

# 時系列モデルを用いたインターネット広告の効果予測

本橋 永至 総合研究大学院大学 統計科学専攻 博士後期課程3年

## 1 はじめに

企業がインターネットを用いて消費者とコミュニケーションをとる手段は、バナー広告やモバイル広告等のインターネット広告、企業が直接所有するサイト、SNS (Social Networking Service) 等のソーシャル・メディアなど様々である。インターネットが十分に普及した今日、これらの手段を適切に利用することにより安価で効率的なマーケティングを展開することが可能になってきた。特に、国内のインターネット広告費は年々拡大し続けており、今やテレビ広告費に次ぐ規模となった。

インターネット広告に対する期待が高まるにつれ、その効果の予測および適切な効果測定を行うための手法が求められてきている。本研究では、バナー広告のクリック率を予測するモデルを状態空間モデルをベースに構築し、実際の配信データを用いてモデルの有用性を示す。状態推定には、あらゆる非線形もしくは非ガウス型の状態空間モデルに適用可能な粒子フィルタを用いる。

## 2 モデル

本研究で提案するモデルは、バナー広告がインターネット上に存在する多数のサイトに配信されることを前提とし、観測期間中の第  $t$  日にあるカテゴリーのサイトに配信されたバナー広告のクリック数  $y_t$  ( $t = 1, \dots, T$ ) は  $IMPS_t$ ,  $\pi_t$  をパラメータとする二項分布

$$y_t \sim \text{binomial}(IMPS_t, \pi_t) \quad (1)$$

に従うと仮定する。ここで、 $IMPS_t$ ,  $\pi_t$  はそれぞれ第  $t$  日に配信されたバナー広告のインプレッション数およびクリック率である。インプレッション数  $IMPS_t$  は広告主側で比較的容易に操作が可能なため所与として扱う。クリック率  $\pi_t$  は、トレンド成分  $\mu_t$ , 曜日成分  $w_t$ , 祝日効果項  $h_t$ , 誤差項  $v_t$  により説明できると仮定し、ロジスティック関数を用いて

$$\pi_t = \frac{\exp(\mu_t + w_t + h_t + v_t)}{1 + \exp(\mu_t + w_t + h_t + v_t)} \quad (2)$$

と定式化する。まず、トレンド成分  $\mu_t$  は、2階差分のトレンドモデルに従うと仮定し

$$\mu_t = 2\mu_{t-1} - \mu_{t-2} + \delta_t, \quad \delta_t \sim N(0, \sigma_\delta^2) \quad (3)$$

と定式化する。トレンド成分は、テレビや雑誌等の観測されない要因による長期的な変動を捉えていると解釈できる。また、クリック率は曜日によっても異なる傾向があると考えられる。週の周期変動を捉えるために、曜日成分  $w_t$  を周期7の季節成分モデル

$$w_t = -\sum_{j=1}^6 w_{t-j} + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma_\epsilon^2) \quad (4)$$

により表現する。さらに、祝日のクリック率は日曜のそれに近いだろうという期待から祝日効果項  $h_t$  を

$$h_t = I_t(w_{t,\text{sun}} - w_t) \quad (5)$$

と定式化する。ここで、 $I_t \in \{0, 1\}$  は第  $t$  日が月～金の祝日であれば1, そうでなければ0を取る2値変数,  $w_{t,\text{sun}}$  は直前の日曜の曜日成分である。つまり、第  $t$  日が月～金の祝日であれば、式 (2) の指数内部は  $\mu_t + w_{t,\text{sun}} + v_t$  となる。最後に、誤差項  $v_t$  は平均0, 分散  $\sigma_v^2$  の正規分布に従うと仮定する。

## 3 実証分析

### 3.1 データの概要

インターネットマーケティング事業者によって配信されたバナー広告のインプレッション数とクリック数がサイトごとに日単位で記録された

データを用いて実証分析を行う。観測期間は、2011年4月1日から9月30日までの6か月間である。

分析には、配信方法がリターゲティング広告 (RTA), インプレッション保証型広告 (CPM), 地域特定型広告 (AREA) の3つのパターンを使用する (表1)。いずれのパターンもインターネット上の様々なサイトに配信されているが、本研究ではインプレッション数が多い「パソコン」、「ゲーム」、「ニュース」、「スポーツ」のいずれかのカテゴリーに含まれるサイトを分析対象とする。

表1: 分析対象広告の要約

	広告カテゴリー	配信方法
パターン1	パソコン関連	RTA
パターン2	飲食品関連	CPM
パターン3	自動車関連	AREA

### 3.2 分析結果

本研究で提案したモデルの有用性を示すために、前半3か月間のデータをパラメータの推定期間とし、後半3か月間のデータをクリック率の予測精度の検証期間とする。クリック率の予測値にちょうど1週間前のCTR (実際のインプレッション数に対するクリック数の比率) を用いる方法を比較対象とし、本研究のモデルとこの方法の予測誤差を比較する。

表2は、本研究のモデルから得られた将来1週間の長期予測による予測値とCTRの観測値から計算された3か月分の予測誤差の標準偏差をパターンごとカテゴリーごとに表したものである (括弧内の数値は、比較対象から得られた予測誤差の標準偏差)。本研究のモデルから得られたすべての予測誤差が対応する比較対象よりも小さいため、すべてのデータにおいて本研究のモデルが比較対象よりも予測に優れていると言える。

表2: 予測誤差の標準偏差

	パソコン	ゲーム	ニュース	スポーツ
パターン1	0.000523 (0.000691)	0.000372 (0.000548)	0.000430 (0.000656)	0.000875 (0.001159)
パターン2	0.001290 (0.001599)	0.002618 (0.003171)	0.001316 (0.001843)	0.001438 (0.001744)
パターン3	0.000277 (0.000374)	0.000578 (0.000714)	0.001790 (0.002650)	0.001997 (0.002741)

図1は、パターン3の「ニュース」におけるCTRの観測値と本研究のモデルおよび比較対象から得られた予測値を時系列で表したものである。比較対象では、誤差の影響で予測が大きく外れているのが所々に見られるのに対して、本研究のモデルは、短期的な変動を必要以上に追従することなく長期的な傾向を適切に捉えることで精度の高い予測を実現している。

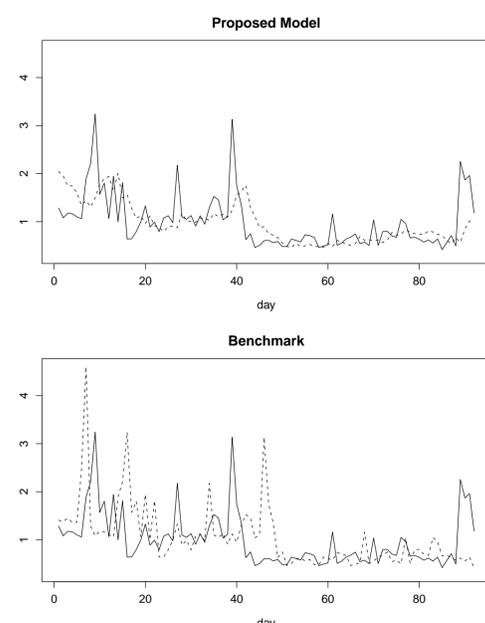


図1: CTRの観測値 (実線) と本研究のモデル (上段) および比較対象 (下段) から得られた予測値 (点線)。縦軸は、検証期間中のCTRの平均を基準とする比率。