IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 24 December 2023, accepted 25 January 2024, date of publication 20 February 2024, date of current version 4 March 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3367770

RESEARCH ARTICLE

User Experiments on the Effect of the Diversity of Consumption on News Services

ATOM SONODA[1], FUJIO TORIUMI[1], AND HIROTO NAKAJIMA[2]
[1]School of Engineering, The University of Tokyo, Bunkyo-ku, Tokyo 113-0033, Japan
[2]Nikkei Inc., Tokyo 100-8066, Japan

Corresponding author: Atom Sonoda (asonoda@g.ecc.u-tokyo.ac.jp)

This work involved human subjects or animals in its research. The authors confirm that all human/animal subject research procedures and protocols are exempt from review board approval.

ABSTRACT On many online platforms, more items are being added every day, requiring users to select efficiently the desired item from among these candidates. In similar environments, recommender systems have been introduced, allowing users to select from among recommended items. While the immediate advantages of recommender systems, such as presenting users with pertinent information efficiently and boosting page views, are well recognized, concerns have been raised that these systems might constrict users' choices and amplify echo chambers. Despite these concerns, the detrimental impacts of recommender systems on user behavior have not been thoroughly investigated. To address the societal challenges stemming from the limitation in the breadth of information to which users are exposed, a novel recommender system offering a wider array of choices is essential. In this study, we examine the user experience of the Nikkei Electronic Version, a major news delivery service in Japan, from the perspective of diversity. Using the language of article titles, we evaluate the diversity of recommended results and click logs presented to users. In addition to an analysis of the actual service, we conduct user experiments with participants recruited from among users of the service to investigate the effect of recommendation algorithms specialized in increasing diversity, separate from the actual service. We propose a recommender system based on the concept of extracting articles as far apart as possible within the language space. Through static experiments and user experiments, we show that the proposed recommender system is effective in diversifying the recommendation list. Furthermore, we show that when candidate articles that are likely to be clicked can be properly extracted, the proposed recommender system is effective in diversifying the articles that users click. Our research is aimed at improving long-term satisfaction by recommending content that users can enjoy in the short term while simultaneously ensuring diversity.

INDEX TERMS Diversity, user engagement, online news, recommender systems.

I. INTRODUCTION

In the digital realm, countless articles and videos are uploaded daily, making it challenging for users to identify pertinent content. Recommender systems, which are increasingly utilized in various sectors, simplify this information search by tailoring content based on users' past browsing and search behaviors.

Various recommendation techniques exist, such as collaborative filtering, rank-based learning, content-based filtering, and hybrid approaches. Both online and offline evaluation methods can assess these techniques. Deployed extensively on platforms like social media, news websites, and video streaming services, recommender systems strive to present content aligned with users' preferences. Academically, there are many in-depth studies about the impacts of these recommendations, addressing issues like popularity biases and analyzing the value of the recommender system for service providers and users, respectively.

The associate editor coordinating the review of this manuscript and approving it for publication was Fabrizio Messina.

As recommender systems become more prevalent, there is a growing concerns about diminishing information choices. Emphasizing content based mainly on user interests has raised fears about the "filter bubble" [1]. This leads to reduced exposure to diverse opinions and new information because of a heavy reliance on recommendations. This not only strengthens group consensus but can also sideline alternative perspectives or fresh insights. It presents challenges both societally and individually. Therefore, it is necessary for the recommender system to simplify information discovery and preserve diversity.

From a business perspective, it is important to diversify recommender systems. Sole dependence on metrics like clicks and likes might pave the way for clickbait-like issues, potentially undermining the long-term user experience and raising the risk of users leaving the service [2], [3], [4]. Providing a diverse array of articles can boost user satisfaction and ensure continued engagement. However, implementing a diversity-focused recommender system has its technical and business challenges, causing hesitance in its real-world adoption.

Few studies have delved into user experiments related to recommender systems that enhance diversity in news services. The primary deterrent is the technological complexity inherent in designing and executing such systems. While conventional recommender systems offer articles consistent with users' reading patterns and keywords, diversity-driven ones must suggest articles from diverse genres and perspectives. Essential to this is a technology that can accurately identify these genres and viewpoints to provide precise recommendations. Adopting a diversity-centric recommendation strategy may initially decrease click-through rates and user engagement. However, over time, such systems could boost user satisfaction by introducing them to a wider range of opinions and fresh insights. Yet, the potential short-term decline in click-through rates, which could negatively impact business, might deter many companies from considering a diversity-focused recommender system.

Concerning the connection between recommender systems and diversity, Nguyen et al. [2] analyzed diversity variations in a movie recommender system. Furthermore, Anderson et al. [5], [6] investigated service persistency rates based on diversity within a music distribution service. Gomez-Uribe and Hunt [7] introduced the Effective Catalog Size as a metric to gauge diversity, illustrating that a personalized recommender system results in a broader range of items being consumed in the media. Although these studies leveraged data from both real-world and research-oriented services, to our understanding, no user experiments have specifically focused on the diversity of information consumed by users in a controlled setting.

In this research, we carried out user experiments and examined user logs provided by Nikkei Inc., a leading economic-focused daily boasting the world's largest circulation and recognized as a prominent online media platform in Japan [8]. We conducted detailed user experiments and meticulously analyzed user logs, leading to the introduction of a unique recommender system — Semantic Volume Based Recommendation (SVR). SVR, utilizing linguistic attributes, identifies unique articles for recommendation. Distinct from traditional diversity-driven recommendation techniques, SVR can sequentially produce a varied list of articles. Although computational expense is a concern for large-scale real services, SVR sidesteps the need for combination computations, ensuring efficient calculations. Our comprehensive research, integrating data analysis and user tests, validates the effectiveness of SVR in enhancing recommendation list diversity. We further delved into the impact of selecting clickable articles and observed shifts in user behavior upon the integration of diversity-focused recommendations.

In conclusion, the proposed SVR efficiently recommends diverse items, keeping computational complexity under real-service constraints, with observed user selections showing increased diversity.

II. RELATED WORK

A. PROLIFERATION OF RECOMMENDER SYSTEMS

Multiple methodologies like collaborative filtering [9], [10], [11], ranking learning [12], content-based filtering [13], and hybrid systems [14] enrich recommendation algorithms. As online media continue to flood users with articles and videos, recommendation systems are increasingly essential in a variety of domains for information discovery. The Netflix Prize contest has significantly shaped research by standardizing evaluation through a single metric [15]. Recommender systems are used not only by news sources [16], [17] and Social networks [18], but also by the travel [19], [20] and retail industries [21], [22]. These systems are even being incorporated into recommendation for multiple roles (e.g. buyers, sellers, and distributors) within multi-sided markets such as food delivery apps [23].

B. IMPACT OF RECOMMENDER SYSTEMS ON USER EXPERIENCE

Research on recommender systems has been extensive, with immediate user responses, such as clicks and likes, being widely studied [24], [25], [26]. There have also been studies on the feedback loop between recommendations and user behavior in streaming services [27], [28], as well as on the effect of personal and situational characteristics on user behaviors on recommender systems [29]. Structural and probabilistic models have been employed in marketing and information retrieval to model user environments [30], [31], while insight into the mechanisms of user experience with recommender systems has been explored in the field of human computer interaction [32], [33]. Research has been conducted to understand user behavior by controlling for potential confounding factors through the use of a simulation construction and field experiments [3], [34], [35].

C. RECOMMENDER SYSTEMS AND DIVERSITY

The behaviors exhibited when utilizing services featuring recommender systems are the result of a confluence of factors, including user preferences, algorithmic suggestions, and temporal effects. Numerous studies investigate the significant impact of item popularity, focusing either on understanding or mitigating its inherent biases [7], [14], [36], [37]. Some research suggests that such biases may not adversely affect evaluations [38], [39]. In the field of information retrieval, the MMR technique aims to minimize redundancy and deliver useful information, particularly when dealing with large sets of overlapping documents [40]. Research has been conducted to differentiate the influences of human behavior and algorithmic recommendations on online music consumption [41]. Moreover, a study was conducted to investigate the influence of recommendation algorithms on content diversity in user consumption by analysing the relationship between diversity in music distribution services and user engagement, as demonstrated through, for example, service sustainability rates [5], [6]. In addition, other studies have assessed the fluctuations in diversity in movie recommender systems [2]. However, few studies have analyzed the diversity of user behavior to grasp the impact of recommender systems on the long-term user experience. No studies have carried out user experiments to analyze this.

III. DIVERSITY EVALUATION METHODS

The goal of this research is to diversify a user's selection of articles by augmenting the range of articles recommended to them. Consequently, a metric for appraising the heterogeneity of the news article set is first established. There are several studies on the definition of the variety of items consumed by users, where the most widely used functions are the Gini coefficient [27], [42] and entropy [43], [44], which are dependent on the frequency of each item consumed. These capture the extent to which users consume dissimilar items, while disregarding the resemblance between these items. For instance, a user who views an article on Federal Reserve policy and an article on Bank of Japan policy with the same frequency and a user who views an article on the World Cup and an article on international politics with the same frequency are both considered highly analogous if the articles are in different categories. In such a case, the former would be classified as equally varied according to these metrics, despite the news viewed being less comparable. Therefore, the diversity metric should consider not only the frequency of consumption, but also similarities between the items consumed.

Therefore, in defining the diversity of a set of news articles, this study evaluates the diversity of linguistic expressions in the article titles, because linguistic diversity is particularly important for news articles due to their nature as text-based media. In addition, we focus on article titles because it is estimated that users often make decisions based on an article's title article when selecting news and because the article titles contain important information due to the characteristics of economic news. Figure 1 shows the news list screen.

FIGURE 1. An example of the news list interface. When a user chooses to read a news article, they select it from the news list screen. The news list screen features thumbnail images and headings. It is assumed that the heading has a substantial influence on the user's determination to pick an article.

In this study, an assessment of the diversity of the news article set was conducted through the following steps:

- Article Vectorization
- Evaluation of the diversity of the vector set

We made vectors for each article by embedding each article title into word embedding space. Moreover, this study employed the Generalist-Specialist score (GS-score) [5] and the average pairwise distance (APD) [2] to gauge the diversity of the vector set. These metrics were based on a vector representation of the items, thus enabling an evaluation of diversity by considering similarities between items.

A. ARTICLE VECTORIZATION

In this study, we vectorized articles by using vectors in the word embedding space derived from the article title. The process of vectorizing an article can be divided into the following steps:

- Separating words with spaces
- Vectorizing each word in the article title
- Vectorizing the word set

The process encompasses three stages. Initially, the article title is separated with spaces, after which each word in the article title is vectorized using word2vec [45]; finally, the word set is vectorized through a Term Frequency Inverse Document Frequency (TF–IDF) weighted average to generate a vector representing the article.

Morphological analysis is used to divide a Japanese sentence into words, referred to as segmentation. For the segmentation of article titles, MeCab [46] was used as the morphological analysis software, and mecab-ipadic-NEologd [47] was used as the system dictionary. With these two techniques, the words contained in the article titles were segmented, and a word set was created.

Word2vec is one method for indicating words in distributed representations. With a distributed representation of words, we can obtain expressions that reflect the context and meaning of words. We used the Japanese Wikipedia entity vector [48] as a pretrained model of word2vec, as it is a distributed representation vector of words and entities that are articles on Japanese Wikipedia, learned from the full text of Japanese Wikipedia.

The TF-IDF weighted average is a method for obtaining a vector representing a set of words by giving special consideration to words deemed particularly important within a document set. Calculating the TF-IDF for a set of titles enables the weights of each word in each title to be obtained. The vector of each article title was calculated by taking the weighted average of the vectors of each respective word contained in the article title using the TF-IDF as a weight.

B. GENERALIST-SPECIALIST SCORE

By incorporating components into the word embedding space, as described in the previous section, the heterogeneity of a given group of articles can be assessed. We introduce the GS-score as a metric for vector divergence. In essence, vectors of varied article collections are scattered in space, while vectors of less varied article collections are closer in space. Consequently, the GS-score measures the mean value of the cosine similarity between vectors of articles. The larger the heterogeneity, the lower this likeness is on average, so the greater the heterogeneity, the smaller the GS-score.

The GS-score of user u_i can be determined as follows,

$$\vec{\mu}_i = \frac{1}{J_i}\sum_j \vec{a}_j \tag{1}$$

$$GS\,(u_i) = \frac{1}{J_i}\sum_j \frac{\overrightarrow{aj}\cdot\vec{\mu}_i}{\|\vec{a}_j\|\cdot\|\vec{\mu}_i\|} \tag{2}$$

where $\vec{a}_j$ is a feature vector created for the article a_j clicked by the user u_i, and J_i is the number of articles regarding user u_i. In (1), the vector mean of the article set is calculated to represent the interest vector $\vec{\mu}_i$ of the user u_i. In (2), the cosine similarity between $\vec{\mu}_i$ and the vector of each article related to the user u_i is added up and divided by the article number J_i to calculate the GS-score. The GS-score takes a value between -1 and 1, where a user with a GS-score of 1 has clicked on extremely homogeneous articles, while a user with a GS-score of -1 has clicked on extremely diverse articles.

C. AVERAGE PAIRWISE DISTANCE

Similar to the GS-score, we introduce the APD as an index for assessing vector diversity [49]. The APD is calculated by averaging the distances between vectors of the articles. The distance between two articles a_i and a_j is derived using the Euclidean distance, as follows.

$$d(a_i, a_j) = \sqrt{\sum_k [a_{i,k} - a_{j,k}]^2} \tag{3}$$

TABLE 1. GS-score and APD for each vector set.

	GS-score	APD
Set 1 (red diamond)	0.828	1.61
Set 2 (yellow cross)	0.908	1.61
Set 3 (green star)	0.828	1.33

where $a_{i,k}$ represents the k^{th} element of article vector $\vec{a}_j$. Subsequently, the pairwise distance is calculated by forming two groups of articles in the collection, which does not permit any duplicates. This is referred to as the pairwise distance (PD) and its mean is calculated as the APD. Intuitively, the vectors of diverse article sets are spread across the space, while vectors of article sets having low diversity are close in the space. As the Euclidean distance increases on average with increasing diversity, the APD increases as the diversity increases.

D. COMPARISON OF GS-SCORE AND PAIRWISE DISTANCE PROPERTIES

We would like to evaluate which of the GS-score and APD is more intuitive as a measure of diversity. In comparison to the APD, the GS-score benefits from an faster processing speed, as fewer combinations of articles are necessary for calculation and it can be processed as a matrix calculation, as it relies on cosine similarity. On the other hand, GS score is calculated based on the distance from the center of gravity, so there is a concern whether it can correctly capture diversity. Consequently, we compared the characteristics of the GS-score and the APD as methods for assessing diversity.

For an intuitive understanding, three sets of vectors containing four elements were prepared.

- Set 1 (red diamond): ((0,1),(1,0),(2,1),(1,2))
- Set 2 (yellow cross): $((1+\frac{1}{\sqrt{2}},1+\frac{1}{\sqrt{2}}),(1-\frac{1}{\sqrt{2}},1+\frac{1}{\sqrt{2}}),(1-\frac{1}{\sqrt{2}},1-\frac{1}{\sqrt{2}}),(1+\frac{1}{\sqrt{2}},1-\frac{1}{\sqrt{2}}))$
- Set 3 (green star): ((2,1),(2,1),(0,1),(0,1))

Each vector is a point on a circumference of magnitude $r = 1$ situated at $(x, y) = (1, 1)$. Each point and a circle of radius $r = 1$ situated at $(x, y) = (1, 1)$ are visualized in Figure 2. The GS-score and APD are computed for each of these vector sets, and Table 1 exhibits the corresponding values.

For sets 1 and 2, the points are on a circle of radius $r = 1$ centered at $(x, y) = (1, 1)$, and the relationship rotated by 45 degrees. Intuitively, the diversity indices are expected to have the same value. However, according to Table 1, the APD has the same value, but the GS-score has a different value. In addition, set 3 is a point on the same circle, but biased on $y = 1$. Intuitively, one would expect the diversity index to have a smaller value than sets 1 and 2. However, according to Table 1, the APD is smaller for set 3 than for sets 1 and 2, but the GS-score has the same value as that of set 1.

Based on the above, we are of the opinion that the APD is a more effective measure for gauging diversity as it is more comprehensible. In this research, the APD is utilized to assess diversity.

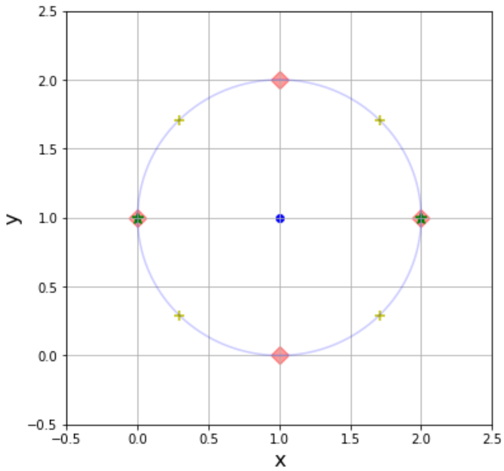

FIGURE 2. Comparison of GS-score and APD.

IV. RECOMMENDER SYSTEM TO IMPROVE DIVERSITY

In this study, we consider a recommender system that generates a variety of articles from a list of potential articles recommended by a preceding recommender system implemented in a news distribution service provided by Nikkei Inc. Previous research that sought to enhance diversity in a book recommender system [49] predominantly adopted the approach of measuring the diversity of the list and then refining it. However, the process of generating a list from candidate articles, assessing the diversity of the list, and then modifying it gives rise to a combinatorial explosion, making the formation of the list and the determination of diversity unfeasible. Henceforth, we suggest the SVR as a method for improving diversity based on the distribution of vectors. The formation of a recommendation list utilizing SVR can be executed promptly, as there is no requirement to measure and adjust the diversity of the list.

A. RECOMMENDATION ALGORITHM

We propose SVR as an approach for increasing diversity based on vector distributions. SVR is modeled on maximizing semantic volume, a technique for sentence summarization proposed by Yogatama et al [50]. An illustration of the concept of Maximizing Semantic Volume is presented in Figure 3. SVR seeks to identify news articles whose associated vectors are maximally distant from one another from among a given set of candidate news articles. This is accomplished by maximizing the supervolume of the space comprising the vectors of the selected news article set. To this end, we select articles that are as distant from each other as possible within the scope of the desired number of articles to be extracted.

The SVR implementation process is as follows:

1) Vectorize the documents.
2) Calculate the center of mass of the document set.
3) Locate the document vector farthest from the center of mass and add it to the recommendation list.
4) Select the document vector farthest from the first document vector and add it to the recommendation list.
5) Establish a reference vector from the two document vectors in the recommendation list and identify the most distant document vector. Include it in the recommendation list.
6) Compute the distance of each document extracted from the remainder of the document collection to the set of document vectors in the recommendation list, and select the farthest document vector. Include it in the recommendation list.
7) Repeat step 6 until the number of articles in the recommendation list reaches the target number of articles.

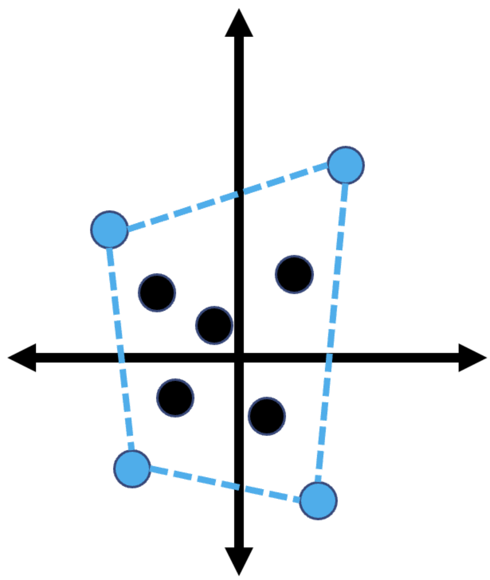

FIGURE 3. A prototypical example of nine article vectors projected onto a two-dimensional semantic space. Contemplate a scenario wherein there are four articles to be isolated. The rating function is optimized by selecting the four articles indicated by the azure dashed lines as authoritative articles. This is because it amplifies the two-dimensional region.

B. RECOMMENDATION PROCEDURE

This research targets news services that utilize existing personalized recommender systems, the particulars of which are withheld for commercial reasons. The recommendation listings for this news service are produced by collaborative filtering based on prior user browsing records and a logic that considers the editor's evaluation of newsworthiness. In this research, the data provider allows us to retrieve these recommendation lists through an API. We aim to propose a variety of articles from these recommendation listings from a set of articles that are in congruence with the user's interests by extracting a variety of articles using SVR. Specifically, the recommendation lists accessible via the API are processed as input L for Algorithm1, resulting in the final SVR-derived recommendation list R.

Algorithm 1 Semantic Volume Based Recommendation (SVR)

Require: Item List L
Ensure: Recommended Items R
Initialize $R \leftarrow \emptyset$
{Embedding}
for each item i in L **do**
  $V(i) \leftarrow \text{Embed}(i)$
end for
{Centroid Calculation}
$C \leftarrow \text{CalculateCentroid}(V(L))$
{Item Selection and Add Recommendation List}
$f_1 \leftarrow \text{FindFurthestItem}(C, V(L))$
$R \leftarrow R \cup \{f_1\}$
$f_2 \leftarrow \text{FindFurthestItem}(f_1, V(L) \setminus \{f_1\})$
$R \leftarrow R \cup \{f_2\}$
while $|R| <$ Target Number of Items **do**
  $B \leftarrow \text{DefineBasisVectors}(R)$
  $f \leftarrow \text{FindFurthestItem}(B, V(L) \setminus R)$
  $R \leftarrow R \cup \{f\}$
end while
return R

C. ANALYSIS OF RECOMMENDATION RESULTS

We tested whether the SVR-based recommendation method would facilitate the presentation of a broad spectrum of news articles. As illustrated in Table 2, the list of articles recommended to a particular user on April 1, 2022, included pieces concerning the Russian invasion of Ukraine, as well as information on increasing interest rates and prices due to distribution disruptions. For the sake of brevity, we compared a case in which five articles were extracted from this list of 20 articles using SVR with a case in which five articles were selected randomly. Table 3 displays the results of extracting five articles through the implementation of SVR. It was confirmed that only one news article related to Russia was extracted. In addition, the result of randomly extracting five articles is shown in Table 4. This table demonstrates that Russia-related articles were duplicated, and several brief articles containing market-related data were included in the table. Moreover, the APD value for the entire 20 articles was 12.27, for the proposed method was 15.09, and for random sampling was 12.02. As mentioned above, by combining the qualitative evaluation through visual inspection and the quantitative evaluation through APD, it was confirmed that the proposed SVR method was able to extract a variety of articles.

V. OFFLINE EXPERIMENT

The goal of this research is to propose a recommender system that improves the diversity of articles clicked by users and to clarify the conditions that trigger users to click on a wide range of articles. To validate the recommendation method based on SVR, presented in Section IV, we performed offline experiments with user behavior logs from a news distribution service provided by Nikkei Inc. We evaluated it using the APD outlined in Section III-C. The results revealed that the proposed SVR method can indeed improve the variety of articles viewed by users.

TABLE 2. List of articles extracted using an existing method. The original Japanese title has been translated into English.

Article Title
Bain considers acquisition of Toshiba, agreement with largest shareholder, going private.
Russian war death toll is enormous, exceeding that of U.S. "Iraq and Afghanistan."
U.S. to release an additional 1 million barrels per day of oil reserves.
Japan Airlines' Cargo dilemma floated in the wake of Ukraine invasion
Putin does not know the truth about the war situation and economic sanctions, and the military is misinformed.
NY Dow continues to fall \$550, profit taking in consumer-related stocks, etc.
Chicago PMI rises 6.6 points in March, above forecast.
Metaverse, halfway to rules, sexual harassment and trademark infringement.
Food price hikes, 90% of which penetrate into store prices, due to logistics and raw material price hikes.
Semiconductor equipment development building, Gakugei University attracts vocational school, divorce decline in China.
Mortgage loans in April, difference in fixed rates Mitsubishi UFJ Bank partially lowered.
Mitsubishi Motors to compete with South Korean fansites in Indonesia.
Nomura revises commissions: "per trade" or "balance-linked".
Omicron derivative "BA.2" mainstream worldwide, majority in the U.S.
China's economy shrinks for first time in 5 months as "invasion" cools.
JFTC strengthens screening of giant IT acquisitions, establishes new department specializing in market analysis.
Russia redeploys troops, offensive in southeastern Mariupol.
Kunio Noji, special advisor to Komatsu (1) 3/11
U.S. consumer expenditure prices rise 6.4% in February, first level in 40 years and 1 month.
OPEC Plus leaves the pace of crude oil production increase unchanged, emphasizing cooperation with Russia.

A. OFFLINE EXPERIMENTAL PROCEDURE

We adopt user behavior and impression logs of the news distribution service provided by Nikkei Inc. from March 2022, which were analyzed in the authors' previous study [51]. Among these logs, we use data from 2,000 randomly sampled users who joined the service in December 2021 and who viewed an article at least once in March.

To evaluate the recommender system, we conduct an offline experiment using impression data and click logs for

TABLE 3. List of articles by proposed method.

Article Title
Russia redeploys troops, offensive in southeastern Mariupol.
Kunio Noji, special advisor to Komatsu (1) 3/11
Food price hikes, 90% of which penetrate into store prices, due to logistics and raw material price hikes.
Nomura revises commissions: "per trade" or "balance-linked".
Bain considers acquisition of Toshiba, agreement with largest shareholder, going private.

TABLE 4. List of articles by random selection.

Article Title
Chicago PMI rises 6.6 points in March, above forecast.
U.S. consumer expenditure prices rise 6.4% in February, first level in 40 years and 1 month.
JFTC strengthens screening of giant IT acquisitions, establishes new department specializing in market analysis.
Russian war death toll is enormous, exceeding that of U.S. "Iraq and Afghanistan."
Putin does not know the truth about the war situation and economic sanctions, and the military is misinformed.

each user. For this purpose, a virtual recommendation list is created by evaluating the articles included in the impression data using the recommendation algorithm to be evaluated and by extracting articles with high scores. The articles included in this virtual recommendation list were actually displayed to the user because they are included in the impression data. The clicked articles in these recommendation lists are treated as click logs in the offline experiment. By analyzing the recommendation lists and click logs created in this way, we can analyze the click rate and diversity of the recommendation algorithms under evaluation. For business reasons, we cannot disclose the number of clicks, impressions, click rates, etc. for each user during the period under analysis. Therefore, the values of the parameters of the recommender system necessary to explain the experimental procedure are expressed as N_1, N_2.

In this offline experiment, we compare the following four recommendation algorithms:

- Recommendation V-0 Random selection
- Recommendation V-1 Sorted by the number of clicks
- Recommendation V-2 Extraction by SVR
- Recommendation V-3 Extraction by SVR from the list of top clicks

For Recommendation V-0, we randomly extracted N_1 items from the list of items featured in the impression log for each user as a baseline for comparison. In Recommendation V-1, we arranged items according to the number of clicks and selected the top N_1 items from the list of items featured in each user's impression log. Because the offline experiment cannot leverage the personalized recommendation list information present in existing services, we decided to recommend articles with a high number of clicks as an alternative. For Recommendation V-2, we employed SVR to select items from each user's impression log so the N_1 item list would be the most varied. Recommendation V-3 diversified the item list by considering the number of clicks. For the item lists featured in the impression logs for each user, we first sorted the items according to the number of clicks and chose the top N_2 items. For the N_2 item list, SVR was utilized to extract items so that the N_1 item list would be the most varied.

To regulate the parameters of every algorithm, the size of the N_1 recommendation list was normalized. The value of N_1 was set to be less than half the amount of views throughout the duration for the specified user, so that the division of elements in each recommendation list is significantly distinct. In this instance, we verified that the number of clicked items in each recommendation list would also be adequate.

B. OFFLINE ANNALYSIS

We assess the four recommendation strategies outlined in Section V-A to evaluate their performance. Figure 4 illustrates the APD values and the number of clicks for the recommendation lists generated for each recommendation, as well as the items that were clicked. The number of clicks is represented as a percentage of the number of clicks when the recommendation in Section V-0 is set to 1.

We conducted correspondence t-tests to evaluate the recommendation results for the same user group log data. Although there was no significant variation in the APD values of the recommendation list between Recommendation V-0 and Recommendation V-2, there were significant differences at the 1% level for the other combinations. In addition, we could not ascertain a conspicuous difference in the APD values of the click list between Recommendation V-1 and Recommendation V-2, but there were significant differences in the other combinations at the 1% level.

Regarding the diversity of the recommendation list, we observe that the APD value is the most prominent and diversity is highest in the case of Recommendation V-3. The diversity of clicked items also reflects that the APD value and diversity are the greatest in the case of Recommendation V-3. It can be argued that by diversifying items using SVR and considering the ease of clicking, a wide array of items can be presented to users. The highest number of clicks was detected in the case of Recommendation V-1, which presented the items with the highest number of clicks, but clicks were also abundant in the case of Recommendation V-3 compared to the random Recommendation V-0. Although it is essential to maintain the number of clicks to implement the proposed method into actual services, these results demonstrate that the method is at a stage in which it can be contemplated for implementation.

VI. ONLINE EXPERIMENT

We build an experimental service and compare the recommendation methodology by employing SVR with the recommendation technique adopted in existing services

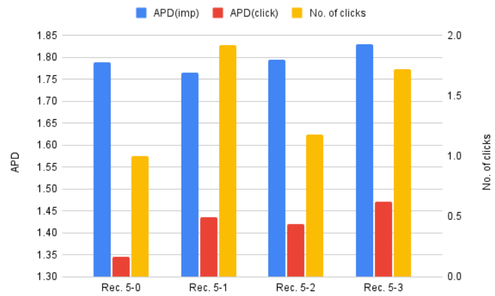

FIGURE 4. Graphs of the diversity metric APD values and the number of clicks for the recommendation and click lists. To compare the four recommendation methods proposed in Section V-A, each result is denoted as RecV-[number]. The APD value for the recommendation list is labeled as APD(imp), and for the click list as APD(click). The APD values are shown on the left as the primary axis. The number of clicks is displayed as a ratio, with the number of clicks in Recommendation V-0 being 1, and is presented on the right as the secondary axis. Recommendation V-3 had the highest APD value for the click list and was able to achieve the most diverse clicks.

through user experiments, and we examine whether the recommendation technique improves the diversity of articles browsed by users. For this purpose, we evaluate the diversity of recommendation lists and click lists using the APD proposed in Section III-C. We also analyze the impact on user engagement, such as the number of clicks and the number of login days. Then, we compare the diversity evaluation with that based on the entropy of the affiliations of the reporters who authored the articles.

A. ONLINE EXPERIMENTAL PROCEDURE

In this study, we conducted a user experiment to assess the efficacy of the recommender system. Participants consisted of registered users of a news distribution service provided by Nikkei Inc. In the experiment, we constructed a website that simulated the news list screen of the actual news distribution service, provided it to each participant with a different URL, and collected logs. The experimental period spanned two weeks, from March 8, 2022, to March 21, 2022. The total number of participants in the experiment was 413, of whom 192 responded to the pre- and post-surveys and were provided with a valid Nikkei ID. Participants were briefed on the experiment and consent to participate was obtained. They were also informed that they could withdraw from the experiment at any time. The experiment was reviewed by the Ethics Committee of the University of Tokyo. The number of participants was significantly lower than the anticipated experimental participants. This is attributed to the fact that the news service targeted in this study is a high-cost subscription service, and there were fewer users in the survey company's monitor pool with IDs than expected. Furthermore, due to the nature of the service provided by a newspaper company specializing in economic news, the news service exhibits a characteristic where the number of updates and clicks on news articles are biased toward weekdays. Considering that the number of weekdays in a two-month period is about 40, a threshold for the number of clicks was identified based on the distribution of the number of clicks in the target user group during this period, and we extracted users above a certain threshold, resulting in 47 users. For business reasons, specific click values cannot be disclosed.

In this study, we compared the proposed method, which uses SVR to improve the diversity of the article set, with an existing method built on current news distribution services. Participants were randomly assigned to an experimental and a control group to receive recommendations by one of the two recommendation methods. The number of participants in each group is shown in Table 5. There was no significant difference in the dropout rates between the experimental and control groups.

B. RECOMMENDATION PROCEDURE

The news service under evaluation here implements a personalized recommender system. For business reasons, the specific algorithm has not been disclosed to us. The recommender system utilizes collaborative filtering based on past user browsing logs, in conjunction with logic that considers the editor's appraisal of newsworthiness, to generate a recommendation list.

In our experiment, we were issued with a daily list of recommendations by Nikkei Inc. The control group received 20 recommendations from the existing system and presented them as they were. Meanwhile, the experimental group was given a list of 50 recommendations from the existing system, of which 20 were extracted and presented using SVR.

C. ANALYSIS OF DIVERSITY OF RECOMMENDATION LISTS AND CLICK LISTS

We evaluated the diversity of the recommendation and click lists during the experiment using APD. Figure 5 shows the APD values of the recommendation and click lists for the experimental and control groups, respectively. There was a significant difference between the experimental and control groups at the 1% level. The results for the user group limited to users with a sufficient number of past clicks are also shown, where there was a significant difference between the experimental and control groups at the 5% level.

Confirming the diversity of the recommendation lists, the APD values of the experimental group are greater, both when constrained by the number of past clicks and when not limited by the number of past clicks, indicating that the proposed method is capable of recommending a variety of articles during the experimental period, similar to the findings of Section IV-C. Moreover, when the diversity of the click list is assessed, the APD of the experimental group is larger and more varied when the list is limited by the number of past clicks. Conversely, when the number of past clicks is not limited, the APD values of the experimental group is smaller and less diverse, implying that if the number of past clicks is sufficient and the existing recommender system is effective in extracting candidate articles, the proposed method using SVR can improve the diversity of articles clicked by users. From this result, it can be concluded that the proposed method

TABLE 5. The number of participants in the experiment. Approximately half the participants in the experiment provided valid Nikkei IDs in response to the pre- and post-surveys.

	No. of valid participants	No. of valid participants with sufficient past clicks	No. of invalid participants
Experimental group	99	22	106
Control group	93	25	115

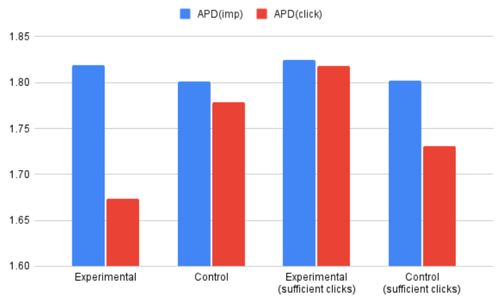

FIGURE 5. The APD values for the recommendation list and click list during the experimental period.

using SVR can encourage users to click on a variety of articles in an environment in which candidate articles that are likely to be clicked are successfully extracted, while simply recommending a variety of articles will not result in clicks.

D. ACCURACY AND ENGAGEMENT METRICS

In Section VI-C, we confirmed that the proposed method using SVR can encourage users to click on a variety of articles when the number of past clicks is sufficient. In this section, we evaluate the impact of the proposed method on short-term user engagement metrics, such as the number of clicks and the number of login days, as well as the accuracy of recommendations, for a group of users with a sufficient number of past clicks. If these user engagement metrics, including the number of clicks and login days, are found to be decreasing, it could be problematic for business implementation, even if diversity is improved. This situation would make it challenging to implement the proposed method in actual services.

Table 6 presents the precision, recall, and F1-SCORE for both the experimental and control groups. Since all users view the recommended results, the recall is 1. No significant decrease was observed in precision or F1-score when SVR was used. Figure 6 shows the mean number of clicks and login days for the experimental and control groups, respectively. No significant differences were observed in either the number of clicks or the number of login days, indicating that there was no clear deterioration in user engagement when the proposed method was introduced, and that the proposed method is of a suitable quality to contemplate its deployment in actual services. In practice, it is important to balance the trade-off between engagement metrics and diversity, and to appropriately select the size of recommendation lists L and R in Algorithm 1, as described in Section IV-B.

TABLE 6. Precision, recall, and F1-SCORE for the experimental and control groups.

	precision	recall	F1-SCORE
Experimental Group	0.134	1.0	0.237
Control Group	0.139	1.0	0.244

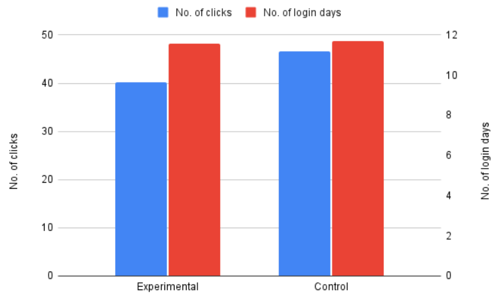

FIGURE 6. Mean number of clicks and login days for experimental and control groups.

E. ANALYSIS OF DIVERSITY OF REPORTER AFFILIATIONS

The news articles of the news services in this study contain the affiliation information of the reporters who wrote them as meta-information. We evaluate the diversity of the affiliation information of the reporters for the recommendation and click lists during the experimental period for a group of users with a sufficient number of past clicks. To evaluate this diversity, we draw on existing research regarding the diversity of items consumed by users, utilizing entropy as a measure [43], [44].

Figure 7 shows the mean entropy of the affiliations of the reporters for the experimental and control groups. Analysis of the diversity of the recommendation lists reveals that the control group yielded a higher entropy value at the 1% level of significance, implying that the affiliations of the reporters who wrote the articles are more diverse in the control group. In contrast, when assessing the diversity of the click list, there was no significant difference between the experimental and control groups. Unlike the evaluation by APD in Section VI-C, the control group had a higher diversity of recommendation lists, likely because the control group is presented as-is as a result of the existing recommender system, which facilitates the selection of items equally from multiple departments according to the editor's intention. On the other hand, the diversity of the click list did not improve, even when a recommendation list with a high diversity of reporter affiliations was given. In other words, simply presenting articles by reporters of various affiliations does not necessarily result in clicks.

This result implies that the algorithm of equalizing the departments from which articles are written according to the editor's intention is unable to change users' browsing behavior. Not only diversity based on the language vector of article titles, but also macro diversity with respect to large labels, such as the affiliation of the reporter who wrote the article, is important when discussing diversity in user behavior when consuming news. In comparison to the improvement in diversity based on the linguistic vector of article titles, it is inferred that improving macro diversity requires a mechanism that considers the extent of similarities between each label.

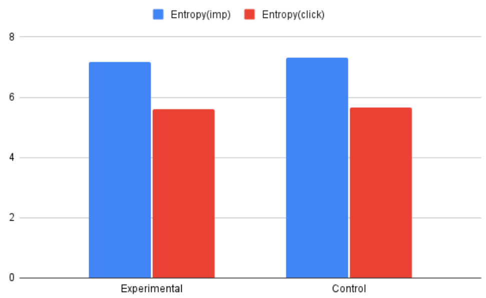

FIGURE 7. Mean entropy of the affiliations of the reporters for the experimental and control groups. The value of entropy for the recommendation list is denoted as entropy(imp), and that for the click list as entropy(click).

VII. EVALUATION OF MOVIELENS DATA

To demonstrate the effectiveness of the recommender system using SVR beyond news data, an offline experiment was conducted on the MovieLens 1M dataset [52]. As a result, we confirmed that the SVR enables users to select a variety of items.

A. OFFLINE EXPERIMENTAL PROCEDURE

We focus on the user review data and movie information contained in the MovieLens 1M dataset. These files include 1,000,209 reviews for 3,900 movies by 6,040 members who joined MovieLens in the year 2000. To evaluate the recommender system, an offline experiment using each user's review log is conducted. 75% of the review logs are used for building the recommendation system, and the remaining 25% are used as test data. Thereafter, the targeted recommendation algorithms evaluate items, and a virtual recommendation list is created by extracting items with high scores. Items in this virtual recommendation list that have actual review data are treated as click logs in the offline experiment. By analyzing these recommendation lists and click logs, the click-through rate and diversity of the evaluated recommendation algorithms are analyzed.

In this offline experiment, the following three recommendation algorithms are compared:

- Recommendation VII-0 Random extraction
- Recommendation VII-1 Recommendation using SVD (Singular Value Decomposition) [53]

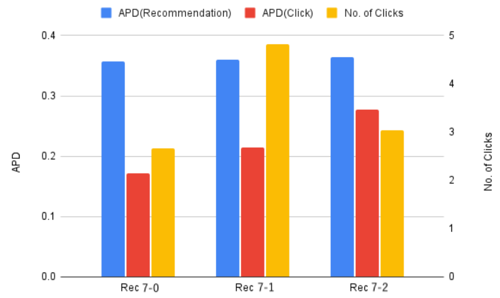

FIGURE 8. Graphs of the diversity metric APD values and the number of clicks for the recommendation and click lists. To compare the four recommendation methods proposed in Section VII-A, each result is denoted as RecVII-[number]. The APD value for the recommendation list is labeled as APD(Recommendation), and for the click list as APD(click). The APD values are shown on the left as the primary axis. The number of clicks is presented on the right as the secondary axis. Recommendation V-2 had the highest APD value for the click list and was able to achieve the most diverse clicks.

- Recommendation VII-2 Extraction of diverse items using SVR from a recommendation list generated by SVD

For Recommendation VII-0, as a baseline for comparison, 25 items were randomly extracted from a list of items not included in each user's training data. In Recommendation VII-1, 25 items were extracted by evaluating items using Singular Value Decomposition (SVD) with the Python library surprise [54], used for building and analyzing recommender systems. In Recommendation VII-2, 100 items were initially extracted using SVD, and then 25 items were further extracted using SVR. To control for conditions across algorithms, the length of the recommendation lists was standardized to 25 items.

B. OFFLINE ANALYSIS

We assess the three recommendation strategies outlined in Section VII-A to evaluate their performance. Figure 8 illustrates the APD values and the number of clicks for the recommendation lists generated for each recommendation, as well as the items that were clicked.

These results indicate that diversifying the recommendations obtained through SVD by using SVR allows us to provide a more diverse range of items to the user. However, the number of clicks has decreased more significantly compared to the news article case in Section V. This is presumed to be due to the broader range of item choices in movie reviews compared to news services, and the smaller absolute number of clicks, making the impact of missing a single item more significant. From these results, it is necessary to consider factors such as the absolute number of clicks, but the increase in diversity of selected items through the proposed method is confirmed to not be limited to news services.

VIII. CONCLUSION

In this study, we analyzed user behavior on the news service provided by Nikkei Inc. and an online experiment on an experimental website prepared for this study, which concerns the diversity of articles clicked. The contributions of this study are as follows. To improve the diversity in recommendations, we proposed a recommender system, SVR, which is designed to extract articles as far apart as possible from the language space. Through static and user experiments, we evaluated the SVR recommender system for improving diversity, and we showed that it can recommend a wide variety of articles through qualitative and quantitative assessments. We also showed that the proposed method using SVR is effective in improving the diversity of articles clicked by users in an environment in which clickable article candidates are well extracted.

From the results of this study, it was found that to encourage users to click on diverse articles, it is not enough simply to present diverse articles; rather, it is necessary to recommend articles that attract users' interests and are diverse in nature. For future work, we would like to develop a recommender system that considers not only the linguistic diversity of article titles but also diversity in terms of the affiliations of the reporters who wrote the articles.

REFERENCES

[1] E. Pariser, *The Filter Bubble: What Internet is Hiding From You*. London, U.K.: Penguin, 2011.

[2] T. T. Nguyen, P.-M. Hui, F. M. Harper, L. Terveen, and J. A. Konstan, "Exploring the filter bubble: The effect of using recommender systems on content diversity," in *Proc. 23rd Int. Conf. World Wide Web*, Apr. 2014, pp. 677–686.

[3] A. J. B. Chaney, B. M. Stewart, and B. E. Engelhardt, "How algorithmic confounding in recommendation systems increases homogeneity and decreases utility," in *Proc. 12th ACM Conf. Recommender Syst.*, Sep. 2018, pp. 224–232.

[4] X. Zhao, Z. Zhu, and J. Caverlee, "Rabbit holes and taste distortion: Distribution-aware recommendation with evolving interests," in *Proc. Web Conf.*, Apr. 2021, pp. 888–899.

[5] I. Waller and A. Anderson, "Generalists and specialists: Using community embeddings to quantify activity diversity in online platforms," in *Proc. World Wide Web Conf.*, May 2019, pp. 1954–1964.

[6] A. Anderson, L. Maystre, I. Anderson, R. Mehrotra, and M. Lalmas, "Algorithmic effects on the diversity of consumption on spotify," in *Proc. Web Conf.*, Apr. 2020, pp. 2155–2165.

[7] C. A. Gomez-Uribe and N. Hunt, "The Netflix recommender system: Algorithms, business value, and innovation," *ACM Trans. Manage. Inf. Syst.*, vol. 6, no. 4, pp. 1–19, Jan. 2016.

[8] Nikkei Inc. *Nikkei Media Data*. Accessed: Oct. 7, 2023.

[9] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, and J. Riedl, "GroupLens: An open architecture for collaborative filtering of netnews," in *Proc. ACM Conf. Comput. Supported Cooperat. Work (CSCW)*, 1994, pp. 175–186.

[10] S. Rendle, C. Freudenthaler, Z. Gantner, and L. Schmidt-Thieme, "BPR: Bayesian personalized ranking from implicit feedback," 2012, *arXiv:1205.2618*.

[11] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, "Item-based collaborative filtering recommendation algorithms," in *Proc. 10th Int. Conf. World Wide Web*, Apr. 2001, pp. 285–295.

[12] S. Rendle, C. Freudenthaler, and L. Schmidt-Thieme, "Factorizing personalized Markov chains for next-basket recommendation," in *Proc. 19th Int. Conf. World Wide Web*, Apr. 2010, pp. 811–820.

[13] P. Lops, M. D. Gemmis, and G. Semeraro, "Content-based recommender systems: State of the art and trends," in *Recommender Systems Handbook*. Boston, MA, USA: Springer, 2011, pp. 73–105.

[14] H. Steck, "Training and testing of recommender systems on data missing not at random," in *Proc. 16th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Jul. 2010, pp. 713–722.

[15] X. Amatriain and J. Basilico, "Recommender systems in industry: A Netflix case study," in *Recommender Systems Handbook*. Boston, MA, USA: Springer, 2015, pp. 385–419.

[16] G. Zheng, F. Zhang, Z. Zheng, Y. Xiang, N. J. Yuan, X. Xie, and Z. Li, "DRN: A deep reinforcement learning framework for news recommendation," in *Proc. World Wide Web Conf. (WWW)*, 2018, pp. 167–176.

[17] J. Liu, P. Dolan, and E. R. Pedersen, "Personalized news recommendation based on click behavior," in *Proc. 15th Int. Conf. Intell. User Interfaces*, Feb. 2010, pp. 31–40.

[18] Z. Wang, J. Liao, Q. Cao, H. Qi, and Z. Wang, "Friendbook: A semantic-based friend recommendation system for social networks," *IEEE Trans. Mobile Comput.*, vol. 14, no. 3, pp. 538–551, Mar. 2015.

[19] D. R. Fesenmaier, K. W. Wöber, and H. Werthner, *Destination Recommendation Systems: Behavioural Foundations and Applications*. Cabi, 2006.

[20] M. Grbovic and H. Cheng, "Real-time personalization using embeddings for search ranking at Airbnb," in *Proc. 24th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Jul. 2018, pp. 311–320.

[21] T. Chatzidimitris, D. Gavalas, V. Kasapakis, C. Konstantopoulos, D. Kypriadis, G. Pantziou, and C. Zaroliagis, "A location history-aware recommender system for smart retail environments," *Pers. Ubiquitous Comput.*, vol. 24, no. 5, pp. 683–694, Oct. 2020.

[22] B. Smith and G. Linden, "Two decades of recommender systems at Amazon.com," *IEEE Internet Comput.*, vol. 21, no. 3, pp. 12–18, May 2017.

[23] Y. Wang, L. Tao, and X. X. Zhang, "Recommending for a multi-sided marketplace with heterogeneous contents," in *Proc. 16th ACM Conf. Recommender Syst.*, Sep. 2022, pp. 456–459.

[24] Z. Zhao, L. Hong, L. Wei, J. Chen, A. Nath, S. Andrews, A. Kumthekar, M. Sathiamoorthy, X. Yi, and E. Chi, "Recommending what video to watch next: A multitask ranking system," in *Proc. 13th ACM Conf. Recommender Syst.*, Sep. 2019, pp. 43–51.

[25] S. Zhang, L. Yao, A. Sun, and Y. Tay, "Deep learning based recommender system: A survey and new perspectives," *ACM Comput. Surv.*, vol. 52, no. 1, pp. 1–38, Jan. 2020.

[26] P. Covington, J. Adams, and E. Sargin, "Deep neural networks for Youtube recommendations," in *Proc. 10th ACM Conf. Recommender Syst.*, Sep. 2016, pp. 191–198.

[27] R. Zhou, S. Khemmarat, and L. Gao, "The impact of Youtube recommendation system on video views," in *Proc. 10th ACM SIGCOMM Conf. Internet Meas.*, Nov. 2010, pp. 404–410.

[28] C. Hansen, R. Mehrotra, C. Hansen, B. Brost, L. Maystre, and M. Lalmas, "Shifting consumption towards diverse content on music streaming platforms," in *Proc. 14th ACM Int. Conf. Web Search Data Mining*, Mar. 2021, pp. 238–246.

[29] R. P. Karumur, T. T. Nguyen, and J. A. Konstan, "Personality, user preferences and behavior in recommender systems," *Inf. Syst. Frontiers*, vol. 20, no. 6, pp. 1241–1265, Dec. 2018.

[30] A. Ansari, Y. Li, and J. Z. Zhang, "Probabilistic topic model for hybrid recommender systems: A stochastic variational Bayesian approach," *Marketing Sci.*, vol. 37, no. 6, pp. 987–1008, Nov. 2018.

[31] F. Sanna Passino, L. Maystre, D. Moor, A. Anderson, and M. Lalmas, "Where to next? A dynamic model of user preferences," in *Proc. Web Conf.*, Apr. 2021, pp. 3210–3220.

[32] B. Xiao and I. Benbasat, "E-commerce product recommendation agents: Use, characteristics, and impact," *MIS Quart.*, vol. 31, no. 1, p. 137, 2007.

[33] B. P. Knijnenburg, M. C. Willemsen, Z. Gantner, H. Soncu, and C. Newell, "Explaining the user experience of recommender systems," *User Model. User-Adapted Interact.*, vol. 22, nos. 4–5, pp. 441–504, Oct. 2012.

[34] D. Holtz, B. Carterette, P. Chandar, Z. Nazari, H. Cramer, and S. Aral, "The engagement-diversity connection: Evidence from a field experiment on spotify," in *Proc. 21st ACM Conf. Econ. Comput.*, Jul. 2020, pp. 75–76.

[35] J. Zhang, G. Adomavicius, A. Gupta, and W. Ketter, "Consumption and performance: Understanding longitudinal dynamics of recommender systems via an agent-based simulation framework," *Inf. Syst. Res.*, vol. 31, no. 1, pp. 76–101, Mar. 2020.

[36] P. Cremonesi, Y. Koren, and R. Turrin, "Performance of recommender algorithms on top-N recommendation tasks," in *Proc. 4th ACM Conf. Recommender Syst.*, Sep. 2010, pp. 39–46.

[37] A. H. Jadidinejad, C. Macdonald, and I. Ounis, "The Simpson's paradox in the offline evaluation of recommendation systems," *ACM Trans. Inf. Syst.*, vol. 40, no. 1, pp. 1–22, Jan. 2022.

[38] R. Cañamares and P. Castells, "Should I follow the crowd? A probabilistic analysis of the effectiveness of popularity in recommender systems," in *Proc. 41st Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, 2018, pp. 415–424.

[39] E. Mena-Maldonado, R. Cañamares, P. Castells, Y. Ren, and M. Sanderson, "Popularity bias in false-positive metrics for recommender systems evaluation," *ACM Trans. Inf. Syst.*, vol. 39, no. 3, pp. 1–43, Jul. 2021.

[40] J. Carbonell and J. Goldstein, "The use of MMR, diversity-based reranking for reordering documents and producing summaries," in *Proc. 21st Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, Aug. 1998, pp. 335–336.

[41] Q. Villermet, J. Poiroux, M. Moussallam, T. Louail, and C. Roth, "Follow the guides: Disentangling human and algorithmic curation in online music consumption," in *Proc. 15th ACM Conf. Recommender Syst.*, Sep. 2021, pp. 380–389.

[42] D. Fleder and K. Hosanagar, "Blockbuster culture's next rise or fall: The impact of recommender systems on sales diversity," *Manage. Sci.*, vol. 55, no. 5, pp. 697–712, May 2009.

[43] M. De Choudhury, S. Counts, and M. Czerwinski, "Identifying relevant social media content: Leveraging information diversity and user cognition," in *Proc. 22nd ACM Conf. Hypertext Hypermedia*, Jun. 2011, pp. 161–170.

[44] L. Qin and X. Zhu, "Promoting diversity in recommendation by entropy regularizer," in *Proc. 23rd Int. Joint Conf. Artif. Intell.*, 2013.

[45] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," 2013, *arXiv:1301.3781*.

[46] T. Kudo, K. Yamamoto, and Y. Matsumoto, "Applying conditional random fields to Japanese morphological analysis," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2004, pp. 230–237.

[47] S. Toshinori, "Neologism dictionary based on the language resources on the web for Mecab," Tech. Rep., 2015.

[48] M. Suzuki, K. Matsuda, S. Sekine, N. Okazaki, and K. Inui, "A joint neural model for fine-grained named entity classification of Wikipedia articles," *IEICE Trans. Inf. Syst.*, vol. E101.D, no. 1, pp. 73–81, 2018.

[49] C.-N. Ziegler, S. M. McNee, J. A. Konstan, and G. Lausen, "Improving recommendation lists through topic diversification," in *Proc. 14th Int. Conf. World Wide Web (WWW)*, 2005, pp. 22–32.

[50] D. Yogatama, F. Liu, and N. A. Smith, "Extractive summarization by maximizing semantic volume," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2015, pp. 1961–1966.

[51] A. Sonoda, F. Toriumi, H. Nakajima, and M. Gouji, "Analysis and modeling of behavioral changes in a news service," in *Proc. IEEE/WIC/ACM Int. Conf. Web Intell. (WI)*, Dec. 2018, pp. 73–80.

[52] F. M. Harper and J. A. Konstan, "The MovieLens datasets: History and context," *ACM Trans. Interact. Intell. Syst.*, vol. 5, no. 4, pp. 1–19, Jan. 2016.

[53] S. Zhang, W. Wang, J. Ford, F. Makedon, and J. Pearlman, "Using singular value decomposition approximation for collaborative filtering," in *Proc. 7th IEEE Int. Conf. E-Commerce Technol. (CEC)*, Jul. 2005, pp. 257–264.

[54] N. Hug, "Surprise: A Python library for recommender systems," *J. Open Source Softw.*, vol. 5, no. 52, p. 2174, Aug. 2020.

ATOM SONODA received the B.Eng. and M.Eng. degrees from The University of Tokyo, Japan, in 2017 and 2019, respectively, where he is currently pursuing the D.Eng. degree. His research interests include news analysis and recommender systems.

FUJIO TORIUMI received the B.E., M.E., and D.E. degrees in engineering from Tokyo Institute of Technology, Tokyo, Japan, in 1999, 2001, and 2004, respectively. In 2004, he joined the Graduate School of Information Science, Nagoya University, Aichi, Japan. In 2012, he joined the School of Engineering, The University of Tokyo, Tokyo, as an Associate Professor, where he has been a Professor, since 2021. His current research interests include computational social science and artificial intelligence for society. He is a member of ACM, the Information Processing Society of Japan (IPSJ), Japan Society of Artificial Intelligence (JSAI), The Institute of Electronics, Information and Communication Engineers (IEICE), and The Society of Socio-Informatics (SSI), and a Board Member of Japan Institute of Law and Information Systems. Since 2018, he has been an Area Editor of *New Generation Computing*. Since 2016, he has been the main Chair of the Workshop on Application of Big Data for Computational Social Science. He was the Sponsor-Chair of PRIMA2018, PC of www2016, and WI2016-2022.

HIROTO NAKAJIMA received the B.S., M.S., and Ph.D. degrees in theoretical physics from The University of Tokyo, in 2004, 2006, and 2009, respectively. He is currently the Chief Scientist of Nikkei Inc., the largest economic media group in Japan. In 2017, he joined the Nikkei Inc., where he launched the Innovation Laboratory. He leads many research and development projects in the laboratory which serves the whole group, including newspapers, magazines, TVs, financial data services, and the market index Nikkei 225. He has been a spokesman for the lab's activities at international media conferences in Asia, Africa, Europe, and the Middle East. He was awarded by Japan Press Net for his contribution to data journalism, in 2019.