

EC サイトにおける購買行動データの学習に基づく クーポン効果分析モデルに関する研究

情報数理応用研究

5215C005-0
指導教員

遠藤海太郎
後藤正幸

A Study of Coupon Effect Maximization Model Based on Purchase Behavior Data Learning on EC Site

Kaitaro ENDO

1 研究背景・目的

EC サイト市場の規模拡大に伴い、ユーザのアクセスログデータが大量に蓄積されるようになった。これにより蓄積された膨大な量のデータを活用する Web マーケティング技術の重要性が高まっており、その施策の一つとして、サイトにアクセスしているユーザに対する割引クーポンの発行が実用化されている。一般に、クーポンの発行によりユーザの購買意欲が刺激され、売上の向上が期待できる。

しかし、サイトにアクセスした全てのユーザに対して無条件にクーポンを発行した場合、定価でも購入する意思のあったユーザに対しても割引をしてしまうことになり、売上に対する機会損失が懸念される。以上から、「購買を迷っていたが、クーポンが表示されることで購買を決心するようなユーザ」を対象としたクーポンの発行が売上増加の観点から重要であると考えられ、購買確率とクーポン発行との関係性のモデル化が求められる。

選択肢集合の選択確率を、その選択肢の持つ効用によって表現するためのモデルとして、ロジットモデル [1] がある。ここで、選択肢集合を「商品を購入する」、「購買しない」とし、効用変数としてクーポン発行の有無とその他の購買行動情報を用いると、EC サイトにおけるユーザの選択確率モデルを構築することができる。

一方で、EC サイトにおける購買行動は、ユーザごとに異なると考えられる。たとえば、複数の商品を時間をかけて吟味するユーザや購買する商品がアクセスする前から決まっているユーザでは、データとして表される行動は大きく異なる。また、同一ユーザであってもアクセスした状況によって、常に同じ購買確率になるとは限らない。すなわち、EC サイトを対象とした分析を行う場合には、ユーザの異質性を前提として、ユーザの行動情報とクーポン発行の関係性をモデル化する必要があると言える。特にアイテムの閲覧と購買には密接な関係があると考えられ、モデルを構築する上で重要な購買行動情報と言える。

そこで本研究では、アイテムの閲覧と購買を共起の事象として、ユーザの異質性を考慮するために潜在クラスモデル [2] とクーポンの発行やその他のユーザの行動情報による購買確率への効果を分析するためにロジットモデルの 2 つのモデル統合した分析モデルの提案を行う。これにより、EC サイトに蓄積されている膨大な量のアクセスログデータを用いて、ユーザごとのクーポンの効果を定量的に評価することができる。さらに実際の EC サイトから提供されたデータを用いて、提案モデルの有効性を示す。

2 準備

2.1 分析対象データ

本研究では、EC サイトの購買行動解析およびユーザへの販促を事業とする Emotion Intelligence 株式会社が保有

している実データ（大手総合通販カタログサイトのアクセスログデータ）を対象とし、ユーザのクーポン効果分析モデルの構築を行う。アクセスログデータにはユーザ ID、ページ遷移開始および終了時間、アクセスページに関する情報のほかに、購買の有無、クーポン発行の有無や購買金額、EC サイトにアクセスしてから離脱するまでの一連の行動に対して付与させるセッション ID などの情報が付与されている。たとえば、1 ユーザが時間を空けて EC サイトに 2 回アクセスした場合、セッション数は 2 であると考えられる。表 1 に本研究で用いる対象 EC サイトのデータ概要を示す。なお、本研究で用いたデータは EC サイトにおけるアイテムの一覧情報を保有していないため、全セッションで閲覧されたアイテムをアイテムテーブルとして分析に用いる。

表 1: 対象 EC サイトにおけるデータ概要

総セッション数		164,238
総 PV 数		2,330,988
PV	Average	14.19
	Median	6
滞在時間 (ms)	Average	1,367,243
	Median	374,134
平均購買金額 (円)	Average	10,169
	Median	7,527
CVR		0.029

表 1 における CVR(Conversion Rate) とは総セッション数のうち、購買したセッションが占める割合を表す。すなわち、本研究が対象とする EC サイトでは、全セッションのうち、約 2.9 %において商品が購買されていることを表している。

2.2 Aspect Model

マーケティングサイエンスの分野において、ユーザの嗜好の異質性を考慮しながらユーザのアイテムの購買傾向をモデル化する様々な潜在クラスモデルが提案されている。その一つである Aspect Model [2] は大量の購買データに対しても容易な適用が可能であり、以下のように定式化される。いま、 I 人からなるユーザ集合を $U = \{u_i : 1 \leq i \leq I\}$ 、 N 個のアイテム集合を $A = \{a_n : 1 \leq n \leq N\}$ 、 $Z = \{z_l : 1 \leq l \leq L\}$ を L 個の潜在クラス集合とする。また、同じ購買傾向を持つユーザは同じ潜在クラスに所属すると仮定する。図 1 に Aspect Model におけるグラフィカルモデルを示す。

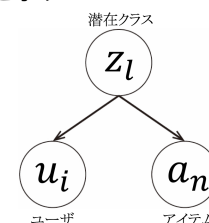


図 1: Aspect Model におけるグラフィカルモデル

このモデルにより、ユーザの購買傾向および購買されたアイテムの特徴を表現することが可能となる。いま、ユーザ u_i がアイテム a_n を購買する事象 (u_i, a_n) の確率は、パラメータ $P(z_l)$, $P(u_i|z_l)$, $P(a_n|z_l)$ を用いて、式 (1) で表される。

$$P(u_i, a_n) = \sum_{l=1}^L P(z_l)P(u_i|z_l)P(a_n|z_l). \quad (1)$$

ここに、 $P(z_l)$, $P(u_i|z_l)$, $P(a_n|z_l)$ にそれぞれ多項分布を仮定する。このモデルのパラメータは全学習データに対して、式 (2) の対数尤度関数を極大にするように推定される。

$$L = \sum_{i=1}^I \sum_{n=1}^N \delta(u_i, a_n) \log P(u_i, a_n). \quad (2)$$

ただし、 $\delta(u_i, a_n)$ はユーザ u_i がアイテム a_n を購買するとき 1、それ以外で 0 をとるインジケータ関数である。式 (2) における潜在クラスは直接的に観測できないため、潜在クラスを含むパラメータは探索的手法の 1 つである EM アルゴリズム [3] を用いて推定を行う。

2.3 二項ロジットモデル

EC サイトにアクセスしたユーザの各セッションにおける購買確率を求めするための基本的アプローチとしては、確率的選択モデルの一つである二項ロジットモデル [1] がある。セッション s ($1 \leq s \leq S$) における購買確率 P_s を二項ロジットモデルで表すと、式 (3) で表される。

$$P_s = \frac{\exp(V_s)}{\exp(V_s) + 1}. \quad (3)$$

ただし、 V_s はセッション s における購買に関する効用で、式 (4) で表すような共変量 \mathbf{x}_s の一次関数とする。

$$V_s = \beta^T \mathbf{x}_s + \epsilon_s. \quad (4)$$

ここで、 $\mathbf{x}_s = (x_1^s, x_2^s, \dots, x_K^s)^T$ はモデル作成に用いる K 個の効用変数からなる K 次元効用ベクトル、 $\beta = (\beta_1, \beta_2, \dots, \beta_K)^T$ はパラメータベクトル、 ϵ_s はランダム効用である。また、 T は転置を表す。上記のモデルにおけるパラメータベクトル β は、学習データから最尤推定法によって推定することができる。

3 提案モデル

3.1 提案モデル概要

本研究の目的は、EC サイトにおけるアクセスログデータを用いて、セッションごとのクーポン効果を定量的に評価するための分析モデルを構築することである。そこで、本研究では前述の二手法 (Aspect Model と二項ロジットモデル) をベースとし、それらを組み合わせたモデルを提案する。具体的には、ユーザのセッションにおいて、ロジットモデルによって表現されるユーザの購買行動とアイテム閲覧履歴との間に潜在クラスを仮定し、潜在クラスごとにクーポンの効果を表現するモデルを提案する。このモデルは、効用と購買・非購買の関係およびアイテム閲覧履歴の共起を事象と捉え、所属するアイテム閲覧履歴の生起確率および、効用と購買・非購買の生起確率の関係にそれぞれ多項分布とロジットモデルを仮定した潜在クラスモデルと位置付けることができる。

3.2 提案モデルの定式化

本節では以下のような新たなモデルを定式化する。いま、 r_s をセッション s が購買したとき 1、それ以外で 0 を取る二値変数とする。アイテム閲覧履歴は N 個のアイテムに対して、 $\mathbf{y}_s = (y_1^s, y_2^s, \dots, y_N^s)^T$ で表されるものとし、バイナリのベクトル変数とする。ここで、 y_n^s はセッション s において n 番目のアイテムを閲覧したとき 1 を取り、それ以外で 0 を取るものとする。

セッション s におけるデータを $(r_s, \mathbf{y}_s)^T$ とする。また、総セッション数 S 個のデータセットを $\{(r_s, \mathbf{y}_s)\}_{s=1}^S$ と表し、それぞれのデータには多項分布を仮定する。

さらに、潜在クラス集合を $Z = \{z_1, z_2, \dots, z_L\}$ とし、セッション s が属する潜在クラスを $w_s \in Z$ で表す。二項ロジットモデルに適用する、 K 個の効用変数からなる K 次元効用ベクトルを $\mathbf{x}_s = (x_1^s, x_2^s, \dots, x_K^s)^T$ とする。

いま、潜在クラス $z_l \in Z$ における各アイテムの閲覧確率を表す多項分布のパラメータ $P(a_n|z_l)$ は式 (5) を満たすものとする。

$$\sum_{n=1}^N P(a_n|z_l) = 1. \quad (5)$$

セッション s における r_s, \mathbf{y}_s の同時確率 $P(r_s, \mathbf{y}_s)$ は以下のように定式化できる。

$$P(r_s, \mathbf{y}_s) = \sum_{l=1}^L P(z_l)P(r_s|z_l)P(\mathbf{y}_s|z_l) \\ = \sum_{l=1}^L P(z_l)P(r_s|z_l) \frac{(\sum_{n=1}^N y_n^s)!}{N} \prod_{n=1}^N \left(P(a_n|z_l) \right)^{y_n^s}. \quad (6)$$

式 (6) における $P(r_s|z_l)$ を以下では P_{ls} と記述し、二項ロジットモデルを用いて以下のように定義する。

$$P(r_s|z_l) = P_{ls} = \frac{\exp(V_{ls})}{\exp(V_{ls}) + 1}. \quad (7)$$

$$V_{ls} = \beta_l^T \mathbf{x}_s + \epsilon_{ls}^l. \quad (8)$$

ただし、 V_{ls} は潜在クラス z_l におけるセッション s の効用、 $\beta_l = (\beta_1^l, \beta_2^l, \dots, \beta_K^l)^T$ は潜在クラス z_l における効用パラメータベクトル、 ϵ_{ls}^l はランダム効用を表す。提案モデルのグラフィカルモデルを図 2 に示す。

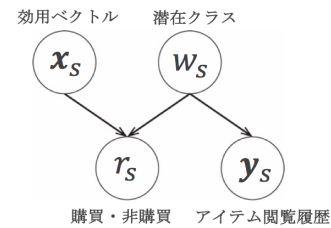


図 2: 提案手法におけるグラフィカルモデル

3.3 EM アルゴリズムによるパラメータ推定

本研究では、EM アルゴリズム [3] に基づき、パラメータを推定する。アクセスログデータに対する対数尤度関数 LL は式 (9) のように示される。

$$LL = \sum_{s=1}^S \log P(r_s, \mathbf{y}_s). \quad (9)$$

ここで、式 (9) には直接観測できない潜在クラス z_l が含まれているため、EM アルゴリズムを適用し、対数尤

度関数 LL が極大となるようなパラメータを求める．EM アルゴリズムは 2 つのステップからなる．まず，E-step では，パラメータが与えられた下で，事後確率を計算する．次に M-step は，E-step で求めた事後確率を用いて，対数尤度関数 LL が極大となるパラメータを求める．この 2 つのステップを繰り返し行うことで，パラメータの推定を行う．E-step では与えられた $\mathbf{x}_s, r_s, \mathbf{y}_s$ (以下では閲覧履歴データと呼ぶことにする) および，パラメータを用いて，潜在クラス z_l の事後確率 $P(z_l|\mathbf{x}_s, r_s, \mathbf{y}_s)$ を計算する．事後確率は式 (10) によって求められる．

< E-step >

$$P(z_l|\mathbf{x}_s, r_s, \mathbf{y}_s) = \frac{P(z_l)P(r_s|\mathbf{x}_s, z_l)P(\mathbf{y}_s|z_l)}{\sum_{l=1}^L P(z_l)P(r_s|\mathbf{x}_s, z_l)P(\mathbf{y}_s|z_l)}. \quad (10)$$

M-step では式 (10) から求めた事後確率を固定し，パラメータ $P(z_l), P(a_n|z_l), P(r_s|\mathbf{x}_s, z_l)$ の更新を行う．

< M-step >

$$P(z_l) = \frac{\sum_{s=1}^S P(z_l|\mathbf{x}_s, r_s, \mathbf{y}_s)}{\sum_{l=1}^L \sum_{s=1}^S P(z_l|\mathbf{x}_s, r_s, \mathbf{y}_s)}. \quad (11)$$

$$P(a_n|z_l) = \frac{\sum_{s=1}^S P(z_l|\mathbf{x}_s, r_s, \mathbf{y}_s)y_n^s}{\sum_{n=1}^N \sum_{s=1}^S P(z_l|\mathbf{x}_s, r_s, \mathbf{y}_s)y_n^s}. \quad (12)$$

ここで， $P(r_s|\mathbf{x}_s, z_l)$ は次のように推定される．セッション s における購買・非購買 r_s と効用変数ベクトル \mathbf{x}_s が与えられた下で，対数尤度関数 LL_{logit} は潜在クラスの効用パラメータベクトル $\beta_l = (\beta_1^l, \beta_2^l, \dots, \beta_K^l)^T$ を用いて，式 (13) によって求められる．

$$\begin{aligned} LL_{logit} &= \log \prod_{s=1}^S P_s^{r_s} (1 - P_s)^{1-r_s} \\ &= \sum_{s=1}^S \left[r_s \left(\sum_{l=1}^L \beta_l^T \mathbf{x}_s \right) - \log \left\{ \exp \left(\sum_{l=1}^L \beta_l^T \mathbf{x}_s \right) + 1 \right\} \right]. \end{aligned} \quad (13)$$

ここで，セッション s における購買確率 $P(r_s|\mathbf{x}_s, z_l)$ は式 (14) によって定義する．

$$\begin{aligned} P_s &= \sum_{l=1}^L P(z_l)P_l \\ &= \sum_{l=1}^L P(z_l) \frac{\exp(V_{ls})}{\exp(V_{ls}) + 1}. \end{aligned} \quad (14)$$

3.4 予測問題への応用

上記の提案モデルでは，各セッションにおいて，EC サイトを離脱するまでのすべてのアクセスログデータを用いて，アイテム閲覧履歴と購買・非購買が定式化されている．しかし，実務上ではユーザがサイトにアクセスしている途中で，閲覧アイテムに対するクーポンを発行した方が，購買意欲の向上に貢献しやすいと考えられる．そこで，セッション s における総閲覧ページ数のうち， t 番目のアイテムを閲覧したタイミングでクーポンを発行するというアプローチを考える．

提案モデルにおいて， t 番目のアイテムを閲覧した段階までをセッションデータとして学習を行うことで，購買確率の予測モデルを構築することが可能である．

いま， t 番目のアイテムを閲覧した段階までのセッションデータにより推定されたパラメータを， $\hat{P}(z_l), \hat{P}(a_n|z_l), \hat{P}(r_s|\mathbf{x}_s, z_l)$ とし， \mathbf{y}_s^* を新たなセッション s^* におけるアイテム閲覧履歴とする．このとき，新たなセッション s^* に対する購買確率の予測値 P_{s^*} は以下の式 (15) で表される．

$$\begin{aligned} P_{s^*} &= \sum_{l=1}^L \hat{P}(z_l|\mathbf{y}_s^*)\hat{P}_{ls^*} \\ &= \sum_{l=1}^L \frac{\hat{P}(z_l)\hat{P}(r_s^*|\mathbf{x}_s^*, z_l)\hat{P}(\mathbf{y}_s^*|z_l)}{\sum_{l=1}^L \hat{P}(z_l)\hat{P}(r_s^*|\mathbf{x}_s^*, z_l)\hat{P}(\mathbf{y}_s^*|z_l)} \hat{P}_{ls^*}. \end{aligned} \quad (15)$$

式 (15) において， \hat{P}_{ls^*} は式 (8) より $x_k^{s^*}$ の関数であることが分かる．したがって，推定された効用パラメータベクトル β を用いて， k 番目の効用変数 $x_k^{s^*}$ が 0 と 1 の場合での購買確率の増減を調べることができる．いま，クーポンの発行の有無に関する効用変数が $x_k^{s^*}$ であるとし， $x_k^{s^*} = 1$ のとき「クーポンを発行する」， $x_k^{s^*} = 0$ のとき「クーポンを発行しない」を意味するものとする．このとき，推定された効用パラメータベクトル $\hat{\beta}_l$ を用いて，クーポンを発行した場合の予測購買確率 $P_{s^*}(x_k^{s^*} = 1)$ と発行しなかった場合の予測購買確率 $P_{s^*}(x_k^{s^*} = 0)$ を式 (15) より求めることが可能である．

これらの予測値の比を求めることで，クーポンを発行したことに対する効果が大きいセッションに対して，クーポンを発行することができる．また，式 (16) によってクーポンを発行したことによる効果が表されるものとする．

$$\frac{P_{s^*}(x_k^{s^*} = 1)}{P_{s^*}(x_k^{s^*} = 0)} > \xi. \quad (16)$$

ただし， $\xi > 1$ は閾値を表し，EC サイト運営者がその値を設定するものとする．

4 提案モデルを用いた購買行動データの解析

提案手法の有効性を確認するため，実際の大手総合通販カタログサイトのアクセスログデータを提案モデルにより分析し，さらにクーポンを発行すべきかについて考察する．

4.1 分析条件

二項ロジットモデルにおける効用変数は，事前分析より表 2 に示す 8 つの変数を用いた．なお，「閲覧時間」，「PV」，「カートボタンクリック数」は正規化した値を用いた．また，事前実験より潜在クラス数は $L = 10$ とした．ただし「アクセス時間帯」においては 4～8 時を基準として分析を行った．

表 2: 本研究で使用する変数と定義域

変数	定義域
滞在時間	0 以上の実数 (ms)
PV	0 以上の実数
閲覧端末	SP=0, PC=1
カートボタンクリック数	0 以上の実数
アクセス時間帯	{0,1}
配信メールからアクセス	{0,1}
広告ページからアクセス	{0,1}
クーポン発行	{0,1}

4.2 分析結果

表 3 に推定された効用パラメータベクトルの結果を示す．まず，クラス z_4 とクラス z_7 に着目すると，これら

表 3: 効用パラメータベクトル β_l の推定結果

潜在クラス	滞在時間	PV	閲覧端末	カートボタン クリック数	アクセス時間帯					配信メール	広告ページ からアクセス	クーポン発行	混合割合 $P(z_l)$	クーポン 発行割合
					0~4	8~12	12~16	16~20	20~24					
1	0.609	0.158	0.213	0.439	0.863	0.576	0.802	0.105	0.727	0.298	0.464	0.676	0.156	0.076
2	0.427	0.968	0.737	0.077	0.302	0.550	0.789	0.808	0.885	0.499	0.429	0.024	0.083	0.046
3	0.188	0.389	0.332	0.859	0.548	0.957	0.613	0.775	0.241	0.721	0.308	0.566	0.090	0.046
4	0.768	0.748	0.932	0.245	0.766	0.661	0.752	0.032	0.666	0.486	0.173	0.842	0.120	0.064
5	0.450	0.429	0.055	0.364	0.965	0.820	0.529	0.911	0.071	0.372	0.954	0.185	0.100	0.057
6	0.216	0.607	0.241	0.239	0.633	0.162	0.976	0.179	0.430	0.655	0.587	0.351	0.046	0.023
7	0.228	0.416	0.951	0.362	0.019	0.876	0.544	0.415	0.220	0.187	0.605	0.948	0.171	0.092
8	0.280	0.464	0.271	0.731	0.098	0.279	0.020	0.034	0.762	0.929	0.705	0.709	0.029	0.016
9	0.218	0.558	0.503	0.391	0.409	0.153	0.287	0.970	0.556	0.925	0.043	0.767	0.192	0.103
10	0.238	0.631	0.225	0.324	0.073	0.327	0.968	0.566	0.455	0.949	0.282	0.389	0.013	0.0062

はクーポン発行に対する効用が高い潜在クラスであることが分かる。したがって、これらのクラスへの所属確率が高いセッションに対して、クーポンの発行を実施することで、購買確率を向上させることが可能となるものと思われる。また、推定されたパラメータのうち、 $P(z_1)$ に着目すると、表 3 においてクラス z_9 への混合割合が 19.2%と最も高いことが分かる。クーポンの発行割合は所属する潜在クラスのうち、実際にクーポンが発行された割合を表しており、やはりクラス z_9 の値が高い。

4.3 考察

それぞれのクラスに関して、詳細に分析した結果を以下に示す。まず、クラス z_7 では、閲覧端末の係数が他のクラスと比較して大きな値を取っていることが分かる。この効用変数は閲覧端末が PC であるとき 1 をとり、スマートフォン (以下, SP) のとき 0 をとるような二値の変数である。したがって、閲覧端末の係数が大きいということは、SP と比較して PC の方が効用が大きくなるということを示している。このクラスに対しては EC サイトに PC でアクセスした場合に特に購買に結びつきやすいと考えられる。また、クラス z_9 のアクセス時間帯に着目すると、16~20 時の時間帯で、購買への効用が大きくなること分かる。

次に、クラス z_2, z_5 に着目すると、クーポン発行に対する効用パラメータが他のクラスと比較して小さい値をとっていることが分かる。すなわち、クーポンを発行したとしても、購買確率の上昇が小さいことを示している。したがって、これらのクラスでは、クーポンの発行による効果が小さいものと考えられる。

一方、現在、実施されているクーポン施策に従った場合のクーポン発行割合は、クラス z_4, z_7 のクーポン発行割合がそれぞれ 6.4%, 9.2%となっており、これらのクラスにより多くのクーポンを発行することが対象 EC サイトにおいては望ましいと考えられる。

次に、クーポン発行に対する効用の大きいクラス z_4, z_7 と効用が小さいクラス z_2, z_5 の購買行動に着目してみる。表 4 に各潜在クラスにおける購買行動について比較した結果を示す。ただし、各セッションは最も所属確率の高い潜在クラスに所属するものとして、各クラスの統計量を比較した。また、平均閲覧アイテム金額は閲覧された 1 アイテムあたりの平均金額、平均購買回数は 1 セッションあたりの購買回数の平均値を表す。クーポン効果は推定したパラメータベクトル β_l におけるクーポン発行に対する効用を表す。

まず、クーポンに対する効用が小さいクラス z_2, z_5 に着目すると、クラス z_2 は表 4 のクラスの中で 2 番目に平均閲覧アイテム金額が高く、平均購買回数は最も大きいことから、クーポンを発行しなくても購買する確率が高いクラスであると考えられる。また、クラス z_5 は購買金額が低く、平均購買回数も低いことから、購買意欲の低い

クラスであることが考えられる。

次に、クーポンに対する効用が大きいクラス z_4, z_7 に着目すると、クラス z_7 は購買金額および平均購買回数も大きいことから、クーポンを発行することによって、大きく購買確率を向上させることができると考えられる。また、クラス z_4 は他のクラスと比較して、総閲覧時間が短く、1 ページあたりの滞在時間も短いという特徴を有し、購買金額の平均値と中央値の差が大きいことが分かる。したがって、短時間の間にアイテムを閲覧するユーザが多く、高額購買のユーザを含むクラスであると考えられる。

表 4: 各潜在クラスにおける購買行動の分析

潜在クラス		2	4	5	7
PV	Average	24.33	16.57	20.26	17.22
	Median	15	10	12	10
閲覧時間 Dwell (min)	Average	23.76	18.89	22.82	20.21
	Median	12.70	7.51	9.11	7.92
Dwell/PV (min)	Average	0.98	1.14	1.13	1.17
	Median	0.85	0.75	0.76	0.79
購買金額 (¥)	Average	8,143	12,508	7,288	10,756
	Median	6,901	6,975	5,594	8,024
平均閲覧アイテム金額 (¥)		3,232	2,360	3,881	2,991
平均購買回数		0.071	0.049	0.028	0.051
クーポン効果		0.024	0.842	0.185	0.948

5 まとめと今後の課題

本研究では EC サイトの閲覧履歴データを用いて、ユーザごとのクーポン効果を定量的に評価するための分析モデルを提案した。また、提案モデルの有効性を大手総合通販カタログサイトの閲覧履歴データを用いた分析により示した。また、クーポン発行の効果が大きいクラスの購買行動を明らかにした。

現状では、アイテム閲覧履歴を用いているが、よりクーポンの効果を表現するモデルを構築するために適切な変数の追加を検討する必要がある。したがって、提案モデルに使用する変数の追加は今後の課題とする。

参考文献

- [1] J.H. Aldrich, F.D. Nelson, "Linear probability logit and probit models," *Sage Publications*, 1984.
- [2] Hofmann T, "Probabilistic latent semantic indexing," *Proc. of the 22nd Annual ACM Conference on Research and Development in Information Retrieval*, pp. 50-57, 1999.
- [3] A.P. Dempster, N.M. Laird, D.B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of Royal Statistical Society. Series B*, vol. 39, no. 1, pp. 1-38, 1977.