

# RFM 分析モデルに基づいた初心者向けのデータ解析システム

佐藤 有人<sup>†</sup> 井上 潮<sup>‡</sup>

<sup>†</sup> 東京電機大学大学院工学研究科 〒120-8551 東京都足立区千住旭町 5

<sup>‡</sup> 東京電機大学工学部 〒120-8551 東京都足立区千住旭町 5

E-mail: <sup>†</sup> 16kmc15@ms. dendai. ac. jp, <sup>‡</sup> inoueu@mail. dendai. ac. jp

## あらまし

マーケティング分野で広く用いられる RFM 分析モデルを基に、購買情報等のビッグデータから解析しやすいデータセットの抽出やデータの可視化を行い、初心者でも容易にデータ解析が行えるシステムを実現した。このシステムの特徴は、既存の RFM 分析を拡張することにより高精度な分析が可能であること、プログラミングが不要な対話型インターフェースで操作できるウェブベースのシステムとなっていることである。

**キーワード** データベース, ビッグデータ, マーケティング, RFM 分析

## 1. はじめに

近年、ID 付き POS データ等の普及により、顧客一人ひとりの購買行動を把握することができるようになった。また、顧客ごとの購買履歴から性別や年齢別の顧客の嗜好を分析でき、経営戦略を立てる上で有利な情報の獲得が可能である。このようなデータを活用する業種・業務、用途は広がってきており、小売業や卸売業などの企業を筆頭に解析および解析データの活用が行われている。ID 付き POS データをはじめとする購買情報のビッグデータを分析する際には主に Python や R 言語が利用されているが、膨大なデータから有益な情報を獲得するのは難しい。

情報工学系の学生においては、データ解析に必要なマーケティング分野の知識が不足している。さらに、分析を手助けしてくれる既存システムのほとんどが高価で高機能であることから学生のデータ解析使用には敷居が高い。そのような学生向けに基本的な分析モデルがあらかじめ用意されており、簡単にデータを可視化できるシステムを提供することで学生のデータ解析への敷居を低くすることができる。

本研究では、購買情報のビッグデータから、マーケティング分野で広く用いられる RFM 分析機能を取り入れ、プログラミングが不要なウェブベースのシステムを実現した。

RFM 分析は、顧客のこれまでの購買行動・購買履歴から、優良顧客のセグメンテーション等を行う分析手法で、顧客の最終購買日(Recency)、購買頻度(Frequency)、総購買金額(Monetary)の 3 つの観点から指標化する。各指標の数値に重み付けした上で合算したスコアが高い顧客が優良顧客と呼ばれる(図 1)。

本論文では、このシステムの概要について紹介する。

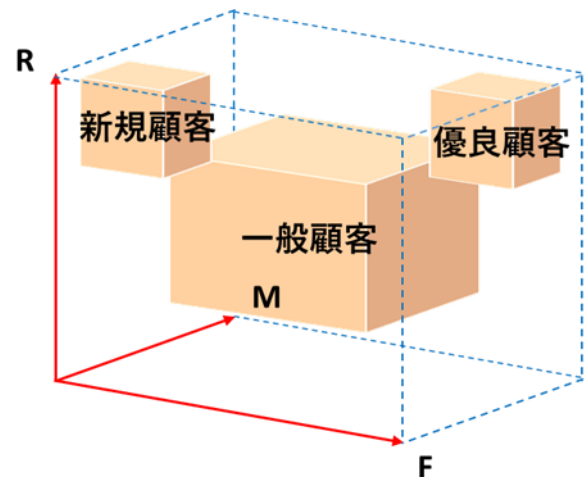


図 1. RFM 分析を利用した顧客のセグメンテーション[3]

## 2. 関連研究と既存ツール

高萩ら[1]は、マーケティング分野のデータ分析技術取得のための 7 種類の分析を行えるウェブ・ベース・システムを開発した。クリップボードに保存しておけるサイズのデータに限られるが、マーケティング分野における一般的な分析を網羅しており、学生が分析方法を学ぶ上では十分な機能である。

徳永ら[2]は、RFM 分析を企業側からではなく、ユーザ側に焦点をあてて一般消費者が自分自身の購買行動をレシートログから分析できるサービスの実装を行った。自分自身の購買行動を振り返ることができるだけでなく、SNS 型のウェブアプリケーションとなっているため友人の購買行動との比較等も可能となってお

り、消費者が自身の購買行動を見なおすことのできる機能を提供している。

既存ツールに関しては、マーケティング向けの分析ツールの大半が商用である。どのツールも同じような機能を持ち、マーケティング分析で用いられる代表的な、RFM分析、デシル分析、回帰分析などの主要な分析が行える。また、回数制限のある使いきり版、30日のお試し版、月額制、永久版等、用途によって使える回数、期間が分かれており、それぞれ価格も異なる。商用であることから各企業におけるマーケティングのプロが使うことを想定した多種多様な機能が実装されており、経験のないユーザが使うには機能が複雑すぎる。そのため、学生がこれらの分析ツールを使用するには敷居が高いといえる。

### 3. 研究目的と技術課題

#### 3.1. 研究目的

本研究の目的は、初心者のためのデータ解析システムの開発である。

具体的な目標としては、分析モデルの選択とパラメータ設定をユーザが行うことで、様々な分析モデルに基づいたデータセットの抽出からデータの可視化までを半自動で行うウェブアプリケーションの構築である。

#### 3.2. 技術課題

本研究では、対象としている解析初心者の特徴がそのまま課題となるため、3つの問題点について述べる。1つ目としてはマーケティング分野の知識が不足していること。2つ目はR言語やPythonといった分析向きのプログラミング言語の習熟度が低いこと。3つ目はシステムを十分に使いこなすことができないことである。

従って、解決が必要な技術的課題は、知識不足でも理解が可能となる解析結果の出力方法、プログラミングが不要または最小限の操作体系、必要かつ十分な機能の提供である。

#### 3.3. 課題の解決方針

課題を解決するために、3つの手法を用いる。1つ目は深い知識がなくても結果の理解が容易な基本的な分析モデルを複数用意しておき、ユーザに好きな分析モデルを選択してもらうこと。2つ目はユーザの操作を分析モデルの選択および、それぞれの分析処理を行う上で必要なパラメータの設定のみの対話型インターフェースを実装すること。最後に解析処理の各工程において、ユーザが行うべき事項をナビゲートすることである。

対話型のインターフェースを実装することで、ユー

ザの操作も単純で済み、1つ1つの処理工程についてのナビゲートも行うことができ、ユーザはどのような順番で処理を行っていけば分析が可能なかの把握が容易となる。

### 4. システムの構成と機能

本システムはJava言語を用い、データベースとして使用しているMySQLとデータをやり取りして分析を行う。処理ステップを図2に示す。

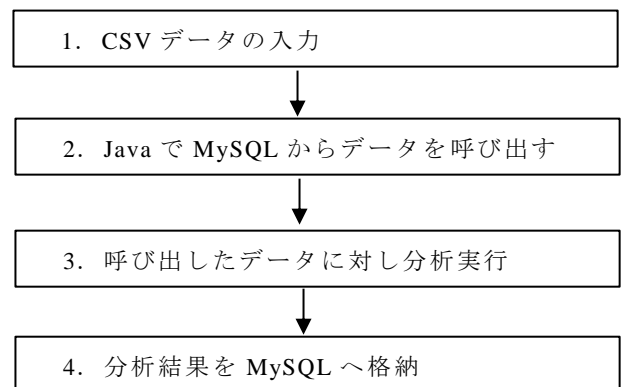


図 2. システムの概要

本システムの特徴的な機能として、拡張RFM分析を実装している。拡張RFM分析は、既存のRFM分析で用いる顧客の最終購買日(Recency)、購買頻度(Frequency)、総購買金額(Monetary)の3つの指標のほかに、顧客の来店周期、購買金額の上昇率を加えたものである。

ID付きPOSデータを用いたマーケティング分析において、使用するデータにはデータ観察期間と呼ばれるデータの蓄積期間が存在している。既存のRFM分析ではデータ期間の最初から来店を続けている顧客については優良顧客として抽出されるが、似たような頻度で来店しているにもかかわらずデータ観察期間の途中から出現した顧客については、どうしても購買頻度(Frequency)の値が小さくなり、優良顧客として抽出されない場合がある。しかし、来店周期の指標が増えることで、データ観察期間途中から出現した顧客についても来店回数だけではなく、来店周期でも顧客のランクを判定することができるため、優良顧客を多く抽出できると考えている。

また、購買金額の上昇率とは、一般の顧客から優良顧客になっていく過程でみられる購買金額の上昇を割合にしたものである。その購買金額の上昇率が図3のように右肩上がりの場合は優良顧客になる可能性が高いと判断できる。また、企業側で重要なこととして優良顧客を増やすことがあげられるが、優良顧客を増や

すための施策をすべての顧客に対して闇雲に行うことはコストがかかりすぎる上、効率が良いとはいえない。従って、優良顧客になりうる顧客の抽出もマーケティング分析において重要事項の1つといえる。

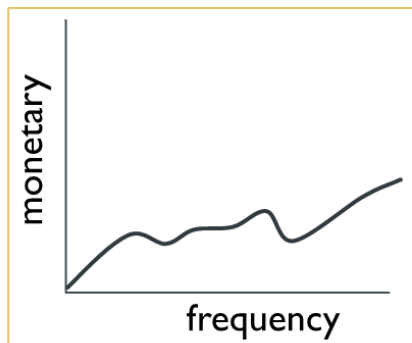


図 3. 購買金額上昇率の例

## 5. 評価

### 5.1. 評価方法

本システムの有効性を確認するため、拡張 RFM 分析の評価を行う。提案した拡張 RFM 分析が既存の RFM 分析よりも高精度な分析が可能であるかの検証が目的である。

評価には、Chen ら [5] の研究で用いられたオンラインショップの購買履歴データ [6] を使用した。このデータは 2010 年 12 月から約 1 年間の購買履歴で、8 属性 54 万件の公開データである。データ観察期間における全顧客数は 4317 人、そのうち優良顧客は 912 人で、RFM 総合スコアが 10 より大きいと定義した顧客である。RFM 総合スコアの算出方法は以下に示す。

$$\text{総合スコア} = R * 2 + F + M > 10$$

本システムで購買履歴データに対する RFM 分析を実行するためにユーザが指定する入出力項目を表 1 に、実行結果を図 4 に示す。

Recency はユーザに基準となる日付を指定させ、その日から顧客が最後に会計を行った日までの日数の差、Frequency は総来店回数、Monetary は、顧客ごとの総購買金額をそれぞれ 1~3 の 3 段階にランク分けを行った結果である。ランク分けの際の 3 つの指標のそれぞれの値の閾値はユーザが任意で設定可能である。

拡張 RFM 分析は表 2 に示す入出力項目を指定して実行する。評価は顧客の来店周期、購買金額上昇率の 2 つの指標それぞれについての準優良顧客の抽出および、2 つの指標を同時に用いた際の準優良顧

客の抽出を行った。ここでの準優良顧客とは、優良顧客になる可能性のある顧客を指す。

CustomerID	MaxInvoiceDate	MinInvoiceDate	DistinctInvoiceNo	TotalInvoiceNo	SumAmount	SumQuantity	Recency	Frequency	Monetary
	2011-12-09 10:26:00	2010-12-01 11:52:00	3710	135080	1447682.12	269562	3	3	3
12346	2011-01-18 10:17:00	2011-01-18 10:01:00	2	2	0.00	0	1	1	1
12347	2011-12-07 15:52:00	2010-12-07 14:57:00	7	182	4310.00	2458	3	3	3
12348	2011-09-25 13:13:00	2010-12-16 19:09:00	4	31	1797.24	2341	2	2	3
12349	2011-11-21 09:51:00	2011-11-21 09:51:00	1	73	1757.55	631	3	1	3
12350	2011-02-02 16:01:00	2011-02-02 16:01:00	1	17	334.40	197	1	1	1
12352	2011-11-03 14:37:00	2011-02-16 12:33:00	11	95	1545.41	470	2	3	3
12353	2011-05-19 17:47:00	2011-05-19 17:47:00	1	4	89.00	20	1	1	1
12354	2011-04-21 13:11:00	2011-04-21 13:11:00	1	58	1079.40	530	1	1	2
12355	2011-05-09 13:49:00	2011-05-09 13:49:00	1	13	459.40	240	1	1	2
12356	2011-11-17 08:40:00	2011-11-18 09:50:00	3	59	2811.43	1591	3	2	3
12357	2011-11-06 16:07:00	2011-11-06 16:07:00	1	131	6207.67	2708	2	1	3
12358	2011-12-08 10:26:00	2011-07-12 10:04:00	2	19	1168.06	248	3	1	3
12359	2011-12-02 11:21:00	2011-01-12 12:43:00	6	254	6245.53	1612	3	3	3
12360	2011-10-18 15:22:00	2011-05-23 09:43:00	3	129	2642.06	1165	2	2	3
12361	2011-02-25 13:51:00	2011-02-25 13:51:00	1	10	189.90	91	1	1	1
12362	2011-12-06 15:40:00	2011-03-17 10:30:00	13	274	5154.58	2212	3	3	3
12363	2011-08-22 10:18:00	2011-04-11 13:20:00	2	23	552.00	408	1	1	2
12364	2011-12-02 10:22:00	2011-08-19 15:24:00	4	85	1313.10	1506	3	2	3
12365	2011-02-21 14:04:00	2011-02-21 13:51:00	3	23	320.69	173	1	2	1
12367	2011-12-05 16:48:00	2011-12-05 16:48:00	1	11	168.90	173	3	1	1
12370	2011-10-19 14:51:00	2010-12-14 12:58:00	4	167	3545.69	2353	2	2	3

図 4. システム実行例

表 1. RFM 分析の入出力項目

入力項目	
基準の項目	カラム名(顧客 ID)
R, F の項目	カラム名(購買日)
M の項目	カラム名(購買金額)
基準の日付	yyyy/mm/dd
R, F, M の基準	各ランクの閾値
出力項目	
抽出された各顧客の集計情報	顧客 ID 直近の購買日 総来店回数 総購買金額 RFM 各スコア

表 2. 拡張 RFM 分析の入出力項目

入力項目	
基準の項目	カラム名(顧客 ID)
R, F の項目	カラム名(購買日)
M の項目	カラム名(購買金額)
基準の日付	yyyy/mm/dd
出力項目	
抽出された各顧客の集計情報	顧客 ID 直近の購買日 総来店回数 総購買金額 来店周期 購買金額上昇率

## 5.2. 来店周期

まずは、来店周期が1週間以内であり、データ観察期間にトータル4回以上来店している顧客を準優良顧客と定義し、抽出を行った。トータル4回以上来店という条件は、全顧客の半数以上が2, 3回の来店しかしておらず、準優良顧客として抽出するには範囲が広すぎると判断したためである。

分析の結果、条件を満たす顧客は38人抽出できた。この人数は優良顧客の4.2%であり、5:25の放送に則ると利益の25%を左右する重要な顧客といえる。

## 5.3. 購買金額上昇率

購買金額上昇率は来店回数における購買金額の回帰直線の傾きが正となる顧客を、準優良顧客とした。傾きが正であることで、購買金額が上昇傾向にあることがわかる。

分析の結果、上昇傾向を見せている顧客を120人抽出することができた。

## 5.4. 評価結果

最後に、5.2, 5.3で述べた来店周期、購買金額上昇率の条件を同時に満たす顧客の抽出を行った。

分析の結果、2つの指標の条件を同時に満たす顧客は3人しか存在せず、来店周期と購買金額上昇率は両方とも必要な指標だといえる。これらの評価パラメータおよび評価結果は表3, 表4に示す。

表3. 評価用パラメータ

顧客分類	RFM分析	拡張RFM分析
全顧客	データ内の全顧客数	
優良顧客	RFM総合スコアが10より大きい	
①準優良顧客 (来店周期)		周期が1週間以内 (トータル4回以上来店)
②準優良顧客 (購買金額上昇率)		回帰直線の傾きが正
準優良顧客 (①∧②)		①と②の条件をどちらも満たす

表4. 評価結果

顧客分類	RFM分析	拡張RFM分析
全顧客	4317	
優良顧客	912	
①準優良顧客 (来店周期)		38
②準優良顧客 (購買金額上昇率)		120
準優良顧客 (①∧②)		3

## 6. おわりに

本論文では、解析初心者や一般的に解析に用いられるPythonやR言語といったプログラミング言語をあまり使用した経験のない学生でもRFM分析を簡単に行える半自動の分析システムについて述べた。

このシステムの特徴は、基本的な分析モデルを複数備えていること、それらの各分析モデルはユーザが自由に設定可能なパラメータ設定のみで利用できること、システムが対話的にユーザを分析手順やパラメータの設定等の分析の流れをナビゲーションするインターフェースとなっていることである。

さらに、分析モデルとして、顧客ごとの来店周期や購買金額の上昇率といった、RFM分析には元々なかった指標を組み込んだ拡張RFM分析機能を実装した。

今後は利用者のさらなる操作性、理解を助けるようなデータの可視化機能をはじめとする、システムの拡張性を高めて、プラグインのような形で利用者が新たな分析モデルを自由に追加できるようにする、すなわち利用者が自身の分析スキルの習熟度に合わせたカスタマイズが可能とする計画である。

## 参考文献

- [1] 高萩栄一郎, 生田目崇, "マーケティング・データ分析のためのウェブ・ベース・システムの開発", 情報科学研究所 所報 No. 79(2012)
- [2] 徳永清輝, 松本真佑, 中村匡秀, "RFMに基づく一般消費者向けレシートログ分析サービスの実装", 信学技報
- [3] Albert ウェブサイト, <https://www.albert2005.co.jp/>
- [4] (有) オズシステム, "RFM分析いいお客様 Pro", [http://www.ozsystem.jp/frdpro/pro\\_top.htm](http://www.ozsystem.jp/frdpro/pro_top.htm)
- [5] Daqing Chen, Sai Liang Sain, and Kun Guo, Data mining for the online retail industry: A case study of RFM model-based customer segmentation using data mining, Journal of Database Marketing and Customer Strategy Management, Vol. 19, No. 3, pp. 197-208, 2012.
- [6] Online Retail Data Set <https://archive.ics.uci.edu/ml/datasets/Online+Retail>