1B-04

ECサイトの商品特性を考慮した2次元確率表による購買予測

西村 直樹[†]　鮭川 矩義[†]　高野 祐一[‡]　岩永 二郎[§]　水野 眞治[†]
[†] 東京工業大学　[‡] 専修大学　[§] 株式会社NTTデータ数理システム

1 はじめに

インターネットの普及に伴い，現在では商品販売やサービスをウェブサイト上で提供するEC（Electronic Commerce：電子商取引）サイトを多くの企業が運営するようになった．ECサイトにおける顧客の閲覧や購買の予測は実務上の重要な課題であり，本研究では，ECサイトに来訪する顧客の購買商品を予測することを分析課題とする．

岩永ら[1]は，アクセスログのデータから閲覧商品に対する「関心度」と「忘却度」を数量的に定義し，これら2種類の特徴量から購買につながる商品を予測する手法を提案した．しかし，この手法では全顧客・全商品に対して単一の確率表を参照して購買予測をしており，ECサイトの商品特性が十分に考慮されていない．

そこで本研究では，多種多様な商品を扱い，幅広い年齢層の顧客を抱えるECサイトを想定して，その商品特性を考慮した2次元確率表の作成方法を提案する．

2 既存手法

本研究では，関心度を表す特徴量は「当該商品に対する閲覧回数」とし，$I = \{1, 2, \ldots, 20\}$とした．忘却度を表す特徴量は「当該商品の最終閲覧日からの経過日数」とし，$J = \{1, 2, \ldots, 28\}$とした．このとき，関心度が$i \in I$で忘却度が$j \in J$の商品が購買される確率（実績購買確率）p_{ij}は過去データから計算することができる．

購買確率は商品に対する関心度が高いほど増加し，商品に対する忘却度が高いほど減少することが期待される．しかし，実績購買確率ではそのような単調性が満たされない場合がある．そこで，先行研究[1]では関心度と忘却度に対する単調性制約の下で，確率表のセル$(i, j) \in I \times J$のデータ数によって重み付けられた残差2乗和が最小となるように購買確率x_{ij}を推定する問題を，以下の凸2次最適化問題として定式化した：

$$
\begin{aligned}
\text{最小化}_{x_{ij}}：\ & \sum_{i \in I} \sum_{j \in J} c_{ij}^2 (x_{ij} - p_{ij})^2 \\
\text{制約条件}：\ & x_{i_1 j} \le x_{i_2 j} \quad (i_1 < i_2\ (\in I),\ j \in J), \\
& x_{i j_1} \ge x_{i j_2} \quad (i \in I,\ j_1 < j_2\ (\in J)), \\
& 0 \le x_{ij} \le 1 \quad (i \in I,\ j \in J).
\end{aligned}
$$

ここで，c_{ij}はセル(i, j)の実績購買確率p_{ij}の計算に用いたデータ数とする．そして，顧客が閲覧した商品の中から，確率表を参照して購買確率が上位の商品を購買商品として予測する．

上記の凸2次最適化問題は単調回帰とみなすこともできる．また，この手法は，2種類の数値属性からなる領域と事象の生起を関連付ける2次元数値相関ルール（文献[2]などを参照）とも関連が深い．

3 提案手法

本研究ではECサイトの商品特性を考慮した確率表の作成方法として，「顧客と商品の類型化」と「凹/凸性制約」を提案する．

3.1 顧客と商品の類型化

顧客や商品の多様性を考慮した方法として，顧客や商品の類型$k \in K$に対応させて複数の確率表を作成することを考える．例えば，男性と女性とで購買傾向が異なるとすれば類型を$K = \{$ 男性, 女性 $\}$と設定し，商品分類によって購買傾向が異なるとすれば$K = \{$ Tシャツ, 時計, $\ldots$, 財布 $\}$のように設定する．

このように複数の確率表を作成することは，顧客や商品の多様性を表現できるという利点がある．しかし，類型の個数$|K|$が多くなると各類型に割り当てられるデータ数が減少し，過剰適合が生じて逆に予測精度が悪化する可能性がある．過剰適合を軽減するために，本研究では確率表間の乖離を抑制する制約条件を課した定式化を提案する．具体的には，類型kの確率表のセル(i, j)の購買確率を変数x_{ijk}とし，補助変数$\hat{x}_{ij}$を導入する．そして，x_{ijk} $(k \in K)$が一定範囲内に収ま

Purchase Prediction Using Two-Dimensional Probability Table for E-Commerce Sites
Naoki NISHIMURA[†]　Noriyoshi SUKEGAWA[†]　Yuichi TAKANO[‡]　Jiro IWANAGA[§]　Shinji MIZUNO[†]
[†] Tokyo Institute of Technology　[‡] Senshu University
[§] NTT DATA Mathematical Systems Inc.

るように以下の制約条件を追加する：

$$\frac{1}{1+\lambda}\hat{x}_{ij} \le x_{ijk} \le (1+\lambda)\hat{x}_{ij} \quad (i \in I,\ j \in J,\ k \in K).$$

ここで，λ は確率表間の乖離の度合いを調節するパラメータである．

各類型の確率表を求める問題は，以下の凸 2 次最適化問題として定式化できる：

$$\begin{aligned}
\underset{x_{ijk},\hat{x}_{ij}}{\text{最小化}}：\ & \sum_{i\in I}\sum_{j\in J}\sum_{k\in K} c_{ijk}^2 (x_{ijk}-p_{ijk})^2, \\
\text{制約条件}：\ & \frac{1}{1+\lambda}\hat{x}_{ij} \le x_{ijk} \le (1+\lambda)\hat{x}_{ij} \\
& \qquad (i \in I,\ j \in J,\ k \in K), \\
& x_{i_1jk} \le x_{i_2jk} \\
& \qquad (i_1 < i_2\ (\in I),\ j \in J,\ k \in K), \\
& x_{ij_1k} \ge x_{ij_2k} \\
& \qquad (i \in I,\ j_1 < j_2\ (\in J),\ k \in K), \\
& 0 \le x_{ijk} \le 1 \quad (i \in I,\ j \in J,\ k \in K).
\end{aligned}$$

ここで，p_{ijk} は類型 k の確率表のセル (i,j) の実績購買確率とし，c_{ijk} は p_{ijk} の計算に用いたデータ数とする．

3.2 凹/凸性制約

商品の購買確率は関心度に対して単調に増加していくが，関心度が増えるにつれてその効果が薄れ，購買確率の増加率は徐々に 0 へと近づいていくことが予想される．同様に，忘却度に対する購買確率の減少率も徐々に 0 へ近づいていくと予想される．

そこで本研究では，購買確率を表す関数の傾きが関心度に対して単調に減少し，忘却度に対して単調に増加することを表す以下の線形制約を提案する：

$$\begin{aligned}
& x_{i-1,j} - x_{i-2,j} \ge x_{ij} - x_{i-1,j} \\
& \qquad (i \in I\backslash\{1,2\},\ j \in J), \\
& x_{i,j-1} - x_{i,j-2} \le x_{ij} - x_{i,j-1} \\
& \qquad (i \in I,\ j \in J\backslash\{1,2\}).
\end{aligned}$$

上記の制約条件は凹関数/凸関数の特徴付けとも一致するため，本研究では凹/凸性制約と呼ぶこととする．

4 数値実験

経営科学系研究部会連合協議会主催，平成 25 年度データ解析コンペティションで提供されたファッション EC サイトのデータを用いて，各顧客の購買商品を予測する数値実験を行なった．図 1〜3 に推定した確率表の一例を示す．結果の詳細については当日報告する．

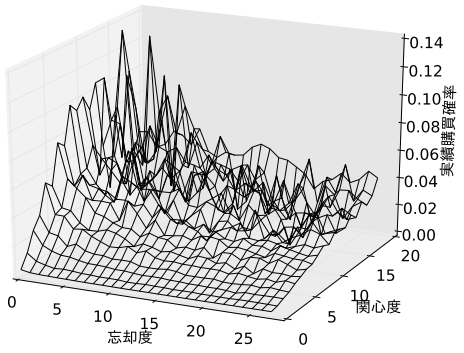

図 1: 実績購買確率

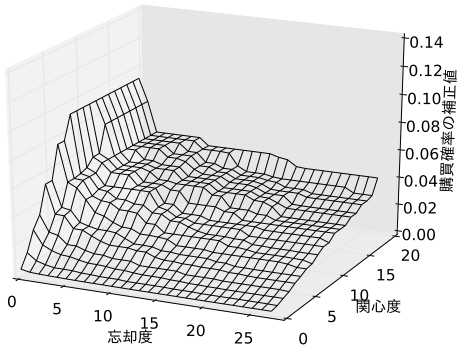

図 2: 単調性制約により補正した購買確率

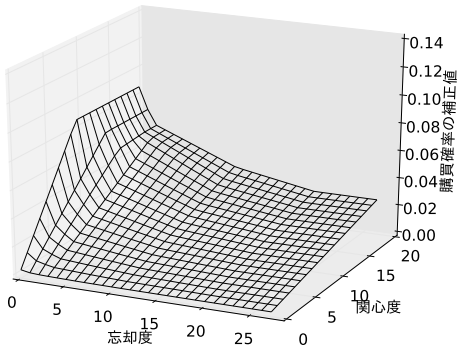

図 3: 単調性制約と凹/凸性制約により補正した購買確率

参考文献

[1] 岩永二郎, 鍋谷昴一, 梶原悠, 五十嵐健太: 関心度と忘却度に基づくレコメンド手法 —単調性制約付きレコメンドモデルの構築—. オペレーションズ・リサーチ, Vol.59, No.2, pp.72–80 (2014).

[2] 加藤直樹, 羽室行信, 矢田勝俊: データマイニングとその応用. 朝倉書店 (2009).