

決定木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モデルの提案

土井 千章^{†1} 川崎 仁嗣^{†1} 中川 智尋^{†1} 片桐 雅二^{†1}
稲村 浩^{†1} 太田 賢^{†1}

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予測する、アクティブユーザ推測モデルの提案を行う。提案するモデルは、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。構築したモデルの評価結果から、利用開始から14日分の利用履歴を用いた場合、70.3%の精度でアクティブユーザと非アクティブユーザの予測ができることを確認した。先行研究ではこれらの顧客の発見は早ければ早いほど有用であるにも関わらず、モデル構築に用いるデータの期間による予測精度の違いについては明らかにされていなかった。本研究では利用できるデータの期間と予測精度の関係を実データを用いて明らかにする。さらに、顧客の利用状況からクラスタを構築し、クラスタ別にメッセージを送信する利用促進施策を実験的に実施し、その効果および有用性の評価を試みる。

Building an Active Users Prediction Model for Social Network Service using Decision Tree

Chiaki Doi^{†1} Satoshi Kawasaki^{†1} Tomohiro Nakagawa^{†1} Masaji Katagiri^{†1}
Hiroshi Inamura^{†1} Ken Ohta^{†1}

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言われており^[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、全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][5]}。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、そうではない顧客、それぞれに合わせた施策等の実施が可能になる。また、これらのサービスから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)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程発見できるのかは、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予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、顧客一人一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とする。

^{†1}(株)NTTドコモ

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共に、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、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予測精度の評価、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予測精度は、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予測を行うとともに、K-means 法を用いて利用状況に基づいて顧客の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、結びつ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つ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つきを探索して

抽出した。顧客間の結びつきの強さに注目することにより約 8 割程度の予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、顧客同士の結びつきが弱いサービスでは、顧客間の結びつきが算出できず、離反の予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及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「みんなレポ」^[8]サービスの利用状況データを用いて検討を行う。「みんなレポ」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として投稿し、他者とシェアを行うアプリケーションである。

本稿では、19842 人分の 8 か月間の操作履歴データを用い、ユーザ毎に表 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。

表 1 使用データ

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

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。それ以外の N 日以上サービスを利用していない、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。また、基準日は、

アクティブユーザ/非アクティブユーザを判定するため、全データ 8 か月の中の初日から数えて 5 か月分のデータの最終日とした。

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有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のユーザ数が同数になるようにランダムサンプリングした計 14220 人分のデータを用いてモデルを構築する。

モデルの構築に用いる手法は、一般的に分類問題に対して用いられる決定木のランダムフォレスト法^[10]と、サポートベクターマシン^[11]、単純ベイズ^[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。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予測に有効であると考えられる。本稿では、最も良く用いられる手法の一つである RBF カーネルを採用した。また、単純ベイズは、独立仮定とベイズの定理に基づいた機械学習法であり、多値判別等にも使用される手法である。手法ごとの予測精度の比較を行うため、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。評価尺度としては、再現率 R_{average} を用いることにする。ここで再現率 R_{average} とは、アクティブユーザおよび非アクティブユーザに対し、それぞれ交差検証により再現率 R_{active} 、 $R_{\text{non-active}}$ を得、それらを平均した値である。アクティブユーザの再現率 R_{active} は以下の式(1)で、非アクティブユーザの再現率 $R_{\text{non-active}}$ は式(2)で算出され、再現率 R_{average} は以下の式(3)で算出される。

$$R_{\text{active}} = X_{\text{active}} / Y_{\text{active}} * 100 \quad (1)$$

$$R_{\text{non-active}} = X_{\text{non-active}} / Y_{\text{non-active}} * 100 \quad (2)$$

$$R_{\text{average}} = (R_{\text{active}} + R_{\text{non-active}}) / 2 \quad (3)$$

ここで、 X_{active} ($X_{\text{non-active}}$) はアクティブユーザ (非アクティブユーザ) と推定され、かつ実際にアクティブユーザ (非アクティブユーザ) であった人数を表し、 Y_{active} ($Y_{\text{non-active}}$) は実際にアクティブユーザ (非アクティブユーザ) である人数を表す。

評価結果を図 1 に示す。Base は、単純にアクティブユーザか非アクティブユーザの予測確率を考慮した場合であり、RF はランダムフォレスト法、NB は単純ベイズ、SVM はサポートベクターマシンの再現率を示す。

評価結果より、ランダムフォレスト法を用いたモデルの再現率が高いことが確認できた。利用開始から 3 日分と 7 日分のデータを用いて作成したモデルを用いると約 6 割程度の精度で予測ができることが分かった。14 日分のデータを用いると 7 割程度の精度で予測できることを確認した。14 日分と 30 日分のデータを用いたモデルの再現率の差は、7.1%であり、30 日分のデータを用いると、約 8 割の精度で予測ができることが分かった。

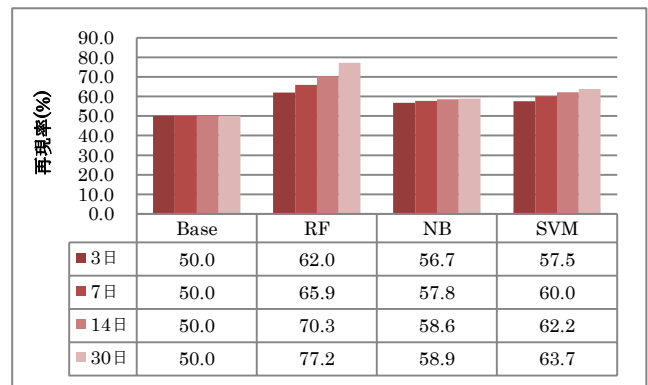


図 1 使用日数、手法別 再現率(R)

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「みんなレポ」サービスの全顧客 19842 人分の内、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 14143 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内、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。

表2 因子分析結果

	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

	因子名	人数(人)
第一因子	オピニオンリーダタイプ	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タ

	1 番目	2 番目	人数(人)
クラスタ1	第二因子	対応なし	7045
クラスタ2	積極的に投稿	第一因子 オピニオンリーダタイプ	4787
クラスタ3	投稿・バッジ収集	第三因子 投稿から友達探しタイプ	1582
クラスタ4	集タイプ	第四因子 気になる投稿をチェックタイプ	527

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、第一因子のオピニオンリーダタイプと第二因子のログイン・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かったが、バッジ数に特徴が見られた。第三因子のコメント・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。第四因子のウィッシュ・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た上位2つの因子でその顧客の利用状況を表す。特徴が最も表れた因子別の人数を表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最も多く支配的となった。さらにより深く顧客を理解するため、本稿では実験的にログイン・投稿・バッジ収集タイプのユーザを二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を持つ顧客は、7045人であり、二番目に特徴が表れた因子を保持しているのは、6896人であった。

Longら^[4]が分類した5つのクラスタは、サービス利用開始から退会までの傾向を表した2つのクラスタと3つの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クラスタであった。利用開始から退会までの傾向を表したクラスタは、登録はしたが退会するまでにサービスの面白さを発見できなかったと考えられる”come-and-leave users”と全ての機能を少しずつ使用するが最終的にはサービスを退会する”a-bit-of everything users”であった。各機能の利用状況に特化したクラスタは、インスタントメッセージ機能のみを利用する”IM-active-only users”と友人を探すために利用する”friends finders”とゲームを主に利用する”App players”であった。関連研究^[4]では、インスタントメッセージやゲーム等1つの機能に特化したクラスタが作成されているのに対し、本稿で提案したユーザクラスタは、いいねやフォロー等複数機能の利用状況が表現されたクラスタが作成された。

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もっと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基に、第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、第3章のアクティブユーザ予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、第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。

表 5 利用促進施策スケジュール

	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

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。本稿で対象とするユーザは、実際に施策を実施する場合にはアクティブユーザを推定する必要があることを考慮し、サービスの登録日から 14 日分のデータから 3.2 節で示したアクティブユーザ推測モデルを用い、アクティブユーザと推測された顧客とした。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、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に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つけ、フォ

ロー機能を利用するクラスタのため、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。クラスタ 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タのため、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。

表 6 グループ別メッセージ 例

グループ	週	クラスタ	メッセージ
グループ 2	第一週	-	みんなのレポぜひ見て♪ 春色のモノ、集まってきています！
	第二週		最近、どこに行った？ おでかけた場所、レポしてね♪
	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		今週末どこにおでかけ？ 連休中に見つけたステキな所は？
グループ 3	第一週	クラスタ 1	次に投稿するレポは何？ みんなに詳しく教えてあげてね♪
		クラスタ 2	新しいお題が出たよ 賛気気分になったモノ教えてね♪
		クラスタ 3	検索窓も使って見て♪ 気になるワードでレポ検索してね
		クラスタ 4	みんなレポ見て何買った？ つられて買ったモノもレポしてね
	第三週	クラスタ 1	最近食べた旨いモノは？
		クラスタ 2	どんなモノ買った？食べた？
		クラスタ 3	人気のレポ、もう見た？ タイムラインから検索してみて♪
		クラスタ 4	レポ見て買ったモノは？あなたの感想ぜひレポしてね♪

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。クラスタを用いた顧客への施策効果について図 2 を用いて説明する。効果検証は、顧客の利用状況が確認できるログイン数と投稿数の増加比率を確認する。図 2 は、投稿数増加比率を利用促進施策が実施された各週と利用促進施策実施前 1 週間の投稿数を比較し、さらにグループ 1 との差分を算出した結果である。各グループの投稿数増加比率から、クラスタ別にメッセージを送信したグループ 3 の平均増加比率が 22.8%と最も投稿数が増加していることを確認した。全員に同じメッセージを送信したグループ 2 の平均増加比率は、18.5%であり、グループ 3 と比較して約 4.3%の差が見られた。これにより、投稿数の増加には、メッセージを顧客のタイプに合わせた内容を送信する施策が有効である可能性を確認した。

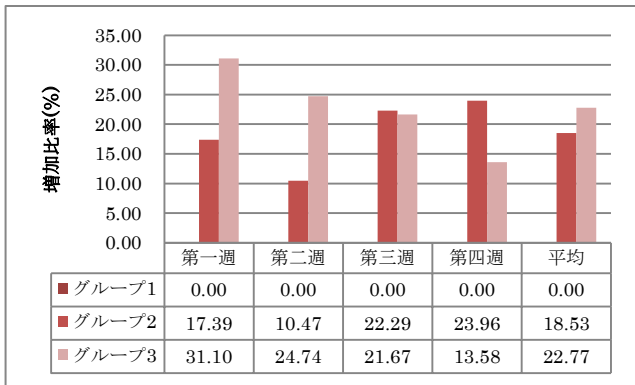


図2 投稿数増加比率

表7 媒体別投稿数増加比率(平均)

	ノーティフィケーション増加比率	メッセージR増加比率
グループ3	26.39%	19.16%

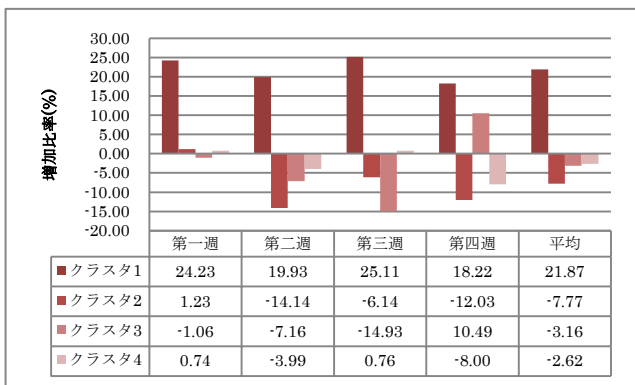


図3 クラスタ別ログイン数増加比率

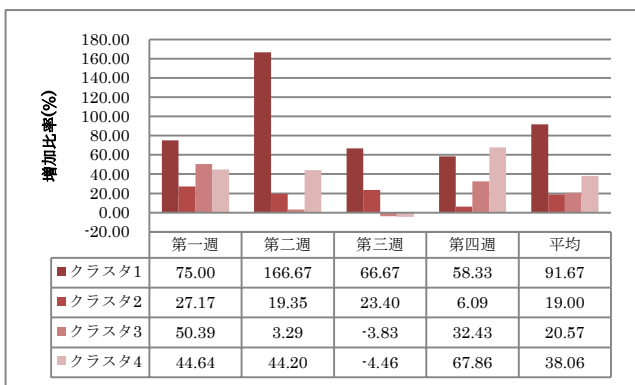


図4 クラスタ別投稿数増加比率

また、顧客へのメッセージ配信に用いる媒体別の効果は、グループ3の第一週、第三週にノーティフィケーション、第二週、第四週にメッセージRでメッセージを送信した場合の投稿数の増加比率を利用促進施策前と比較する。表7に示す各媒体別の投稿数増加比率の平均は、ノーティフィケーションの場合が26.39%の増加、メッセージRの場合が

19.16%の増加であり、ノーティフィケーションの増加比率が7.23%、メッセージRを上回った。

次にクラスタ別に異なるメッセージを送信したグループ3の利用状況を図3、図4に示す。図3にグループ1と比較したグループ3のログイン数増加比率をクラスタ別に示す。増加比率は、利用促進施策前1週間と利用促進施策期間のログイン数の総数を用いて算出した。ログイン数の増加比率は、クラスタ1に高い傾向が見られたが、クラスタ2, 3, 4についてはグループ1よりもログイン数が少ない傾向が見られた。図4にグループ1と比較したグループ3の投稿数増加比率をクラスタ別に示す。増加比率は、図3と同じく利用促進施策前1週間と利用促進施策期間の投稿数の総数を用いて算出した。増加比率は、特にクラスタ1, 4に高い傾向が見られた、これらの結果より、利用促進施策の影響は、クラスタ別に差があることが考えられる。また、クラスタ2, 3, 4についてはログイン数が増加していないが、投稿数は増加している。このことから、本利用促進施策は、投稿数の増加に有効であったと考えられる。

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、機械学習の一手法であるランダムフォレスト法を用いて構築した。構築したモデルの評価結果よりサービスの登録日から14日分の利用履歴を用いた場合、70.3%の精度でアクティブユーザ/非アクティブユーザの予測ができることが確認できた。さらに、14日分と30日分のデータを用いたモデルの再現率の差は、7.1%であり、30日分のデータを用いると約8割程度の精度で予測できることも確認した。

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ノーティフィケーションを用いて顧客へ送信した。その結果、投稿数の増加には、メッセージRよりもノーティフィケーションが7.23%有効であった。また、投稿数の増加率は、メッセージを送信しなかったグループ1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3とでは、約23%の差があることを確認した。これにより、投稿数の増加には、メッセージを顧客に送信する施策が有効であり、メッセージは、ユーザのタイプに合わせた内容を送信することが有効である可能性を確認した。

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、反応しやすい施策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頂いた。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) EM.Rogers: 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), IEEE Computer Society, pp.210-214 (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 2014-1-22).
- 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 2014-2-24).
- 9) Borko, F.: Handbook of Social Network Technologies and Applications, Springer (2010).
- 10) Breiman, L.: Machine Learning. Kluwer Academic Publishers (2001).
- 11) 金明哲 : R によるデータサイエンス, 森北出版株式会社 (2007).
- 12) 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).
- 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), Vol4, pp.818-821 (2009).
- 14) Ngonmang, B., Viennet, E. and Tchente, M.: Churn prediction in a real online social network using local community analysis, IEEE/ACM Conference on Advances in Social Networks Analysis and Mining (ASONAM), pp.282-288 (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 2014-5-14)