N recommendation list. As the size of test set varies for different users, precision@N may be underestimated for some users who generally consumed fewer n items. Lastly, **nDCG@N** is defined as a discounted cumulative gain (DCG) of items in the top-N recommendation list normalized by the ideal discounted cumulative gain (IDCG), which is obtained by computing DCG for items in the test set sorted descendingly by their consumption frequency n_{ij} in the test set. As nDCG considers multiple levels of relevance, it is more sensitive to the relevance of higher ranked items.

For each algorithm, we first computed scores for each user u_i in each session and averaged the results across all users. Then, we averaged the scores across all sessions to obtain the average performance for each algorithm. The values for all metrics range from 0 (worst) to 1 (best). In the experiment, we set $N = 5$ for the *all-item* evaluation setup and $N = 3$ for the *novel item-only* evaluation setup. Lastly, to evaluate the statistical differences between the performance of different algorithms (RQ2) and the context-specific performance across different groups (RQ3), we performed Kruskal-Wallis H test with Dunn's multiple comparison test at the significance levels of 0.01 and 0.05.

5.4 Results & Discussion

Table 3 displays the average recall@5, precision@5, and nDCG@5 across all 146 experiment sessions for the eight algorithms. Overall, there is a significant difference in the performance of different algorithms (p <0.01). Particularly, **MixtureTW** significantly outperformed the other algorithms in all metrics (p <0.01), e.g., +74.2% of **NMF** in nDCG@5. The average mixture weights π_i of **MixtureTW** and **Mixture** across all sessions was 0.667 (S.D. = 0.168), indicating that the individual exploitation component (θ^I) was generally more important than the population exploration component (θ^P). In addition, the mean decay factor λ for **MixtureTW** was 0.812 (S.D. = 0.046), suggesting a strong recency bias. The superior performance of the multinomial mixture models over the latent-factor models in our experiment is consistent with the results from the prior research [22].

Surprisingly, the popular sequential recommender model, **FPMC**, was not as effective as **MixtureTW** and **Mixture**. This might be due to the highly repetitive nature of food consumption in the MFP dataset. For **FPMC**, the item sequence information from the first-order Markov chains, which is treated as equally important as user preferences, may not be very helpful for the highly repetitive dataset. Amongst the latent-factor models, **NMF** performed better than **HPF** and **LDA**. The relative performance of the latent-factor models in our experiment differs from those reported in the previous studies [15, 22]. This inconsistency could possibly occur due to: (1) the differences in the experiment setup; and (2) the differences in data characteristics and user behavior. Lastly, the simple **Personal** baseline performed fairly competitively (-5.4% of **MixtureTW** in nDCG@5). Again, this might be explained by the heavily repetitive nature of the food consumption data.

Table 3: Performance (mean ± standard deviation) of different algorithms in generating complete recommendations (repeat + novel) across all sessions. The best results are in bold.

Method	recall@5	precision@5	nDCG@5
MixtureTW	**0.389** ± 0.029	**0.352** ± 0.039	**0.465** ± 0.038
Mixture	0.370 ± 0.026	0.337 ± 0.035	0.446 ± 0.034
FPMC	0.355 ± 0.027	0.321 ± 0.034	0.414 ± 0.034
NMF	0.185 ± 0.009	0.176 ± 0.014	0.267 ± 0.012
HPF	0.099 ± 0.005	0.102 ± 0.010	0.143 ± 0.002
LDA	0.071 ± 0.013	0.060 ± 0.005	0.083 ± 0.009
Global	0.054 ± 0.004	0.052 ± 0.004	0.073 ± 0.004
Personal	0.366 ± 0.026	0.333 ± 0.035	0.440 ± 0.034

While **MixtureTW** was the most effective in predicting next-day food consumptions, compared to the other methods, it performed rather poorly in predicting novel consumptions. Specifically, we examined the recommended lists of food items used in the main evaluation (Table 3) and measured the average recall@3, precision@3, and nDCG@3 of the subsets of *novel item-only* recommendations. As shown in Table 4, most algorithms failed to correctly identify a few novel food items from a large set of approximately 23K items in the training set – a much more challenging setup than the typical novel recommendation task, most of the time.

Table 4: Performance (mean ± standard deviation) of different algorithms in generating novel item-only recommendations across all sessions. The best results are in bold.

Method	recall@3	precision@3	nDCG@3
MixtureTW	0.00464 ± 0.001	0.00788 ± 0.002	0.00732 ± 0.001
Mixture	0.00463 ± 0.001	0.00790 ± 0.002	0.00732 ± 0.001
FPMC	0.00286 ± 0.001	0.00576 ± 0.001	0.00560 ± 0.001
NMF	**0.03486** ± 0.005	**0.05970** ± 0.010	**0.07354** ± 0.012
HPF	0.00409 ± 0.001	0.00702 ± 0.002	0.00657 ± 0.001
LDA	0.00405 ± 0.001	0.00694 ± 0.002	0.00655 ± 0.001
Global	0.00463 ± 0.001	0.00790 ± 0.002	0.00732 ± 0.001
Personal	0.00013 ± 0.000	0.00022 ± 0.000	0.00021 ± 0.000

Summary of Findings: The effectiveness of the eight recommender algorithms in the just-in-time implicit food recommendation greatly varied from 0.465 - 0.073 according to the nDCG@5 metrics. The multinomial mixture models (**MixtureTW** and **Mixture**), which explicitly considered the balance of the repeat and novel consumptions, were the most effective methods in predicting the individuals' next-day food consumptions. Moreover, **MixtureTW** significantly outperformed **Mixture** by incorporating the recency bias in decaying consumption count data over time. The state-of-the-art sequential recommender (**FPMC**) and the general recommender systems (**NMF**, **HPF**, and **LDA**) all performed poorly overall. Lastly, all algorithms were not effective in predicting next-day novel consumptions. The results are generally consistent with prior research [4, 22] and emphasize the challenging nature of the just-in-time food recommendation task.

6 RQ3: CONTEXT-AWARE EVALUATION

Do diverse groups of users equally receive the same benefits from the recommendations generated by the best algorithm, i.e., **MixtureTW**? Figure 9 displays nDCG@5 of **MixtureTW** for different genders, age groups, regions of residence, days of the week, weekdays and weekends, and meal occasions. The best result is highlighted in a darker shade for each group. As we can see, there is a clear bias between the algorithm performance and the propensity to reconsume of the users in different contexts and demographic groups, previously presented in sections 4.
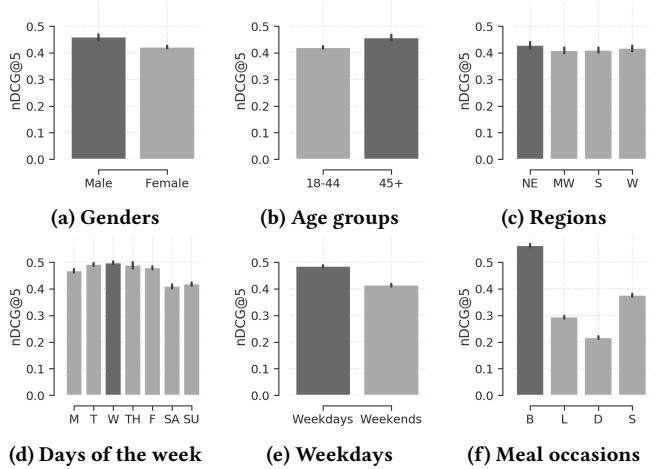

(a) Genders (b) Age groups (c) Regions

(d) Days of the week (e) Weekdays (f) Meal occasions

Figure 9: MixtureTW's performance for different demographic groups and contexts. The groups with the highest nDCG@5 are highlighted.

For different genders and age groups, the differences in the nDCG@5 were significant across all between-group comparisons (p <0.01). Overall, **MixtureTW** was able to predict the daily eating habits of males 9.03% better than females. The average nDCG@5 for males is 0.459 (S.D. = 0.195), whereas the average nDCG@5 for females is 0.421 (S.D. = 0.195). Next, **MixtureTW** was 8.81% more effective in predicting the food consumptions of older users than younger users. The average nDCG@5 is for users at least 45 years old is 0.457 (S.D. = 0.199), whereas the average nDCG@5 for users between 18 and 44 is 0.420 (S.D. = 0.194). Moreover, there were no significant differences between the nDCG@5 of the users in different regions of residence.

In terms of temporal lifestyle factors, there was a significant difference in the nDCG@5 between weekdays and weekends (p <0.01). That is, **MixtureTW** was 17.15% more effective in predicting food consumption during weekdays than weekends. The average nDCG@5 for weekdays is 0.485 (S.D. = 0.021), whereas the average nDCG@5 for weekdays is 0.414 (S.D. = 0.017). Next, there was a significant difference in the nDCG@5 of different days of the week (p <0.01). Specifically, the differences were significant (p <0.01) for all weekday-weekend pairs (except for Monday-Sunday being significant with p <0.05). Amongst the weekday pairs, Monday had a significantly lower nDCG@5 than Wednesday (p <0.01) and Thursday (p <0.05). However, there were no significant differences between the nDCG@5 for the other weekday pairs (e.g.,

Monday vs. Tuesday) and the weekend pair (Saturday vs. Sunday). Wednesdays had the highest nDCG@5 (mean = 0.498; S.D. = 0.011), which is 21.76% higher than that of the lowest group, i.e., Saturdays; nDCG@5 = 0.409 (S.D. = 0.019).

Lastly, there was a significant difference between the nDCG@5 of different meal occasions (p <0.01). Furthermore, the differences in the nDCG@5 across all meal pairs were also significant (p <0.01). Specifically, **MixtureTW** was significantly more effective in predicting food consumptions during breakfast (nDCG@5 = 0.563; S.D. = 0.250) than the other meals. On the other hand, dinner was the most challenging meal to predict (nDCG@5 = 0.217; S.D. = 0.188). The performance gap between breakfast and dinner is 159.45%, the highest amongst any between-group differences.

Summary of Findings: We observed significant algorithmic bias in the context-aware evaluation of **MixtureTW**. Overall, **MixtureTW** was more effective in predicting daily food consumption patterns of the users in the meal occasions, temporal lifestyle factors, and demographic groups with higher average fraction of repeat consumptions as shown in sections 4.1, 4.2, and 4.3, respectively. Specifically, male and older-adult users unevenly received greater benefits from **MixtureTW**'s recommendations than their counterparts. Interestingly, although there was a significant difference in the repeat consumption tendency amongst different regions (shown in section 4.3), there was no algorithmic bias across regions.

7 SUMMARY & IMPLICATIONS

In this section, we first briefly summarize the main findings of our RQs. Then, we discuss the implications of those findings in regards to the performance of the just-in-time food recommender systems and their practicality in the just-in-time health interventions.

- **RQ1**: Repeat food consumption is highly ubiquitous, recency biased, and significantly differ across different contexts and demographic groups.
- **RQ2**: Most state-of-the-art recommender systems are not as effective as the best algorithm – the time-weighted mixture model (**MixtureTW**), in the just-in-time implicit food recommendation task.
- **RQ3**: The performance of **MixtureTW** is significantly biased in favor of the users with high repeat consumption tendency, which is manifested in diverse contexts and demographic groups.

Implications for food recommender systems: To further increase the effectiveness of the just-in-time food recommender systems, several technical improvements can be made. First, additional research should investigate other state-of-the-art temporal models [19], which may better capture the dynamics of repeat consumptions and the recency bias. Next, the lists of recommended items generated by most algorithms in this study comprise independently selected food items, ignoring the *complementary* nature of food consumption [30]. Thus, incorporating such item-item complementarity when generating the recommended lists may help improve the performance of the novel items prediction. Lastly, the data preprocessing steps used in this work may not be sufficient in reducing data sparsity in the food consumption data. Therefore, other techniques, such as biclustering, should be further investigated.

Implications for just-in-time interventions: Food recommender systems is a potential facilitator of the just-in-time healthy eating interventions where specific food items are adaptively recommended tailored to individuals. The presence of algorithmic bias against the users in certain contexts (e.g., weekends) and demographic groups (e.g., young adults), who are less likely to adopt healthy eating behaviors [2], may adversely affect the overall success of the interventions. Next, the highly recurring and the recency-biased natures of food consumption emphasize the importance of habit formation [35] as another facilitator of sustained healthy eating lifestyle. The designs of just-in-time interventions may incorporate both the food recommender systems and the habit formation mechanisms to improve a long-term success of the interventions. Lastly, the surprising effectiveness of the personal favorite heuristics suggests that a simple rule-based algorithm is a good enough alternative to more sophisticated algorithms (e.g., **FPMC**, **HPF**, etc.), especially in the population-scale interventions where computational resources and efficiency are likely ones of the technical constraints.

8 LIMITATIONS & FUTURE WORK

We recognize that the demographic distributions and repeat consumption behavior of the MFP users used in this study, the majority of whom were young female adults on weight-loss dieting, may not be representative of those of the general public. Particularly, one can surmise that some MFP users might have a higher propensity to reconsume than the general public due to their strict dietary regimen. As a result, this may affect the generalizability of our findings about repeat consumption patterns. Next, since the food consumption data used in the study were self-reported on a daily basis, they were likely to contain some inaccuracy, omission, and incompleteness, especially those from around the holiday periods. Even though we have addressed most of these issues in the data cleaning step, they might still have some impacts on the food recommender results. Our just-in-time food recommendation study only began to uncover initial insights into the performance of many state-of-the-art algorithms in predicting daily eating habits of individuals. The fact that only few algorithms manage to outperform a simple personal favorite heuristics underscores the challenging nature of the task. We discussed a few potential algorithmic improvements in the implications. Next, our offline evaluation was not able to answer other important questions, particularly regarding the balance of repeat and novel items in the recommendations. During the actual about-to-eat moment, would the users be more likely to adopt the previously consumed items than the novel but substitutable items in the recommendations? Is it at all useful to recommend such items? Therefore, it is also crucial to conduct an online food recommendation evaluation to answer these questions.

9 CONCLUSION

We present a large-scale computational study of repeat food consumption and just-in-time food recommendation. The findings reveal the pervasive and significantly different patterns of repeat consumptions across meal occasions, temporal lifestyle factors, and demographic groups. Next, the experimental results demonstrate

the effectiveness of the time-weighted mixture model, which explicitly models the exploration-exploitation and the temporal dynamics of consumptions, in predicting next-day food consumptions over existing state-of-the-art sequential recommender and latent-factor based algorithms. Lastly, the results of the context-aware evaluation show significant algorithmic bias of the food recommender system towards specific groups of users. Overall, our study establishes an important first step in the just-in-time healthy eating interventions through the characterization and prediction of repeat food consumptions.

10 ACKNOWLEDGEMENT

REFERENCES

[1] Sofiane Abbar, Yelena Mejova, and Ingmar Weber. 2015. You Tweet What You Eat: Studying Food Consumption Through Twitter. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15*. ACM Press, 3197–3206. https://doi.org/10.1145/2702123.2702153

[2] Palakorn Achananuparp, Ee-Peng Lim, and Vibhanshu Abhishek. 2018. Does Journaling Encourage Healthier Choices? Analyzing Healthy Eating Behaviors of Food Journalers. In *Proceedings of the 2018 International Conference on Digital Health - DH '18*. ACM Press, 35–44. https://doi.org/10.1145/3194658.3194663

[3] Palakorn Achananuparp and Ingmar Weber. 2016. Extracting Food Substitutes From Food Diary via Distributional Similarity. arXiv:cs.CY/1607.08807

[4] Ashton Anderson, Ravi Kumar, Andrew Tomkins, and Sergei Vassilvitskii. 2014. The Dynamics of Repeat Consumption. In *Proceedings of the 23rd International Conference on World Wide Web - WWW '14*. ACM Press, 419–430. https://doi.org/10.1145/2566486.2568018

[5] Austin R. Benson, Ravi Kumar, and Andrew Tomkins. 2016. Modeling User Consumption Sequences. In *Proceedings of the 25th International Conference on World Wide Web - WWW '16*. ACM Press, 519–529. https://doi.org/10.1145/2872427.2883024

[6] Rahul Bhagat, Srevatsan Muralidharan, Alex Lobzhanidze, and Shankar Vishwanath. 2018. Buy It Again: Modeling Repeat Purchase Recommendations. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining - KDD '18*. ACM Press, 62–70. https://doi.org/10.1145/3219819.3219891

[7] David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent Dirichlet Allocation. *Journal of machine Learning research* 3, Jan (2003), 993–1022.

[8] Munmun De Choudhury, Mrinal Kumar, and Ingmar Weber. 2017. Computational Approaches Toward Integrating Quantified Self Sensing and Social Media. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing - CSCW '17*. ACM Press, 1334–1349. https://doi.org/10.1145/2998181.2998219

[9] Yi Ding and Xue Li. 2005. Time Weight Collaborative Filtering. In *Proceedings of the 14th ACM International Conference on Information and Knowledge Management - CIKM '05*. ACM Press, 485. https://doi.org/10.1145/1099554.1099689

[10] Michael D Ekstrand, Mucun Tian, Ion Madrazo Azpiazu, Jennifer D Ekstrand, Oghenemaro Anuyah, David McNeill, and Maria Soledad Pera. 2018. All The Cool Kids, How Do They Fit In?: Popularity and Demographic Biases in Recommender Evaluation and Effectiveness. In *Proceedings of the 1st Conference on Fairness, Accountability and Transparency (Proceedings of Machine Learning Research)*, Sorelle A Friedler and Christo Wilson (Eds.), Vol. 81. PMLR, 172–186.

[11] Peter Forbes and Mu Zhu. 2011. Content-Boosted Matrix Factorization for Recommender Systems. In *Proceedings of the Fifth ACM Conference on Recommender Systems - RecSys '11*. ACM Press, 261. https://doi.org/10.1145/2043932.2043979

[12] Jill Freyne and Shlomo Berkovsky. 2010. Intelligent Food Planning: Personalized Recipe Recommendation. In *Proceedings of the 15th International Conference on Intelligent User Interfaces - IUI '10*. ACM Press, 321–324. https://doi.org/10.1145/1719970.1720021

[13] Jeff Galak, Justin Kruger, and George Loewenstein. 2011. Is Variety the Spice of Life ? It All Depends on the Rate of Consumption. *Judgment and Decision Making* 6, 3 (2011), 230–238.

[14] Mouzhi Ge, Mehdi Elahi, Ignacio Fernaández-Tobías, Francesco Ricci, and David Massimo. 2015. Using Tags and Latent Factors in a Food Recommender System. In *Proceedings of the 5th International Conference on Digital Health 2015 - DH '15*. ACM Press, 105–112. https://doi.org/10.1145/2750511.2750528

[15] Prem Gopalan, Jake M. Hofman, and David M. Blei. 2015. Scalable Recommendation with Hierarchical Poisson Factorization. In *Proceedings of the Thirty-First Conference on Uncertainty in Artificial Intelligence - UAI '15*. 326–335.

[16] Mitchell L. Gordon, Tim Althoff, and Jure Leskovec. 2019. Goal-Setting and Achievement in Activity Tracking Apps: A Case Study of MyFitnessPal. In *Proceedings of the Web Conference 2019 - WWW '19*. https://doi.org/10.1145/3308558.3313432

[17] Morgan Harvey, Bernd Ludwig, and David Elsweiler. 2013. You Are What You Eat: Learning User Tastes for Rating Prediction. In *Proceedings of the 20th International Symposium on String Processing and Information Retrieval - SPIRE '13*. 153–164. https://doi.org/10.1007/978-3-319-02432-5

[18] Marion M. Hetherington, Ali Bell, and Barbara J. Rolls. 2000. Effects of Repeat Consumption on Pleasantness, Preference and Intake. *British Food Journal* 102, 7 (2000), 507–521. https://www.emeraldinsight.com/doi/abs/10.1108/00070700010336517

[19] Komal Kapoor, Karthik Subbian, Jaideep Srivastava, and Paul Schrater. 2015. Just in Time Recommendations: Modeling the Dynamics of Boredom in Activity Streams. In *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining - WSDM '15*. ACM Press, 233–242. https://doi.org/10.1145/2684822.2685306

[20] Adwait Khare and J. Jeffrey Inman. 2006. Habitual Behavior in American Eating Patterns: The Role of Meal Occasions. *Journal of Consumer Research* 32, 4 (2006), 567–575. https://doi.org/10.1086/500487

[21] Yehuda Koren. 2009. Collaborative Filtering with Temporal Dynamics. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '09)*. ACM, 447–456. https://doi.org/10.1145/1557019.1557072

[22] Dimitrios Kotzias, Moshe Lichman, and Padhraic Smyth. 2018. Predicting Consumption Patterns with Repeated and Novel Events. *IEEE Transactions on Knowledge and Data Engineering* (2018), 371–384. https://doi.org/10.1109/TKDE.2018.2832132

[23] Yelena Mejova, Sofiane Abbar, and Hamed Haddadi. 2016. Fetishizing Food in Digital Age: #foodporn Around the World. In *Proceedings of the Tenth International AAAI Conference on Web and Social Media - ICWSM '16*. 250–258.

[24] Ferda Ofli, Yusuf Aytar, Ingmar Weber, Raggi al Hammouri, and Antonio Torralba. 2017. Is Saki #delicious?: The Food Perception Gap on Instagram and Its Relation to Health. In *Proceedings of the 26th International Conference on World Wide Web - WWW '17*. ACM Press, 509–518. https://doi.org/10.1145/3038912.3052663

[25] Steffen Rendle, Christoph Freudenthaler, and Lars Schmidt-Thieme. 2010. Factorizing Personalized Markov Chains for Next-basket Recommendation. In *Proceedings of the 19th International Conference on World Wide Web - WWW '10*. ACM Press, 811. https://doi.org/10.1145/1772690.1772773

[26] Jonathan W. Roberti. 2004. A Review of Behavioral and Biological Correlates of Sensation Seeking. *Journal of Research in Personality* 38, 3 (jun 2004), 256–279. https://doi.org/10.1016/S0092-6566(03)00067-9

[27] Markus Rokicki, Christoph Trattner, and Eelco Herder. 2018. The Impact of Recipe Features, Social Cues and Demographics on Estimating the Healthiness of Online Recipes. In *Proceedings of the 12th International AAAI Conference on Web and Social Media - ICWSM '18*.

[28] Anish Das Sarma, Sreenivas Gollapudi, Rina Panigrahy, and Li Zhang. 2012. Understanding Cyclic Trends in Social Choices. In *Proceedings of the fifth ACM international conference on Web search and data mining - WSDM '12*. 593–602. https://doi.org/10.1145/2124295.2124367

[29] Donna Spruijt-Metz and Wendy Nilsen. 2014. Dynamic Models of Behavior for Just-in-Time Adaptive Interventions. *IEEE Pervasive Computing* 13, 3 (2014), 13–17.

[30] Chun-Yuen Teng, Yu-Ru Lin, and Lada A. Adamic. 2012. Recipe Recommendation Using Ingredient Networks. In *Proceedings of the 3rd Annual ACM Web Science Conference on - WebSci '12*. ACM Press, 298–307. https://doi.org/10.1145/2380718.2380757

[31] Christoph Trattner and David Elsweiler. 2017. Food Recommender Systems: Important Contributions, Challenges and Future Research Directions. *CoRR* abs/1711.02760 (nov 2017). http://arxiv.org/abs/1711.02760

[32] Christoph Trattner and David Elsweiler. 2017. Investigating the Healthiness of Internet-Sourced Recipes: Implications for Meal Planning and Recommender Systems. In *Proceedings of the 26th International Conference on World Wide Web - WWW '17*. ACM Press, 489–498.

[33] Pengfei Wang, Jiafeng Guo, Yanyan Lan, Jun Xu, Shengxian Wan, and Xueqi Cheng. 2015. Learning Hierarchical Representation Model for Next Basket Recommendation. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval - SIGIR '15*. ACM Press, 403–412. https://doi.org/10.1145/2766462.2767694

[34] Ingmar Weber and Palakorn Achananuparp. 2016. Insights from Machine-Learned Diet Success Prediction. In *Proceedings of Pacific Symposium on Biocomputing (PSB)*.

[35] Wendy Wood and David T. Neal. 2009. The Habitual Consumer. *Journal of Consumer Psychology* 19, 4 (2009), 579–592. https://doi.org/10.1016/j.jcps.2009.08.003