

重要顧客特定に関する購買行動特徴抽出方法への CMOS アニーリングの適用

伊藤 瑞生^{†1}, 小池 和弘^{†2}, 小埜 和夫^{†1}, 山岡 雅直^{†1}

キーワード: 購買行動, 特徴抽出, 最適化問題

1. はじめに

インターネット通販では、ポイント付与や魅力的な商品の提案など、顧客の購買意欲を高めるための販売施策が一般的に行われている[1]。販売施策を行う場合、すべての顧客に同じ施策を行うのではなく、購入した商品分野と購買行動の特徴別に顧客をグルーピングし、それぞれのグループにあった販促を行うことで高い効果が期待できる。さらに、多数ある購買行動の特徴の中で、重要顧客に結びつく特徴を抽出することで、その特徴に焦点を当てた販促が可能となる。本報告では、日立製作所が開発する疑似量子計算機技術「CMOS アニーリング」[2]を用いて、多数の購買行動の特徴から重要顧客をよく説明する特徴の組み合わせを抽出し、その有効性について検討した。

2. 問題設定

本研究で使用する顧客データには、各顧客の購買行動に関するデータとして、過去に顧客が商品を購入した頻度や購入金額、購入した商品種類数などの情報が格納されている。これらの特徴を購買行動特徴と呼ぶ。またそれぞれの購買行動特徴に対して各顧客が有する具体的な数値を特徴量と定める。すなわち顧客 i に対して、購買行動特徴 n に関する特徴量 $a_{i,n}$ が紐づけられている。さらに顧客データには、顧客 i がインターネット通販事業にとってどの程度重要かを示す指標である重要顧客度 C_i が設定されている。

次に各特徴量と重要顧客度を関係づけるため、購買行動特徴 n について特徴量 $a_{i,n}$ を入力とし、重要顧客度を出力とする関数 $f_n(a_{i,n})$ を作成する。以降、関数 $f_n(a_{i,n})$ を特徴量関数と呼ぶ。

特徴 n に対して、抽出するならば1、抽出しないならば0となるバイナリ変数 x_n を決定変数とする。図1に本研究で使用した顧客データおよび決定変数の関係を示した。

上記の特徴量、特徴量関数、決定変数を用いて、顧客 i に対し、下記のような状態ベクトルを定める。

$$v_i := [x_1 f_1(a_{i,1}), x_2 f_2(a_{i,2}), \dots, x_n f_n(a_{i,n}), \dots, x_m f_m(a_{i,m})] \quad (1)$$

以降、この状態ベクトルを用いて数理モデルを作成する。

	特徴1	...	特徴 n	...	特徴183	重要顧客度
顧客1	$a_{1,1}$...	$a_{1,n}$...	$a_{1,183}$	C_1
...						...
顧客 i	$a_{i,1}$...	$a_{i,n}$...	$a_{i,183}$	C_i
...						...
顧客100	$a_{100,1}$...	$a_{100,n}$...	$a_{100,183}$	C_{100}
バイナリ変数	x_1	...	x_n	...	x_{183}	

図1 顧客データと決定変数の関係

3. 数理モデル

重要顧客を最もよく説明する特徴の組み合わせを抽出する問題を数理計画問題として定式化する。目的関数は、状態ベクトル v_i を用いて以下のように定式化できる。

$$\min. \sum_i (\|v_i\|^2 - N C_i^2)^2 \quad (2)$$

N は抽出する特徴の個数を示す定数である。(2)式は抽出する特徴の個数を考慮した上で、各特徴量で構成された状態ベクトルと重要顧客度との差が最も小さくなるように最小化問題を定式化している。(2)式は(1)式より以下のように書き直すことができる。

$$\min. \sum_i \left(\sum_n (x_n f_n(a_{i,n}))^2 - N C_i^2 \right)^2 \quad (3)$$

ここで x_n はバイナリ変数のため、以下が成り立つ。

$$x_n^2 = x_n \quad (4)$$

ゆえに、(3)式は、最終的に以下のように書き直すことができる。

$$\min. \sum_i \left(\sum_n x_n (f_n(a_{i,n}))^2 - N C_i^2 \right)^2 \quad (5)$$

(5)式が数理計画問題の目的関数となる。制約条件は以下となる。

$$\sum_n x_n = N \quad (6)$$

(6)式はすべての購買行動特徴から N 個の特徴の組み合わせを抽出することを示す。上記の目的関数と制約条件から、

^{†1} 株式会社日立製作所

^{†2} アスクル株式会社

重要顧客を最もよく説明する購買行動特徴の組み合わせを抽出する問題は、線形制約付きの2次計画問題として定式化される。

4. 実験結果

本研究では、アスクルが通販で取り扱う主要な商材分野である医療品、資材、事務用品についての顧客データを使用した。商材分野により特徴の種類が異なるため、各商材分野ごとに購買行動特徴を抽出する必要がある。

医療品分野では、顧客10,000名分のデータを使用し、183種類の特徴から100種類の特徴を抽出した。資材分野、事務用品分野も同様に114種類から50種類、100種類から50種類をそれぞれ抽出した。抽出する特徴の個数は、アスクルが各商材分野の販売施策に利用することを想定し、抽出前の特徴数に対して半数程度となるように設定した。

5. 評価

本実験では、CMOSアニーリングを用いて抽出した特徴について、優良顧客を良く表現できているかを評価するため、アスクルの顧客データを利用したLTV(Life Time Value)予測モデルを使用した。LTV予測モデルとは、顧客が企業にもたらす予想収益の総額を予測するモデルである[3]。顧客の生涯価値を正確に予測することで、企業はより効果的な顧客獲得戦略を立てることができ、マーケティング予算をより価値の高い顧客に集中させることが可能になる。すべての特徴を用いて作成した従来のLTV予測モデルと、CMOSアニーリングにより抽出した特徴を用いて作成したLTV予測モデルの精度を比較した。

モデルの予測性能を評価する指標として、モデルの予測値と実際の値の差異を表すGamma Devianceを使用した。Gamma Devianceはガンマ分布を仮定する回帰モデルにおいて、予測誤差を図る指標として一般的に用いられている。ガンマ分布は連続型確率分布の一種で指数分布を一般化させたものであり、連続的で非負の値を取るためLTVの分布の近似に適している。Gamma Devianceは以下の式で計算される。

$$\frac{2}{n} \sum_i^n \left(-\log \left(\frac{w_i}{\hat{w}_i} \right) + \frac{w_i - \hat{w}_i}{\hat{w}_i} \right) \quad (7)$$

ここで、 w_i は各観測値、 \hat{w}_i はモデルによって予測された値、 n はデータポイントの総数を表す。Gamma Devianceの値が小さいほど、モデルの予測が実際のデータに適合しており、モデルの予測性能が高いことを意味する。

購買行動特徴を抽出した各分野ごとに、すべての特徴を用いて作成した従来のLTV予測モデルと、CMOSアニーリングにより抽出した特徴を用いて作成したLTV予測モデルにおけるGamma Devianceを求めた。

各Gamma Devianceの値を表1に示す。医療品分野では、

従来のLTV予測モデルのGamma Devianceは0.0937であったのに対し、本実験で抽出した特徴を用いて作成したLTV予測モデルのGamma Devianceは0.0816であり、従来と比較しモデルの精度は13%改善した。資材分野、事務用品分野では、それぞれ2.7%の悪化、8.8%の改善が認められた。

表2 LTV予測モデルのGamma deviance

	従来	本研究
医療品	0.0937	0.0816
資材	0.1110	0.1140
事務用品	0.2422	0.2210

医療品、事務用品分野で精度が改善した理由として、留意された購買行動特徴の中から、重要顧客度とは無関係であり重要顧客を説明する上でノイズとなる特徴を除外できたためと考えられる。抽出前の特徴の中には、重要顧客度に関係ない特徴量がランダムに分布する特徴が存在するため、そのような特徴を抽出しないことが予測モデルの精度改善に繋がる。一方、精度が悪化した資材分野では、CMOSアニーリングによる特徴の抽出から重要顧客をよく説明する特徴が漏れてしまった可能性がある。

本報告書では、CMOSアニーリングを用いて多数ある購買行動特徴の中から重要顧客をよく説明する特徴の組み合わせを抽出する手法を提案した。抽出して絞り込んだ購買行動特徴を使用してLTV予測モデルを構築した結果、従来より少ない特徴によって、全特徴を使用して作成したLTV予測モデルと同等程度の精度を達成できることを確認した。このことは本報告で提案した手法により、重要顧客に結びつく特徴の組み合わせを所望の数抽出できたことを意味する。講演では以上の内容についてより詳細な発表を行う。

参考文献

- [1]K. Koike, D. Sagawa, and K. Tanaka: "Customer clustering and customer-specific product recommendation by distance from core customer group in a multidimensional space of purchase behavioral features", The 37th Annual Conference of the Japanese Society for Artificial Intelligence, p. 1U4IS1a05-1U4IS1a05 (2023)
- [2]T. Okuyama, T. Sonobe, K.-i. Kawarabayashi and M. Yamaoka: "Binary optimization by momentum annealing", Physical Review E, 100, 1, p. 012111 (2019); 他.
- [3]D. C. Schmittlein, R. A. Peterson: "Customer Base Analysis: An Industrial Purchase Process Application", Marketing Science, Vol. 13, pp. 41-67 (1994)