Electronic Preprint for Journal of Information Processing Vol.30

Regular Paper

Content-based Stock Recommendation Using Smartphone Data

Kohsuke Kubota[1,a)] Hiroyuki Sato[1] Wataru Yamada[1] Keiichi Ochiai[1] Hiroshi Kawakami[1]

Received: August 5, 2021, Accepted: February 4, 2022

Abstract: The number of investors holding risky assets in Japan is much lower than that in western countries even though it is an effective way for building investor assets. Although Japanese investment companies offer a service to invest in points through coalition loyalty programs instead of actual currency to address this situation, the problem still persists. One reason for this is that novice investors do not know in which stocks to invest. One possible solution is recommending stocks; however, we still face the cold-start problem because there is no transaction history for novice investors. In this study, we propose a novel content-based recommendation approach that utilizes touchpoint information, *e.g.*, payment and app usage data, on smartphones in daily life. This approach employs user-weighted recency, frequency, and monetary, called UW-RFM and a complementary module to comply with Japanese guidelines that prohibit presenting only a small number of companies and establishing a minimum number of companies to be presented. We conduct an online evaluation to validate the effectiveness of the proposed approach in an actual investment service. The evaluation results show that the proposed approach motivates users to invest more, *i.e.*, 0.352 more clicks on the recommendation area and 3,016 points (yen), than the baseline method that does not consider touchpoint information.

Keywords: stock recommendation, propensity score matching, smartphone data

1. Introduction

In Japan, holding risky assets such as stocks and investment trusts is much less common than in the United States and the euro area [34]. According to a report by the Bank of Japan [34], risky assets account for 13% of household financial assets in Japan, which is lower than the 44.8% for investors in the United States and 25.9% in the euro area.

Investing in risky assets can be an effective way for individual investors to build their assets in the long run [7]. This is because stock prices are thought to follow a random walk with positive drift, and risky assets are thought to have a higher expected return rate than savings in long-term investments [7]. Japanese Financial Authorities encourage Japanese people to use iDeco [*1,*2], which is a private-pension plan governed by the Defined Contribution Pension Act, in order to generate long-term assets. Thus, investing in risky assets can be an effective means of long-term asset building for individual investors, although it does not always guarantee the principal. On the other hand, it is important for individual investors to invest in the stocks of companies because it leads to growth in the economy as a whole [15]. If individual investors increase their investment in stocks and the supply of funds to corporations expands, this will lead to economic growth on the whole [15].

Japanese companies provide investment services that allow users to invest using reward points earned through coalition loyalty programs (CLPs) to encourage them to hold risky assets [6] [*3,*4,*5]. According to a survey [32], 25.4% of corporate employees would begin investing if there were a function that allowed them to invest their pocket change or reward points. This survey suggests that investment using reward points may urge individuals to invest. However, the survey also shows that 20.1% of respondents opened accounts but did not trade because they did not know which stocks to buy [32]. This survey indicates that it is difficult for individual investors to select which stocks to buy. Therefore, investors need support in terms of investment participation and in selecting stocks. One possible solution is recommending stocks. However, we still face the cold-start problem [1], [38] because there is no transaction history for them.

To address the cold-start problem, we focus on the fact that existing studies have shown that individual investors tend to invest in stocks with which they are familiar [5], [14], [17], [24], [33]. Considering this characteristic, it is possible to promote user investment behavior by presenting companies with which they are familiar. Although the existing studies [5], [14], [17], [24] have analyzed this tendency using their transaction history, the effectiveness for recommendation, especially for the cold-start problem, has not been revealed. In addition, Prast et al. [33] has shown that presenting companies that appear in advertisements in women magazines, which women are likely to be familiar with through questionnaires, promotes the decision-making process of

[1] NTT DOCOMO, INC., Chiyoda, Tokyo 100–6150, Japan
[a)] kousuke.kubota.xt@nttdocomo.com

[*1] https://www.ideco-koushiki.jp/english/
[*2] https://www.fsa.go.jp/singi/singi_kinyu/market_wg/siryou/20190412/02.pdf
[*3] https://froggy.smbcnikko.co.jp/promo/lp_account/
[*4] https://www.rakuten-sec.co.jp/web/lp/point_investment/02/
[*5] https://www.sbineomobile.co.jp/lp/point01/

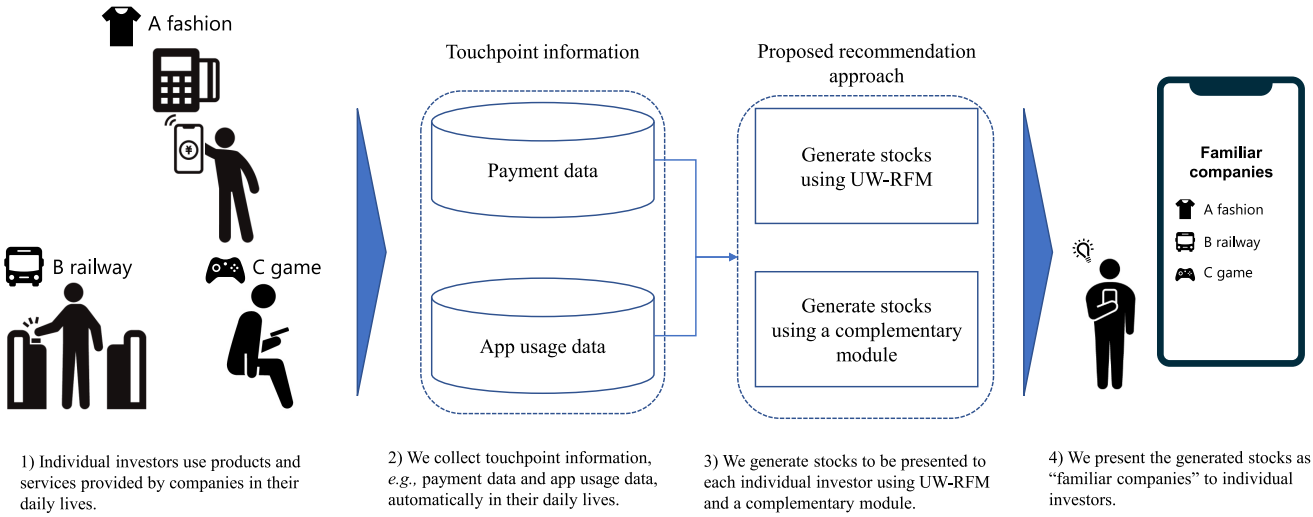

Fig. 1 Overview of content-based recommendation approach.

women. However, the degree of familiarity of users with companies may differ from user to user, and the effectiveness of presenting stocks based on the degree of familiarity of each user with companies has not been clarified. We consider a novel approach that identifies and presents user-familiar companies on their use of those company products or services in their daily lives in order to understand the degree of familiarity with companies that differs from user to user. Here, we define the companies that they use in their daily lives as companies with a touchpoint (CTs). To actualize the idea of presenting stocks based on their CTs, we need to know those companies.

A simple method for understanding individual investor CTs is to use a questionnaire survey; however, there are three challenges facing these methods. First, the number of choices is excessively large, and the response process is burdensome when users are asked to select their CTs in a questionnaire survey. Since there are approximately 4,000 companies listed on the Tokyo Stock Exchange, searching for CTs from the entire list of listed companies is burdensome. This leads to users withdrawing from the survey. Second, the apparent dissociation between listed company names and the brand names that users are familiar with usually results in missing answers. For example, the company that operates the apparel brand UNIQLO is listed as FAST RETAILING CO., LTD., and users maybe unaware of the correspondence between the listed company and the brand name product.

Moreover, when we present stocks to users through an investment service, we need to adhere to the "Guidelines Concerning Advertising, etc." [25]. According to these guidelines, in the case of presenting stocks selected by the service on its website, the service is required to avoid excessive sales of a small number of stocks to an unspecified number of investors. This is because the formation of fair prices may be impaired by large-scale purchases of the presented stocks. Although there is no clear standard for the guideline, it is necessary to consider stock bias as much as possible. We define this problem as the mass recommendation sales problem. In addition, these guidelines state that service providers of investment services should display five or more stocks with the basis for selecting those stocks when displaying them on their websites [25]. Therefore, we should generate N ($N \geq 5$) stocks with a recommender system for an investment service based on the current guidelines in Japan. We define the problem as a display problem.

In this study, we propose a novel content-based stock recommendation approach that utilizes payment data and app usage data obtained on users' smartphones in daily life, which we define as touchpoint information. **Figure 1** shows an overview of the proposed content-based recommendation approach. This approach consists of two modules: user-weighted recency, frequency, and monetary (UW-RFM) calculating the degree of investor familiarity, which we define as the familiarity score, and a complementary module that compensates for the lack of stocks for users who do not have the predetermined number of stocks.

This approach deals with the above challenges in the following manner. (1) This approach addresses the cold-start problem using touchpoint information obtained from investors' smartphones in daily life. (2) Compared to a questionnaire, this approach does not require investors to list explicitly the CTs because touchpoint information is automatically accumulated in their daily lives. (3) This approach enables the identification of listed companies whose brand name products and listed companies are dissociated and listed companies producing the products users use without being aware. We achieve this identification by using a mapping table that maps CTs to listed companies. This results in no omissions because CTs are automatically estimated from touchpoint information. (4) This approach enables expansion of the recommendation set and considers reducing stock bias using multiple input data. (5) Any user can secure a predetermined number of stocks.

We evaluate the performance of the proposed approach using real-world datasets collected from an investment service. We conduct an observational study using propensity score matching to verify the effectiveness of the proposed approach. The results show that the proposed approach promotes user investment behavior.

The contributions of the paper are given below.

- We propose a novel content-based stock recommendation

approach that utilizes touchpoint information obtained from users' smartphones in daily life.

- We implement a novel stock recommender system employing UW-RFM and a complementary module in an investment service and we verify the effectiveness of the proposed approach. Specifically, we conduct an observational study using propensity score matching and find that the number of monthly clicks to the recommendation area increases by 0.352 times and the number of monthly stock purchase points increases by 3,016 with the collected data on the investment service.

2. Preliminaries

Our goal is to identify the top N companies that individual investors are more familiar with among the CTs from their smartphones. To generate company stock recommendations, we use payment data, such as credit card and reward point data, and app usage data from users' smartphones. These types of data relate to product purchases or services and app usage data, and enable us to understand investor CTs. Here, we regard this problem as identifying the top N companies for each user based on the familiarity score using his/her touchpoint information and relevance calculated using the touchpoint information of the other companies.

We formally define the problem below:

Definition 1 (Payment Data and App Usage Data) Each payment data, pay, has user u, timestamp t, service s that is the store where the user make a payment, settlement amount p, and service provider c. Since service s represents a payment store, if service s consists of two stores, such as a restaurant in a commercial facility, two records are generated where u, t, s, and p are the same and c is different. App usage data app consists of user u, timestamp t, package name s representing the application used, app usage time a, and application provider c. In other words, these are represented as $pay = (u, t, s, p, c)$ and $app = (u, t, s, a, c)$, respectively.

Definition 2 (Payment Sequence and App Usage Sequence) We define payment sequence x_{pay} as a series of payment data pay arranged in ascending order in a time series, and app usage sequence x_{app} as a series of app usage data app arranged in ascending order in a time series. In other words, these are represented as $x_{pay} = \{pay_{t_0}, pay_{t_1}, \cdots, pay_{t_l}\}$ and $x_{app} = \{app_{t_0}, app_{t_1}, \cdots, app_{t_m}\}$, respectively.

Definition 3 (Problem Identifying the Top N Companies) Given payment sequence x_{pay} and app usage sequence x_{app}, the goal is to identify the top N companies for each user based on his/her familiarity score and relevance calculated using the other sequences of x_{pay} and x_{app}. Specifically, we aim to find a function that calculates a score that expresses the familiarity score with provider company c and generates N companies. If the number of companies is less than N, we aim to find a function that calculates the relevance between user u and company c, and generates N companies. This process allows us to identify the top N companies for each user.

3. Proposed Approach

In this section, we describe the proposed content-based stock recommendation approach in detail. This approach comprises UW-RFM to deal with the mass recommendation sales problem and a complementary module to deal with the display problem. We describe UW-RFM in detail after we explain naive RFM and explain the complementary module. **Figure 2** shows a flowchart for the stock recommender approach.

In this paper, we define each element of RFM as given hereafter. Recency (R) represents the value for the period from the last usage date to the time R is calculated for the usage cycle of the target company for each user. R is calculated as

$$\mathrm{R} = \frac{usage\_inteval}{((total\_days + 1) - usage\_interval) \;/\; total\_usage\_num} \tag{1}$$

where $usage\_interval$ is the number of days from the last usage date to a certain date, $total\_days$ is the total number of days in the aggregation period, and $total\_usage\_num$ is the number of times a user used the company during the aggregation period ($total\_usage\_num \geq 1$). User R for a company with no usage at all is set to a large value. The lower the R value, the more recently the user has used the target company. Frequency (F) represents the number of times the user used the company within a certain period. F is calculated as

$$\mathrm{F} = total\_usage\_num \tag{2}$$

The higher the F value, the more often the user used the company product or service. Monetary (M) represents the average payment to the target company of each user within a certain period. M is calculated as

$$\mathrm{M} = \frac{total\_amount}{total\_usage\_num} \tag{3}$$

where $total\_amount$ is the usage amount of the user, which represents the usage amount in the case of payment data or the usage

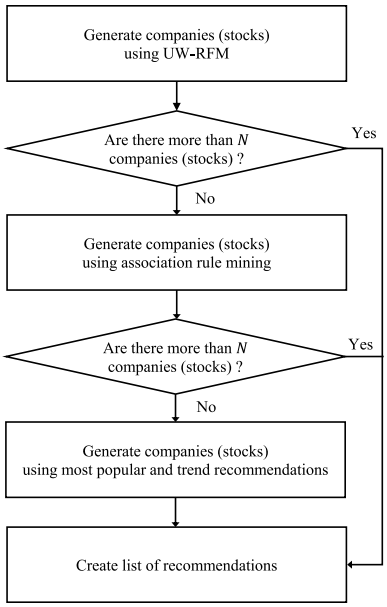

Fig. 2 Flowchart for proposed stock recommendation approach.

time in the case of app usage data. The higher the M value, the more the user pays to the target company. For the app usage data, we use the usage time, which we define as the average usage time to the target company of each user within a certain period.

3.1 Naive RFM

RFM scoring is a standard scoring method that employs usage tendencies of users [9], [30]. The most common scoring method is to sort users in descending order (best to worst) [30]. Specifically, this method divides users into q groups by ordering them in terms of the R, F, and M variables and assigns a $q - i + 1$ ($i = 1, 2, \cdots, q$) score to the user group with the highest rank and a score of 1 to the user group with the lowest rank [9], [30]. Thus, the familiarity score of each user with a company in this method is calculated as

$$familiarity\_score_{u,c} = \sum_{d \in D} (R\_score_{u,c,d} + F\_score_{u,c,d} + M\_score_{u,c,d}) \quad (4)$$

where $familiarity\_score_{u,c}$ is the familiarity score of user u with company c, D is a set of data sources, d is each dataset such as payment data and app usage data, $R\_score_{u,c,d}$ is the R property score ($1 \leq R\_score_{u,c,d} \leq q$), $F\_score_{u,c,d}$ is the F property score ($1 \leq F\_score_{u,c,d} \leq q$), and $M\_score_{u,c,d}$ is the M property score ($1 \leq M\_score_{u,c,d} \leq q$). Hereafter, we define this method as naive RFM.

3.2 UW-RFM

If we adopt naive RFM for stock recommendation in this paper, we still face the mass recommendation sales problem. Specifically, if a company is used by many users and the familiarity scores of most users with that company is high, the recommendation approach based on using naive RFM will include that company in the recommendation list of most users.

In addition, even if a user does not use a certain company, a minimum score of one is assigned to the user score for that company. This may lead to the recommendation of stocks of companies that the user does not use at all.

In order to address the mass recommendation sales problem, we propose UW-RFM, which assigns a higher score to companies with fewer users by utilizing the number of users per company. The basic idea behind UW-RFM is to consider the ratio of users using the target company to the total number of users in calculating R, F, and M values. This method assigns companies with a small number of users a relatively higher score than that of companies with many users. This assignment makes it easier for companies with a small number of users to be selected as recommendation candidates. In addition, we assign a score of zero to the companies that users do not use at all.

Specifically, we perform scoring as indicated hereafter. First, for each company, we assign a zero score to a company that users do not use at all. Second, we sort the users based on Z (Z can be one of R, F, or M). The users are sorted in ascending order of the R value and in descending order of the F and M values. Third, we divide users into q groups among all users except those with no usage at all, which are referred to as corporate users. When the ratio of corporate users to all users is r_Z, we assign a score of $(q - i + 1)/r_Z$ ($i = 1, 2, \cdots, q$) to the users belonging to the i-th group from the top group. We calculate the scores for each of R, F, and M using the above procedure and then sum the calculated scores for the three elements as the final score. After these processes, the familiarity score of each user with a company in this method is calculated as

$$familiarity\_score_{u,c} = \sum_{d \in D} (R\_score_{u,c,d}/r_R + F\_score_{u,c,d}/r_F + M\_score_{u,c,d}/r_M) \quad (5)$$

As a result, by dividing by r_Z, the companies with a small number of users are assigned relatively higher scores than those of the companies of many users. This allows companies that are less frequently used by all users to be added to the user recommendation list, which mitigates the mass recommendation sales problem.

3.3 Complementary Module

If the number of companies generated using UW-RFM is less than N, we use the complementary module to address the display problem. We utilize the following methods as the complementary module. We use association rule mining with an *apriori* algorithm [3] to select companies with high relevance as recommendations based on the user CTs. The process of generating companies using association rule mining is described below [2]. Let $C = \{c_1, c_2, \cdots, c_m\}$ be a set of companies, and D be a set of transactions, where each transaction T is a set of companies for each user such that $T \subseteq C$. An association rule is an implication of the form $X \Rightarrow Y$, where $X \subset C$, $Y \subset C$ and $X \cap Y = \emptyset$. Given a set of transactions D, we generate all association rules that are greater than the user-specified minimum support referred to as *minsup* and minimum confidence referred to as $minconf$. Here, *minsup* is the lower bound on the threshold for the fraction of transactions in D that contains $X \Rightarrow Y$ and $minconf$ is the lower bound on the threshold for the fraction of transactions in D that contains X also contains Y.

After we use association rule mining, we select companies to recommend to the users whose number of companies is less than N. We select companies using the two methods below depending on the user attribute information. (1) We select companies with the highest number of unique users based on whole users or gender and age or region of residence, which we refer to as the most popular recommendation. (2) We select companies with the highest rate of increase in the number of unique users based on whole users or gender and age or region of residence, which we refer to as the trend recommendation. The rate of increase in the number of unique users for a company represents the difference between the number of unique users at t and the number of unique users at $t - 1$ divided by the number of unique users at $t - 1$. We use a combination of these methods to generate stocks.

4. Experiments

In this section, we describe experiments using a real-world dataset collected using an investment service to evaluate the effectiveness of the proposed UW-RFM. This service allows users

Fig. 3 Example screen presented to user.

to invest small amounts of money using reward points. We implemented the proposed approach as a recommender system, and we conducted evaluation experiments from three perspectives: (1) We evaluated whether the stocks generated using UW-RFM were less biased toward a few specific stocks than the stocks generated by the baseline naive RFM in order to evaluate whether the proposed UW-RFM reduced the mass recommendation sales problem. (2) We evaluated whether or not the proposed UW-RFM promoted user investments. (3) We evaluated how investment behavior changed when the number of a user's own CTs that were included in the recommendation list was changed.

4.1 Dataset

Figure 3 shows an example of the screen presented to the user. We recorded the number of clicks on the service and the point data used to purchase stocks in the service between August 21, 2020, and September 30, 2020.

We used cashless payment data and reward point data collected between April 1, 2020, and June 30, 2020, and app usage data collected from May 1, 2020, to July 31, 2020, as input for the recommender system. Furthermore, we used a mapping table made by hand that maps CTs to listed companies. As a result, we secured 753 companies as candidates for the recommendation system. These represented approximately one-fifth of the companies listed on the Tokyo Stock Exchange. These data included 14,248 users, who agreed to use the data when they started using the service.

4.2 Parameter Setting

We describe parameters for the stock recommender approach implemented on the investment service. We set the number of candidate stocks for each user, N, to 10 because we retained five stocks to address the display problem even if there were companies that have gone bankrupt or been delisted. When the user views the investment service display screen as shown in Fig. 3, or the user refreshes the screen, five randomly selected stocks are presented from 10 stocks. In this setting, we used the same list of candidate stocks for each user from August 21, 2020, to September 30, 2020.

We set parameters of the proposed approach as indicated hereafter. We divided users into five groups of 20 percentile for each of R, F, and M in UW-RFM. That is $q = 5$. In the association rule mining of the complementary module, the lengths of X and Y were one, $minsup$ was 0.1, $minconf$ was 0.2. We used the collected dataset to check the available attribute information of users and found that there were two groups of users: those who had all of the information on gender, age, and region of residence and those who were missing all of the information on gender, age, and region of residence. Thus, we generated 10 stocks to complement shortage and supplemented the user shortage with the list of stocks in order from top to bottom according to user attribute information. The following is a list of stocks for the former user group: $\{c_{most,ga}, c_{most,re}, c_{most,ga}, c_{most,re}, c_{trend,ga}, c_{trend,re}, c_{most,ga}, c_{most,re}, c_{most,ga}, c_{most,re}\}$. Here, $c_{most,ga}$ is the stock generated by the most popular recommendation by gender and age, $c_{most,re}$ is the stock generated by the most popular recommendation by region of residence, $c_{trend,ga}$ is the stock generated by the trend recommendation by gender and age, and $c_{trend,re}$ is the stock generated by the trend recommendation by region of residence. The following is a list of stocks for the latter user group: $\{c_{most,all}, c_{most,all}, c_{most,all}, c_{most,all}, c_{trend,all}, c_{trend,all}, c_{most,all}, c_{most,all}, c_{most,all}, c_{most,all}\}$. Here, $c_{most,all}$ is the stock generated by the most popular recommendation by all, and $c_{trend,all}$ is the stock generated by the trend recommendation by all.

Table 1 Catalog coverage and Gini coefficient for evaluating stock bias.

Method	Catalogue Coverage	Gini Coefficient
Naive RFM	0.811	0.868
UW-RFM	**0.862**	**0.772**

4.3 Evaluation of Bias in Recommendations

We evaluated whether or not UW-RFM mitigated the mass recommendation sales problem with a real-world dataset collected using an investment service. Specifically, we evaluated how many stocks UW-RFM recommended out of all the stocks that were recommended from the data, and evaluated the bias in the frequency of stocks presented to the entire user population.

We used catalog coverage [22] and the Gini coefficient [20] as metrics in this evaluation. Catalog coverage is the percentage of stocks that can be recommended to users out of all the stocks recommended from the data. The higher the catalog coverage value, the more stocks can be recommended. The Gini coefficient is a metric for distributional inequity and represents the bias in the frequency of stocks presented to all users. The lower the Gini coefficient value, the less biased the recommendation candidates are among the stocks.

We compared our UW-RFM to naive RFM as a baseline method. Naive RFM was performed in the following manner. First, we divided users into five 20 percentile groups for R, F, and M. Second, we assigned a score of $6 - i$ ($i = 1, 2, \cdots, 5$) to the users belonging to the i-th group from the top group. Third, we calculated the scores for each of R, F, and M using the above procedure and then sum the calculated scores of the three elements as the final score. We used a complementary module such as association rule mining, the most popular recommendation, and trend recommendation for UW-RFM and naive RFM because there were users who did not have 10 CTs.

Table 1 gives the results of the catalog coverage and Gini coefficient for evaluating stock bias. As shown in Table 1, UW-RFM

Table 2 Catalogue coverage, and Gini coefficient for UW-RFM as q is varied by 2, 5, 8.

q	Catalogue Coverage	Gini Coefficient
2	0.862	0.770
5	0.862	0.772
8	0.862	0.772

achieved 0.051 higher catalog coverage and 0.096 lower Gini coefficient than naive RFM. These results show that UW-RFM reduces the bias of stocks better than naive RFM and mitigates the mass recommendation sales problem.

In order to validate that the recommendation lists do not change significantly as q is varied, we checked recommendation lists for a user when q was varied by 2, 5, and 8. As a result, all the recommendation lists were the same. Therefore, we qualitatively confirmed that the recommendation list for a user does not change significantly as q is varied from a few examples.

To quantitatively confirm that if the recommendation lists change significantly across all users, we calculated the catalogue coverage and Gini coefficient for UW-RFM as q is varied. **Table 2** shows the catalogue coverage, and Gini coefficient for our proposed approach as q is varied by 2, 5, 8. As shown in Table 2, there is almost no change in the catalogue coverage, while the Gini coefficient tends to increase as q increases but does not change significantly. This result shows that the change in q does not considerably change the recommendation list. Since the optimal q for promoting a user's investment behavior was not clear, we set q to five based on Miglautsch [30] in the subsequent experiments.

4.4 Effect on User Investment Behavior

We evaluated whether or not presenting stocks of CTs inferred from data on users' smartphones promoted user investment behavior. In order to evaluate this effect, a Randomized Controlled Trial (RCT) is desirable because it can remove selection bias by randomly assigning users to a treatment group and control group [18]. In this study, an RCT means randomly assigning users of an investment service presented with stocks of CTs and with stocks of randomly selected companies. However, the stocks must be presented with the rationale for the stock selection when it is displayed in order to deal with the display problem, and it is not possible to present the stocks of randomly selected companies, making it difficult to conduct the RCT.

In order to address this challenge, we adopted a framework of quasi-experimental design [40] to evaluate causality based on observational data. As shown in Fig. 2, we generated N stocks using UW-RFM and a complementary module. In the process of generating N stocks, two groups of users are formed: those who generate N stocks using only UW-RFM and those who generate N stocks using only the complementary module. This means that we have two user groups: one group presented only with stocks of their own CTs, and the other group presented only with stocks complemented by other user data of CTs. In the context of this situation, we considered presenting only the user's own CTs for treatment. Here, we define the former user group as the treatment group and the latter user group as the control group. We compared the treatment group with the control group in order to estimate the effect of the treatment. The number of users in the treatment group and that in the control group were 4,658 and 490, respectively.

Table 3 Balancing statistics on covariates after matching between two groups.

Variable	SMD
NA	0
Male	0.048
Female	−0.048
Age 20s or younger	0.089
Age 30s	−0.158
Age 40s	0.066
Age 50s	0.045
Age 60s	0.002
Age 70s or older	0.011
Account usage period	−0.078
Service usage period	−0.162
Point balance	−0.020
Median absolute SMD	0.048
Maximum absolute SMD	0.162

In order to reduce the difference between the treatment group and the control group, we adopted propensity score matching [16], [37] [*6]. We adopted the logistic regression as done in previous studies [4], [26], [31]. We performed nearest neighbor matching to extract pairs from the treatment group and the control group. Moreover, in this study, since the number of users in the control group was small compared to the number of users in the treatment group, we adopted matching with replacement [42], which means that control units that look similar to many treatment units can be used multiple times. Since control units were selected as a match using matching with replacement, we weighted the units of the control group selected for multiple matching by the inverse of the matching frequency when we analyzed the effect of user investment behavior [42].

In this study, we used the covariates listed in **Table 3** to calculate the propensity score. Specifically, we used the following two categories of covariates for each user: demographic information (age and gender) and investment-related behavior (period since account was opened, period using point investment function, and point balance) [*7]. We used the point balance on August 20, 2020, the day before the presentation to the user, as a covariate.

Table 3 gives balancing statistics on covariates, which are inputs to the logistic regression after matching. In order to evaluate whether or not the treatment group is indistinguishable from the control group, we calculated the standardized mean difference (SMD) [42]. SMD is defined as the difference between the mean of the treatment group and that for the control group divided by the standard deviation within the treatment group. All covariates with an absolute SMD lower than 0.25 means that the two groups are indistinguishable or balanced [4], [26], [42]. As shown in Table 3, the absolute SMD for all covariates were well balanced within 0.162, which is less than 0.25, and half of the covariates were within 0.048. Therefore, we adequately eliminate the selection bias caused by potential differences between the treatment

[*6] In this study, we use the MatchIt package in R (version 4.1.0) [23].
[*7] We controlled for NA because previous studies have recommended explicitly [4], [42].

Table 4 Estimation of effect of treatment on numbers of monthly clicks on recommendation area and monthly purchase points for stocks.

Outcome	Estimate (95% CI)	p-value
Number of monthly clicks on recommendation area	**0.352** (0.118–0.586)	$p < 0.01$
Number of monthly purchase points for stocks	**3,016** (1,049–6,549)	$p < 0.001$

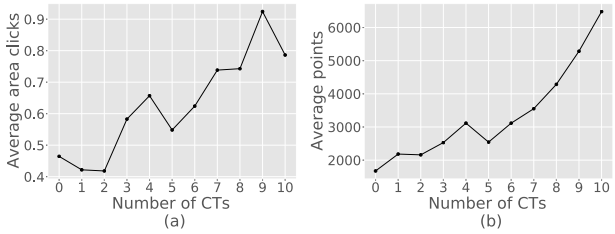

Fig. 4 (a) Average number of monthly clicks on recommendation area and (b) average number of monthly purchase points of stocks when the number of the user's own CTs are included in the recommendation list.

group and the control group in using the data after matching.

Table 4 gives an estimation of the effect of treatment on the numbers of monthly clicks on the recommendation area and monthly purchase points for stocks. After adjustment using propensity score matching to ensure the validity of between-group comparisons, the results showed a significant increase of 0.352 monthly clicks on the recommendation area (t-test for regression coefficient: $p < 0.01$) and 3,016 monthly purchase points for stocks on average (t-test regression coefficient: $p < 0.001$). These results indicate that presenting CTs based on data collected from user smartphones, such as payment data and app usage data, can promote user investment behavior.

If unobserved covariates cause a selection bias of treatment effects between two groups, these results may not be valid. Since we controlled for many significant covariates, it is unlikely that there are unobserved covariates that would cause a sharp change in the treatment assignment. A sensitivity analysis [36] of the results showed that the difference between the two groups was statistically significant (Wilcoxon signed-rank test at $p = 0.05$) even when the observed covariates differed in the odds of treatment for two identical matched users by up to 1.442 times for the number of monthly clicks and up to 2.476 times for the number of monthly purchase points for stocks. These results indicate that the revealed treatment effects are robust compared to the level shown in a previous study [4].

4.5 Effect of Changes in the Number of CTs in the Recommendation List

We evaluated how the investment behavior changed when the number of the user's own CTs in the recommendation list changed. **Figure 4** shows the average numbers of monthly clicks on the recommendation area and monthly purchase points when the number of the user's own CTs included in the recommendation list changed. As shown in Fig. 4, the average numbers of monthly clicks on the recommendation area and monthly purchase points increased as the number of CTs in the recommendation list increased.

However, the results in Fig. 4 can be subject to selection bias. Users who have more reward points at the time of treatment have possibly more stock purchase points and clicks that lead to purchases than users who have fewer. Users with more CTs tend to have more reward points than those with fewer, since they use more services and can collect more reward points easily. As a result, the causal effect of how the investment behavior changed can be overestimated by simply comparing the average area clicks and average points for each number of CTs. To mitigate the effect of this selection bias and to identify the causal effect, we applied a causal inference using propensity score, which can reduce the difference in user groups and compare the average clicks and average purchase points. Specifically, we adopted propensity score weighting for multiple treatments using generalized boosted models (GBMs) [28], [29] [*8]. When n is the number of CTs for each user, we divided users into three groups: low familiarity group, i.e., users with $0 \leq n \leq 3$, medium familiarity group, i.e., users with $4 \leq n \leq 6$; and high familiarity group, i.e., users with $7 \leq n \leq 10$. The number of users in the low familiarity group, medium familiarity group, and high familiarity group were 2,448, 2,245, and 9,550, respectively. We estimated the effect of changes in the number of CTs with three groups.

Table 5 Balancing statistics on covariates after weighting between each group. SMD of each variable is calculated for each combination, and the maximum value is shown.

Variable	SMD
NA	0.087
Male	0.030
Female	0.021
Age 20s or younger	0.020
Age 30s	0.042
Age 40s	0.013
Age 50s	0.035
Age 60s	0.015
Age 70s or older	0.001
Account usage period	0.023
Service usage period	0.035
Point balance	0.024
Median absolute SMD	0.024
Maximum absolute SMD	0.087

We checked the effects on the three treatment groups by comparing each combination of low familiarity group versus medium familiarity group, medium familiarity group versus high familiarity group, and low familiarity group versus high familiarity group. Since the purpose of the evaluation is to analyze how the investment behavior of users changes according to the number of CTs presented in the recommendation list, we analyze the effect on the medium familiarity group and high familiarity group based on the low familiarity group.

Table 5 gives balancing statistics on covariates, which are inputs to the GBMs after weighting between each group. As shown in Table 5, the absolute SMD for all covariates was well balanced within 0.087, which is less than 0.25, and half of the covariates

[*8] In this study we use the twag package in R (version 1.6) [10], [35].

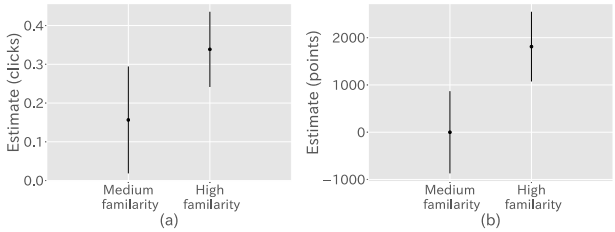

Fig. 5 Estimated ATE for medium and high familiarity group based on low familiarity group in terms of (a) number of monthly clicks on the recommendation area and (b) number of monthly purchase points of stocks.

were within 0.024. Therefore, we adequately eliminate the selection bias caused by potential differences among groups using the data after weighting.

Figure 5 shows the results of estimated average treatment effect (ATE) for the medium and high familiarity group on numbers of monthly clicks on the recommendation area and monthly purchase points of stocks based on the low familiarity group. The ATE of the medium familiarity group on the low familiarity group was a significant increase of 0.156 monthly clicks on the recommendation area (t-test for regression coefficient; $p < 0.1$). On the other hand, there was no effect in the ATE of the medium familiarity group on the low familiarity group in terms of the number of monthly purchase points. In addition, the ATE of the high familiarity group on the low familiarity group was a significant increase of 0.339 monthly clicks on the recommendation area (t-test for regression coefficient; $p < 0.001$) and 1,811 monthly purchase points of stocks (t-test for regression coefficient: $p < 0.001$). These results indicate that users in the user groups with a higher number of CTs in the recommendation list are more likely to exhibit positive investment behavior such as more clicks and purchase stock points. Based on these results, we see that the user investment behavior is promoted as the number of their own CTs included in the recommendation list is increased.

5. Discussion and Limitation

Table 4 shows that presenting CTs to users based on data collected from their smartphones, such as payment data and app usage data, can promote their investment behavior. This result has important implications for the long-term asset formation of individual investors. The monthly increase in investment of 3,016 points (yen) results in an increase of 1,466,699 yen over 30 years of operation at a yield of 1.92%. This result is calculated using a simulator provided by the Financial Services Agency [*9] and we adapt the weighted average yield of companies listed on the First Section of the Tokyo Stock Exchange as of March 2021 [*10]. This amount corresponds to $1{,}466{,}699 \div 27{,}090 \fallingdotseq 54.142$, which is almost 54 months based on the monthly household income shortfall of 27,090 yen for an elderly single-person household [*11]. Therefore, the proposed approach has a great impact on the problem of insufficient funds for retirement by promoting the asset formation of individual investors.

Moreover, we can see that the proposed approach, which addresses the cold-start problem and can be applied to users with no stock trading data, promotes investment behavior. Since the number of CLP users in Japan is very high (72% of the total population, or 90.3 million people [*12,*13]), the proposed approach can be applied to a large number of people.

This study has several limitations and future directions. Firstly, we estimate the companies familiar to the users from the company touchpoint information on their smartphones. However, this approach cannot present the stocks of the companies that do not provide opportunities for cashless payments, and the companies that do not offer apps. This limitation could be addressed using a hybrid approach that combines the proposed approach using touchpoint information and collaborative filtering [43] using user stock trading data.

Next, this approach aims to encourage user investment behavior and does not take into account the returns and risks of stocks presented to a user. Promoting user investment behavior will be the first step for users to achieve long-term asset formation because users will pay attention to news and economic trends related to the companies they invest in by starting to invest, which will lead to the acquisition of financial literacy necessary for long-term assets building. In addition, because the proposed approach can be used in conjunction with existing methods that support investor decision making by predicting stock prices and trade timing [11], [12], [27], [46], this hybrid approach has the potential to promote long-term asset building for a user while taking losses into account.

Additionally, this approach is based on a batch process that learns from the records of a certain period. Therefore, we need to retrain the model when we acquire new records. If we shorten the update interval of the model, we can recommend the stocks of CTs a user used yesterday. In our paper, since the main focus of our proposal was to show the effectiveness of recommending stocks based on payment data and app usage data, we experimented with a single period of data to train the model. However, the approach based on real-time processing may further promote a user's investment behavior.

Moreover, to deal with the different implications of RFM among the companies, we make the values of R_score, F_score, and M_score of service provider and application provider c for each dataset d in Eq. (5) fall within the range of 1 to q. As a result, we confirmed the effect of clicks and purchase points. However, the above procedure alone does not fully eliminate the differences in the meaning of RFM analysis for each type of business. Therefore, we believe that examining the weights for each type of business is a future issue.

Finally, the estimation of effects based on observational studies has limitations. Propensity score matching approximates RCTs by controlling for observed covariates; however, there may be unobserved covariates that affect the treatment probability. Studies using causal inference always have this limitation [26]. Although the display problem prevents us from conducting an RCT

[*9] https://www.fsa.go.jp/policy/nisa2/moneyplan_sim/index.html
[*10] https://www.jpx.co.jp/english/markets/statistics-equities/monthly/b5b4pj0000043saq-att/05_yield2103.pdf
[*11] https://www.stat.go.jp/data/kakei/2019np/gaikyo/pdf/gk02.pdf
[*12] https://www.stat.go.jp/english/data/jinsui/tsuki/index.html
[*13] https://www.jftc.go.jp/houdou/pressrelease/2020/jun/200612_4.pdf

immediately, we would like to design an RCT so that it can be conducted in compliance with the guidelines [25].

6. Related Work

In this section, we relate this study to research on the relationship between investment behavior and familiarity, research on stock recommender systems, and research on scoring loyalty using RFM analysis, and explain the position of this study.

6.1 Relationship Between Investment Behavior and Familiarity

Several studies have reported that individual investors tend to prefer to invest in companies with which they are familiar from various perspectives. Specifically, some studies have reported that individual investors tend to select stocks of companies to invest in based on geographic proximity [14], [24]. Other studies have reported that individual investors have a preference for investing in the stock of their employer company and for owning stock of companies in the industry in which they work [5], [17]. Although the existing studies [5], [14], [17], [24] have analyzed the tendency to invest in companies with which they are familiar using their trading history, the effectiveness for recommendation, especially for the cold-start problem addressed in this paper, has not been revealed. In addition, Prast et al. [33] has shown that presenting companies that appear in advertisements in women magazines, which women are likely to be familiar with through questionnaires, promotes the decision-making process of women. However, the degree of familiarity of users with companies may differ from user to user, and the effectiveness of presenting stocks based on the degree of familiarity of each user with companies has not been clarified. This study is different in that it evaluates the effect of presenting stocks selected individually for each user based on their familiarity with companies using data automatically accumulated in their daily lives, such as payment data and app usage data.

6.2 Stock Recommender Systems

Many previous studies have attempted to assist users in making decisions regarding the purchase and sale of stocks. Previous studies have constructed trading strategies that contribute to individual investor decision support by predicting stock prices and trading timing based on trading data [11], [12], [27], [46], news analysis [21], [39], and sentiment analysis [19], [44]. Taghavi et al. [45] also proposed a technique to recommend the most profitable stocks according to investor preferences regarding investment style such as emphasizing long-term returns or aggressive risk taking.

This study differs from these previous studies in that it aims to promote investment behavior by presenting users with stocks of companies whose products and services they usually use and fostering interest in the stocks of those companies themselves. Previous studies and this study can be used in combination to promote investment behavior because these studies can recommend stocks from different perspectives.

6.3 Scoring Loyalty Based on RFM Analysis

Many previous studies have proposed methods for quantifying and utilizing customer loyalty based on RFM analysis. RFM analysis is a method that aggregates the purchasing behavior of customers based on R, F, and M [8]. In direct marketing targeting specific customers, RFM analysis identifies the customers to be contacted [13], [41].

In this study, we calculate a company-familiarity score that is generated based on user usage of company services, and use this score as the familiarity with the company stock. This study differs from previous studies because it relates familiarity with a company stock to familiarity with its service.

7. Conclusion

In this paper, we proposed a novel content-based stock recommendation approach that utilizes touchpoint information obtained from users' smartphones in daily life. We implemented a novel stock recommender system using UW-RFM and a complementary module in an investment service and we verified the effectiveness of the proposed approach. Specifically, we conducted an observational study using propensity score matching. The results showed that the average numbers of monthly clicks on the recommendation area and monthly purchase points significantly increased by 0.352 times and 3,016 points, respectively, by presenting the stock of their CTs. In addition, we evaluated how the investment behavior changed when the number of the user's own CTs in the recommendation list changed. The evaluation results showed that the ATE of the medium familiarity group and the high familiarity group on the low familiarity group was positive in terms of the numbers of monthly clicks on the recommendation area and monthly purchase points of stocks. From these results, we conclude that estimating the CTs that users use in their daily lives, which can be obtained from their smartphones, and presenting the company stocks promotes their investment behavior.

References

[1] Acilar, A.M. and Arslan, A.: A collaborative filtering method based on artificial immune network, *Expert Systems with Applications*, Vol.36, No.4, pp.8324–8332 (2009).

[2] Agrawal, R., Imielinski, T. and Swami, A.: Database mining: A performance perspective, *IEEE Trans. Knowl. Data Eng.*, Vol.5, No.6, pp.914–925 (1993).

[3] Agrawal, R., Srikant, R., et al.: Fast algorithms for mining association rules, *Proc. 20th Int. Conf. Very Large Data Bases, VLDB*, Vol.1215, pp.487–499 (1994).

[4] Althoff, T., Jindal, P. and Leskovec, J.: Online actions with offline impact: How online social networks influence online and offline user behavior, *Proc. 10th ACM International Conference on Web Search and Data Mining*, pp.537–546 (2017).

[5] Benartzi, S.: Excessive extrapolation and the allocation of 401(k) accounts to company stock, *J. Finance*, Vol.56, No.5, pp.1747–1764 (2001).

[6] Berman, B.: Developing an effective customer loyalty program, *California Management Review*, Vol.49, No.1, pp.123–148 (2006).

[7] Brealey, R.A., Myers, S.C., Allen, F. and Mohanty, P.: *Principles of corporate finance*, Tata McGraw-Hill Education (2012).

[8] Buckinx, W. and Van den Poel, D.: Customer base analysis: partial defection of behaviourally loyal clients in a non-contractual FMCG retail setting, *European Journal of Operational Research*, Vol.164, No.1, pp.252–268 (2005).

[9] Bult, J.R. and Wansbeek, T.: Optimal selection for direct mail, *Marketing Science*, Vol.14, No.4, pp.378–394 (1995).

[10] Burgette, L., Griffin, B.A. and McCaffrey, D.: Propensity scores for multiple treatments: A tutorial for the mnps function in the twang

package, *R package*, Rand Corporation (2017).
[11] Cho, V.: MISMIS – A comprehensive decision support system for stock market investment, *Knowl. Based Syst.*, Vol.23, No.6, pp.626–633 (2010).
[12] Chou, S.-C.T., Yang, C.-C., Chan, C.-H. and Lai, F.: A rule-based neural stock trading decision support system, *IEEE/IAFE 1996 Conference on Computational Intelligence for Financial Engineering* (*CIFEr*), pp.148–154, IEEE (1996).
[13] Colombo, R. and Jiang, W.: A stochastic RFM model, *Journal of Interactive Marketing*, Vol.13, No.3, pp.2–12 (1999).
[14] Coval, J.D. and Moskowitz, T.J.: Home bias at home: Local equity preference in domestic portfolios, *The Journal of Finance*, Vol.54, No.6, pp.2045–2073 (1999).
[15] De Long, J.B. and Summers, L.H.: Equipment investment and economic growth, *Q. J. Econ.*, Vol.106, No.2, pp.445–502 (1991).
[16] Dehejia, R.H. and Wahba, S.: Propensity score-matching methods for nonexperimental causal studies, *Review of Economics and Statistics*, Vol.84, No.1, pp.151–161 (2002).
[17] Døskeland, T.M. and Hvide, H.K.: Do individual investors have asymmetric information based on work experience?, *J. Finance*, Vol.66, No.3, pp.1011–1041 (2011).
[18] Duflo, E., Glennerster, R. and Kremer, M.: Chapter 61 Using Randomization in Development Economics Research: A Toolkit, *Handbook of Development Economics*, Schultz, T.P. and Strauss, J.A. (Eds.), Vol.4, pp.3895–3962, Elsevier (2007).
[19] Eickhoff, M. and Muntermann, J.: Stock Analysts vs. The Crowd: A Study on Mutual Prediction, *PACIS*, p.144 (2015).
[20] Fleder, D.M. and Hosanagar, K.: Recommender systems and their impact on sales diversity, *Proc. 8th ACM Conference on Electronic Commerce, EC '07*, pp.192–199, Association for Computing Machinery (2007).
[21] Geva, T. and Zahavi, J.: Empirical evaluation of an automated intraday stock recommendation system incorporating both market data and textual news, *Decision Support Systems*, Vol.57, pp.212–223 (2014).
[22] Herlocker, J.L., Konstan, J.A., Terveen, L.G. and Riedl, J.T.: Evaluating collaborative filtering recommender systems, *ACM Trans. Inf. Syst. Secur.*, Vol.22, No.1, pp.5–53 (2004).
[23] Ho, D.E., Imai, K., King, G. and Stuart, E.A.: MatchIt: Nonparametric Preprocessing for Parametric Causal Inference, *Journal of Statistical Software*, Vol.42, No.8, pp.1–28 (2011) (online), available from ⟨https://www.jstatsoft.org/v42/i08/⟩.
[24] Huberman, G.: Familiarity breeds investment, *The Review of Financial Studies*, Vol.14, No.3, pp.659–680 (2001).
[25] Japan Securities Dealers Association: [Guidelines Concerning Advertising, etc.] Koukoku nado ni kansuru shishin (Japanese only) (2016), available from ⟨https://www.jsda.or.jp/about/jishukisei/web-handbook/103_koukoku/files/koukokushishin_1710.pdf⟩.
[26] Li, A., Wang, A., Nazari, Z., Chandar, P. and Carterette, B.: Do podcasts and music compete with one another?, Understanding users' audio streaming habits (2020).
[27] Lu, C.-J., Lee, T.-S. and Chiu, C.-C.: Financial time series forecasting using independent component analysis and support vector regression, *Decis. Support Syst.*, Vol.47, No.2, pp.115–125 (2009).
[28] McCaffrey, D.F., Griffin, B.A., Almirall, D., Slaughter, M.E., Ramchand, R. and Burgette, L.F.: A tutorial on propensity score estimation for multiple treatments using generalized boosted models, *Statistics in medicine*, Vol.32, No.19, pp.3388–3414 (2013).
[29] McCaffrey, D.F., Ridgeway, G. and Morral, A.R.: Propensity score estimation with boosted regression for evaluating causal effects in observational studies, *Psychol. Methods*, Vol.9, No.4, pp.403–425 (2004).
[30] Miglautsch, J.R.: Thoughts on RFM scoring, *Journal of Database Marketing & Customer Strategy Management*, Vol.8, No.1, pp.67–72 (2000).
[31] Miroglio, B., Zeber, D., Kaye, J. and Weiss, R.: The effect of ad blocking on user engagement with the web, *Proc. 2018 World Wide Web Conference*, pp.813–821 (2018).
[32] Mitsubishi UFJ Financial Group, Inc.: Research of Financial Literacy Survey of 10,000 Individuals – Differences between "those who have experienced investing" and "those who have not" (2018), available from ⟨https://www.tr.mufg.jp/shisan-ken/english/pdf/financial_literacy_01.pdf⟩.
[33] Prast, H., Rossi, M., Torricelli, C. and Druta, C.: Do women prefer pink?, The effect of a gender stereotypical stock portfolio on investing decisions (2014).
[34] Research and Statistics Department Bank of Japan: Flow of Funds – Overview of Japan, the United States, and the Euro area (2020), available from ⟨https://www.boj.or.jp/en/statistics/sj/sjhiq.pdf⟩.
[35] Ridgeway, G., McCaffrey, D., Morral, A., Burgette, L. and Griffin, B.A.: Toolkit for Weighting and Analysis of Nonequivalent Groups: A tutorial for the twang package, *Santa Monica, CA: RAND Corporation* (2017).
[36] Rosenbaum, P.R. et al.: *Design of observational studies*, Vol.10, Springer (2010).
[37] Rosenbaum, P.R. and Rubin, D.B.: The central role of the propensity score in observational studies for causal effects, *Biometrika*, Vol.70, No.1, pp.41–55 (1983).
[38] Schafer, J.B., Frankowski, D., Herlocker, J. and Sen, S.: Collaborative filtering recommender systems, *The Adaptive Web*, pp.291–324, Springer (2007).
[39] Schumaker, R.P. and Chen, H.: Textual analysis of stock market prediction using breaking financial news: The AZFin text system, *ACM Trans. Information Systems* (*TOIS*), Vol.27, No.2, pp.1–19 (2009).
[40] Shadish, W.R., Cook, T.D., Campbell, D.T., et al.: *Experimental and quasi-experimental designs for generalized causal inference*, Houghton Mifflin (2002).
[41] Sohrabi, B. and Khanlari, A.: Customer lifetime value (CLV) measurement based on RFM model (2007).
[42] Stuart, E.A.: Matching Methods for Causal Inference: A Review and a Look Forward (2010).
[43] Su, X. and Khoshgoftaar, T.M.: A survey of collaborative filtering techniques, *Advances in Artificial Intelligence*, Vol.2009 (2009).
[44] Sun, Y., Fang, M. and Wang, X.: A novel stock recommendation system using Guba sentiment analysis, *Personal and Ubiquitous Computing*, Vol.22, No.3, pp.575–587 (2018).
[45] Taghavi, M., Bakhtiyari, K. and Scavino, E.: Agent-based computational investing recommender system, *Proc. 7th ACM Conference on Recommender Systems, RecSys '13*, pp.455–458, Association for Computing Machinery (2013).
[46] Wen, Q., Yang, Z., Song, Y. and Jia, P.: Automatic stock decision support system based on box theory and SVM algorithm, *Expert Systems with Applications*, Vol.37, No.2, pp.1015–1022 (2010).

Kohsuke Kubota received his B.E. and M.E. degrees from Kyoto University in 2016 and 2018, respectively. He has joined NTT DOCOMO, INC. since 2018. He is interested in applying machine learning and causal inference into large-scale membership services, including loyalty programs and mobile payment.

Hiroyuki Sato received his B.E. and M.E. degrees from the University of Tokyo in 2012 and 2014, respectively. He has joined NTT DOCOMO, INC. since 2015. His research interests include data science, artificial intelligence, and mobile and ubiquitous computing.

Wataru Yamada received his M.E. degree from The University of Tokyo. He has joined NTT DOCOMO, INC. since 2012 and have worked on HCI field. His interest includes wearable computing, ubiquitous computing, and output devices. He is also currently a Ph.D. student in The University of Tokyo. He is a member of IPSJ and ACM.

Keiichi Ochiai received his B.E. and M.E. degrees from Chiba University in 2006 and 2008, respectively. He has joined NTT DOCOMO, INC. since 2008. He received Ph.D. degree at the University of Tokyo in 2017. He is a project research associate at Graduate School of Engineering, the University of Tokyo. His research interest includes data science, artificial intelligence, and mobile and ubiquitous computing. He is a member of IPSJ, DBSJ, and ACM.

Hiroshi Kawakami received his B.E. M.E. and Ph.D. degrees from Niigata University in 1992, 1994, and 2004, respectively. He has joined NTT DOCOMO, INC. since 1994. He has been engaged in the development of mobile communication systems and services. His research interests include data science based on large-scale membership data.