

コンシューマ・サービス論文

# ランダムフォレストを用いたソーシャルネットワークサービス向けのアクティブユーザ推測と利用促進への応用

土井 千章<sup>1,a)</sup> 片桐 雅二<sup>1</sup> 川崎 仁嗣<sup>1</sup> 中川 智尋<sup>1</sup> 稲村 浩<sup>1</sup> 太田 賢<sup>1</sup>

受付日 2014年12月21日, 採録日 2015年5月21日

**概要:** ソーシャルネットワークサービスの利用履歴を用いて, 対象顧客が今後もサービスを利用し続ける可能性が高いアクティブユーザか, 近いうちに利用を停止する可能性が高い非アクティブユーザかを推測するとともに顧客の利用状況から構築したクラスタを用いて, アクティブユーザに対して利用促進を目的とした利用促進施策を実施し, その効果および有用性の評価を試みる. 実データを用いた評価結果により, 利用履歴からアクティブユーザと非アクティブユーザの推測ができることを確認する. さらに推測されたアクティブユーザに対し, 実サービスを用いた利用促進施策の試行結果から投稿数の増加率を評価する. 投稿数の増加率を評価した結果を用い, 利用促進のためのメッセージを送信しなかったグループとユーザクラスタ別に異なるメッセージを送信したグループでは, 大きな差があり有効であることを確認する.

**キーワード:** ランダムフォレスト, ソーシャルネットワーキングサービス, 顧客維持

## Active Users Prediction for Social Network Service Using Random-Forest and Its Application for User Retention

CHIAKI DOI<sup>1,a)</sup> MASAJI KATAGIRI<sup>1</sup> SATOSHI KAWASAKI<sup>1</sup>  
TOMOHIRO NAKAGAWA<sup>1</sup> HIROSHI INAMURA<sup>1</sup> KEN OHTA<sup>1</sup>

Received: December 21, 2014, Accepted: May 21, 2015

**Abstract:** This paper proposes an active users prediction model which predicts customer status that moves from an active status to inactive in the future. For the purpose of promoting the use of active users, utilization promotion by using the cluster constructed from the customer's usages are executed and evaluated the effectiveness of user cluster constructed from the customer's service usage. The evaluation result by using real data shows that can predict of active users and inactive users from the usage history. Furthermore, the evaluation result through personalized promotion for active users shows that send a different message for each user cluster increase ratio of post.

**Keywords:** random forest, social networking service, customer retention

### 1. はじめに

多くのサービスにとって既存顧客の維持は重要な課題である. 新規顧客開拓は既存顧客の維持と比較して数倍コストがかかるといわれており [1], 様々な業界で既存顧客の維持を目的とした施策が行われている. しかし, すべての顧

客に対して施策を行うのは多大なコストがかかるため, より施策の効果を期待できる顧客の選定が重要である.

先行研究として, ソーシャルネットワークサービス (以下 SNS) の利用状況を用い, サービスから近い将来離反する可能性が高い顧客か否かを事前に予測する手法が提案されている [2], [3], [4], [14]. これらの手法を用いることによって, 離反防止やサービスの利用促進を目的とし, 離反する可能性が高い顧客 (非アクティブユーザ), 可能性が低い顧客 (アクティブユーザ), それぞれに合わせた施策等

<sup>1</sup> 株式会社 NTT ドコモ  
NTT DOCOMO, INC., Yokosuka, Kanagawa 239-8536, Japan

<sup>a)</sup> chiaki.doi.tf@nttdocomo.com

の実施が可能になる。また、これらのサービスから非アクティブユーザ、アクティブユーザの発見は一般的に早ければ早いほど有用である。実際の課題への適用を考える場合には、どの程度の段階でどれほどの顧客を発見できるのかは、有用性を考える上で重要である。しかしながら、先行研究では予測に用いる利用状況データの期間の長さや予測精度の関係については論じられていない。そこで、本稿では、実証的にこれを評価し、議論する。具体的には、各顧客の利用状況からアクティブユーザ/非アクティブユーザかを推測するアクティブユーザ推測モデルを提案する。提案するアクティブユーザ推測モデルは、先行研究 [4] にてその有用性が確認されている一般的に SNS で取得可能な投稿数、閲覧数やフォロー数等のサービス利用状況を入力データとし、機械学習の一手法であるランダムフォレスト法を用いて構築する。モデル構築に用いる利用状況データの日数と推測精度の関係を明らかにするため、各ユーザの利用開始から 3 日、7 日、14 日、30 日、45 日、60 日分のデータを用いてモデルを構築し、精度の評価を行う。

一方で、サービスの利用促進を目的として顧客に行う具体的な施策内容は、顧客 1 人 1 人の特性が異なることから、個々の顧客の利用状況や趣味嗜好に合わせた施策である方が高い利用促進効果を期待できる。そのためには各顧客のデモグラフィック情報や利用状況を用いて顧客を理解し、顧客に適した施策を検討して実施する必要がある。

既存顧客の維持を目的とする際には、今後もサービスを利用する可能性が高いアクティブユーザをターゲットに考える場合と、近いうちに利用を停止する可能性が高い非アクティブユーザをターゲットにする場合とで、適切な施策が異なると筆者は考える。既存顧客維持にはどちらのユーザへの施策実施も重要であると考えられるが、本稿では特にアクティブユーザを対象として取り上げ、サービスの利用促進をねらいとした施策を実施し、既存顧客の維持を図ることにする。

各顧客を理解するため、顧客が離反するか否かを判定するとともに、顧客の利用状況を用いて顧客クラスタを作成し、分類する手法が提案されている [4]。この手法を用いることにより、所属する顧客クラスタに応じて異なる施策を実施することができる。しかしながら、先行研究では実際に施策を適用した結果の報告はなされておらず、その効果については明らかにされていない。そのため、筆者らは実験的に施策を実施することで、所属するクラスタに応じた施策の有用性評価を試みる。具体的には因子分析を用いて全顧客の利用状況からユーザクラスタの作成を行い、アクティブユーザに対して、所属する顧客クラスタごとに設計した利用促進施策を実施して、実施しなかった場合、均一な施策を実施した場合と比較し、その効果を確認する。

以降 2 章で関連研究について述べ、3 章ではアクティブユーザ推測モデルの構築と推測精度の評価、4 章ではユー

ザクラスタへの分類について述べる。5 章では 3 章で述べたアクティブユーザ推測モデルと 4 章で述べたユーザクラスタを用いた利用促進施策について説明し、利用促進施策の効果を検証する。6 章ではまとめと今後の課題について述べる。

## 2. 関連研究

Ma らは、決定木の一手法である ID3 法を用い、顧客がサービスから離反するか否かを判定するモデルを構築する方法を提案した [13]。SNS の利用状況ではないが、場所、顧客の仕事の事業規模、利用回数、サービスの満足度を説明変数として用いた。しかし、異なる機械学習の手法を用いた場合の推測精度は、評価されていない。

また、Long らは、顧客のオンラインソーシャルネットワークの 1 カ月分の利用状況から、機械学習の一手法である決定木を用いて離反者の推測を行うとともに、K-means 法を用いて離反者の利用状況に基づいてクラスタリングを行った [4]。使用された利用状況データは、ログイン数や投稿数、結び付きのある友人数等一般的な SNS で取得可能なデータに加え、3 つの各ゲームに対する使用回数、インスタントメッセージの利用状況（ログイン回数、ログイン日数、メッセージ数等）である。サービス利用開始から退会までの傾向を表す 2 つのクラスタと、インスタントメッセージやゲーム等各機能の利用状況に特化した 3 つのクラスタが得られたことが報告され、クラスタごとに異なる施策を実施できることが言及されている。しかしながら、これらの結果を用いて、クラスタごとに異なる施策を実施する等の効果検証は行われていない。

Ngonmang らは、ソーシャルネットワークでの顧客間の結び付きと離反の関係性に着目し、コミュニティの抽出とサポートベクターマシンを用いた離反の予測を行う手法を提案した [14]。コミュニティは、各顧客をノードとして扱い、グリーディサーチを用いてノード間の結び付きを探索して抽出した。顧客間の結び付きの強さに注目することにより約 8 割程度の推測精度で離反者を予測できることが確認された。しかし、この手法では、顧客同士の結び付きが弱いサービスでは、顧客間の結び付きが算出できず、離反の推測が行えない可能性がある。

## 3. アクティブユーザ推測モデル

本章では、アクティブユーザ推測モデルの構築手法および構築したモデルの推定精度について述べる。

### 3.1 使用データ

本稿では、株式会社ドコモ・インサイトマーケティング [7] から提供されている「みんレポ」[8] サービスの利用状況データを用いて検討を行う。「みんレポ」は買ったもの、食べたもの、行った場所の写真や感想をレポートとして投

表 1 使用データ  
Table 1 Dataset.

データ項目	内容
ログイン数	アプリケーションを1回以上起動した日数
投稿数	レポの投稿数
閲覧数	レポの閲覧数
いいね数	レポに対する“いいね”ボタンの押下回数
いいね取得数	自分のレポに対する“いいね”取得数
ウィッシュ数	レポに対する“ウィッシュ”ボタンの押下回数
ウィッシュ取得数	自分のレポに対する“ウィッシュ”取得数
フォロー数	フォローしている人数
フォロワー数	フォローされている人数
コメント数	レポに対してつけたコメント数
コメント取得数	自分のレポに対してつけたコメント数
バッジ取得数	バッジの取得数
登録日	みんなレポサービスを使用開始した日付
最終利用日	みんなレポサービスを最後に使用した日付

稿し、他者とシェアを行うアプリケーションである。

本稿では、19,842人分の8カ月間の操作履歴データを用い、ユーザごとに表1に示すデータ項目を、対象期間を指定して算出し利用する。

### 3.2 アクティブユーザ推測モデルの構築

本節では、アクティブユーザと非アクティブユーザの定義、および機械学習の手法を用いてアクティブユーザ推測モデルを構築する手法を説明する。

アクティブユーザは、サービスの登録日から最終利用日の間隔日数がN日以上であり、サービスの最終利用日から設定した基準日まで30日以上経過していない顧客と定義する。基準日は、全データ8カ月の中の初日から数えて5カ月分のデータの最終日とする。それ以外のサービスを利用した期間がN日未満である、もしくは最終利用日から基準日まで30日以上経過している顧客を非アクティブユーザと定義する。

説明変数として3.1節で示した登録日、最終利用日以外のログイン回数、投稿数、閲覧数、いいね数、いいね取得数、ウィッシュ数、フォロー数、フォロワー数、コメント数、コメント取得数、バッジ取得数を用いる。また、モデル構築に用いる学習データとして、利用する期間による推定精度の違いを明確に評価するために、使用開始日から3日、7日、14日、30日、45日、60日分のデータセットを用意した。アクティブユーザと非アクティブユーザのユーザ数に偏りがあると推定精度に影響を及ぼす可能性があるため、アクティブユーザと非アクティブユーザが同数になるようにランダムサンプリングした計14,220人分のデータを用いてモデルを構築する。これは、実際の母集団よりもアクティブユーザの人数の比率が高くなっているため、

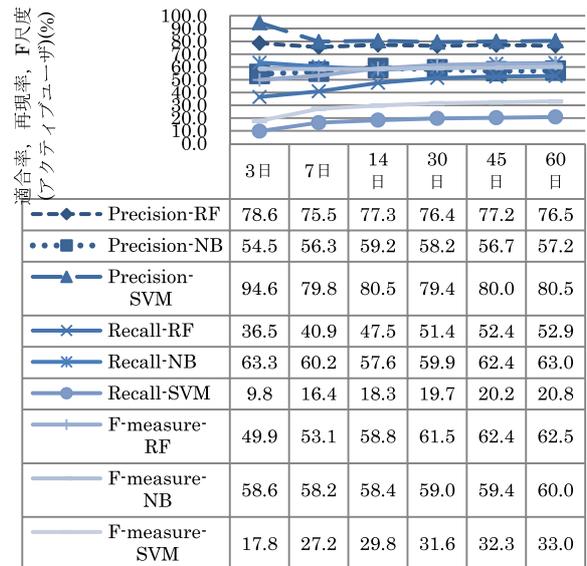


図1 使用日数、手法別の適合率、再現率、F尺度 (アクティブユーザ)  
Fig. 1 Active user's precision, recall and f-measure by each method and period of use.

アクティブユーザの特徴が強調されたモデルが構築されることを示す。しかし、本研究ではアクティブユーザの推定に主眼を置いているため、アクティブユーザ推測モデルの構築後にモデルの補正等は行わない。

モデルの構築に用いる手法は、先行研究[4]と同様に一般的に分類問題に対して用いられる決定木の手法からランダムフォレスト法[10]と、単純ベイズ法[11]、サポートベクターマシン[12]とする。ランダムフォレストは、説明変数をランダムサンプリングして作成された複数の決定木を用いて目的変数の推定を行う手法である。弱学習機を複数組み合わせることによって、精度の高いモデルの構築が実現できる。決定木の深度、抽出する説明変数の個数については、Breiman[10]の基準を用いた。単純ベイズ法は、独立仮定とベイズの定理に基づいた手法であり、多値判別等にも使用される。また、サポートベクターマシンは2値判別を行う手法であり、アクティブユーザまたは、非アクティブユーザの推測に有効であると考えられる。本稿では、最もよく用いられる手法の一つであるRBFカーネルを採用した。

先行研究[4]では、決定木以外の手法が評価されていなかった。そのため、本研究では手法ごとの推測精度を評価して使用すべき手法を明確にするため、これらの手法を用いてモデルを構築する。

### 3.3 推定精度の評価

本節では、3.2節で述べたアクティブユーザ推測モデルの評価を行う。モデルの評価は、10分割交差検証を用い、9割をモデルの構築、1割をモデルの評価に使用する。評価尺度は、それぞれ交差検証結果より得た適合率と再現率とF尺度を用いる。図1、図2にアクティブユーザと

非アクティブユーザの使用日数別および手法別に適合率 (Precision), 再現率 (Recall), F 尺度 (F-measure) を示す。手法は, RF はランダムフォレスト法, NB は単純ベイズ法, SVM はサポートベクターマシンを表す。

本研究では, アクティブユーザを推定することを目的としており, アクティブユーザの推定精度に着目する。アクティブユーザを用いて説明すると, 適合率は, 推定したアクティブユーザの中にどの程度アクティブユーザが含まれていたかを示す。再現率は, 全アクティブユーザのうち, どの程度アクティブユーザを推定できたかを示す。F 尺度は, 適合率と再現率の調和平均を求めた値であり, F 尺度の値が高いモデルは推定性能が高いことを示す。本研究では F 尺度を推定精度としてアクティブユーザ推定モデルを評価する。

アクティブユーザの推定については, 3 日, 7 日分のデータを用いた場合は, 単純ベイズ法で構築したモデルの推定精度が最も高かった。14 日以上分のデータを用いた場合は, ランダムフォレスト法で構築したモデルの推定精度が, 単純ベイズ法で構築したモデルの推定精度よりも高いことを確認した。また, 45 日, 60 日分のデータを用いてランダムフォレスト法で構築したモデルの推定精度の値の差は 0.1% であり, 推定精度の飽和を確認した。これは, 60 日以上分のデータを用いてモデルの構築を行っても, 推定精度がほぼ変化しないことを示す。そのため, 本研究ではアクティブユーザ推定モデルの構築に 60 日までのデータを用いて評価を行った。

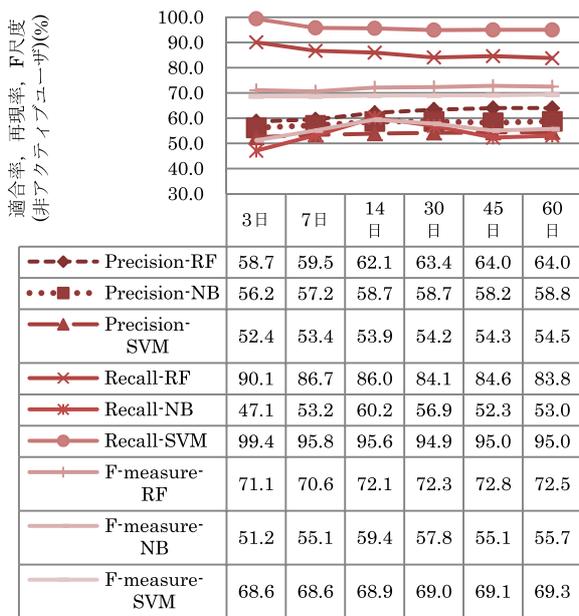


図 2 使用日数, 手法別の適合率, 再現率, F 尺度 (非アクティブユーザ)

Fig. 2 Non-active user's precision, recall and f-measure by each method and period of use.

#### 4. ユーザクラスタの構築

本章では, 因子分析を用いたユーザクラスタの構築について述べる。

##### 4.1 使用データ

ユーザクラスタの構築には, 3.1 節にて説明した「みんレポ」サービスの全顧客 19,842 人分のうち, 利用状況の特徴を抽出するため, 基準日数 (M 日) 以上サービスの利用がある 14,143 人分のデータを用いる。基準日数 (M 日) は, 全顧客のサービス利用期間内のログイン回数分布を考慮して設定した。

使用する説明変数は, 表 1 に記載のデータ項目のうち, 登録日, 最終利用日以外の 12 種類とする。ユーザクラスタの構築には, 各顧客の上記 12 種類の説明変数を算出し, 利用日数で除算して正規化したデータを用いる。

##### 4.2 ユーザクラスタの構築

ユーザクラスタの構築のために, 4.1 節で述べた 12 種類のデータ項目を用いて因子分析を行い, 4 つの因子を抽出した。抽出した因子数は, 作成したスクリープロットから固有値の落ち込みを確認し, 4 つの軸に決定した。表 2 に因子分析の結果を示す。

各因子の特徴を, 因子負荷量を用い, 主観的に命名したものを表 3 に示す。投稿数に特徴が表れたのは, 第 1 因子のオピニオンリーダータイプと第 2 因子のログイン・投稿・バッジ収集タイプであった。オピニオンリーダータイプは, フォロワー数や投稿に対して他者からの反響が確認できる, いいね取得数やウィッシュ取得数, コメント取得数に特徴が表れた。一方で, ログイン・投稿・バッジ収集タイプは, 他者からの反響を表す説明変数に特徴が見られな

表 2 因子分析結果

Table 2 Result of factor analysis.

	第一因子	第二因子	第三因子	第四因子
コメント取得数	0.87		0.34	0.29
フォロワー数	0.80	0.19	0.32	0.13
ウィッシュ取得数	0.84	0.28	0.13	0.26
いいね取得数	0.86	0.28	0.20	0.25
バッジ取得数	0.20	0.91	0.11	
ウィッシュ数	0.26			0.84
コメント数	0.23		0.75	
閲覧数	0.40	0.26	0.72	0.33
フォロワー数	0.32	0.14	0.32	0.29
いいね数	0.19	0.19	0.14	0.44
ログイン数	0.16	0.81	0.11	
投稿数	0.46	0.46	0.38	0.33

表 3 因子名と人数

Table 3 Number of customers in 4 factors.

	因子名	人数(人)
第一因子	オピニオンリーダータイプ	38
第二因子	ログイン・投稿・バッジ収集タイプ	13941
第三因子	コメント・閲覧・フォロータイプ	93
第四因子	ウィッシュ・いいねタイプ	71

表 4 利用促進施策対象クラス

Table 4 Target cluster of utilization promotion.

	1 番目	2 番目	人数(人)	
クラスタ 1	第二因子 ログイン・投稿・バッジ収集タイプ	対応なし	7045	
クラスタ 2		第一因子	オピニオンリーダータイプ	4787
クラスタ 3		第三因子	コメント・閲覧・フォロータイプ	1582
クラスタ 4		第四因子	ウィッシュ・いいねタイプ	527

かったが、バッジ数に特徴が見られた。第 3 因子のコメント・閲覧・フォロータイプは、コメント数、レポの閲覧数、フォロー数に特徴が表れた。第 4 因子のウィッシュ・いいねタイプからは、投稿に対して行うウィッシュ数、いいね数に対して特徴が表れた。

顧客は各因子の特徴を複合的に持っていると考えられるが、利用状況の理解を容易にするため、ここでは単純化して特徴が表れる上位の因子でその顧客の利用状況を表すことにする。特徴が最も表れた因子により顧客を分類した結果を表 3 に示す。因子別の人数では、ログイン・投稿・バッジ収集タイプが大半を占め、支配的となった。さらに深く顧客を理解するため、本稿では実験的に最上位因子がログイン・投稿・バッジ収集タイプのユーザを、2 番目に説明力の強い因子を用い 4 つのクラスタに分類することにした。表 4 に各クラスタ別の人数を示す。ログイン・投稿・バッジ収集タイプのみで利用状況が表され他の因子の寄与が認められない顧客は、7,045 人であり、2 番目に特徴が表れた因子を保持しているのは、6,896 人であった。

本稿では以降、全ユーザの大半を占めたログイン・投稿・バッジ収集タイプの因子に最も特徴が表れたユーザを取り上げることとし、表 4 に示したクラスタをもとに、5 章で利用促進施策の効果を検証する。

## 5. 利用促進施策

本章では、3 章のアクティブユーザ推測モデルを用いてアクティブユーザと判定された顧客に対して、4 章で述べたユーザクラスタ別に異なる利用促進施策を実施することに対する効果検証の試みについて述べる。

### 5.1 利用促進施策について

本節では、利用促進施策の概要について説明する。本利

表 5 利用促進施策スケジュール

Table 5 Schedule of utilization promotion.

	第一週	第二週	第三週	第四週
グループ 1 配信なし	-	-	-	-
グループ 2 平常通り	N	N	N	N
グループ 3 クラスタ別に配信	N	M	N	M

M: メッセージ R を用いた配信

N: ノーティフィケーションを用いた配信

用促進施策は、アクティブユーザのサービスの利用促進を目的としてユーザクラスタ別に異なる施策を実施する。

利用促進施策は、アクティブユーザに対し、表 5 に示すようにサービスの利用を促進するメッセージをメッセージ R [15] とアプリケーションのノーティフィケーションを用いて顧客へ通知する。送信頻度は、週に 1 回とし、4 週間実施する。

利用促進施策の効果を検証するため、メッセージを配信しないグループ、通常配信している内容を配信するグループ、クラスタ別に検討されたメッセージを配信するグループの 3 グループに分けて施策を実施する。本稿で施策を実施する対象者は、アクティブユーザである。そのため、サービスの登録日から 14 日分のデータを用い、3.2 節で示したアクティブユーザ推測モデルによってアクティブユーザと推測された顧客を対象とした。各グループの人数は、アクティブユーザと推測された顧客から各クラスタの顧客数がグループ間で同数になるようにランダムサンプリングを行い、514 人とした。各グループを構成するクラスタごとの人数は、クラスタ 1 が 118 人、クラスタ 2 が 280 人、クラスタ 3 が 93 人、クラスタ 4 が 23 人であった。なお、本利用促進施策の主眼は、投稿数を増加させることである。そこで、これらの施策を実施する 3 つのグループ間に差はないという仮説を事前に検定するため、顧客の利用促進施策 1 週間前から 3 週間前の投稿数に対して有意水準  $p = 0.05$  で t 検定を実施し、有意差がないことを確認した。

### 5.2 メッセージ

本節では、利用促進施策で顧客へ送信するメッセージについて説明する。

メッセージを送信する対象は、グループ 2 とグループ 3 に属する顧客である。メッセージは、表 6 に示すように投稿や閲覧、検索を促進する内容とした。グループ 2、グループ 3 のどちらに送信するメッセージも週ごとに異なる内容とした。クラスタ別のメッセージは、クラスタ 1 には、投稿を促すメッセージとした。クラスタ 2 は積極的に投稿し、他者からの投稿に対する反応も多いクラスタのため、投稿に関する内容を提案する“お題”機能へ誘導や投稿を促すメッセージとした。クラスタ 3 は興味のある投稿を見

表 6 グループ別メッセージ例

Table 6 Example of messages for each group.

グループ	週	クラス	メッセージ
グループ 2	第一週		みんなのレボゼひ見て♪春色のモノ、集まってきています!
	第二週		最近、どこ行った? お出かけした場所レボしてね♪
	第三週	-	最近なに買った? 増税を意識したモノは何かある?
	第四週		今週末どこにお出かけ? 連休中に見つけた素敵なお店は?
グループ 1	第一週	クラス1	次に投稿するレボは何? みんなに詳しく教えてあげてね♪
		クラス2	新しいお題が出たよ。贅沢気分になったモノ教えてね♪
		クラス3	検索窓も使ってみて♪ 気になるワードでレボを検索してね
		クラス4	みんなレボ見て何買った? つられて買ったモノもレボしてね
	第二週	クラス1	買ったお店の名前、行った場所も みんなに教えてあげてね
		クラス2	みんなにあなたのレボについて、詳しく教えてあげてね♪
		クラス3	検索機能もぜひ使ってね
		クラス4	みんなのレボを見て、「思わず買ってしまった」「行ってしまった場所」はありませんか? レボを参考に買ったものは?
	第三週	クラス1	ホワイトデーらしいモノどんなモノ買った? 食べた?
		クラス2	最近食べた旨いモノは? レボ投稿でみんなと仲良くなるよ
		クラス3	人気のレボ、もう見た? タイムラインから検索してみよう
		クラス4	レボ見て買ったモノは? あなたの感想ぜひレボしてみよう
	第四週	クラス1	三連休、あなたの過ごし方を教えてくださいね♪ 観光した場所の入場料や買ったモノのボリューム感などもみんなに教えてあげてね
		クラス2	三連休、あなたの過ごし方を教えてくださいね♪ あなたが行った場所、買ったモノの情報を詳しく教えてあげてね♪
		クラス3	地図画面からレボを検索したことはありますか? ? ぶだんの活動エリアや、旅行してみたい場所、まったく馴染みのない土地... などなど、ぜひ、地図画面を開いて、レボを探してみてください
		クラス4	地図画面からレボを見たことはありますか? ? ぶだんの活動エリアや、旅行した場所... などなど、ぜひあなたのレボも地図画面に載せてください

つけ、フォロー機能を利用するクラスのため、より興味のある投稿や顧客を見つけることができる「検索窓」機能についてのメッセージとした。クラス4は、気になる投稿に対していいねやウィッシュを行うクラスのため、影響を受けた投稿に関して投稿を促すメッセージとした。

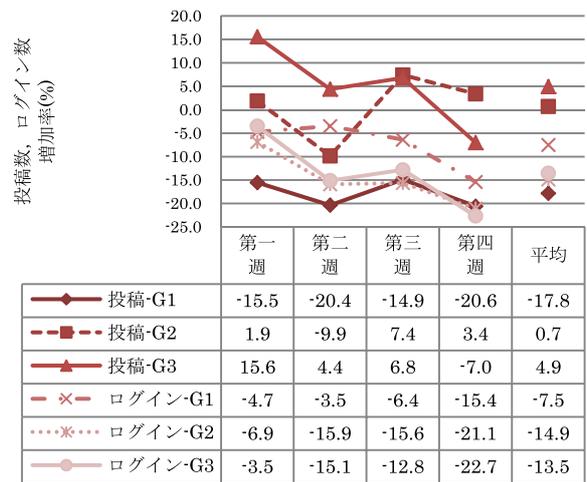


図 3 投稿数, ログイン数増加率  
Fig. 3 Increase ratio of post and login.

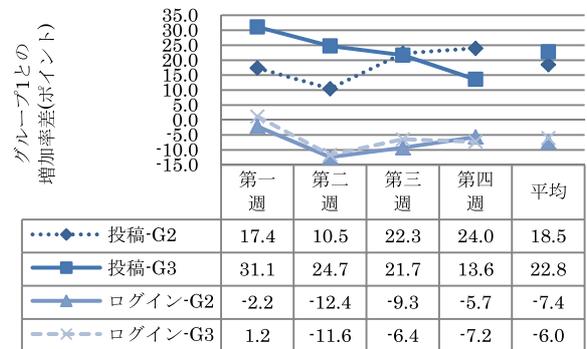


図 4 グループ 1 に対する投稿数, ログイン数増加率差  
Fig. 4 Difference of increase ratio of post for group 1.

### 5.3 効果検証

本節では、利用促進施策の効果検証結果について述べる。効果検証は、本利用促進施策の主眼である投稿数と、顧客のログイン数の増加率を確認する。図 3 は、投稿数とログイン数の増加率を利用促進施策が実施された各週と利用促進施策を実施した1週間前の投稿数およびログイン数を比較して算出した結果であり、図 4 はグループ 1 とグループ 2 およびグループ 3 の差分を算出した結果である。図内では、グループを G と表記する。利用促進施策を実施した1週間前に、全グループに対してノーティフィケーションによる利用促進施策を実施していたことに注意する。

図 3 の各グループの投稿数増加率から利用促進施策を実施していないグループ 1 の投稿数増加率は低くなっていることが読み取れる。これは、利用促進施策を実施しないと投稿数が減少することを示している。利用促進施策を実施していないグループ 1 とグループ 3 の投稿数増加率の差は平均 22.8 ポイントであり、グループ 1 とグループ 2 の投稿数増加率の差は平均 18.5 ポイントであった。グループ 2 とグループ 3 を比較して約 4.3 ポイントの差が見られたことにより、投稿数の増加には顧客のタイプに合わせた内容のメッセージを送信する施策が有効である可能性を確認

表 7 媒体別投稿数増加率 (平均)

Table 7 Increase ratio of post by each medium.

	ノーティフィケーション 増加率	メッセージ R 増加率
グループ 3	26.4%	19.2%

した。

ログイン数については、全体的に増加率が時間経過ともなって低下しており、グループ 2 とグループ 3 においては、グループ 1 よりも増加率が低い結果となった。この結果より、ログイン数と投稿数は異なる傾向を持つことが明らかになった。

次に、表 7 に媒体別の投稿数増加率を示す。顧客へのメッセージ配信に用いる媒体は、グループ 3 の第 1 週、第 3 週にノーティフィケーション、第 2 週、第 4 週にメッセージ R を用いた。各媒体別の投稿数増加率の平均値はノーティフィケーションの場合が 26.4%、メッセージ R の場合が 19.2% であり、ノーティフィケーションを用いた場合、投稿数増加率が 7.2 ポイント高かった。また、ログイン数増加率においては、メッセージ R を用いた第 2 週と第 4 週のログイン数増加率の平均値がノーティフィケーションを用いた第 1 週と第 3 週の平均値と比較して低かった。これらの結果より、本研究で対象としたサービスにおいて利用促進施策に用いる媒体はノーティフィケーションが有用であることが明らかになった。

次にクラスタ別の投稿数増加率を確認する。図 5 にクラスタ別の投稿数増加率を示す。図内では、クラスタを CI と表記する。グループ 1 の投稿数増加率から、利用促進施策を実施しないと全クラスタの投稿数増加率が低下することを確認した。施策を実施したグループと実施していないグループの投稿数増加率を比較するため、図 6 にグループ 1 と比較したクラスタ別投稿数増加率の差を示す。グループ 2 とグループ 3 を比較し、グループ 3 のクラスタ 1、クラスタ 2 においては投稿数増加率差の平均値は高いが、クラスタ 3、クラスタ 4 は、低い傾向が見られた。また、グループ 3 のクラスタ 1 においては、第 1 週から第 4 週まで、クラスタ 2 においては、第 1 週から第 3 週までグループ 2 の対応するクラスタよりも投稿数増加率差が高かった。これは、クラスタ 1、クラスタ 2 は、積極的に投稿を行うクラスタであり、投稿に関する内容の提案や投稿を促進するメッセージを送ることで投稿数に影響を与えることができたためと考えられる。一方で、クラスタ 3 は興味のある投稿を見つけることによって自身の投稿数の増加も見込めるのではないかと筆者らは考え、“検索窓”機能についてのメッセージを送信したが、グループ 2 のクラスタ 3 で確認できるように具体的な投稿に関する内容を送信した方が効果的であった。クラスタ 4 は気になる投稿に関して投

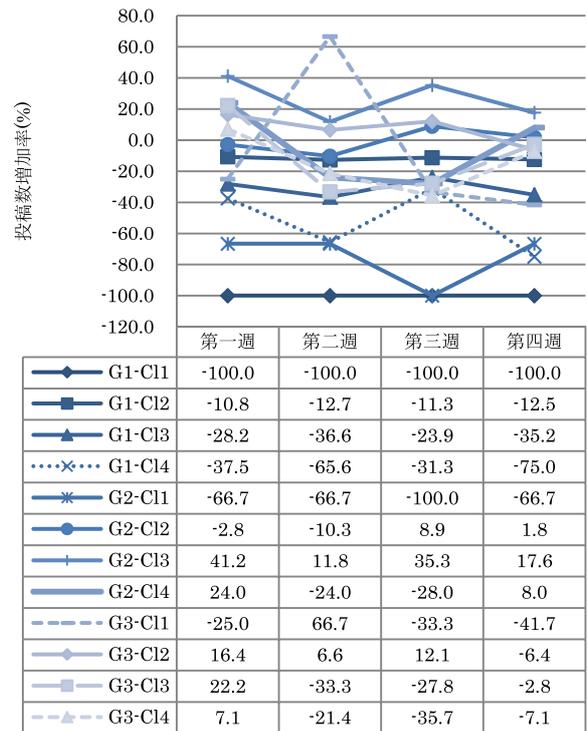


図 5 クラスタ別投稿数増加率

Fig. 5 Increase ratio of post for each cluster.

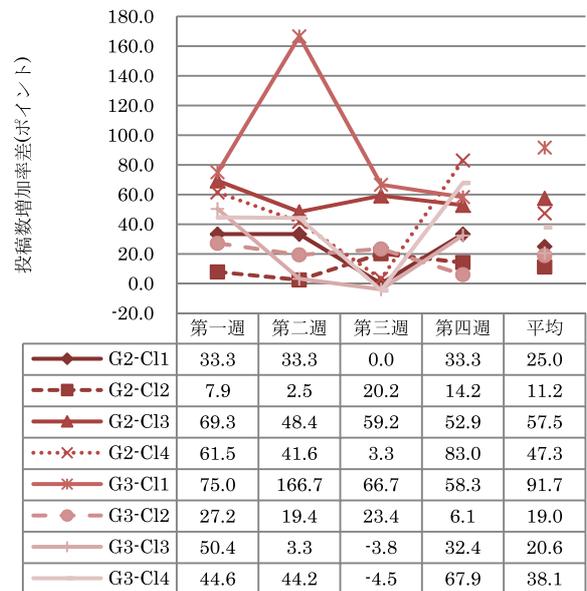


図 6 グループ 1 とのクラスタ別投稿数増加率差

Fig. 6 Difference of increase ratio of post for each cluster between group 1.

稿を促すメッセージを送信したが、投稿数を増加させるには、具体的な投稿内容を送信した方が効果的であることが分かった。これらの結果より、投稿を促進するメッセージを各クラスタに合わせて設定し、送ることによって投稿数増加の効果が得られる可能性を実証できた。

次にクラスタごとのログイン数の増加率を確認するため、図 7 にクラスタ別のログイン数増加率を示す。ログイ

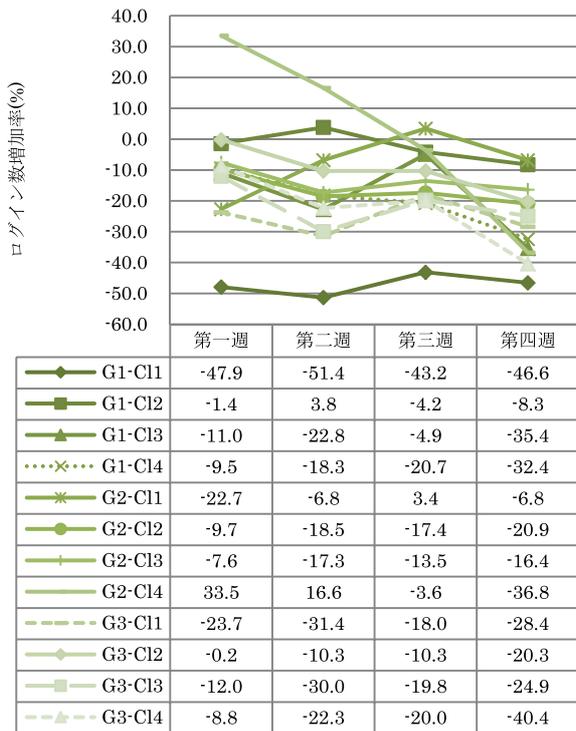


図 7 クラスタ別ログイン数増加率  
Fig. 7 Increase ratio of login for each cluster.

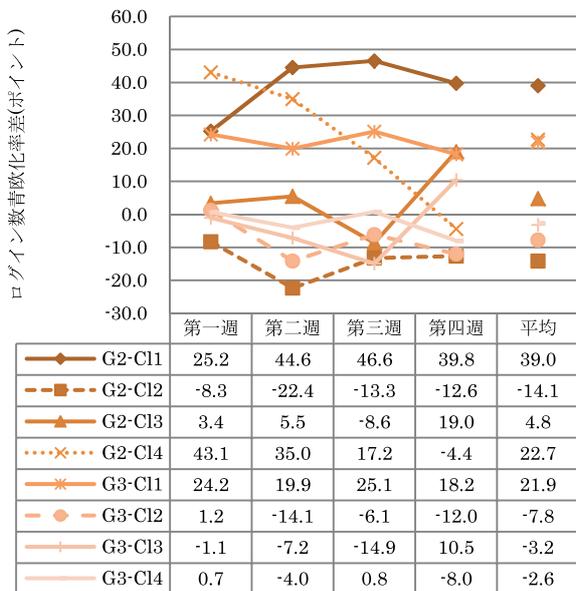


図 8 グループ 1 とのクラスタ別ログイン数増加率差  
Fig. 8 Difference of increase ratio of login for each cluster between group 1.

数の増加率は、クラスタ別に確認しても施策実施 1 週間前と比較してグループ 2 のクラスタ 1 とクラスタ 4 以外、低下していることを確認した。図 8 にグループ 1 と比較したクラスタ別ログイン数増加率の差を示す。グループ 2 とグループ 3 のクラスタ 1 においては、利用促進施策を実施することによってログイン数の増加率差の平均が向上していることを確認した。しかしながら、全体的にログイン

数の増加率差は低い傾向が見られた。

## 6. おわりに

本研究では、ソーシャルネットワークサービスの利用履歴を用いて、対象顧客が今後サービスを利用する可能性が高いアクティブユーザか近いうちに利用を停止する可能性が高い非アクティブユーザかを推定するアクティブユーザ推測モデルの提案を行った。アクティブユーザ推測モデルは、機械学習の手法であるランダムフォレスト法、単純ベイズ法、サポートベクターマシンを用いて構築し、各手法で構築されたモデルに対して F 尺度を用いて評価を行った。アクティブユーザの推定については、3 日、7 日分のデータを用いた場合は単純ベイズ法で構築したモデルの推定精度が最も高かった。14 日以上以上のデータを用いた場合は、ランダムフォレスト法で構築したモデルの推定精度が、単純ベイズ法で構築したモデルの推定精度よりも高いことを確認した。また、45 日、60 日分のデータを用いてランダムフォレスト法で構築したモデルの推定精度の値の差は 0.1% であり、推定精度の飽和を確認した。本研究では、14 日分のデータを用いてランダムフォレスト法で構築したモデルを採用したが、アクティブユーザ/非アクティブユーザは、サービスを使用している期間内、かつサービスを利用してからより短い期間で推測できることが望ましい。そのため、実運用では、複数の日数分のデータを用いて、任意で設定された推定精度が担保される日数のデータで構築されたモデルを採用することが有用であると考えられる。

また、顧客の利用状況に合った利用促進施策を実施するため、因子分析を用いてユーザクラスタを作成した。利用促進施策は、アクティブユーザに対し、ユーザクラスタ別に用意したメッセージをメッセージ R とアプリケーションの通知機能を用いて顧客へ送信した。その結果、利用促進施策を実施していないグループ 1 とグループ 3 の投稿数増加率の差は平均 22.8 ポイントであり、グループ 1 とグループ 2 の投稿数増加率の差は平均 18.5 ポイントであった。グループ 2 とグループ 3 を比較して約 4.3 ポイントの差が見られたことにより、投稿数の増加には顧客のタイプに合わせた内容のメッセージを送信する施策が有効である可能性を確認した。

さらに、利用促進施策の実施より投稿数の増加率に影響が出やすいクラスタとそうでないクラスタが明らかになった。今回の利用促進施策では、クラスタ別に施策を実施することにより一時的に投稿数の増加率を向上させることができることを確認した。しかしながら、投稿数の増加率が時間経過とともに単調減少する傾向が確認されており、継続的な投稿数の増加は課題である。

今後も既存顧客の維持を目的とした利用促進施策をクラスタ別に継続して行い、反応しやすいクラスタや施策内容を明確にしていきたい。また、本稿で提案したアクティブ

ユーザ推測モデルを用いて、非アクティブユーザと推定された顧客の利用促進方法を検討していきたい。

謝辞 本研究を進めるにあたり、株式会社ドコモ・インサイトマーケティングから「みんレポ」サービスのデータ提供や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] 池田謙一, 唐沢 穰, 工藤恵里子, 村本由紀子: 社会心理学, 有斐閣 (2010).
- [2] Rogers, E.M.: *Diffusion of Innovations*, Free Press (1971).
- [3] Oentaryo, R., Lim, E., Lo, D., et al.: Collective Churn Prediction in Social Network, *Proc. IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM2012)*, pp.210-214, IEEE Computer Society (2012).
- [4] Long, X., Yin, W., An, L., et al.: Churn Analysis of Online Social Network Users Using Data Mining Techniques, *Proc. International Multi Conference of Engineers and Computer Scientists (IMECS2012)*, IMECS, pp.551-556 (2012).
- [5] Zhang, G.: Customer Segmentation Based on Survival Character, *International Conference on Wireless Communications, Networking and Mobile Computing (WICOM2007)*, pp.3391-3396 (2007).
- [6] Katz, E. and Lazarsfeld, F.: *Personal Influence: The Part Played by People in the Flow of Mass Communications*, Transaction Publishers (2005).
- [7] 株式会社ドコモ・インサイトマーケティング(オンライン), 入手先 (<http://www.dcm-im.com/>) (参照 2014-01-22).
- [8] みんレポ紹介と基本的な操作, 株式会社ドコモ・インサイトマーケティング(オンライン), 入手先 (<http://minrepo.com/help/>) (参照 2014-02-24).
- [9] Borko, F.: *Handbook of Social Network Technologies and Applications*, Springer (2010).
- [10] Breiman, L.: *Machine Learning*, Kluwer Academic Publishers (2001).
- [11] 照井伸彦: ベイズモデリングによるマーケティング分析, 東京電機大学出版局 (2008).
- [12] 金 明哲: Rによるデータサイエンス, 森北出版株式会社 (2007).
- [13] Ma, H., Qin, M. and Wang, J.: Analysis of the Business Customer Churn Based on Decision Tree Method, *The 9th International Conference on Electronic Measurement & Instruments (ICEMI 2009)*, Vol.4, pp.818-821 (2009).
- [14] Ngonmang, B., Viennet, E. and Tchunte, M.: Churn Prediction in a Real Online Social Network Using Local Community Analysis, *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM2012)*, pp.282-288, IEEE Computer Society (2012).
- [15] メッセージサービスとは, 株式会社NTTドコモ(オンライン), 入手先 (<https://www.nttdocomo.co.jp/service/customize/messagefr/about/>) (参照 2014-05-14).



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, 大規模データ

を用いた行動分析の研究に従事。



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R&D総務部情報企画担当部長。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。



川崎 仁嗣 (正会員)

株式会社NTTドコモ先進技術研究所勤務。平成20年筑波大学システム情報学研究科博士前記課程修了。同年(株)NTTドコモ入社。モバイルコンピューティング, 端末セキュリティ, 分散システムに関する研究に従事。



中川 智尋 (正会員)

株式会社NTTドコモ先進技術研究所勤務。2000年東京大学大学院工学系研究科電子情報工学専攻修士課程修了。同年(株)NTTドコモ入社。現在, 同社先進技術研究所勤務。入社以来, モバイルコンピューティング, 端

末セキュリティ, コンテキストウェア・コンピューティングの研究に従事。



稲村 浩 (正会員)

株式会社 NTT ドコモ先進技術研究所勤務。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 6 年から 7 年にカーネギーメロン大学計算機科学科にて訪問研究員。平成 10 年より NTT

ドコモ。平成 22 年慶應義塾大学大学院開放環境科学専攻後期博士課程単位取得退学。博士(工学)。



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。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。