

コンシューマ・サービス論文

店舗チェックイン履歴情報を用いた優良顧客推定手法と 来店促進への応用

土井 千章^{1,3,a)} 石井 暁¹ 荒木 尊士¹ 稲村 浩² 太田 賢¹ 重野 寛³ 片桐 雅二¹

受付日 2016年9月30日, 採録日 2017年2月27日

概要: 本研究では店舗チェックイン履歴のみを保持する顧客から潜在的な優良顧客を発見する手法を提案する。この手法を用いることによって購買履歴を保持していない顧客の優良顧客レベルを推定でき、新規顧客獲得施策や来店促進施策に用いることができる。購買履歴と店舗チェックイン履歴を使用し、機械学習手法を用いて優良顧客モデルを獲得することで店舗チェックイン履歴のみから対象顧客の対象店舗における優良顧客レベルを推定する。推定精度および来店促進施策を実施して得られた対象店舗への来店率を用いて優良顧客推定手法の有効性と実用性を示す。

キーワード: 優良顧客推定, 来店促進, ランダムフォレスト, チェックイン履歴

Estimating Value of Customer through Store Check-in Histories and its Application for Visitor Promotion

CHIAKI DOI^{1,3,a)} AKIRA ISHII¹ TAKASHI ARAKI¹ HIROSHI INAMURA² KEN OHTA¹
HIROSHI SHIGENO³ MASAJI KATAGIRI¹

Received: September 30, 2016, Accepted: February 27, 2017

Abstract: This paper proposes a method to estimate value of customers based on store check-in histories. The proposed method enables to distinguish possible loyal customers whose purchase histories are not available. Machine learning process is employed for model acquisition. The outcomes of estimation are able to improve efficiency of visitor promotions and new customer acquisition events. The result of actual visitor promotion trial confirms effectiveness of the proposed method.

Keywords: value of customer, visitor promotion, random forest, check-in history

1. はじめに

近年、クレジットカードや電子マネー、ポイントカード等から取得した購買履歴情報と他社サービスのサービス利用履歴情報の分析結果に基づいた施策が多くの企業によって実施されている。自社データと他社データを組み合わせて

分析することによって、自社データでは得られなかった顧客の関心や実行動向を知ることができる。American Express International は、各顧客のユーザ属性とクレジットカードにより決済された店舗ごとの購入総額を保持していた。しかし、クレジットカードによる決済では商品個別の購買履歴情報を取得していないため、Facebook や Foursquare の「いいね」や「チェックイン」情報を結びつけて分析することによって、顧客の関心や実行動向に基づいたクーポンの提供を実現した [1]。ポイントカードの Ponta [2] や T ポイント [3] は、複数店舗のポイントカード利用履歴から潜在顧客を可視化して業種を横断した店舗間の相互送客を行っている [4], [5]。また、購買を行った店舗の情報を保持するポイントカードやチェックインを行うサービスでは、現在

¹ 株式会社 NTT ドコモ
NTT DOCOMO, INC., Yokosuka, Kanagawa 239-0847, Japan

² 公立はこだて未来大学
Future University Hakodate, Hakodate, Hokkaido 041-8655, Japan

³ 慶應義塾大学
Keio University, Yokohama, Kanagawa 223-0061, Japan

a) chiaki.doi.tf@nttdocomo.com

の位置情報だけではなく過去の位置情報を蓄積することができる。これらの蓄積した位置情報から顧客の生活圈やライフスタイルを推定し、顧客に合った情報の配信を行うことができる [6]。

優良顧客を見つけ出すことは、ターゲティングするうえで様々な施策の効率を向上させることができるため、重要な課題である。具体的には、顧客へのアプローチの最適化、ランクアップ施策や休眠顧客の活性化施策等の対象顧客選定に用いることができる [7], [8], [9]。そのために広く用いられる購買履歴情報を用いた顧客分析手法として RFM モデルがある。この手法では、購買の履歴から Recency (最終購買日), Frequency (購買頻度), Monetary (購入金額) を指標として用いて顧客の優良度合いを判定し、既存顧客の優良顧客レベルを推定することができる [9]。クレジットカードや電子マネー、ポイントカード等から取得した購買履歴情報を保持する顧客に対しては、この手法を用いることで優良顧客レベルを容易に付与することができる。しかしながら、現実的には購買履歴情報を保持していない顧客は多く存在し、その中に多数の優良顧客が潜在している。

チェックインというアクションが導入されているサービスとして Shopkick [10] や楽天チェック [11], ショプリエ [12] 等があり、これらのサービスでは来店ポイント付与やクーポン提示等の施策と組み合わせて用いられている。Shopkick は、1,500 万人以上のユーザに利用されており、今後もチェックインを用いた O2O サービスは拡大していくと考えられ、チェックイン履歴のあるユーザの増大も見込まれる。これらのサービスで取得可能なチェックイン履歴は、主に購買行動が行われている場所で記録されており、顧客が使用する可能性が高い近隣の店舗を推薦し、顧客の購買行動を誘導することにも活用できる。また、このチェックイン履歴情報を活用することで、購買情報を保持していない顧客に対して優良度合いを推定することができれば、既存顧客の維持だけではなく新規顧客の獲得にも応用でき、有用であると考えられる。しかしながら、筆者の知る限りでは、対象店舗内外のチェックイン履歴情報から優良顧客度合いを推定し、実サービスを用いた評価は行われていない。そこで、チェックイン履歴情報は特定店舗の購買状況と相関があるという仮説を設定し、その妥当性について実証を行う。これらが明らかになると、チェックイン履歴情報が優良顧客推定に有用である可能性が示唆でき、実用に向けた知見を積み上げることができる。

本研究では、店舗チェックイン履歴のみを保持する顧客から潜在的な優良顧客を推定する手法を提案する。具体的には、クレジットカードにより決済された対象店舗の購買履歴情報と店舗チェックイン履歴情報を使用し、機械学習の手法を用いて店舗チェックイン履歴情報のみから対象顧客の対象店舗における優良顧客レベルを推定する。評価として推定精度を示すととも到来店促進施策を実施し、それ

に対する対象店舗への来店率を用いて優良顧客推定手法の有効性と実用性を示す。これらの手法は、店舗チェックイン履歴を取得している O2O サービスへ適用可能であると考える。

以降 2 章で関連研究について述べ、3 章では本研究で用いているデータについて説明する。4 章では、優良顧客レベルの定義と優良顧客推定手法について述べる。5 章では、4 章で述べた優良顧客推定手法の効果を確認するために実施した来店促進施策について説明し、提案手法の効果を検証する。6 章では考察を述べ、7 章で本研究についてまとめる。

2. 関連研究

チェックイン履歴を用いて、顧客と情報提供者のニーズにあった情報配信を行うことを目的とした研究がすでに行われている。林ら [13] は、チェックイン履歴から曜日・時間帯を考慮して習慣的な行動か非習慣的な行動かを定量的に把握することで顧客にあった配信内容を選択可能にする方法を提案した。この手法を用いることで、本人の習慣に沿った行動をしている場合は顧客の行動傾向にあった情報、非習慣的な行動をしている状態では意外性を考慮した情報の配信等顧客の状態に合わせて配信する情報を出し分けることができる。また、習慣的な行動が類似した傾向にある他者の履歴を考慮することで、配信する情報を選別することができる。しかしながら、訪れた場所で行った行動については考慮されておらず、店舗側の新規顧客の開拓というニーズにおいてこの手法では他者との習慣的な行動の類似性から「店舗 A に訪れそうな顧客」は推定できても、店舗側が求めている「店舗 A における利益率の高い優良顧客」は考慮できない。そこで、本研究ではチェックイン履歴を用いて店舗 A における優良顧客を推定する方法を提案する。

購買履歴情報を用いて顧客の優良度合いの分析・予測を行うことの有効性については多くの研究で明らかにされている [7], [14]。購買履歴情報以外のデータを用いて優良顧客レベルを推定する研究についても報告を見つけることができる。一例として、大畑ら [15] は、スーパーマーケットの購買履歴情報と、店舗内の動線データから抽出した訪問パターンやエリアを用いて高額購買顧客と低額購買顧客の分類を行う方法を提案した。彼らは、決定木の手法の 1 つである C4.5 を用いることで、訪問パターンやエリアから購買顧客の優良度合いを推定した。本稿でも同様の手法を利用できるが、単一の決定木ではなく Random Forest 法等の決定木を弱学習器として用い集団学習を行う手法を用いることで推定精度の向上が期待できる。そのため、本研究では先行研究と提案手法を比較し、より高い精度で優良顧客度合いを推定する方法を提案する。また、既存手法は実際に店舗で購買行動が行われている場合に有効であり既存顧客の優良顧客度合いは推定することができるが、新規

顧客には適用することができない。そのため、本研究では対象店舗以外での顧客の行動にも着目し、新規顧客の優良顧客度合いを推定可能にする。

3. 使用データ

本章では、本研究で使用する店舗チェックイン履歴情報を取得した(株)NTTドコモの「ショッぷらっと」[16]サービスとデータについて述べる。

3.1 「ショッぷらっと」サービス

「ショッぷらっと」サービスは、(株)NTTドコモより提供された無料のアプリケーションをダウンロードすることで使用できる。本アプリケーションは、チェックインスポットでチェックインを行うことによって商品券やギフトクーポン等に交換できるポイントの取得やクーポンや店舗のお知らせ情報等を受信できる。チェックインは、以下の2つのアクションを行うことで完了する。①アプリケーションを起動してチェックインスポットに移動。②画面上で自動表示されるポイントのタグを下に引っ張る動作を行う。

3.2 購買履歴情報と店舗チェックイン履歴情報

本稿では、東急カード株式会社のクレジットカード[17]によって決済された購買履歴情報と「ショッぷらっと」サービスで取得された店舗チェックイン履歴情報を用いて検討を行う。支払いは、現金での支払いも考えられるが、本稿ではクレジットカードのみを対象とする。本検討に用いる購買履歴情報と店舗チェックイン履歴情報は、顧客より利用の許諾を得たデータである。購買履歴情報のデータ項目を表1に示す。購買履歴情報は、東京都渋谷区にある複合商業施設である渋谷ヒカリエ ShinQs [18]で行われた購買行動に限定する。購買履歴は、36,522 トランザクションを

使用した。テナントは、渋谷ヒカリエ ShinQs 内のテナントを示す。

表2に店舗チェックイン履歴情報のデータ項目を示す。チェックインスポットは都内を中心に百貨店やスーパーマーケット、レストラン等に設置されている。チェックインスポットは、各テナントに存在するわけではないことに注意する。チェックインスポットは、購買履歴情報と店舗チェックイン履歴情報の両方を保持する顧客が1回以上チェックインを行ったスポットとした。チェックインスポットは96のスポットを対象とし、全2,345,976回分のチェックイン履歴を用いた。

4. 優良顧客推定手法

本章では、優良顧客レベルの付与、優良顧客推定手法および手法の推定精度について述べる。

4.1 優良顧客レベル

本節では、RFM分析を用いた優良顧客レベルの算出方法について説明する。RFM分析では、Recency (最終購買日)、Frequency (累計購買回数)、Monetary (累計購買金額)を指標にして顧客の優良顧客レベルを算出する。

本研究では、購買履歴情報と店舗チェックイン履歴情報の双方を保持している329名分の、基準日から2年分の購買履歴情報を用いて、指標別に顧客を3分割に等数分割を行い、優良顧客レベルが低い順に1から3のスコアを付与した。各指標におけるスコアの合計値の低い順にCl.1~Cl.7とし、これを優良顧客クラスタ(Cl)とする。たとえば、最も優良度合いが高いCl.7は、Recency, Frequency, Monetaryの各スコアが3でスコアの合計値が9になった顧客のクラスタであり、逆に最も優良度合いが低いCl.1に所属する顧客は、スコアの合計値が3である。図1に各優良顧客クラスタの分布を示す。

本稿では、Cl.1, Cl.2に所属する顧客を一般顧客、Cl.6, Cl.7に所属する顧客を最優良顧客と呼ぶことにし、これらを優良顧客レベルとする。顧客*c*の優良顧客レベル*l_c*は式

表1 購買履歴情報
Table 1 Purchasing history data.

データ項目	内容
ユーザID	顧客のID
日時	クレジットカードによって決済が行われた日時
店舗	渋谷ヒカリエ ShinQs
テナント	店舗内にあるテナント名
金額(総額)	1度の決済で支払われた支払総額

表2 店舗チェックイン履歴情報
Table 2 Store check-in history data.

データ項目	内容
ユーザID	顧客のID
日時	チェックインが行われた日時
チェックインスポット	チェックインが行われた場所

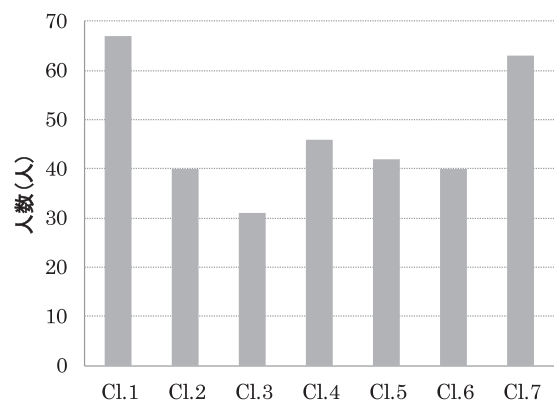


図1 各優良顧客クラスタの分布
Fig. 1 Distribution of customer loyalties.

(1) で表す.

$$l_c \in \{\text{最優良顧客, 一般顧客}\} \quad (1)$$

本検討では、優良顧客レベルの高い顧客を対象店舗に送客可能とすることを目指している. そこで、各優良顧客レベルの特徴が現れている顧客データを用いることで各優良顧客レベルの推定精度向上が期待できると考えた. そのため、本研究では各優良顧客レベルに所属する顧客の Recency, Frequency, Monetary の各値の差が大きいこと、各優良顧客レベルに所属する人数が多いことを要件として優良顧客レベルの推定に使用する学習データを選定した. 各優良顧客レベルに所属する人数は同数であることが望ましいため、人数を考慮して ① 一般顧客を Cl.1, 最優良顧客を Cl.7, ② 一般顧客を Cl.1, Cl.2, 最優良顧客を Cl.6, Cl.7, ③ 一般顧客を Cl.1, Cl.2, Cl.3, 最優良顧客を Cl.5, Cl.6, Cl.7 の 3 パターンより選択した. 優良顧客レベル間の距離 Dis は、式 (2) を用いて算出する. 距離は Recency, Frequency, Monetary の 3 次元で表現されて各次元で値域が異なるため、① との各次元の差を基準として正規化を行う. 優良顧客クラス cl における Recency の平均値を R_avg_{cl} , Frequency の平均値を F_avg_{cl} , Monetary の平均値を M_avg_{cl} とする. ① の一般顧客を Cl.1, 最優良顧客を Cl.7 とした場合の Recency の距離を RB_Abs (式 (6)), Frequency の距離を FB_Abs (式 (7)), Monetary の距離を MB_Abs (式 (8)) とする. num は対象とする優良顧客クラスに所属する人数である. 各優良顧客レベル間の距離を示す Dis は、値が大きいほど最優良顧客と一般顧客の特徴が異なっていることを示す. Dis は ① が 130.0, ② が 143.7, ③ が 143.1 であったため、② の一般顧客を Cl.1, Cl.2, 最優良顧客を Cl.6, Cl.7 が最善の組合せであると考えた. そのため、本研究では優良顧客レベルの推定に Cl.3, Cl.4, Cl.5 は用いず、Cl.1, Cl.2, Cl.6, Cl.7 を用いることとする.

$$Dis = (R_Abs + F_Abs + M_Abs) / 3 \times \sum_{cl} num_{cl} \quad (2)$$

$$R_Abs = |R_avg_{cl(\text{最優良顧客})} - R_avg_{cl(\text{一般顧客})}| / RB_Abs \quad (3)$$

$$F_Abs = |F_avg_{cl(\text{最優良顧客})} - F_avg_{cl(\text{一般顧客})}| / FB_Abs \quad (4)$$

$$M_Abs = |M_avg_{cl(\text{最優良顧客})} - M_avg_{cl(\text{一般顧客})}| / MB_Abs \quad (5)$$

$$RB_Abs = |R_avg_{cl.7} - R_avg_{cl.1}| \quad (6)$$

$$FB_Abs = |F_avg_{cl.7} - F_avg_{cl.1}| \quad (7)$$

$$MB_Abs = |M_avg_{cl.7} - M_avg_{cl.1}| \quad (8)$$

4.2 優良顧客推定手法

本節では、店舗チェックイン履歴情報を用いて優良顧客

レベルの最優良顧客と一般顧客を推定する手法を説明する. 店舗チェックインの状況を示す顧客 c の説明変数 i_c として、式 (9) に示すように各チェックインスポットでの顧客 c のチェックイン回数を正規化した値を用いる. チェックイン回数に閾値を設定することで優良顧客レベルを推定する方法が考えられるが、顧客や店舗ごとのチェックイン回数の最小値や最大値には差があり、これらの値の差が優良顧客の推定精度に影響を与える可能性があると考えた. そこで、チェックイン回数を用いた場合、顧客ごとに正規化したチェックイン回数を用いた場合、店舗ごとに正規化したチェックイン回数を用いた場合の 3 通りの推定精度を事前に算出した. 最も推定精度が高かったチェックイン回数を顧客ごとに正規化する場合と最も推定精度が低かったチェックイン回数を用いた場合では推定精度 (F 値, Random Forest 法を採用) に 12.6% の影響を与えることが分かった. そのため、本手法ではチェックイン回数を顧客ごとに正規化して用いる.

ここで h はチェックインスポット, H は対象のチェックインスポット数, $\alpha_{c,h}$ は顧客 c のチェックインスポット h におけるチェックイン回数を示し、正規化したチェックイン回数 $n_{c,h}$ は式 (10) を用いて算出する. s_c は顧客 c の全チェックイン回数であり、式 (11) を用いて算出する.

$$i_c = (n_{c,1}, \dots, n_{c,H}) \quad (9)$$

$$n_{c,h} = \alpha_{c,h} / s_c \quad (10)$$

$$s_c = \sum_{h=1}^H \alpha_{c,h} \quad (11)$$

本検討では顧客を最優良顧客と一般顧客に分類する形とした. このため、4.1 節で述べた優良顧客レベル l_c を学習時には教師情報として与えるとともに、推定時には最優良顧客と一般顧客のいずれか一方を推定結果として出力する.

優良顧客レベルの推定に用いる機械学習手法としては、先行研究 [15] で用いられた C4.5 と一般的に分類問題に対して用いられる機械学習手法から Random Forest 法, Logistic 回帰法およびサポートベクタマシン (以下 SVM) を候補とする. Random Forest 法 [19] は、説明変数をランダムサンプリングすることによって作成した複数の決定木を用いて目的変数を推定する手法である. サンプリングする説明変数の個数や決定木の深度については Breiman の基準を用いる. Logistic 回帰法 [20] は、2 値判別や発生確率を予測する場合に用いられる手法である. SVM [21] は、2 値判別を行う手法であり、カーネル関数は RBF カーネルを使用する.

評価は、10 分割交差検証によって行うこととし、データの 9 割を学習、1 割を評価に使用する. 交差検証には、優良顧客レベル l_c が最優良顧客 (Cl.6, Cl.7) もしくは一般顧客 (Cl.1, Cl.2) のデータのみを用いた. 顧客クラスターの Cl.3, Cl.4, Cl.5 のデータは学習データとして用いていないことに注意する.

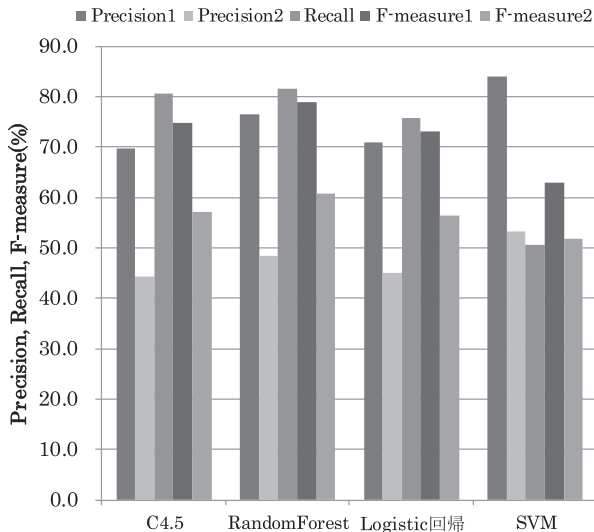


図 2 手法別優良顧客レベル (最優良顧客) の推定精度

Fig. 2 Accuracy rate of each method (loyal customers).

次に本提案手法の推定性能を評価する。尺度は、適合率、再現率および F 値を用いる。前述のとおり、交差検証においては Cl.3, Cl.4, Cl.5 のデータを用いていないが、実際にはこれらのクラスターに所属する顧客は存在しており、推定の結果として最優良顧客もしくは一般顧客のいずれかに分類されてしまうことになる。そのため、この分を加味した性能評価を行う。ここでは、交差検証の結果から得られた適合率を p 、上記を加味して補正した適合率を p' と表す。なお、再現率 r において上記は影響を及ぼさないことに注意する。優良顧客レベル l の適合率 p_l は式 (12) を用いて算出する。優良顧客レベル l の推定で真陽性および偽陽性と推定された人数の割合と同様の割合で Cl.3, Cl.4, Cl.5 のデータからも推定されると仮定し、適合率 p'_l は式 (13) を用いて算出する。ここで、 t_l は優良顧客レベル l と推定された顧客数であり、 v_l は優良顧客レベル l と推定された顧客のうち実際に優良顧客レベル l に属する顧客数である。

u_l は優良顧客レベル l に属する顧客数を示し、 W は交差検証に用いた顧客数 (Cl.1, Cl.2, Cl.6, Cl.7 に所属する顧客数) である。 O は 4.1 節で述べた顧客クラスター Cl.3, Cl.4, Cl.5 に所属する顧客の総数である。再現率 r_l は式 (15) を用いて算出する。F 値 f_l は、適合率 p_l と再現率 r_l の調和平均を求めた値であり、式 (16) を用いて算出する。また、補正された適合率 p'_l と再現率 r_l に基づいた F 値 f'_l は、式 (17) を用いて算出する。

$$p_l = v_l / t_l \tag{12}$$

$$p'_l = v_l / (t_l + O \times t_l / W) \tag{13}$$

$$W = \sum_{l \in \{\text{最優良顧客, 一般顧客}\}} u_l \tag{14}$$

$$r_l = v_l / u_l \tag{15}$$

$$f_l = (p_l \times r_l \times 2) / (p_l + r_l) \tag{16}$$

$$f'_l = (p'_l \times r_l \times 2) / (p'_l + r_l) \tag{17}$$

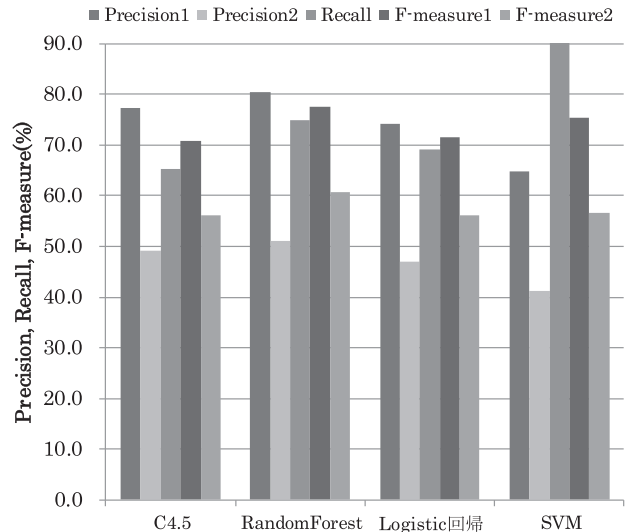


図 3 手法別優良顧客レベル (一般顧客) の推定精度

Fig. 3 Accuracy rate of each method (regular customers).

図 2 に最優良顧客を推定した結果を示す。適合率 p を Precision1, 適合率 p' を Precision2, 再現率 r を Recall と示す。Random Forest 法を用いた場合は F 値 f_l (図内 F-measure1) が 78.9% であり、全顧客の分布を考慮した場合の推定精度を示す F 値 f'_l (図内 F-measure2) は 60.8% であった。図 3 に一般顧客を推定した結果を示す。F 値 f_l (図内 F-measure1) は Random Forest 法を用いた場合は 77.6%, SVM を用いた場合は 75.5% であり、F 値 f'_l (図内 F-measure2) は、Random Forest 法を用いた場合が 60.8%, SVM を用いた場合は 56.6% であった。先行研究 [15] で用いられた手法と比較しても Random Forest 法を用いた場合の推定精度がいずれも高いことを確認した。これらの結果より、機械学習の方式により一定の差はあるものの F 値で 60% 程度でチェックイン履歴のみから最優良顧客の推定が可能であることを確認できる。本研究の目的は、優良顧客レベルの高い顧客を選定し、対象店舗へ送客することである。そのため、最優良顧客の推定において F 値 f'_l が最も高かった Random Forest 法を採用し、5 章で説明する来店促進施策の対象顧客を選定する。

また、実サービス上でのチェックイン履歴情報から推定した優良顧客度合いによる効果は、筆者が知る限り明らかにされていない。そこで、本研究では、これらの推定精度がどの程度来店促進施策に影響を与えるのか実証を行い、有用性を明らかにする。来店促進施策は、「シヨッぷらっと」サービスの利用者から渋谷ヒカリエ ShinQs の最優良顧客を送客し、その効果を確認する。

5. 来店促進施策における有用性の評価

4 章で述べた優良顧客推定手法の有用性を示すために、本手法を用いて推定した最優良顧客、一般顧客に対して来店促進施策を渋谷ヒカリエ ShinQs で実施した。本章では、



図 4 渋谷ヒカリエ ShinQs 配信情報

Fig. 4 Delivered content for Shibuya ShinQs Hikarie trial.

実施した来店促進施策と実施結果について述べる。

5.1 来店促進施策概要

来店促進施策は、本手法を用いて対象とする顧客を選定し、テナントの新規開店および新装開店に関する情報を「ショッぷらっと」アプリケーションの一機能であるノートیفケーション機能を用いて配信し、その効果を観察する。配信した情報を図 4 に示す。

対象顧客は、「ショッぷらっと」サービスを利用しており、店舗チェックイン情報を保持しているが購買履歴情報を保持していない顧客とし、その中から優良顧客推定手法を適用して最優良顧客と一般顧客を選定した。

対象とするチェックインスポットは、購買履歴情報と店舗チェックイン履歴情報の両方を保持する顧客が 1 回以上チェックインを行った 96 カ所のスポットとした。そのため、対象となるチェックインスポットのいずれにおいてもチェックインを行っていない顧客の優良顧客レベルは、本提案手法では推定できない。本評価では、このような推定ができない顧客を推定不可顧客として扱い、施策実施結果を確認する。

さらに「ショッぷらっと」サービスにおける過去 1 年分の店舗チェックイン履歴から、渋谷ヒカリエ ShinQs のチェックインスポットでチェックインを行ったチェックイン履歴情報を保持している顧客を既存顧客、保持していない顧客を新規顧客と定義する。既存顧客と新規顧客を区別して、比較評価を行うことで優良顧客推定手法の有効性を明らかにできる。表 3 に推定を行う対象顧客の人数分布を示す。推定不可顧客は対象店舗内のスポットへ過去 1 年間に 1 度もチェックインを行っていない顧客であるため、既存顧客は存在せず、すべてが新規顧客である。

5.2 渋谷ヒカリエ ShinQs における来店促進施策の効果

本節では、渋谷ヒカリエ ShinQs における 4.2 節で述べ

表 3 推定対象顧客の分布
Table 3 Distribution of target users.

優良顧客レベル	新規/既存	人数(人)
最優良顧客	新規顧客	111
	既存顧客	516
一般顧客	新規顧客	11,397
	既存顧客	2,602
推定不可顧客	新規顧客	2,665

た優良顧客推定手法を用いた利用促進施策の効果測定結果について述べる。効果は、配信情報の閲覧率と対象期間における来店率を用いて確認する。

購買履歴情報と店舗チェックイン履歴情報の両方を保持しておりデータの利用を許諾した 329 名に、来店促進施策を実施するまでに新しく利用の許諾を行った 682 名を加えた 1,011 名に対して施策を実施し、閲覧率と来店率、渋谷ヒカリエ ShinQs の購買行動に限定した購買率を確認する。閲覧率は式 (18) を用いて算出する。d は情報を配信した顧客の人数とし、e は配信した情報を閲覧した人数とする。来店率は式 (19) を用いて算出する。g は配信した情報を閲覧してから対象店舗に来店した人数とする。配信した情報を閲覧せずに来店した場合や、対象店舗に来店した後に配信した情報を閲覧した場合は g の人数には含まれないことに注意する。購買率は式 (20) を用いて算出する。b は配信した情報を閲覧してから来店し、対象店舗でクレジットカード決済により商品を購入した顧客の人数とする。

$$\text{閲覧率} = e/d \tag{18}$$

$$\text{来店率} = g/e \tag{19}$$

$$\text{購買率} = b/g \tag{20}$$

購買履歴を保持している顧客には、購買履歴を用いることで正確な優良顧客レベルを付与することができる。顧客の配信した情報の閲覧、店舗への来店、購買行動の有無を優良顧客レベル別に観察することで、優良顧客レベルの分類に期待できる有効性およびその程度を確認することができる。購買履歴を用いて優良顧客レベルを付与した結果、最優良顧客は 305 名、一般顧客は 334 名であった。また、利用促進施策結果が優良顧客レベルによって有意に異なっていることを確認するため、各顧客の閲覧、来店、購買が観察できた場合を 1、できなかった場合を 0 としてカイ二乗検定を実施する。購買履歴を保持している顧客の効果検証結果を図 5 に示す。最優良顧客は、一般顧客よりも閲覧率が 7pt、来店率が 15pt、購買率が 26pt 高かった。また、有意水準を 5% としてカイ二乗検定を行った結果、各々に対して統計的にも有意に差があることを確認した。これらの結果より最優良顧客は一般顧客よりも来店をして購買を行う顧客が多く所属していることが分かり、RFM 分析の有効性が確認された。

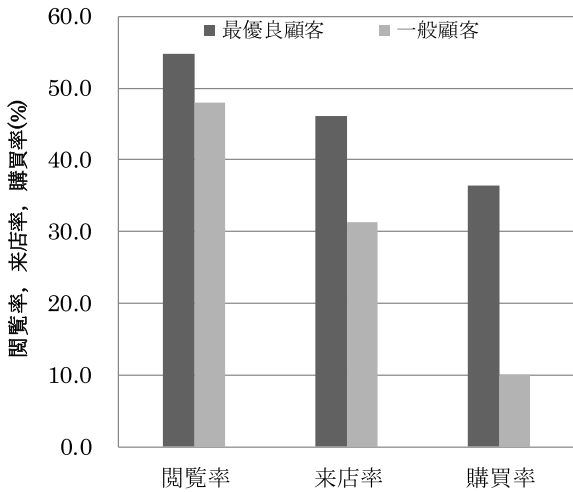


図 5 購買履歴情報保持顧客の閲覧率, 来店率, 購買率

Fig. 5 Access rate, Visit rate and Purchasing rate on customers who have purchasing history data.

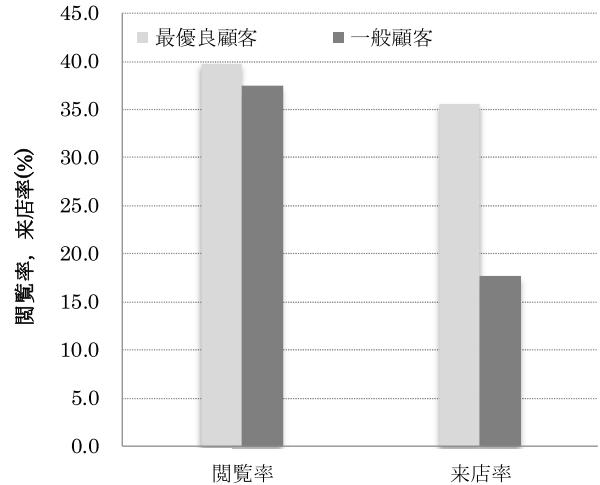


図 7 既存顧客の閲覧率, 来店率

Fig. 7 Access rate and Visit rate on existing customers.

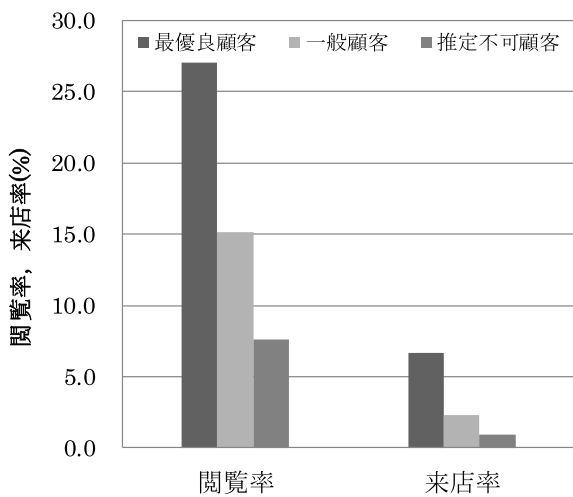


図 6 新規顧客の閲覧率, 来店率

Fig. 6 Access rate and Visit rate on new customers.

次に 4.2 節で述べた優良顧客推定手法を用いて優良顧客レベルを推定した購買履歴情報を保持していない顧客 17,291 名に対する来店促進の効果検証結果を図 6, 図 7 に示す。図 6 は, 新規顧客に対する効果検証結果である。最優良顧客の閲覧率は, 一般顧客と比較すると 12pt, 推定不可顧客は 19pt 高かった。一般顧客の閲覧率は推定不可顧客と比較して 7pt 高く, 最優良顧客, 一般顧客, 推定不可顧客の順で閲覧率が高く各々に対して統計的にも有意な差があることを確認した ($p < 0.05$, カイ二乗検定)。また, 最優良顧客の来店率は, 一般顧客と比較して 5pt, 推定不可顧客では 6pt 高い傾向が見られた。しかしながら, 最優良顧客と一般顧客の来店率については有意な差が確認できなかった ($p < 0.05$, カイ二乗検定)。新規顧客の最優良顧客については, サンプル数が少ないため, 統計的な有意差まで確認できなかった可能性が考えられる。図 7 は, 既存顧客の効果検証結果である。最優良顧客は, 一般顧客と比

較して閲覧率が 3pt, 来店率が 18pt 高く, 新規顧客と同様に最優良顧客の閲覧率, 来店率が高い傾向を確認した。

最優良顧客の閲覧率, 来店率が相対的に他より高かったことから, 店舗チェックイン履歴を用いた優良顧客推定手法の有効性が実証された。さらには, 最優良顧客は, 新規顧客と既存顧客の双方において配信した情報を閲覧しやすく, 情報を配信することによって来店しやすい傾向が得られた。また, 新規顧客は既存顧客と比較して, 最優良顧客でも閲覧率が 13pt 減, 来店率が 29pt 減と統計的にも有意な差 (低い傾向) が確認できた ($p < 0.05$, 二項検定)。これにより, 新規顧客の来店を促進することは, 既存顧客の来店を促進することよりも難しいことも示された。

また, 推定不可顧客は, 閲覧率および来店率が他と比較して低い傾向が観察できる。この推定不可顧客とは, 渋谷ヒカリエ ShinQs における最優良顧客および一般顧客がチェックインを行ったエリアで 1度もチェックインを行っていない顧客であり, 渋谷ヒカリエ ShinQs にとっての優良な顧客にはなりにくいと考えられる。

6. 考察・議論

本研究では, RFM 分析により顧客の優良度の判定を行う際に, テナントの業態については考慮せず対象店舗における全購買を合算する形で用いた。このため現状では, 渋谷ヒカリエ ShinQs 全体に対する最優良顧客を推定している形となるが, 現実的には洋菓子を主に扱うテナントの最優良顧客が, 服飾を主に扱うテナントの最優良顧客であるとは限らない。また, Recency, Frequency, Monetary の各々において, 取扱商材が異なるとその値 (期間, 金額) の意味する優良度合いが大きく異なることが想定される。テナントの業態を考慮することにより, テナントレベルでの最優良顧客の推定が可能になると考える。今後は, テナントの業態を考慮して RFM 分析および最優良顧客の推定

を行いたい。これにより、テナントの業態による推定精度の差や推定精度を向上させる説明変数を明らかにできると考える。

提案手法では、チェックインスポットごとのチェックイン回数を優良顧客の推定に用いた。単純にチェックイン回数の合計値に対して閾値を設けることで、閾値以上であれば優良顧客、閾値以下であれば一般顧客と推定する方法も考えられる。しかしながら、店舗チェックイン履歴は、様々なエリアや店舗のチェックインスポットで記録されており、チェックイン回数の合計値に閾値を設ける方法では対象店舗の優良顧客レベル推定は困難である。一例として、渋谷エリアでチェックイン回数が多い顧客は活動エリアが渋谷エリアであり、横浜エリアにある店舗への送客は困難であると考えられる。本手法を用いることで、これらのエリアの違いも考慮可能であり、対象店舗の優良顧客レベルが推定可能であることを示した。今後の課題として、顧客の行動を表すチェックインを行った時間や曜日、同日中にチェックインを行ったチェックインスポットの順番等他にも優良顧客の推定精度に影響を及ぼす変数が存在すると考える。顧客の行動を表すのに十分なデータ量が蓄積できれば、優良顧客の推定精度の向上や顧客に合ったタイミングでの情報配信が可能になると考えられ、今後明らかにしていきたい。

さらには、本研究では最優良顧客（RFMのいずれのスコアも高い顧客）を推定することを目的にして、優良顧客推定手法の提案を行った。しかしながらRFM分析では、Recency, Frequency, Monetaryの各スコア値を3次元空間として解釈し用いることで、より細かい優良顧客レベルへ分類することができ、たとえば過去に最優良顧客であったが一定期間来店がない顧客（離反顧客）、最優良顧客と離反顧客の中間層を離反する可能性がある顧客等とも分類することができる。このため、本手法をより精緻にすることにより離反する可能性が高い顧客を推定して既存顧客の維持を目的とした施策を実施することや、離反顧客の特徴を分析することによって離反傾向を明らかにできる可能性があるだろう。

また、本研究ではクレジットカードによって決済された購買履歴情報を用いて評価を実施し、現金での支払いは考慮していない。今後はポイントカード等より現金で支払いを行った顧客の購買履歴情報を取得し、現金で支払った顧客も対象としていきたい。

7. おわりに

本研究では、事前に一部顧客の購買履歴と店舗チェックイン履歴情報を用いて優良顧客レベルごとに店舗チェックイン履歴の特徴を学習することで、顧客の店舗チェックイン履歴情報のみから対象店舗における優良顧客レベルを推定する手法の提案を行った。評価結果により、機械学習の

手法を使用することで店舗チェックイン履歴情報から優良顧客を約60%程度の精度で推定できることが確認された。また、実サービス上での推定した優良顧客レベルの有用性を実証するため、「ショッぷらっと」サービスの利用者から渋谷ヒカリエShinQsの最優良顧客を送客する来店促進施策を実施した。施策の効果測定結果より、優良顧客レベルの中で最優良顧客の閲覧率、来店率が相対的に高かったことから店舗チェックイン履歴を用いた優良顧客推定手法の有効性を確認した。これにより、店舗チェックイン履歴情報は特定店舗の購買状況と相関関係があり、これらを用いることで優良顧客レベルを推定できることが明らかになった。

今後は、チェックインを行った時間や連続的な位置情報にも着目し、優良顧客のさらなる推定精度の向上を目指していきたい。本研究は購買履歴と店舗チェックイン履歴を保持していることが前提となる。本手法では、店舗チェックイン履歴が少ない顧客は、多い顧客と比較して、優良顧客レベルの推定精度が低下する可能性が考えられる。しかし、店舗ではなく店舗業種ごとにグルーピングすることにより、優良顧客レベルの推定精度を低下させることなく推定できるかもしれない。今後はこれらのコールドスタート問題への対応も検討していきたい。また、最優良顧客以外にも離反顧客、離反する可能性がある顧客等を対象として、分析および施策等を実施していきたい。

謝辞 本研究を進めるにあたり、東京急行電鉄株式会社、東急カード株式会社および株式会社東急百貨店からデータ提供や実証実験へのご協力、有益なコメントをいただいた。ここに記して感謝する。

参考文献

- [1] 森田哲明：ID情報分析による顧客中心マーケティング：顧客IDをキーとしたビッグデータ活用戦略・施策立案と新たなビジネスの創出、オペレーションズ・リサーチ：経営の科学, Vol.57, No.12, pp.673-682 (2012).
- [2] 共通ポイントPonta, (オンライン), 入手先 (<http://www.ponta.jp/>) (参照 2016-04-09).
- [3] Tサイト, (オンライン), 入手先 (<http://tsite.jp/>) (参照 2016-04-09).
- [4] 中村博之：データサイエンティストの虚像と実像—データ分析をビジネスに結びつけるための人材と体制, 野村総合研究所 ITロードマップセミナー AUTUMN2013 (2013).
- [5] 客ごとにクーポン発行 スリーエフTカードの履歴分析, 日経MJ (流通新聞), 入手先 (<https://messe.nikkei.co.jp/rt/i/news/112860.html>) (参照 2016-04-09).
- [6] 横山隆治, 榎田良輝：リアル行動ターゲティング, 日経BP社 (2015).
- [7] 久松俊道, 朝日弓未, 山口俊和：ドラッグストアのID付きPOSデータを用いた日用品購買パターンの比較分析, オペレーションズ・リサーチ：経営の科学, Vol.57, No.2, pp.63-69 (2012).
- [8] 安岡寛道：「ポイント・会員制サービス」入門, 東洋経済新報社 (2014).
- [9] 上田隆穂, 田島博和, 奥瀬喜之, 斉藤喜一：リテールデータ入門, 中央経済社, pp.105-112 (2014).

- [10] Shopkick, (オンライン), 入手先 <<http://www.shopkick.com/>> (参照 2016-12-15).
- [11] 楽天チェック, (オンライン), 入手先 <<https://check.rakuten.co.jp/>> (参照 2016-12-15).
- [12] ショプリエ, (オンライン), 入手先 <<https://shoplier.jp/>> (参照 2016-12-15).
- [13] 林 亜紀, 松林達史, 澤田 宏: 位置情報を利用した情報配信のための習慣度算出手法, 日本データベース学会和文論文誌, Vol.13, No.1, pp.64-71 (2014).
- [14] 松本 健, 西郷 彰: データ解析コンペティション課題設定部門—EC サイト顧客の顧客セグメントの予測, オペレーションズ・リサーチ: 経営の科学, Vol.58, No.2, pp.68-73 (2013).
- [15] 大畑善裕, 大野麻子, 山崎高弘, 常盤昴一郎: 小売店の内側エリアにおける顧客の行動パターンと購買額に関する分析, 第 14 回情報科学技術フォーラム (FIT2015), pp.297-302 (2015).
- [16] ショップらっと, (オンライン), 入手先 <<http://shoplat.net/>> (参照 2016-03-24).
- [17] 東急カード, (オンライン), 入手先 <<http://www.topcard.co.jp/>> (参照 2016-03-24).
- [18] 渋谷ヒカリエ ShinQs, 入手先 <<https://www.tokyu-dept.co.jp/shinqs/>> (参照 2016-12-15).
- [19] Breiman, L.: *Machine Learning*, Kluwer Academic Publishers (2001).
- [20] David, C.: The Regression Analysis of Binary Sequences, *Journal of the Royal Statistical Society, Series B (Methodological)*, Vol.20, No.2, pp.215-242 (1958).
- [21] Bernhard, B., Isabelle, G. and Vladimir, V.: A training algorithm for optimal margin classifiers, *Proc. 5th Annual Workshop on Computational Learning Theory*, pp.144-152 (1992).



土井 千章 (正会員)

株式会社 NTT ドコモ先進技術研究所勤務。平成 21 年慶應義塾大学理工学研究科博士前期課程修了。同年 (株) NTT ドコモ入社。モバイルコンピューティング, Android アプリケーションのセキュリティ, 顧客理解を目的とした行動モデリングや大規模データを用いた行動予測技術の研究に従事。



石井 暁

株式会社 NTT ドコモコンシューマビジネス推進部勤務。平成 16 年電気通信大学大学院電気通信学研究科博士前期課程修了。同年 (株) NTT ドコモ入社。基地局パラメータ自律設定, O2O サービス: ショップらっと, レシーボンの新規事業創出, 家族の安心安全サービス: イマドコサーチに従事。



荒木 尊士

株式会社 NTT ドコモ先進技術研究所勤務。平成 27 年大阪府立大学大学院工学研究科物質・化学系専攻修士課程修了。同年 (株) NTT ドコモ入社。データマイニングの研究に従事。



稲村 浩 (正会員)

平成 2 年慶應義塾大学大学院理工学研究科修士課程修了。同年日本電信電話 (株) 入社。平成 10 年より NTT ドコモ。平成 28 年より公立はこだて未来大学教授。博士 (工学)。モバイルネットワーク, スマートデバイスのシステムソフトウェアに関する研究開発に従事。電子情報通信学会, ACM, IEEE 各会員。



太田 賢 (正会員)

株式会社 NTT ドコモ先進技術研究所勤務。平成 10 年静岡大学大学院博士課程修了。博士 (工学)。平成 11 年 NTT 移動通信網 (株) 入社。現在, NTT ドコモ先進技術研究所勤務。モバイルコンピューティング, 端末セキュリティ, 分散システムに関する研究に従事。共著「モバイルネットワーク」, 訳書「コンピュータネットワーク第 5 版」等。電子情報通信学会会員。



重野 寛 (正会員)

平成 2 年慶應義塾大学理工学部計測工学科卒業。平成 9 年同大学大学院理工学研究科博士課程修了。現在, 同大学理工学部教授。博士 (工学)。情報処理学会論文誌編集委員, 電子情報通信学会英文論文誌 B 編集委員等を歴任。現在, 情報処理学会マルチメディア通信と分散処理研究会主査, Secretary of IEEE ComSoc APB。ネットワーク・プロトコル, ITS 等の研究に従事。著書「ユビキタスコンピューティング」(オーム社), 「情報学基礎第 2 版」(共立出版) 等。電子情報通信学会, IEEE, ACM 各会員。



片桐 雅二 (正会員)

昭和 61 年早稲田大学大学院理工学研究科博士前期課程修了。日本電信電話，UC Berkeley 訪問研究員，NTT ドコモマルチメディア研究所，ドコモ USA 研究所等を経て，現在，NTT ドコモ先進技術研究所主幹研究員。図形

処理，マルチメディアアプリケーション，データマイニングの研究開発に従事。大阪大学博士（情報科学）。電子情報通信学会，IEEE 各会員。