

A Study on the Detection of Features and the Indicator of the Out of Degree Time Series Data for Customer Classification

ZHANG Xiaoyan

1 研究背景・目的

近年、IoT 技術の発達と普及に伴い、企業が自社製品の様々な使用状況を収集することが可能となっている。本研究では、プリンタの使用ログデータ解析に着目し、その分析方法について提案を行う。現在、多くのプリンタはオンライン利用が可能となり、ユーザの許諾を元に、企業が使用履歴が収集することができる。このログデータには、各ユーザの印刷枚数に加え、ジャム発生回数や通電時間などの情報が含まれている。そのようなログデータを適切な方法で分析すれば、特徴的な使い方をユーザー群が検出でき、ユーザの使用状況に合わせたキメの細かい対応が可能と考えられる。さらに、初期段階の使用状況に基づき将来的な優良顧客を予測できれば、予測に基づく対応や優良顧客に成長しやすいユーザの特性を考慮した施策を検討することにも役立つと考えられる。

本研究では、得られる時系列ログデータから、特殊な使用をしているユーザの発見、すなわち外れ値パターンの検出、及びそれらの情報の活用を焦点を当てる。一般に、それぞれのデータが正常であるか異常であるかが既知となっている学習データを利用できる場合、異常検知モデルの学習は比較的容易である [1]。しかし、本研究で対象とするデータには、そのような異常ラベルは付与されていない。そこで、データセット内の大多数は正常であるという仮定のもとでデータセット内の外れ値を探すことによって、ラベルなしテストデータセットにある異常を見つける One-class SVM [3] 法に着目する。

本研究の対象とするプリンタの使用ログデータは、大多数の通常ユーザは類似した使い方をし、特殊ユーザはそれらと異なった使い方をするという特徴を持っている。ユーザの特徴は使用履歴に反映されているため、One-class SVM 法に基づき、ユーザ間の使い方に対する距離を計算するだけで、そのユーザが特徴的であるかどうかの判断が可能となる [4]。他のユーザと異なる行動を取ると、特徴空間上における他のユーザからの非類似度（以降、距離）は大きくなっていく。その距離を用いて、各ユーザの外れ度合いを求め、他のユーザと異なり、特徴的な使い方をしているユーザを検出することができる。更に、これらの特徴的なユーザが具体的にどのような使い方をしているかを分析することも可能である。

一方、企業により多くの利益を与える優良顧客に成長するであろうユーザを初期に発見することも、ビジネス上、大変有用である。このような将来の優良顧客発見モデルとして、最も基本的な方法は、得られる時系列の特徴量をそのまま用い、分類モデルにより分類することである。これに対し本研究では、個々の特徴から得られた外れ度合いを特徴量として活用し、新たな分類モデルを提案する。これより、個々の特徴量についてより明確に考察でき、モデルの性能を向上させることができることを示す。また、これらの分析結果を用いて適切なマーケティング施策を講じることで、企業はサービスを効率的に改善し、ユーザの満足度を向上させる方法を検討する。本研究の成果により、ユーザの使用頻度を高め、優良顧客の増加への一助となることが期待できる。

2 One-class Support Vector Machine

One-class Support Vector Machine (One-class SVM) は、SVM 法の特異なケースとして、新規性または異常値検知のためによく用いられる [2]。これは、データをカーネル関数によりより高次元な特徴空間に写像し、学習データの大部分を含む最小の超球を見つけることで正常空間を構成する方法である [3]。

まず、観測された学習データを $\mathbf{x}_i (i = 1, 2, \dots, n)$ とする。 \mathbf{w} を識別境界の法線ベクトルとする。また、一般に、データは線形分離不可能であるため、スラック変数 ξ_i を導入する [5]。One-class SVM の目的関数を L_p とすると、それを最小化するための最適化問題は以下のように与えられる。ここで、 ρ は識別関数の切片で、 $\varphi(\cdot)$ は、カーネル関数とする。

$$\min L_p(\mathbf{w}, \boldsymbol{\xi}) = \min \left(\frac{1}{2} \mathbf{w}^T \mathbf{w} - \rho + \frac{1}{vN} \sum_{i=1}^N \xi_i \right),$$

$$s.t. \quad \mathbf{w}^T \varphi(\mathbf{x}_i) - \rho + \xi_i \geq 0, \xi_i \geq 0. \quad (1)$$

この問題に対してラグランジュの未定乗数 α_i, α_j を導入し、対応する双対問題の目的関数を L_d とすると、以下のように表すことができる。

$$\max L_d(\boldsymbol{\alpha}) = \min \left(-\frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j \varphi^T(\mathbf{x}_i) \varphi(\mathbf{x}_j) \right),$$

$$s.t. \quad 0 \leq \alpha_i \leq \frac{1}{vN}, \sum_{i=1}^N \alpha_i = 1. \quad (2)$$

ここで、 $\varphi^T(\mathbf{x}_i) \varphi(\mathbf{x}_j)$ の部分を式 (3) のようなガウシアンカーネル $K(\mathbf{x}_i, \mathbf{x}_j)$ で置き換える。 $\|\cdot\|$ はユークリッド距離を表す。

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp \left(-\frac{1}{2\sigma^2} \|\mathbf{x}_i - \mathbf{x}_j\|^2 \right) \quad (3)$$

このとき、データの外れ度合いは式 (4) で表される。

$$d = \sum_{i=1}^N \alpha_i K(\mathbf{x}_i, \mathbf{x}_j) - \rho \quad (4)$$

3 提案手法

3.1 概要

本研究では、某プリンタメーカーが提供するプリンタの時系列ログデータを用いる。その中で、特殊な使用をしているユーザの発見、及びそれらの情報の活用を焦点を当てる。今回、対象とするデータでは、外れ値かどうかの情報は既知となっていない。そこで、特殊な使用パターン（外れ値パターン）のみを検出する手法 One-class SVM を適用する。これにより、特徴的なユーザの発見が期待できる。さらに、特徴的な使い方をしているユーザ情

報として、ユーザの各特徴量の外れ度合いを活用し、優良顧客を抽出するモデルを構築する。

まず、各ユーザから一定期間の p 種類の特徴量を抽出し、各特徴量の時系列ログデータをベクトル表現した、 $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(p)}$ を説明変数ベクトルとして作成し、それぞれ独立に One-class SVM を適用する。また、 p 種類の特徴量の全期間の時系列データを横に接続した説明変数ベクトル \mathbf{X} と、それらの特徴量のいくつかの統計量で構成した説明変数 \mathbf{X}' も分析対象とする。これらの説明変数を変更することによって、異なる検出結果が得られる。それらの結果をクロス分析することで、どの特徴量が例外的パターンを示しているのかを各ユーザに付与することができる。

さらに、各特徴量の時系列を月ごとに分割し、それぞれの期間に One-class SVM を適用し、外れ度を算出する。それらの外れ度を新たな説明変数として元のデータに追加し、分類モデルにより分類を行う。それにより、高い精度を持つ分類モデルの推定が期待される。

3.2 One-class SVM による分析

本研究では、各説明変数に対する時系列毎にデータセットを作成し、それぞれのデータセットごとに One-class SVM を適用することで、特徴的なユーザの使用パターンの抽出を実現する方法を提案する。また、異なる多様な外れパターンを抽出するために、異なる説明変数で検出した外れ値ユーザの集合でクロス分析し、ある特定の要素の影響を強調した結果を得る方法を示す。その手順は以下の通りである。

手順 1) p 種類の特徴量を持つ時系列データ $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(p)}$ 及び、それらの p 種類の時系列データを接続した説明変数ベクトル \mathbf{X} 、それらの特徴量の統計量で構成した説明変数ベクトル \mathbf{X}' を作成する。

手順 2) それぞれの説明変数ベクトルの要素に対し平均 0、分散 1 の正規化を行う。それらを説明変数として用い、One-class SVM を適用し、外れ値ユーザを検出する。各データセットにおける外れ値ユーザの集合を $OCSVM(\mathbf{x}^{(p)}), OCSVM(\mathbf{X}), OCSVM(\mathbf{X}')$ と定義する。

手順 3) 各集合内の外れ値ユーザを集計し、外れ値ユーザの実際の使用状況の時系列をプロットする。

手順 4) 複数の結果に関する積集合を求める。例えば、 $OCSVM(\mathbf{x}^{(p)}) \cap OCSVM(\mathbf{x}^{(p')})$ が求まった場合、その積集合内の外れ値ユーザは、特徴量 $\mathbf{x}^{(p)}$ と $\mathbf{x}^{(p')}$ の影響が同時に考慮された結果と考える。その積集合内の使用状況の時系列をプロットする。そこで、1つの特徴量のみが考慮される検出法よりさらに特徴的なユーザを検出することができる。

3.3 顧客分類モデルの構築

企業にとって、他のユーザより多く利益を与える優良顧客群を初期段階で検出することには様々なメリットがある。ビジネスチャンスを見逃すことを可能な限り避けるために、優良顧客に成長するであろうユーザを早期に発見し、顧客維持の施策を講じることができる。そこで、本研究では、過去のデータとして得られている学習データに対して、優良顧客と一般顧客というラベルを付与し、これらを用いて分類モデルを構築する。まず、One-class SVM を各特徴量の最初の w 週分の時系列に適用し、外れ度 $d_w^{(p)} = (d_w^{(1)}, d_w^{(2)}, \dots, d_w^{(p)})$ を求める。また、それと別に、時系列を月ごとに分割し、 m 月目の時系列に再び One-class SVM を適用する。各月の外れ度を $d_m^{(p)} = (d_m^{(1)}, d_m^{(2)}, \dots, d_m^{(p)})$ を求める。ここで、4 週分のデータの外れ度は 1 月目の外れ度と等しい。それらの外れ度及び、各特徴量の最初の w 週分の時系列の平均値を説明変数とする。それらのラベルと説明変数をランダム・フォレストで学習し、分類モデルを構築する。そし

て、予め用意されたテストデータでモデルの性能を評価する。これにより、新規顧客が優良顧客に成長するか否かを予測することができ、それらの実際の使用パターンを分析することにより、企業は具体的なマーケティング戦略を講じ、適切なアプローチ方法によって利益に結び付けることが期待できる。

4 One-class SVM の実データ分析への応用

4.1 分析条件

本分析では、某プリンタメーカーが提供するモノクロレーザープリンタの時系列ログデータを使用した。データは 2017 年 10 月から 2018 年 7 月までの週次データである。分析には、実際に使用期間が 24 週以上のユーザ計 33,380 人を対象とした。ただし、25 週目以降の時系列データは分析対象外とする。特徴量は計 149 種類である。本研究では、ユーザの行動が最も関係性が高いことが明らかとなっている 6 つの特徴量を選択した。また、特徴量の統合の観点で、各特徴量の期間中の統計量を用いた。それらの定義は表 1 に示す。

表 1 説明変数の定義

$\mathbf{x}^{(1)}$	週間印刷枚数	$\hat{E}(\mathbf{x}^{(p)})$	平均
$\mathbf{x}^{(2)}$	週間ジャム発生数	$\hat{V}(\mathbf{x}^{(p)})$	分散
$\mathbf{x}^{(3)}$	オンラインユーザ数	$\max(\mathbf{x}^{(p)})$	最大値
$\mathbf{x}^{(4)}$	週間トナー用量	$\min(\mathbf{x}^{(p)})$	最小値
$\mathbf{x}^{(5)}$	週間トナー交換回数	$\text{median}(\mathbf{x}^{(p)})$	中央値
$\mathbf{x}^{(6)}$	週間通電時間		

これらを用い、異常検知のための特徴空間を構成する。全期間の時系列を横に接続した説明変数ベクトルを $\mathbf{X} = (\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \mathbf{x}^{(3)}, \mathbf{x}^{(4)}, \mathbf{x}^{(5)}, \mathbf{x}^{(6)})$ と定義した。各変数は 24 週間の観測値で構成された時系列データであるため、それらは $\mathbf{x}^{(p)} = (x_1^{(p)}, x_2^{(p)}, \dots, x_{24}^{(p)})$ ($1 \leq p \leq 6$) と表される。次に、特徴量の統計量で構成した説明変数ベクトルを $\mathbf{X}' = (\mathbf{X}'^{(1)}, \mathbf{X}'^{(2)}, \mathbf{X}'^{(3)}, \mathbf{X}'^{(4)}, \mathbf{X}'^{(5)}, \mathbf{X}'^{(6)})$ とする。ここで、 $\mathbf{X}'^{(p)} = (\hat{E}(\mathbf{x}^{(p)}), \hat{V}(\mathbf{x}^{(p)}), \max(\mathbf{x}^{(p)}), \min(\mathbf{x}^{(p)}), \text{median}(\mathbf{x}^{(p)}))$ である。

ここでは以下の 3 つの分析により外れ度指標を得る。

- 6 つの特徴量を用いた説明変数ベクトルそれぞれに One-class SVM を適用する。
- 6 つの特徴量で構成された説明変数ベクトルに One-class SVM を適用する。
- 時系列データに対する統計的特徴量（分散、平均、最大値、最小値、中央値）で構成された説明変数ベクトルに One-class SVM を適用する。

4.2 外れ値ユーザの検出結果

各説明変数ベクトルに One-class SVM を適用し、各ユーザの外れ度合いの指標を計算した。その外れ度指標の標本分布から計算した標準偏差を基準に複数の区間に分割し、各区間に属するユーザの実際の使用状況を観測した。なお、外れ度の計算式 (4) により、値が小さければ小さい程、データはより外れ値である。

ここで、外れ値と判定するための閾値は学習済 One-class SVM によって計算された外れ度の標準偏差 σ を基準に決めることとした。外れ度指標が平均から $n\sigma$ ($n \geq 0$) 離れたユーザの人数を表 2 に示す。この結果から、異なる条件の下で異なるユーザが外れ値として検出されることが分かる。

ここで、説明変数 \mathbf{X} 、すなわち、全ての時間の全ての特徴量を用いたときに計算された外れ度が最も小さいユーザ 3 名と値が最も大きい通常ユーザ 3 名の時系列を図 1 に示す。図の横軸は週間数を表し、縦軸は特徴量によって異なる意味合いを持っている。

図 1 より外れ値ユーザの時系列データの変化は激しいことが分かり、それと比べて、通常ユーザの時系列データ

はほぼ不変であることが伺える。また、この結果は、ジャム発生数 $x^{(2)}$ に最も強い影響を受けているということが読み取れる。

表2 異なる説明変数による検出した外れ値ユーザ人数

説明変数	3σ	2σ	σ	0
$x^{(1)}$	828	1,260	1,954	18,835
$x^{(2)}$	1,195	1,772	2,559	5,236
$x^{(3)}$	615	1,128	5,886	10,400
$x^{(4)}$	385	1,139	3,317	17,662
$x^{(5)}$	412	1,397	4,798	13,921
$x^{(6)}$	155	685	3,507	19,181
X	573	1,537	4,526	14,116
X'	715	1,106	3,152	14,828

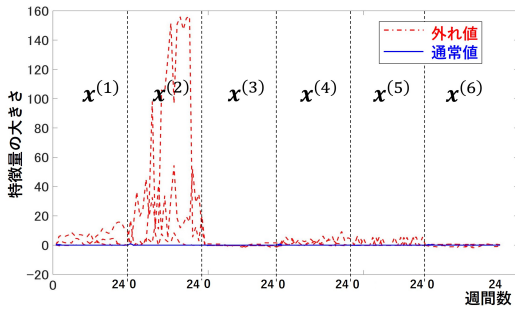


図1 外れ値ユーザと通常ユーザの時系列プロット

4.3 クロス分析結果

複数の説明変数による検出結果の集合 $OCSVM(x^{(p)})$ と $OCSVM(x^{(p')})$ の積集合を求めることにより、特定の要素の影響を強調した結果や、偏った使い方をしているユーザを発見できる。本研究の対象事例において最も重視されている特徴量はトナー交換回数であり、それに関連した使用パターンについて考察すべきと考える。ここでは一例として、印刷枚数 $x^{(1)}$ とトナー交換回数 $x^{(5)}$ に着目した外れ度について考察する。

図2に $x^{(1)}$ と $x^{(5)}$ の外れ値に関する二次元のベクトルを生成し、その散布図を示す。直線は各特徴量の外れ度の 3σ の閾値ラインを表す。ここでは、閾値より外れ度の値が小さい特徴量を外れ値と判断する。

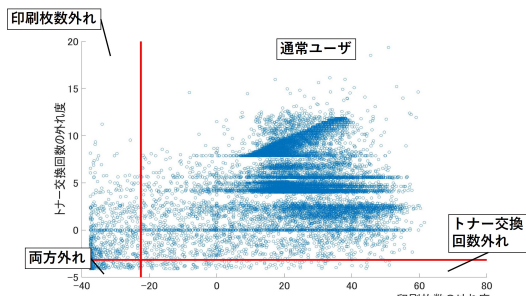


図2 印刷枚数-トナー交換回数散布図

表3には、クロス分析による検出した外れ値と通常ユーザの人数を示す。偏った使い方をしているユーザの人数は太字で表している。また、それぞれの範囲内に外れ度が最上位となった1人のユーザの時系列は図3と図4に示す。

表3 特徴量 $x^{(1)}$ と $x^{(5)}$ のクロス分析結果 (単位: 人)

	$x^{(5)}$ 通常	$x^{(5)}$ 外れ値
$x^{(1)}$ 通常	32,418	134
$x^{(1)}$ 外れ	550	278

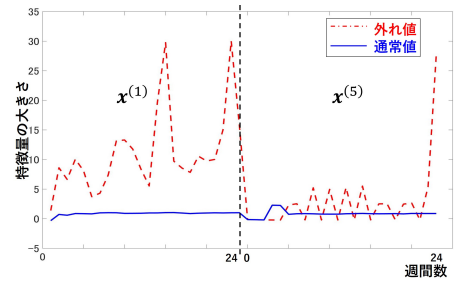


図3 印刷枚数・トナー交換回数の時系列プロット (通常と両方外れ)

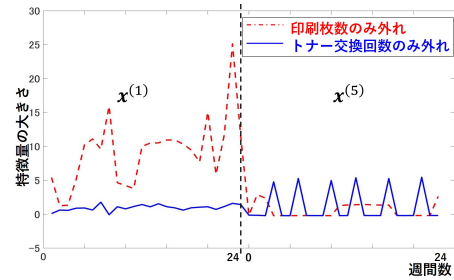


図4 印刷枚数・トナー交換回数の時系列プロット (偏った使い方)

図3と図4で示した通り、外れ値ユーザは通常ユーザより、印刷枚数もトナー交換回数も多い。また、印刷枚数のみが外れ値となったユーザは、印刷枚数が多い割に、トナー交換が少ない。一方、トナー交換回数のみが外れ値となったユーザは、印刷枚数が少ないが、トナーを頻繁に交換する傾向がある。クロス分析の組み合わせにより、それ以外にも様々な外れ値パターンが検出できる。個々のユーザ群の特徴に合わせた施策をカスタマイズすることが企業にとって有意義であると考えられる。

5 顧客分類モデルの構築

5.1 データ分析条件

顧客分類モデルに用いたデータは、前章と同様に6つの特徴量の24週間分の時系列データである。データ提供元企業によると、ユーザがプリンタを購入した後、その消耗品(トナー)を累計4回以上交換すれば、優良顧客と判断できる。そこで、本研究では、トナー交換回数 $N \geq 4$ のユーザを優良顧客と定義し、ラベル $l = 1$ を付与する。それ以外のユーザを一般顧客とし、ラベル $l = -1$ を付与する。 $l = 1$ のユーザは1,695人であり、 $l = -1$ のユーザは31,685人である。両方は不均衡であるため、少数ラベルのデータに対し、オーバーサンプリングを適用して学習を行う。

ここで、プリンタを使用開始から3か月以内の時系列を月ごとに分割し、それぞれに One-class SVM を適用し、期間外れ度 $d_w^{(p)}$ ($w = 4, 8, 12$) と月ごとの外れ度 $d_m^{(p)}$ ($m = 1, 2, 3$) を計算する。ただし、 $d_4 = d_1$ である。

また、各特徴量の最初の w 週分の平均値を $\hat{E}(x_w^{(p)})$ とする。この時、外れ度及び平均値を要素とした。以下のように説明変数を生成する。

- ・1か月分: $X_1'' = (d_4^{(p)}, \hat{E}(x_4^{(p)}))$, ($1 \leq p \leq 6$).
- ・2か月分: $X_2'' = (d_8^{(p)}, d_1^{(p)}, d_2^{(p)}, \hat{E}(x_8^{(p)}))$.
- ・3か月分: $X_3'' = (d_{12}^{(p)}, d_1^{(p)}, d_2^{(p)}, d_3^{(p)}, \hat{E}(x_{12}^{(p)}))$.

それらの説明変数を用いてラベルの分類モデルをランダム・フォレストで学習する。その際、ラベルは6か月後に優良顧客に成長するか否かとする。また、本提案手法の比較対象となる従来手法として、各特徴量の最初の w 週分の時系列そのものをランダム・フォレストで学習した分類器を用いる。

5.2 評価指標

本研究では、予測ラベルの適合率、再現率及びF値を計算し、10分割交差検証を用いたモデルを評価する。

ここで、予測された優良顧客の人数は N とし、実際の優良顧客の人数は M とする。正しく予測された優良顧客人数は R とする。評価式は以下に示す。

$$\text{適合率: Precision} = \frac{R}{N} \quad (5)$$

$$\text{再現率: Recall} = \frac{R}{M} \quad (6)$$

$$F \text{ 値: } F = \frac{2\text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} \quad (7)$$

5.3 分類の精度

最初の1~3ヵ月分の時系列データを学習し、それぞれの予測結果を表4に示す。

表4 分類モデルの予測評価

		1ヵ月	2ヵ月	3ヵ月
従来手法	適合率	0.554	0.611	0.650
	再現率	0.700	0.787	0.816
	F値	0.618	0.688	0.724
提案手法	適合率	0.522	0.609	0.625
	再現率	0.720	0.823	0.881
	F値	0.605	0.700	0.731

情報量の増加と共に、モデルの性能も向上した。そこで、全ての条件において、提案手法の再現率は従来手法より上回った。しかし、従来より多くの顧客が優良顧客として予測されたため、適合率が低下している。ここで、今回の分類問題では、優良顧客をできるだけ漏れなく発見することを企業にとっての最重要の課題としている。すなわち、本事例において、再現率の向上は重要であると考慮しており、提案手法は、従来手法よりも優れると判断することができる。

5.4 特徴量の重要度分析

ここでは、分類モデルに用いた説明変数の各次元の要素を特徴量とし、それらが分類モデルにとっての重要度を数量化し、分析を行う。3ヵ月分のデータを用いた場合の各特徴量の重要度を表5に示す。

表5 各特徴量の重要度(上位10個)

$\hat{E}(\mathbf{x}_{12}^{(5)})$	1.360
$d_3^{(4)}$	0.887
$\hat{E}(\mathbf{x}_{12}^{(1)})$	0.861
$\hat{E}(\mathbf{x}_{12}^{(4)})$	0.745
$d_{12}^{(5)}$	0.621
$d_3^{(6)}$	0.554
$d_2^{(4)}$	0.505
$d_2^{(5)}$	0.483
$d_3^{(3)}$	0.475
$d_1^{(5)}$	0.433

上位10個中、統計量であるトナー交換回数の平均値 $\hat{E}(\mathbf{x}_{12}^{(5)})$ 、印刷枚数の平均値 $\hat{E}(\mathbf{x}_{12}^{(1)})$ とトナー用量の平均値 $\hat{E}(\mathbf{x}_{12}^{(4)})$ の重要度は全体的に上位となり、時系列データを統計量で集約する効果があると考えられる。また、外れ度を表す特徴量も比較的大きな値となっているため、外れ度は優良顧客の判断に有効であるといえる。一方で、トナー交換回数や、トナー用量の重要度も比較的高いため、初期段階のトナー交換回数とその後の顧客の成長に関係することが示唆される。一方で、従来手法では、時系列の各週の値を独立成分として用いたため、重要度の分析は困難であり、分類効率が低下していると考えられる。

以上のように、提案手法は、企業のニーズに応えられ

る分類モデルを構築しただけでなく、特徴量の重要度分析により、企業により多く有用な情報を提供する点が特徴である。

6 考察

本研究では、事前分布が推定しにくいユーザの使用履歴時系列データに対し、One-class SVMを適用することにより、特殊ユーザを抽出することを可能にした。各ユーザは複数の特徴量を持つため、様々な使用パターンが抽出できるように複数の説明変数に対して個別に外れ度指標を生成する必要があると考える。各説明変数の外れ度を組み合わせた分析により、異なる特殊ユーザ群が検出されることが分かった。すなわち、検出結果をクロス分析することで、より多様な使用パターンを考察することが可能であるといえる。本稿では一例として、印刷枚数とトナー交換回数の特徴量を組み合わせた例を示したが、偏りのある使用パターンが検出された。偏りが発生した原因として、各ユーザ群に属するユーザが印刷時、1枚あたりのトナー使用量に大きな差が存在するためだと考える。企業は、その結果からユーザ群の属性を推測し、それぞれに対して適切なアプローチを取ることができる。

また、特殊ユーザ群と通常ユーザ群の実際の時系列データをプロットし、比較することで、新たな知見が得られた。特殊ユーザとは、データセット内の大多数のユーザとは異なる使い方をしているユーザである。本研究の実データ分析では、特殊ユーザは通常ユーザより使用頻度が高く、より活発であることが分かった。そのため、各ユーザの外れ度を新たな特徴量として用いることが、優良顧客分類モデルに役立ったと考えられる。

より高い精度の優良顧客分類モデルを構築するために、外れ度を用いるだけでなく、元の時系列データを統計的に集約することも有効である。また、時系列データを集約することで、どの期間中にユーザの使用状況が最も重要なかがより明らかになった。重要度分析により、企業はマーケティング戦略を展開する際に、どこに重心を置くべきかを明らかにすることができる。

7 まとめと今後の課題

本研究では、顧客分析を目的として、それらの外れ度指標を特徴量として用いた分析法、及び顧客分類モデルを提案した。また、これらの方法の有効性を示すために、プリンタの多次元時系列データにOne-class SVMを適用し、外れ値が検出できることを確認した。次に、クロス分析により、プリンタの時系列データが外れ値になるような原因を確認した。さらに、優良顧客予測モデルの構築及び評価により、一定のデータ量を確保した場合には、従来手法より優れた性能を示したため、提案手法は有効であると考えられる。

今後の課題として、時間の推移を考慮した学習を行うことや、期間が適切な時系列を学習で用いることが挙げられる。また、一般顧客の中で、優良顧客になれる可能性の高いユーザを特定し、優良顧客に転換させるような施策を考えることも課題である。

参考文献

- [1] Chandola, V., Banerjee, A., Kumar, V., "Anomaly detection: A survey," *ACM computing surveys (CSUR)*, Vol. 51, No. 11, Article. 15, 2009.
- [2] Scholkopf, B., Williamson, R. C., Smola, A. J., Shawe-Taylor, J., Platt, J. C., "Support vector method for novelty detection," *Advances in neural information processing systems*, pp. 582-588, 2000.
- [3] 高島泰斗, 香田正人, "1クラスSVMと近傍サポートによる領域判別," *Operations research as a management science research*, Vol. 51, No. 11, pp. 677-682, 2006.
- [4] 山西健司, "データマイニングによる異常検知," 共立出版, 37(3), pp.13-44.
- [5] 後藤正幸, 小林学, "入門パターン認識と機械学習," コロナ社, pp.114-134, 2014.