

マーケティングプロセスの自動化に関する研究

Automation of the marketing process

佐々木 崇*¹
Takashi Sasaki

高木友博*¹
Tomohiro Takagi

*¹ 明治大学理工学研究科基礎理工学専攻

Computer Science Course, Graduate School of Science and Technology, Meiji University

In the current marketing, segmentation, and targeting, positioning is done in order, but the optimal segment against product characteristics cannot be achieved by this way. In this study, product characteristics that help to make the best sale and achieve the segment that is appealed most effectively and with that the segment that is appealed most effectively, and suggest the approach that guarantee the unified optimality of the three. With the evaluation experiment, I have shown that optimal segment and segment can be achieved the high-dimensional characteristic space.

1. はじめに

現在企業は、新たな製品を作る際、フィリップ・コトラーが提唱した STP マーケティング [Kotler 2014] を行っている。STP マーケティングとは市場における顧客のニーズごとにグループ化するセグメンテーション、セグメント化した結果、自社の参入すべき市場セグメントを選定するターゲティング、ターゲットとする市場で、競合他社と自社の関係を把握し、新しいポジションを確立するポジショニングからなり、STP マーケティングにより次に製品や価格を決められる。そのため、STP マーケティングを行うのは企業にとってとても重要なことである。

しかし STP マーケティングには問題点がある。現在のマーケティングは、セグメンテーション、ターゲティング、ポジショニングの順で行われるが、これでは商品特性に対して最適なセグメントを求めることができないという点である。

そのため、ポジショニングとセグメンテーションを同時に行うことでこの問題に取り組む。本研究では、最適な商品特性を求め、その商品を最も効率的に訴求させるセグメントを求める事で、3 者の統一的最適性を保証するアプローチを提案する。実験により、高次元の特性空間で、最適な商品特性とセグメントが求められる事を示した。

以下 2 章で提案手法の構造と計算方法の説明。3 章で、実験の方法と結果、また実験結果に基づく考察を行い、4 章で本論文を締めくくる。

2. 提案手法

2.1 全体像

提案手法について説明する。提案手法の概要図を図 1 で示す。提案手法はポジショニングを行う処理とセグメンテーションを行う処理からなる。ターゲティングは最適なポジションを最も効率的に訴求させるセグメントとする。

初めにポジショニングについて説明する。データからユーザが購入したアイテムの説明文に含まれる単語のユーザ

プロフィールを作成する①の処理と、次にユーザプロフィールと仮のポジションを用いて、売り上げ規模を計算する②の処理がある。

次にセグメンテーションについて説明する。どの単語で分類されるのかわかるため C5.0 [Quinlan 1993] を用いてセグメンテーションを行う (④の処理)。C5.0 を用いるため最適なポジションとユーザプロフィールのスコアを計算し、正例・負例を求める③の処理がある。

これより各処理の具体的な実現方法を示す。

2.2 ポジショニング

(1) アイテムのポジション

アイテムポジションはアイテム説明文に含まれる単語の TFIDF 値で表されるとする。j 番目のアイテム I_j のポジションを $P(I_j)$ とするとき、 I_j のポジションは式(1)で表される。ここで w_i は i 番目の単語の TFIDF 値である。

$$P(I_j) = (w_0, w_1, \dots, w_n) \quad (1)$$

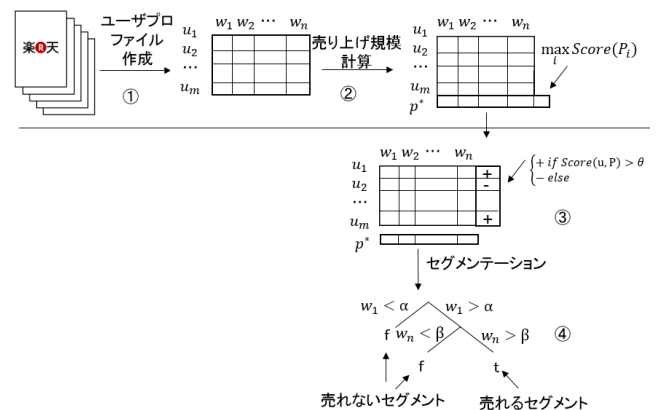


図 1 提案システムの概要図

(2) ユーザプロフィール

ユーザプロフィールはユーザが購入したアイテムで表され

るとする。ユーザプロファイルは、式(1)で求められる購入したアイテムのポジションの TFIDF 値の平均からなる。ユーザ i のユーザプロファイルを UP_i とするとき、ユーザプロファイルは式(2)で表される。ここで n はユーザが買った数である。

$$UP_i = \sum_j^n (P(I_j)/n) \quad (2)$$

(3)属性選択

アイテムのポジション及びユーザプロファイルには多くの単語が存在し、どの単語がユーザやアイテムの特性を表す重要な単語か判断することができない。そこで頻出単語がアイテムやユーザの特性を表す重要な単語と考え、出現頻度上位 20 語を選択する。ユーザ i のユーザプロファイル UP_i における j 番目の単語 w_j の TFIDF 値が 0 かどうかを判定する式(4)を用いて、単語の出現頻度を式(3)で求める。

$$\text{Count}(w_j) = \sum_i^n f(UP_i, w_j) \quad (3)$$

$$f(UP_i, w_j) = \begin{cases} 1 & \text{if } UP_i(w_j) \neq 0 \\ 0 & \text{else otherwise} \end{cases} \quad (4)$$

(4)売り上げ規模

・アイテムのスコア

アイテムの特性と類似性の高い特性を持つユーザには訴求し、類似性の低いユーザには訴求しないと考え、アイテムポジションとユーザプロファイルの類似性をそのユーザのスコアとする。具体的には、式(1)、式(2)でそれぞれ求められるアイテムポジションとユーザプロファイルの距離の逆数とし、式(5)で求められるものとする。

$$\text{Score}(UP_i, P) = \frac{1}{\sqrt{(\sum_j^n (UP_{ij} - P_j))^2 + 1}} \quad (5)$$

・売り上げ規模のスコア

一般的に市場には、自社アイテムだけでなく、競合他社のアイテムも存在するため、その影響を考慮する必要がある。それには対象ユーザにおける売り上げ規模が他社アイテムによって割り引かれると考え、自社アイテムのスコアから他社アイテムから見たスコアを減算することによって求めることとする。また市場全体の売り上げ規模は、すべてのユーザのスコアの合算値とし式(6)で求める。ここで P_c は他社のアイテムのポジションである。

$$\text{Sales}(P) = \sum_i^n (\text{Score}(UP_i, P) - \text{score}(UP_i, P_c)) \quad (6)$$

(5)最適化

企業において最も重要なことは利益を上げることだと考え、最適なポジションは売り上げ規模が最大となるポジションと考える。売り上げ規模が最大となるポジションは式(7)で求める。最適化には滑降シンプレックス法[John 1965]で行う。

$$P^* = \underset{P}{\text{argmax}} \text{Sales}(P) \quad (7)$$

2.3 セグメンテーション

セグメンテーションを一種の分類問題だと考え、root から葉までの枝が 1 個のセグメントとする。すべての true の葉の枝が最適なポジションのアイテムを最も効率的に訴求させるセグメントである。どの単語で分類されるのかを見るため C5.0 を用いた。C5.0 を行うために正負例を求める。今回閾値 θ は 3 つ以上の単語の TFIDF 値が 0.1 以上違う場合負例とするため、0.85 とし式(8)で求める。

$$\begin{cases} \text{true} & \text{if } \text{Score}(u, P) - \text{Score}(i, P) > \theta \\ \text{false} & \text{else otherwise} \end{cases} \quad (8)$$

3. 実験

本章では、提案手法の実験の方法と実験結果、また実験結果に基づく考察を述べる。

3.1 動作実験

人目で判断して正しく動いているかの動作実験を、ダミーデータを用いて他社を考慮しない場合と、考慮した場合で行った。人目で判断するためダミーデータは 2 次元である。

初めに他社を考慮しない場合である。軸は属性 1、属性 2 であり、青い点はユーザを示している。結果は図 2 で、オレンジの点が、求めたアイテムポジションであり、ユーザの密集点の真ん中に来ていることがわかる。人間の感覚と一致しているので人目で正しく動いているのがわかる。赤い線がセグメンテーションの結果で、0.3 以上が買うユーザ、0.3 以下が買わないユーザである。

他社のアイテムを考慮した場合の実験結果を図 3 に示す。赤い点は仮に置いた他社のアイテムを示している。他社のアイテムがあることにより他社のアイテムにユーザを取られるため、図 3 のオレンジの点が図 2 のオレンジの点より他社のアイテムから逃げているところにポジションしていることがわかり、人間の感覚と一致した。

3.2 実データを用いた実験

(1)実験設定

今回は楽天公開データ[1]の楽天市場のデータを用いた。前処理として、デジカメ購入者 206 名のデータを用いて実験を行う。実データの一部分が表 1 である。例えば user1 はフィルムの値が 1.359 で、エレコムが 2.885 でほかの値が 0 の user である。他社のアイテムは購入されたアイテムからランダムで一つ選んだ。

(2)他社のアイテムなし

初めに他社のアイテムなしで実験を行う。ポジションを求めた結果が表 2 である。今回の結果からいえることは、高次元においても正常に動作したことである。また、保証書付きでプリンターにつなげなれ、プレイヤーに対応している商品が売れるというのがわかった。

次にセグメントの結果が表 3 である。結果から以下の条件を満たすユーザが最適なポジションのアイテムを最も効率的に訴求させるセグメントとなった。

- ・保証書>0
 - ・対象<1.68
 - ・セット<1.29
- または、
- ・保証書>0

- ・対象<1.68
- ・セット>1.29
- ・カード<1.42
- ・ポイント<2.15

(3)他社のアイテムあり

仮に他社のアイテムがあったとする.その場合の得られたポジションが表4である.他社のアイテムを考慮した場合,他社を考慮しなかった場合に比べ値がずれていることが分かり,他社を考慮しても正常に動いた.高次元においても最適なポジションを求められ,人間ではできない高次元においてのポジショニング,セグメンテーションを行うことができ,有効である.

保証書 <= 0: f (87)		
保証書 > 0:		
...対象 > 1.680055: f (7)		
対象 <= 1.680055:		
...セット <= 1.290984: t (38)		
セット > 1.290984:		
...カード > 1.404644: f (7)		
カード <= 1.404644:		
...ポイント <= 2.152279: t (8/1)		
ポイント > 2.152279: f (3)		

表4 他アイテムあり 結果

対象	1.68
保証書	1.73
プリンター	1.71
ビデオカメラ	1.69
レンズ	1.59
価格	1.63
プレイヤー	1.73
デジタル	0.62
カメラ	0.57

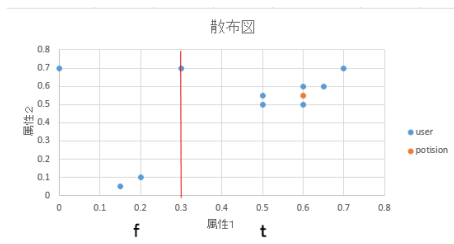


図2 他社を考慮しない場合の結果

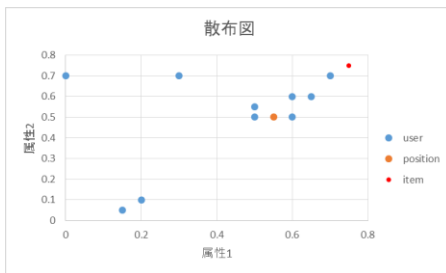


図3 他社を考慮した場合の結果

表1 実データ(一部)

対象	あり	無料	デジカメ	カード	保証書	フィルム	プリンター	ビデオカメラ	エレコム
user1	0	0	0	0	0	1.359387	0	0	2.865629
user2	0	9.325051	4.158883	0	0	0	0	0	2.865629
user3	0	9.325051	0	0	0	0	0	0	2.865629
user4	0	0	0	1.4046435	0	0	0	0	0
user5	0	0	0	1.632427	0	0	0	0	0
user6	0	0	0	3.264855	0	0	0	0	0
user7	0	0	0	0	0	1.359387	0	0	0
user8	1.680055	0	0	0	1.730066	0	1.71732687	1.609438	0
user9	0	0	0	1.632427	0	1.359387	0	1.609438	0
user10	0	0	0	0	0	0	0	0	0

表2 結果

対象	1.59
無料	0.28
カード	0.05
保証書	1.63
フィルム	0.41
プリンター	1.62
ビデオカメラ	1.43
レンズ	1.51
価格	1.54
ポイント	0.42
送料	0.26
プレイヤー	1.63
デジタル	0.68
カメラ	0.66
セット	0.06
液晶	0.47

表3 セグメント結果

4. おわりに

本稿では,マーケティングプロセスの自動化手法を提案した. 評価実験ではダミーデータを用いて,他者を考慮した場合でも最適なポジションと最適なセグメントを求めることが可能であることを示せた. 提案手法は,各ポジションのスコアを計算し,C5.0でセグメンテーションを行うことでどのような単語でセグメンテーションを行えばよいかわかり,新たなユーザを考慮するとき有用だと考えている.

今後の課題としていくつかあり,他社アイテムを複数でも考慮できるようにすること. 属性選択を出現頻度の上位 20 属性としているが,出現頻度が特性を表すのに重要な単語なのかどうか,20 という数字の信ぴょう性などの問題がある. さらにセグメンテーションにおいて正負例の閾値を 0.85 としているがそれが適切な値かどうか. 最後に精度実験を行い,出力したポジション,セグメントがどれくらい有用かどうかを示すことである.

参考文献

[Kotler 2014] フィリップ・コトラー,ゲイリー・アームストロング,著;月谷真紀訳,コトラーのマーケティング入門第4版,ピアソンエデュケーション.

[Quinlan 1993] Quinlan, J. R. C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers.

[John 1965] Nelder, John A.; R. Mead. "A simplex method for function minimization". *Computer Journal* 7:308-313

[1] <http://rit.rakuten.co.jp/opendataj.html>