Recommending Personalized News in Short User Sessions

Elena Viorica Epure*
CRI, Panthéon-Sorbonne
Paris, France
Elena.Epure@malix.univ-paris1.fr

Benjamin Kille*
DAI-Lab, TU Berlin
Berlin, Germany
benjamin.kille@dai-labor.de

Jon Espen Ingvaldsen
IDI, NTNU
Trondheim, Norway
jonespi@idi.ntnu.no

Rebecca Deneckere
CRI, Panthéon-Sorbonne
Paris, France
Rebecca.Deneckere@univ-paris1.fr

Camille Salinesi
CRI, Panthéon-Sorbonne
Paris, France
Camille.Salinesi@univ-paris1.fr

Sahin Albayrak
DAI-Lab, TU Berlin
Berlin, Germany
sahin.albayrak@dai-labor.de

ABSTRACT

News organizations employ personalized recommenders to target news articles to specific readers and thus foster engagement. Existing approaches rely on extensive user profiles. However frequently possible, readers rarely authenticate themselves on news publishers' websites. This paper proposes an approach for such cases. It provides a basic degree of personalization while complying with the key characteristics of news recommendation including news popularity, recency, and the dynamics of reading behavior. We extend existing research on the dynamics of news reading behavior by focusing both on the progress of reading interests over time and their relations. Reading interests are considered in three levels: short-, medium-, and long-term. Combinations of these are evaluated in terms of added value to the recommendation's performance and ensured news variety. Experiments with 17-month worth of logs from a German news publisher show that most frequent relations between news reading interests are constant in time but their probabilities change. Recommendations based on combined short-term and long-term interests result in increased accuracy while recommendations based on combined short-term and medium-term interests yield higher news variety.

CCS CONCEPTS

• **Information systems → Data analytics**; **Content ranking**; **Web log analysis**; *Data stream mining*; *Personalization*;

KEYWORDS

News reading behavior, News reading interests, Recommender system, Personalization, Markov processes, Stationarity analysis

*Both authors contributed equally to this work.

RecSys '17, August 27–31, 2017, Como, Italy

ACM ISBN 978-1-4503-4652-8/17/08…$15.00
https://doi.org/10.1145/3109859.3109894

1 INTRODUCTION

The digital transformation has been a game changer for legacy news media. Keeping readers loyal has become highly competitive for publishers as their new business models rely on advertisement or related revenues [28]. Personalized recommendation has emerged as a popular way to tackle this challenge by automatically suggesting news to online readers. However, creating effective solutions faces particular constraints [15, 27]. Large publishers release hundreds of news daily, implying that they must deal with fast-growing numbers of items that get quickly outdated and irrelevant to most readers. News readers exhibit more unstable consumption behavior than users in other domains such as entertainment. External events affect, e.g. breaking news, affect readers interests [15]. In addition, the news domains experiences extreme levels of sparsity.

Existing news recommender solutions address these challenges to some extent. Most of them suggest fresh and popular news. Some consider the dynamics of news reading behavior. However, they rely on the availability of rich user profiles for personalized recommendations. Still, many publishers lack this knowledge, unlike news aggregators' services [2, 6] or company blogs [31], as readers tend to consume online news without authentication. Cross-device and cross-browser tracking are techniclly challenging endevaors. Thus, in practice, publishers observe relatively short sessions with fewer than ten clicks on average [9, 10, 27].

As per-user models are unsuitable in short sessions, we propose a new approach to recommendation which ensures basic personalization. The approach combines crowd reading behavior over time and the current user session. Thereby, we model crowd reading behavior for different time frames. When providing recommendations, should the models from the same day, from the last weeks, or from the last months be used? In the current work, we analyze the dynamics of the crowd news reading behavior and its effect on recommendations. We show that such design choices can significantly affect the outcome.

This work includes four contributions. First, we extend the existing distinction between short- and long-term reading behavior as we establish medium-term reading behavior. The reading behavior concerns readers' interests, which are linked to news categories. Section 2 conveys a detailed description. Second, we identify reading episodes and derive models specifically reflecting engage reading. Third, we assess the dynamics of the crowd reading behavior, complementary to related studies in news media. Specifically, we focus on the evolution of the relations between news categories rather than on standalone categories. Fourth, we compare three recommendation policies in terms of performance and news variety: (a) a policy based merely on short-term behavior; (b) a policy integrating short- and medium-term behaviors; (c) a policy integrating short- and long-term behaviors. Experiments are conducted with real data from a German news publisher, spanning 17 months. The long period is chosen to account for fluctuations caused by seasonal or other types of effects.

Our findings show that in reading episodes users are likely to read news within the same category. The relations between news categories are stable in time as the most likely target categories from a given source stay the same. However, their priorities, represented by transition probabilities, change every 1 to 4 months. Augmenting the short-term behavior with knowledge about the long- and medium-term behavior improves recommendations. The policy combining short- and long-term interests yields higher accuracy than combining short- and medium-term interests. Contrarily, combining short- and medium-term interests yields higher variety.

Further, the formalization of the proposed approach is presented. The experiments are described in Section 3 and discussed in Section 4, followed be the related work in Section 5. Finally, conclusions are drawn and directions for future works are indicated.

2 DEFINING READING BEHAVIOR IN NEWS

We consider an online environment in which a publisher provides a collection of news articles to interested readers. Let $\mathcal{U} = \{u_m\}_{m=1}^{M}$ represent the increasing set of visitors. Further, let $\mathcal{I} = \{i_n\}_{n=1}^{N}$ represent the increasing collection of news articles. The publisher observes how visitors act on the website. In particular, the publisher keeps track of events whenever a visitor reads an article. Let $\theta = (\theta_u, \theta_i, \theta_t, \theta_c)$ represent such an event where the individual variables refer to visitor, article, time, and context. The publisher observers a sequence of events $\Theta = \{\theta^{(\alpha)}\}_{\alpha=1}^{A}$ such that $\theta_t^{(\alpha)} < \theta_t^{(\alpha+1)}, \forall\alpha$. We represent the reading behavior of individual users as $\Theta_u = \{\theta : \theta_u = u\}$. We assume that visitors have limited time and desire to read articles. In particular, the number of news articles a visitor is willing to read is $X \ll N$.

The publisher tries to engage visitors by providing a small set of suggestions, every time an article is read. Formally, the publisher employs a policy π which takes a given event $\theta^{(\alpha)}$ and automatically produces a ranked list of suggestions $S^{(\alpha)} = \{s_k \,|s_k \in I\}_{k=1}^{K}$. The publisher monitors how visitors react upon the received suggestions in order to drive policy improvement. The policy that produced the suggestions gets credited with the reward R whenever a visitor reads any of the suggested articles:

$$R(\pi, \theta^{(\alpha)}, S^{(\alpha)}) = \begin{cases} 1 & \text{if} \quad \exists\beta > \alpha, \theta^{(\beta)} \in (\Theta_u \cap S^{(\alpha)}) \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

For practical purposes, the publisher disregards future events if the visitor is inactive for more than a specified time τ_u. Publishers can employ a variety of policies $\pi \in \Pi$. Their objective is to find the policy that maximizes the cumulative rewards:

$$\pi^* = \underset{\pi \in \Pi}{\arg\max} \sum_{\theta \in \Theta} R(\pi, \theta, S). \quad (2)$$

We investigate three policies: (a) the short-term news reading; (b) the long- and short-term news reading; (c) the medium- and short-term news reading. These are further described.

Editors put the most recent and significant news at the start of newspapers. Analogously, our first policy suggests articles which have been popular recently. We refer to this policy as *baseline* and it corresponds to the short-term news reading interests. The baseline has associated a list L of fixed size ϵ. As the system observes another event, θ_i is added to L. If adding θ_i exceeds ϵ, the oldest element is dismissed. Thereby, L continuously stores the most recently read articles. At the same time, the more popular an article is, the more often it will appear in L. The baseline policy suggests elements from L, most recently added and different from the article being currently read by the user u. Also, the recommendation has an implicit popularity bias as the probability of choosing a list of articles $S^{(\alpha)}$ depends on $f_{i|L}, i \in S^{\alpha}$, the frequency of article i in L.

The remaining two policies enrich the baseline with information about the news reading process—how readers transition between news categories. A stochastic process models a random system changing over time. Formally, if D is a subinterval of $[0, \infty)$, a continuous-time stochastic process is a set of random variables $\{X_d\}, d \in D$. If we restrict $D = \mathbb{N}_0$, we obtain a discrete-time stochastic process. Let $\Xi = \{\xi\}_{v=1}^{V}$ be the finite set of news categories, corresponding to the random variables of the news reading process. Let $\xi(\theta_i)$ refer to the category assigned to the article i. Hence, we can transform the reading behavior of an individual user u, $\Theta_u = (\theta^{(1)}, \theta^{(2)}, \dots, \theta^{(\alpha)})$ into a sequence of categories: $\Xi_u = (\xi(\theta_i^{(1)}), \xi(\theta_i^{(2)}), \dots, \xi(\theta_i^{(\alpha)}))$. The Markov property characterizes a stochastic process whose current state captures all information necessary to compute the probability of the next state [21]. For the news reading behavior, this is formally defined as:

$$\Pr[X_{d+1} = \xi_v | X_d = \xi_w, X_{d-1} = \xi_{w-1}, \dots, X_0 = \xi_{w_0}] =$$
$$\Pr[X_{d+1} = \xi_v | X_d = \xi_w] = p_{ij}, \forall d \in \mathbb{N}_0, \xi_{w_0}, \dots, \xi_w, \xi_v \in \Xi \quad (3)$$

A transition matrix over all permutations of two categories ($|\Xi| = V$) represents the dynamics of such as system:

$$T = \begin{bmatrix} p_{11} & \dots & p_{1V} \\ p_{21} & \dots & p_{2V} \\ \vdots & \ddots & \vdots \\ p_{V1} & \dots & p_{VV} \end{bmatrix} \quad p_{vw} \geq 0, \sum_{w}^{\xi_w \in \Xi} p_{vw} = 1, \quad \forall v, \xi_v \in \Xi \quad (4)$$

The transition probabilities p_{vw} can be estimated from observations as the ratio between the frequency of transitions from category ξ_v to category ξ_w and the total number of transitions from ξ_v. If the transition matrix T stays constant as the system evolves, the stochastic process is referred to as *First-Order Markov Process.*

We use the defined reading behavior process to create a new recommendation policy on top of the baseline. For each category ξ, there is a separate list L_ξ used to model short-term reading interests. For each event θ, we determine the category $\xi(\theta_i) = \xi_v$. Subsequently, we select the most likely category for the next article: $\xi^* = \arg\max_{p_{vw}} T$ and return suggestions from L_{ξ^*}. We distinguish the medium-term news reading interests from the long-term ones based on the observations used for estimating T. The matrix corresponding to the long-term behavior contains all the

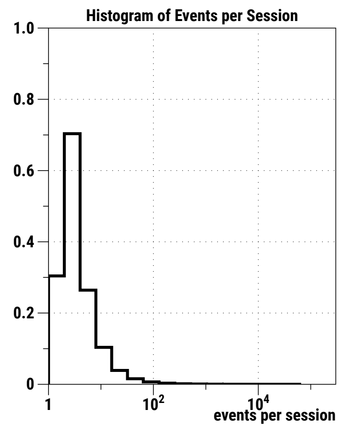

Figure 1: Histogram of events per session (log scale)

existing observations. The matrix corresponding to the medium-term behavior contains observations from a limited period only.

3 EXPERIMENTS

The research questions we address in this paper are:

Q_1 How stable in time is the news reading behavior, modeled as Markov processes over news categories?

Q_2 How do different strategies of combining long-, medium- and short-term news reading interests in a recommender compare in terms of performance and news variety?

3.1 Data Description

The data set is provided by a German news publisher that accommodates on average 3 million visits per week. It consists of ≈ 196.6 million events, which span 475 days from August 2014 to December 2015. Data from June 16 to June 27, 2015 is missing due to technical problems with the logging system. An event is generated and logged when an article is accessed on the news publisher's website. The events are logged with the following information: the user session's identifier, the event's time-stamp, and the article's link and meta-data, which includes the associated categories. The categories are manually assigned by editors, each time an article is uploaded on the website, and are a set of pre-established keywords (see Table 1). We rely on editors' experiences to assign categories in a consistent manner. Most of the user sessions have less than 10 events as shown in Figure 1.

3.2 Data Transformation

Before the experiments, the data has been pre-processed. First, we eliminated events associated to the default session. This session captures all events with cookies disabled such that individual users cannot be tracked. Second, we chronologically sorted the events within sessions. Third, we segmented sessions into reading episodes. Therein, we cut session if events occurred in less than 60 s or longer than 3600 s apart. These thresholds were set after observing and questioning our personal circle about the time spent on article reading. Other sources also suggest an average of 20 min per article[1].

The transition matrices over news categories have been computed as follows. Consecutive events in a reading episode become a transition, the first being the source, the second the target. Multiple transitions emerge if either source or target event have several categories assigned. In this case, the Cartesian product of the source and target category sets is calculated, each obtained element becoming a transition.

3.3 News Reading Dynamics

The first research question aims to assess the dynamics of the news reading behavior modeled as a Markov process over news category. We have analyzed the stationarity of the Markov process using a method based on the χ^2 statistical test [3]. The method starts with computing the transition matrix for the complete period τ and the transition matrices for all consecutive sub-intervals $t \in \tau$. Then, these local transition matrices are tested against the overall transition matrix for statistical difference. We formulate the null hypothesis and the alternative hypothesis:

$H_0 : \forall t \in \tau : p_{vw|t} = p_{vw} \leftrightarrow H_a : \exists t \in \tau : p_{vw|t} \neq p_{vw}$

- p_{vw} denotes the estimated transition probability from state ξ_v to state ξ_w for the entire period τ;
- $p_{vw|t}$ denotes the estimated transition probability from state ξ_v to state ξ_w only for the sub-periods $t \in \tau$;

Given that there are at least two positive values in each row v of the overall transition matrix T, the χ^2 test is computed as follows:

$$Q = \sum_{t \in \tau} \sum_{\xi_v \in \Xi} \sum_{\xi_w \in \Xi_v} z_{v|t} \frac{(p_{vw|t} - p_{vw})^2}{p_{vw}} \approx \mathrm{asy}\chi^2\left(\sum_{\xi_v \in \Xi} (a_v - 1)(b_v - 1)\right) \quad (5)$$

- $z_{v|t}$ denotes the number of observed transitions from state ξ_v in sub-period t, it could be 0;
- $\Xi_v = \{w : p_{vw} > 0, \xi_w \in \Xi\}, \xi_v \in \Xi$; contains all target states observed from state ξ_v for the entire period τ;
- Q has an asymptotic chi-squared distribution (asyχ^2) with the number of degrees of freedom computed as sum over all states $\xi_v \in \Xi$ by considering two terms: a_v is the number of sub-periods t for which transitions from state ξ_v are observed; $b_v = |\Xi_v|$ is the number of positive values in the row v of the transition matrix for the entire period τ.

The test could be also adjusted to assess how much a certain sub-period t differs from the complete period τ. In this case, the outer sum in (Eq. 5) has 2 terms ($|\tau| = 2$): the period t and the transition matrix computed for all the other periods from τ except t.

The first prerequisite to conduct the stationarity analysis on our data set is to decide the magnitude of the sub-periods t: days, weeks or months. Periods spanning days or weeks were excluded. In a descriptive analysis, we observed that breaking news changes the local reading pattern. Also, we randomly selected consecutive days and weeks, and ran the stationarity test. The results showed a significant matrix variance (p-value $p < 0.001$). Consistent with

[1] http://contentmarketinginstitute.com/2016/01/visitors-read-article/

Table 1: News categories and associated codes

Cars, Motor, Traffic	1	Science, Communication	2	Games, Virtual World, Toys	3
Politics, Business, Economics	4	Travel, Tourism, Navigation	5	TV, Radio, Video, Photo	6
General News	7	Professional, Career	8	Computers, Technology	9
Family, Education, Leisure	10	Banking, Finance, Insurance	11	Health, Sports, Nutrition	12
Real Estate, Home, Gardening	13	People, Relationships	14	Fashion, Lifestyle, Culture	15

this, Bickenbach and Bode [3] claim that while more granular sub-samples are preferred, they should not be too small. Otherwise, non-stationarity could emerge from the test, even though the process is Markov [3]. Eventually, the sub-periods t were set to 1 month.

When working with matrices of millions of events, a difference caused by few thousands transitions is not necessarily significant. However, as the Chi-Square test is very sensitive to the number of observations, it could yield non-stationarity [12]. For this reason, we decided to correct the transition frequencies by a factor of 0.001. The values were brought from big-data (millions) to a magnitude equivalent to what has been observed in similar studies (thousands) [3]. After correction, the frequencies still reached tens of thousands, which we consider a representative sample size.

3.4 Recommendation Policy Comparison

The second research question aims to compare the three policies introduced in Section 2 in terms of recommendation performance and ensured news variety. Publishers focus on maintaining the engagement of their readers. Thus, one way to measure performance is through the cumulative rewards $R(\pi, \Theta)$. Frequently, they normalize the rewards by the number of requests to obtain the click-through rate (CTR) = $\frac{1}{|\Theta|} \sum_{\theta \in \Theta} R(\pi, \theta, S)$. More sophisticated evaluation metrics such as the ranking-based ones—normalized discounted cumulative gain or mean reciprocal rank— exist but they cannot be applied in conditions with insufficient user feedback.

Besides monitoring how well readers' preferences are met, we look at the overall variety of suggestions. Having a more diverse set of suggestions can lead to readers experiencing serendipity. Let $S^{(A)}$ refer to all recommendations produced by applying the policy π to the upcoming events Θ. Further, we define the number of times an article i is recommended: $v_i = \sum_{s \in S^{(A)}} \mathbb{I}(s = i)$, where $\mathbb{I}$ refers to the indicator function which returns 1 if the condition applies and 0 otherwise. Finally, let $I^{(K)}$ refer to the K items with highest v_i. We quantify overall news variety as,

$$\delta = 1 - \frac{\sum_{i \in I^{(K)}} v_i}{|S^{(A)}|}. \tag{6}$$

In other words, the number of recommendations subsumed by elements of $I^{(K)}$ is normalized by the total number of recommendations and subtracted from 1. Thus, $\delta \in [0, 1]$ with $\delta = 0$ referring to the case that all readers received the identical K suggestions and $\delta = 1$ signaling that all readers received disjoint sets of suggestions.

We use the same data for this evaluation as has been used for the stationarity analysis in the previous section. This allows us to have a sound and transparent base for comparing the recommender outcomes, especially for changing periods or strongly deviating months. The evaluation follows a sliding-window approach. First, the three news recommendation algorithms are initialized with 10 000 events from September 2014 and the initial transition matrix is computed for August 2014. Subsequently, all events are processed in chronological order. Each recommendation algorithm computes a list of $K = 4$ suggestions. The choice of 4 suggestions corresponds to the number of suggestions displayed on the website of the news publisher. Then, for each user associated with the suggestions, it is checked whether they accessed one of the recommended articles within an hour. Having exceeded the 1 h limit, the suggestions are discarded. Then, the cumulative rewards and news variety are hourly computed. This yields a total of 11 355 measurements.

4 RESULTS AND DISCUSSION

This section presents the results of our experimentation. Section 4.1 is devoted to the stationarity analysis. Section 4.2 shows the observation regarding the recommendation policies.

4.1 News Reading Dynamics

We assessed the stationarity of the news reading behavior, modeled as a Markov process over news categories, for the entire period of August, 2014–December, 2015. Using the Chi-Square test (Equation 5), we obtained $Q = 8709.744$ with df $= 1778$ degrees of freedom, leading to the rejection of H_0 ($p < 0.01$). Thus, the transition probabilities fluctuate during the 17-month period. Further, we have checked whether longer periods exhibit stationary patterns. For getting better insights into how to find these periods, and whether they existed, we used the Chi-Square test (Equation 5) adjusted for computing individual period difference: $|\tau| = 2$; the first sum term is t, the assessed month; the second sum term is the rest of the period τ except this month. Then, the resulting Q values per month were plotted. Thereby, we observed a major structural break from March 2015 to April 2015. Then, the test was re-run for the period August 2014 and March 2015 still rejecting H_0 with $Q = 1124.85$, df $= 938$, $p < 0.01$. However, when Q values for each month of this period were plotted, a potential strong similarity has been observed between September, 2014 and February, 2015. Indeed, the Chi-Square test confirmed this similarity, accepting the null hypothesis with $Q = 700.51$, df $= 645$, $p > 0.05$. We repeated this procedure for the period April, 2015–December, 2015. Finally, the following homogeneous periods have been discovered:

- Sep. 2014–Feb. 2015: $Q = 700.5$, df $= 645$, $p > 0.05$,
- Apr. 2015–Jul. 2015: $Q = 6.56$, df $= 48$, $p > 0.99$,
- Sep. 2015–Oct. 2015: $Q = 14.95$, df $= 55$, $p > 0.99$,
- Nov. 2015–Dec. 2015: $Q = 62.16$, df $= 47$, $p > 0.05$.

Figure 2 is consistent with these findings. The y-axis represents the transition probabilities greater than 0.05 from category 7—General News, to all news categories; the x-axis reflects the month.

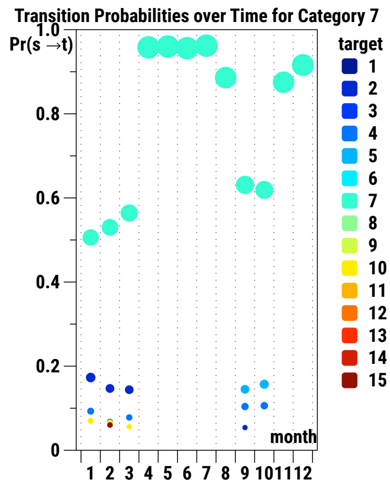

Figure 2: Probabilities per month for transitions from General News—code 7, to all categories in 2015.

The discovered homogeneous periods become visible: the circles for the months within this period line up almost horizontally. Further, we noticed a tendency to read within the same category. The highest transition probability corresponds to category 7, same as the source. Finally, it appears that few target categories are preferred to be read next, after the source category, over the entire period: Science, Communication—code 2; Politics, Business Economics—code 4; Travel, Tourism, Navigation—code 5. However, their priority changes in time. For instance, travel and tourism news are read most frequently together with the general news in the beginning of the year (January–March). We created visualizations for other source categories as well observing similar results (figures omitted for space reason).

Often, in solutions, the model stationarity is taken for granted. However, these results show that news reading behavior has both stable and dynamic parts. The most preferred news categories read after a given category seem to be stable. By contrary, the transitions between these news categories are stationary for limited periods of time, with some singular months with strongly deviating patterns (e.g. August, 2014 and 2015). Consequently, the preceding month appears to be often a suitable base for recommendation, except for the months with strongly deviating reading patterns or starting a homogeneous period. Being able to identify the cause of these changes and predict them could be very useful for adjusting dynamically the recommendation. We contemplate that various factors could influence news reading behavior: changes in human habits induced by seasons, structural changes of news publishers' websites or very important news events spanning longer periods.

4.2 Recommendation Policy Comparison

Figure 3 presents the performance results of the proposed strategies for recommendation: short-term interests only (baseline (B)), short- and long-term interests (transition complete (T_c)), short- and medium-term interests (transition 1 month (T_1)). On the left-hand side, each curve reflects the proportion of per-hour measurements, y-value, exhibiting a maximum response rate of x-value. For instance, baseline intersects the ordinate at ≈ 0.27, meaning that about 27 % of per-hour measurements have associated a 0.0 % response rate. Alternatively, T_c achieves in about 80 % of the cases a response rate up to 10 %. Thus, a more distant curve from the top left corner is preferred. The right-hand side of Figure 3 shows the pair-wise comparison of the recommendation strategies: transition 1 month vs. baseline, transition complete vs. baseline, and transition 1 month vs. transition complete. The x-axis conveys the per-hour measurements in chronological order, the left-most point being the first hour of September 1, 2014, the right-most point the last hour of December 31, 2015. The y-axis values illustrate the difference between the response rates of each pair of measurements. Positive values are shown in blue and negative ones in red.

Both T_1 and T_c achieve higher response rates on average than the baseline. T_c performs similarly to or even better than T_1. Analyzing the granular measurements, we observe that T_1 falls short of T_c in particular during the months: March, May, November and December, 2015. The subpar performance of T_1 could be explained for March—strongly deviating month from February 2015, its base of prediction, and November—the first month of a new homogeneous period so different than October 2015. Nonetheless, May and December, 2015 have associated low response rates even though they were very similar to their preceding months. This outcome could be explained by what happens in the second part of the recommendation, when the short-term interests strategy is used. The dynamic data structures of articles are sensitive to the latest crowd reading behavior. It appears that there is a division among the crowd reading interests, hypothetically caused by the co-existence of multiple strongly influential news, competing in popularity. Another interesting aspect is that sometimes strongly deviating months or the last months of homogeneous periods are a good base for prediction such as March, July and August, 2015. This indicates that even if the overall reading behavior changes, the most likely transition from a given source category stays consistent. Therefore, even for time-variant periods, the target category associated to a source appears stable.

Further, Figure 4 plots the histogram with news variety measurements for T_1 and T_c, hourly computed with Equation 6 and K set to 4. We observe that T_1 varies suggestions better than T_c as shown by the right-most peak in Figure 4. The long-term interests represent the transition probabilities in the long-run between news categories. Aligned with the literature [6, 34], this reading behavior converges to few most preferred categories chosen in recommendations. However, such behavior is not sensitive to local strong trends such as season-induced changes that could temporally modify the most likely transitions. Contrarily, the medium-term interests strategy is able to overcome this, leading to a higher news variety.

4.3 Reflections on News Recommendation

The proposed solution shows that categories provide noticeable improvements despite their simplistic nature and a fair level of personalization when rich user data is missing. Categorical information is readily available, thus it could be implemented as-is in other news recommendations setups.

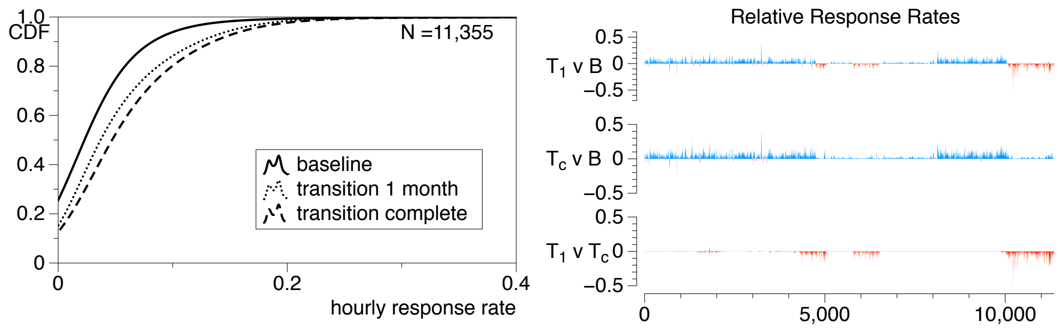

Figure 3: The left-hand side shows the cumulative distribution function (CDF) of the hourly response rates for the compared recommenders. The right-hand side shows the pair-wise comparison. Long and medium-term interests improve the results over short-term interests only. The strategy based on the long and short-term interests yields consistently the best results.

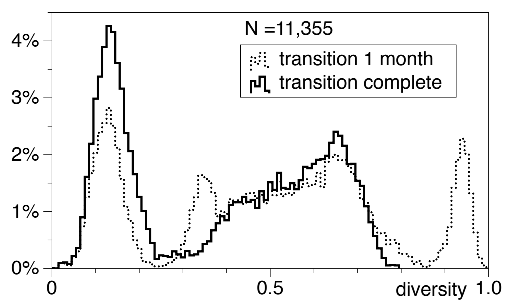

Figure 4: Analysis of the news variety ensured by T_1 and T_c. The two histograms show the distribution across the spectrum. T_1 has a peak of very high variety, which T_c lacks.

Its simplicity ensures scalability and maintainability; the crowd-based models for recommendations and the lack of authentication ensures high privacy. Data sparsity is not an issue as the proposed hybrid solution uses categories, which are limited and mostly stable in time, and the dynamic data structures are pruned to store a fixed number of articles. The proposed analysis of news reading dynamics could support decisions about which data to use for building models.

Our experiments show that by varying the time horizon publishers can trade some accuracy for news variety. Accuracy reflects how well the news recommender anticipates readers' preferences. Variety indicates to what degree readers are exposed to different articles. Which should publishers prioritize? Mark Thompson (CEO of the *New York Times*) [33] lists major challenges that news publishers face as they compete online. Revenue from printed newspapers declines. Publishers have to counteract the downward trend by attracting users online. Hereby, they cannot negotiate as comfortably as they used to for the printed news. Digital social media sites constitute the major players in digital advertising. Thus, successfully harnessing users' attention is crucial to generate online revenues. This suggests focusing on accuracy. Give readers what they want to read to keep their sessions alive. At the same time, news consumption affects society, economy, and politics. Rasmus Kleis Nielsen (Director of Research, *Reuters*) [29] finds that world-wide people prefer personalized news recommendations. Still, he points out that this may create "filter bubbles". These are spaces in which people consume similar news, reinforcing their existing believes and opinions. They represent a serious danger to social, economic, and political discourse as they prevent exposure to deviating opinions. This favors news variety over accuracy. Finding the right balance between both aspects remains a major challenge for publishers to address in the future. Maximizing profits may lead to social divides. Optimizing news variety may yield economic damage to the publisher. Publishers may have to find additional sources of revenue to compensate reduced accuracy. Readers may have to spend money to have access to high-quality journalism.

5 LITERATURE REVIEW

Traditionally, recommender solutions are of grouped into two types [1, 14, 19]. A content-based recommender suggests items similar to previously liked ones. A collaborative recommender suggests items by comparing user preferences. Hybrid solutions are frequently reported to perform best in the news domain [2, 4, 6, 20, 24–26].

5.1 News Reading Interests

News readers' interests are seldom expressed explicitly. Common approaches discover readers' interests from click behavior and articles' categories [23]. In some cases, the categories of the articles are already defined in advance and represented as contextual meta-data [6, 9, 10]. In other cases, categories are discovered automatically and represented either as more granular vocabularies associated to the categories [2, 4, 25], or keywords defining general, well-known topics such as sport or politics [24, 26, 37]. In most of these works, once known, recommenders use the categories as explicit knowledge. Other approaches do not infer users' interests but incorporate them implicitly as built-in features. For instance, Billsus et al. [4] and Li et al. [24] address short-term interests through standard content-based components by recommending articles similar to what users recently read.

Multiple authors distinguish short-term and long-term interests [4, 24, 26]. The long-term interests are considered by most of related works as the users' genuine interests which are less likely to change over time. Contrarily, the short-term interests are discovered from the most recent reading behavior and could represent deviations from the long-term interests triggered by momentary events such as breaking news or interesting reading discoveries. Billsus et al. [4] prioritize the short-term interests model in the formulation of recommendations in Google News as they claim it is more sensitive to changes. In a follow-up work on the Google News system, Liu et al. [26] showed that the short-term interests of an individual user follow closely the short-term interests of the general public. Consequently, in this updated solution, they use the unpersonalized model to address the short-term interests, and the individual user's past reading behavior for predicting categories within the long-term interests. Another mechanism for manipulating the changes in users' preferences is decaying older interests in favor of the newest ones [6, 24].

We distinguish between short- and long-term news reading interests as well, but we implement them differently. The short-term interests are implicitly captured by providing recommendations according to the real-time popularity distribution of the news, most recently read by the crowd. The long-term behavior is explicitly captured with transitions over news categories, dynamically maintained. We also introduce a new type of behavior corresponding to the medium-term interests. Additionally, our models are created per-crowd and from data reflecting engaged reading.

5.2 News Recency, Popularity, and Variety

While some news articles could be relevant even weeks or months after their publication, others get quickly outdated for the majority of users. Thus, recency and popularity are often considered in news recommendation [2, 4, 6, 9, 20, 25].

Billsus et al. [4] report that their recommender takes a list of articles as input, which have been selected in advance by several criteria including recency. Likewise, in the work of Li et al. [24] a probabilistic graphical model is built with recent articles. Furthermore, Das et al. [6] choose to re-build the recommender models every hour in order to present the freshest information to the users. They use a covisitation metric in recommendation that captures the relative popularity (what is the most popular article visited together with the current read article). Similarly, in Yahoo's news aggregator, articles are selected to represent both new and popular events [2]. For this, they consider articles' timestamps and key events identified through the events' count distribution within the selected article set [2]. Liang et al. [25] set a higher weight to the most recent articles from those discovered by using models for short and long-term interests. External knowledge from Twitter can also be used to determine popularity [7, 20, 30].

While popularity and recency are common news features included in recommender systems, variety is mentioned as a future work in some papers [4, 6] or less relevant in some others [2, 9]. The users' long-term interests model used in [4] does not vary the recommendations for users with similar profiles but per user. Das et al. [6] limit the news variety to the cluster of similar users. Li et al. [24] handle news variety explicitly through random walks in the user-item affinity graph created in advance. Likewise, a probabilistic model is proposed in [19] to address this aspect. In contrast, having analyzed the transition matrices extracted from the CLEF NewsREEL 2014 competition data set, Doychev et al. [9] show that users tend to read news from the same category. Moreover, few dominant categories and clicks between these categories account for the majority of items that are read [9]. Driven by similar reasons, Ahmed et al. [2] provide recommendations within the same category—story, but they also focus to some extent on other tangential categories as they claim users are interested in different aspects of an event (e.g. political, economic).

In the short-term reading policy, we consider popularity and recency, without any external sources. As for news variety, we do not prioritize it for categories, but we ensure that two readers are likely to be recommended different articles in a short time window.

5.3 Session-based Recommendation

Session-based recommender systems focus on sessions rather than on complete user profiles. Shani et al. [32] highlight the session-based character of recommender systems deployed in business setups. They move the underlying model from a matrix completion task toward a *Markov Decision Process* (MDP). Businesses encounter session-based recommender systems in domains including music [38], products [18], and news [23]. Deep learning architectures have been applied to session-based recommendation (cf. Hidasi et al. [17], and Tan et al. [22]), achieving promising results. Still, they involve a multitude of parameters to optimize. Our method circumvents the efforts to tune as many parameters albeit sacrificing accuracy to some degree.

In our solution, sessions are analyzed and split in reading episodes in case of too short or long reading times between consecutive events. Related works that approached the session splitting to some degree are [25, 39]. Nonetheless, the former considers fast browsing only, setting a time within 3 s to 250 s, while the later splits in sub-sessions of 30 min, without reasoning on clicks.

5.4 Process Models for Recommendation

Several works model news recommender solutions as probabilistic graphical problems [2, 24, 31]. Li et al. [24] consider the states in the process being both users and articles and the possible transitions are user to article, article to article and article to user. They populate the transition matrix with similarity scores between articles and between users and articles [24]. Ahmed et al. [2] use a transition graph with three types of states: views, clicks, and documents where the views are considered latent variables. The probability of a click to happen is conditioned on the current document, current view, and the previous click [2]. Sahoo et al. [31] choose a Hidden Markov Model where users are the observed variables and the latent classes represent globally preferred consumption patterns per month. Yang et al. [39] propose a topic-aware Markov model for recommending web pages. Segments of sessions—consecutive web page visits, are transformed to temporal states while sequences of articles of the same topic become topical states. The prediction of a page to a user considers the similarity to other users and the probability of observing the user's session containing the predicted page [39].

Though effective, the presented models are difficult to maintain with an increasing number of clicks, users, and items. Categories in our approach are a more stable choice. Also, these solutions rely on much longer sessions. Yang et al. [39] report an average of 3700 clicks per user and others in the news domain [24–26] select only authenticated sessions of minimum 10 clicks. As per-user models are likely unsuitable in short sessions, we aggregate knowledge on news consumption over time and introduce personalization in sessions. Compared with the other works, Sahoo et al. [31] also uses successfully a global model with personalization based on users' cases, but with reported high computational costs.

5.5 News Reading Behavior Analysis

In order to learn more about users' reading behavior and lead the design of the news recommenders, certain works performed high-scale user log analysis over time [11, 24, 26].

Li et al. [24] compared the reading behavior regarding long- and short-term interests. They considered short-term interests being the categories preferred by users within a time window of three days and similarly, the long-term interests were associated with a time period of 15 days. Moreover, three groups of users were created based on the click frequency and separately analyzed. The approach they used for comparison was to plot the Kullback-Leibler (KL) divergence scores, averaged per group, and computed between two consecutive periods for each user. Results showed that long-term goals are quite stable while the short-term ones vary significantly. Liu et al. [26] conducted their analysis by visually projecting the differences between categories' distribution for each user, for each two consecutive months, and also the click distribution for each category per month. The news categories are general topics such as world news, sport, entertainment, and the distribution is represented by the averaged number of clicks per category. They discovered that the current preferences of users change, that the general public interests follow the big events trend (e.g. national news are read more during elections) and, to certain extent, that the individual users' interests follow those of the general public. Esiyok et al. [11] analyzed immediate transitions between categories.

Compared to these works, we assess the dynamics of the news reading behavior through the stationarity of the Markov process across news categories. To our knowledge, we are the first to investigate the evolution of news categories relations over time.

6 CONCLUSION AND FUTURE WORKS

With the digital transformation, news has started being created and delivered by many entities besides news publishers. News consumers are exposed nowadays to increased sources of information and setups. In this highly competitive ecosystem, news publishers have sought ways to engage existing and attract new readers. Hence, personalized news recommendation has become a key element in their strategy. The proposed solutions aimed to ensure the delivery of fresh and interesting news, taking into consideration the dynamics of reading interests.

Still, limited research has tackled news personalization in hybrid recommenders when extensive user profile are unavailable. Most of the news publishers face this issue as the website authentication is a rare habit among online readers and the cookie-identification has its limits. In this paper, we prove that a recommender based on the dynamics of reading behavior, news popularity, and recency can provide a basic level of personalization and comply with the main domain constraints revealed by the related solutions. Specifically, we design and compare three variants of news recommendation centered on short-, medium-, and long-term reading interests. The short-term interests are captured at the article level, by recommending news recently read by the crowd, biased by popularity. The medium- and long-term interests are captured at the news category level, by predicting the next category to be read from the current user session and the Markov process over all categories. Moreover, the dynamics of the news reading behavior modeled as transitions between news categories is assessed over an extensive period.

The proposed research questions are addressed through experimentation with real data from a German news publisher. First, news readers appear to stay loyal to certain categories, which are frequently read together within a reading episode. Nonetheless, the priority for these relations, captured by transitions, changes in months possibly because of seasons, popular events lasting longer time, or structural changes in publishers' websites. Second, augmenting a short-term interests recommendation policy with the long- or medium-term behavior leads to higher response rates. Considering the transitions between news categories from the previous month leads to a higher variety than the long-term transitions.

Several limits could be identified nonetheless. The proposed solution handles cold start well on the user side. Still, it cannot immediately recommend new items until they have been read at least once. Moreover, the dynamic data structures storing articles can have the popularity manipulated artificially, propagating thus a fake filter bubble. Currently, the recommendation of news could be sometimes redundant as the solution checks only the user's most recent read. Parameters are present in the solution: ϵ the size of article buffers per category, the thresholds involved in the identification of reading episodes. Future work should explore other values of these parameters and their effect on recommendations. Online evaluation will show how well this approach performs under real-life conditions. Also, evaluation with additional data sets and with different session-based recommenders is required.

More content and context features could be considered in the future extending the proposed approach. In the current solution, the news categories are manually associated to the articles by editors or journalists. An alternative approach is to use more of the news content data directly. If humans are biased or inconsistent in their tagging practices, algorithms for category assignment could assist them or augment in the back-end the manually attached category set. Solutions for classifying news articles into categories or representing the data items in terms of semantic entities with links to conceptual abstractions have already been proposed [2, 5, 8, 25].

External sources of knowledge such as social media could be explored for identifying the most recent trends and injecting news variety in recommendations by external drivers [13, 16, 36]. Suggestions at the category level in the form of a time-line is another option to be explored [35]. For this, the variety of the recommended categories should be ensured. The literature revealed that there is a lack of agreement about how short- and long-term interests are defined. For instance, long-term is limited to 15 days in [24] and to a month in [26]. However, in order to more easily integrate conclusions emerging from various studies, a sound, well-documented framework should be proposed. Finally, our approach could be applicable to other domains where the item consumption is sequential within the same session, and items have a relevant feature associated or inferred. This could be the case for music or job recommendation.

REFERENCES

[1] Gediminas Adomavicius and Alexander Tuzhilin. 2005. Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions. *Transactions on Knowledge and Data Engineering* 17, 6 (2005), 734–749.

[2] Amr Ahmed, Choon Hui Teo, S V N Vishwanathan, and Alex Smola. 2012. Fair and Balanced : Learning to Present News Stories. In *WSDM*. 333–342.

[3] F. Bickenbach and E. Bode. 2003. Evaluating the Markov Property in Studies of Economic Convergence. *International Regional Science Review* 26, 3 (2003), 363–392. DOI: http://dx.doi.org/10.1177/0160017603253789

[4] Daniel Billsus and Michael J Pazzani. 2007. Adaptive News Access. In *The Adaptive Web*, Peter Brusilovsky, Alfred Kobsa, and Wolfgang Nejdl (Eds.). Springer, Chapter 18, 550–570.

[5] Michel Capelle, Alexander Hogenboom, Alexander Hogenboom, and Flavius Frasincar. 2013. Semantic News Recommendation Using WordNet and Bing Similarities Categories and Subject Descriptors. In *Symposium on Applied Computing*. 296–302.

[6] Abhinandan Das, Mayur Datar, Ashutosh Garg, and Shyam Rajaram. 2007. Google News Personalization: Scalable Online. In *WWW*. 271–280.

[7] Gianmarco De Francisci Morales, Aristides Gionis, and Claudio Lucchese. 2012. From Chatter to Headlines: Harnessing the Real-time Web for Personalized News Recommendation. In *Proceedings of the Fifth ACM International Conference on Web Search and Data Mining (WSDM '12)*. ACM, New York, NY, USA, 153–162. DOI: http://dx.doi.org/10.1145/2124295.2124315

[8] Marco de Gemmis, Pasquale Lops, Cataldo Musto, Fedelucio Narducci, and Giovanni Semeraro. 2015. *Semantics-Aware Content-Based Recommender Systems*. Springer US, Boston, MA, 119–159. DOI: http://dx.doi.org/10.1007/978-1-4899-7637-6_4

[9] Doychin Doychev, Aonghus Lawlor, Rachael Rafter, and Barry Smyth. 2014. An Analysis of Recommender Algorithms for Online News. In *CLEF (Working Notes)*. 825–836.

[10] E. V. Epure, J. E. Ingvaldsen, R. Deneckere, and C. Salinesi. 2016. Process Mining for Recommender Strategies Support in News Media. In *Proceedings of the 10th International Conference on Research Challenges in Information Science (RCIS)*. IEEE.

[11] Cagdas Esiyok, Benjamin Kille, Brijnesh-Johannes Jain, Frank Hopfgartner, and Sahin Albayrak. 2014. Users' Reading Habits in Online News Portals. In *Proceedings of the 5th Information Interaction in Context Symposium (IIiX '14)*. ACM, New York, NY, USA, 263–266. DOI: http://dx.doi.org/10.1145/2637002.2637038

[12] Anna Frankfort-Nachmias, ChavaLeon-Guerrero. 2006. *Social Statistics for a Diverse Society*. Pine Forge Press.

[13] Qi Gao, Fabian Abel, Geert-Jan Houben, and Ke Tao. 2011. Interweaving Trend and User Modeling for Personalized News Recommendation. In *IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology*. IEEE, 100–103. DOI: http://dx.doi.org/10.1109/WI-IAT.2011.74

[14] Florent Garcin, Kai Zhou, Boi Faltings, and Vincent Schickel. 2012. Personalized News Recommendation Based on Collaborative Filtering. In *Proceedings of the IEEE/WIC/ACM International Joint Conferences on Web Intelligence and Intelligent Agent Technology (WI-IAT '12)*. IEEE Computer Society, Washington, DC, USA, 437–441.

[15] Jon Atle Gulla, Bei Yu, Özlem Özgöbek, and Nafiseh Shabib. 2015. Third International Workshop on News Recommendation and Analytics (INRA 2015). In *Proceedings of the 9th ACM Conference on Recommender Systems*. ACM, 345–346.

[16] Anat Hashavit, Roy Levin, Ido Guy, and Gilad Kutiel. 2016. Effective Trend Detection within a Dynamic Search Context. In *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*. ACM, 817–820.

[17] Balázs Hidasi, Alexandros Karatzoglou, Linas Baltrunas, and Domonkos Tikk. 2015. Session-based Recommendations with Recurrent Neural Networks. In *Proceedings of the International Conference on Learning Representations*.

[18] Wenxing Hong, Lei Li, and Tao Li. 2012. Product Recommendation with Temporal Dynamics. *Expert systems with applications* 39, 16 (2012), 12398–12406.

[19] Amin Javari and Mahdi Jalili. 2015. A Probabilistic Model to Resolve Diversity–Accuracy Challenge of Recommendation Systems. *Knowledge and Information Systems* 44, 3 (2015), 609–627. DOI: http://dx.doi.org/10.1007/s10115-014-0779-2

[20] Nirmal Jonnalagedda, Susan Gauch, Kevin Labille, and Sultan Alfarhood. 2016. Incorporating Popularity in a Personalized News Recommender System. *PeerJ Computer Science* 2 (2016), e63.

[21] J.G. Kemény and J.L. Snell. 1960. *Finite Markov Chains*. Van Nostrand.

[22] Yong Kiam Tan, Xinxing Xu, and Yong Liu. 2016. Improved Recurrent Neural Networks for Session-based Recommendations. *arXiv preprint arXiv:1606.08117* (2016).

[23] Lihong Li, Wei Chu, John Langford, and Robert E. Schapire. 2010. A Contextual-Bandit Approach to Personalized News Article Recommendation. In *Proceedings of the 19th International Conference on World Wide Web (WWW '10)*. ACM Press, New York, New York, USA. DOI: http://dx.doi.org/10.1145/1772690.1772758

[24] Lei Li, Li Zheng, Fan Yang, and Tao Li. 2014. Modeling and Broadening Temporal User Interest in Personalized News Recommendation. *Expert Systems with Applications* 41, 7 (2014), 3168–3177.

[25] Ting-Peng Liang and Hung-Jen Lai. 2002. Discovering User Interests from Web Browsing Behavior: An Application to Internet News Services. In *HICSS*. IEEE Computer Society, 203.

[26] Jiahui Liu, Peter Dolan, and Elin Rønby Pedersen. 2010. Personalized News Recommendation Based on Click Behavior. In *Proceedings of the International Conference on Intelligent User Interfaces*. 31–40.

[27] Andreas Lommatzsch. 2014. Real-Time News Recommendation Using Context-Aware Ensembles. In *Proceedings of the 36th European Conference on IR Research (ECIR '14)*. Springer, 51–62. DOI: http://dx.doi.org/10.1007/978-3-319-06028-6_5

[28] Nic Newman. 2017. Overview and Key Findings of the 2016 Report. (2017). http://digitalnewsreport.org/survey/2016/overview-key-findings-2016/.

[29] Rasmus Kleis Nielsen. 2016. People Want Personalised Recommendations (Even as they Worry about the Consequences). In *Digital News Report*.

[30] Owen Phelan, Kevin McCarthy, and Barry Smyth. 2009. Using Twitter to Recommend Real-time Topical News. In *Proceedings of the Third Conference on Recommender Systems (RecSys '09)*. ACM, New York, NY, USA, 385–388. DOI: http://dx.doi.org/10.1145/1639714.1639794

[31] Nachiketa Sahoo, Param Vir Singh, and Tridas Mukhopadhyay. 2012. A Hidden Markov Model for Collaborative Filtering. *MIS Q.* 36, 4 (Dec. 2012), 1329–1356.

[32] Guy Shani, David Heckerman, and Ronen I Brafman. 2005. An MDP-Based Recommender System. *Journal of machine Learning research* (2005).

[33] Mark Thompson. 2016. The Challenging New Economics of Journalism. In *Digital News Report*.

[34] Emily Thorson. 2008. Changing Patterns of News Consumption and Participation. *Information, Communication & Society* 11, 4 (2008), 473–489. DOI: http://dx.doi.org/10.1080/13691180801999027

[35] Giang Tran, Mohammad Alrifai, and Eelco Herder. 2015. Timeline summarization from relevant headlines. In *European Conference on Information Retrieval*. Springer International Publishing, 245–256.

[36] David Vallet, Shlomo Berkovsky, Sebastien Ardon, Anirban Mahanti, and Mohamed Ali Kafaar. 2015. Characterizing and Predicting Viral-and-Popular Video Content. In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*. ACM, 1591–1600.

[37] X. Wu, F. Xie, G. Wu, and W. Ding. 2011. Personalized News Filtering and Summarization on the Web. In *Proceedings of the 23rd International Conference on Tools with Artificial Intelligence*. IEEE, 414–421. DOI: http://dx.doi.org/10.1109/ICTAI.2011.68

[38] Diyi Yang, Tianqi Chen, Weinan Zhang, Qiuxia Lu, and Yong Yu. 2012. Local Implicit Feedback Mining for Music Recommendation. In *Proceedings of the sixth ACM conference on Recommender systems*. ACM, 91–98.

[39] Qingyan Yang, Ju Fan, Jianyong Wang, and Lizhu Zhou. 2010. Personalizing Web Page Recommendation via Collaborative Filtering and Topic-Aware Markov Model. In *Proceedings of the International Conference on Data Mining (ICDM '10)*. IEEE Computer Society, 1145–1150. DOI: http://dx.doi.org/10.1109/ICDM.2010.28