2J3-3

インターネット広告のロバストなCVR推定モデル

Robust Prediction Model of Conversion Rate for Internet Advertisements

河本　哲
Satoshi Kawamoto

株式会社アイモバイル　技術本部
Technical Department, I-mobile Inc.

Predicting CVR of online advertisements is generally hard task because of its sparseness. If we can predict CVR accurately, advertisers will be able to deliver advertisements more efficiently. In this paper, we propose a Hierarchical Bayes Model whose hyperparameter have a constrained expected value calculated from Poisson Mixture Model. This paper also shows the robustness of proposed model compared to Logistic Regression Model.

1. はじめに

インターネット広告の広告効果を測定する指標のひとつにCVR(Conversion Rate) がある。コンバージョンとは、Web上で観測される成果 (例えば商品購入など) を示し、CVR は1クリックあたりのコンバージョンの確率を表したものである。

CVR を精度良く推定することで広告主は効率良く広告を配信することが出来、Web サイト (メディア) の運営者には成果に応じた適正な報酬を保障することが可能になる。しかし、一般的に CVR の推定はコンバージョン数のスパース性のため困難であることが多い。　本研究では、まずユーザー群の広告表示位置 (表示枠) における広告キャンペーンの CVR を確率命題を用いた混合ポアソン分布にて推定し、混合ポアソン分布にて推定された CVR を事前分布の期待値とする制約を設けた階層モデルによって広告 (クリエイティブ) の CVR を推定するモデルを示す。また CVR の推定に多く用いられるロジスティック回帰モデルと提案モデルを比較し、精度とロバスト性の比較を行う。

2. 関連研究

コンバージョンを予測するモデルとして多く提案されているのはロジスティック回帰モデルである [Lee 12]。ロジスティック回帰は計算面で高速であり学習の収束性も良いために多く用いられる。また Feature Hashing による次元圧縮 [Chappele 14] や、特徴量を工夫することで精度向上を図ったり [原 15]、ビジターの異質性をモデルに組み込む [山口 14] といった研究が行われている。

ロジスティック回帰は高速性および学習の収束性を両立しているが、説明変数間に多重共線性が存在する場合、推定精度が低下する課題がある。またロジット変換した際、説明変数と被説明変数にある程度の線形性がないと外れ値の推定精度が低くなる。

3. 提案手法

本章では CVR の推定に用いる特徴量および予測モデルについて述べる。まず最初に、ユーザー群 C が広告表示枠 s に表示されるキャンペーン c にてコンバージョンする確率を複数の特徴量を用い、弱い学習器で粗く推定する。粗く推定された CVR のリストは確率的命題変数として用いられる。確率命題の事後分布を求めて、群 C がキャンペーン c の広告枠 s にてコンバージョンする確率 $\pi(s,c)$ を求める。その後 $\pi(s,c)$ を制約とした階層的な推定により群 C が広告枠 s で、広告クリエイティブ a にてコンバージョンする確率を求める。以降 3.1〜3.5 に、具体的なモデルを示す。

連絡先: 河本　哲，株式会社アイモバイル　技術本部，
Email:kawamoto@i-mobile.co.jp

3.1 特徴量

本研究では、ユーザー群 C のクリック特性 $\boldsymbol{X}$=$(x_0,x_1,x_2,x_3,x_4,x_5,x_6)$ およびコンバージョン特性 $\boldsymbol{Y}$=$(y_0,y_1,y_2,y_3,y_4,y_5,y_6)$ を特徴量として用いる。特徴量の各要素は以下の通りである。いずれも、あるユーザー群 C 内での観測とする。

x_0：枠 s にて c をクリックした回数
x_1：c をクリックした回数
x_2：キャンペーンのカテゴリ Γ をクリックした回数
x_3：枠 s にて Γ をクリックした回数
x_4：枠 s にてクリックした回数
x_5：クリック回数の総数
x_6：枠 s にて a をクリックした回数
y_0：枠 s にて c でコンバージョンした回数
y_1：c でコンバージョンした回数
y_2：Γ でコンバージョンした回数
y_3：枠 s にて Γ でコンバージョンした回数
y_4：枠 s にてコンバージョンした回数
y_5：コンバージョン回数の総数
y_6：枠 s にて a でコンバージョンした回数

3.2 期待値の制約を付与した事前ベータ分布の導入

ベータ分布 $Beta(\alpha,\beta)$ は二項分布に対する共役な事前分布であり、例えばコインを x 回投げそのうち y 回表が出たとき、表が出る確率の事後分布は $Beta(\alpha+y,\beta+x-y)$ となる。ここで、事前分布 $Beta(\alpha,\beta)$ の期待値は π であると仮定すると、$\frac{\alpha}{\alpha+\beta}=\pi$ という制約が付与される。よって、$\alpha=\alpha_0,\beta=\beta_0,E(Beta(\alpha_0,\beta_0))=\pi$ が与えられたとき、表が出る確率の推定値 ζ は

$$
\begin{aligned}
E(\zeta|\beta=\beta_0) &= \frac{\alpha_0+y}{\alpha_0+\beta_0+x} \\
&= \pi+\frac{(1-\pi)y-\pi(1-\pi)x}{\beta_0+(1-\pi)x} \qquad (1)
\end{aligned}
$$

となり、これを周辺化すると

$$
\begin{aligned}
E(\zeta) &= \int_A^B E(\zeta|\beta=\beta_0)\cdot f_\beta(\beta_0)d\beta_0 \\
&= \int_A^B \left\{\pi + \frac{(1-\pi)y - \pi(1-\pi)x}{\beta_0+(1-\pi)x} + \right\} \frac{1}{B-A} d\beta_0 \\
&= \pi + \frac{(1-\pi)y-\pi(1-\pi)x}{B-A}\log\left(\frac{B+(1-\pi)x}{A+(1-\pi)x}\right) \quad (2)
\end{aligned}
$$

となる。但し β は一様分布 $U(A,B)$ に従うものとする。この考え方を確率命題の作成および群 C の s,a における CVR 推定に用いる。

3.3 確率命題の作成

群 C が s,c でコンバージョンする確率は、確率命題 $\boldsymbol{\theta}$=$(\theta_0,\theta_1,\theta_2,\theta_3,\theta_4)$ を用いた混合ポアソン分布によって推定されるとする。$\theta_i (i=0,1,2,...,4)$ をそれぞれ次のように定義する。

$\theta_0 = E(\zeta_0)$
$\theta_1 = E(\zeta_1)$
$\theta_2 = E(\zeta_3)\cdot\frac{E(\zeta_1)}{E(\zeta_2)}$
$\theta_3 = (予定\ CVR)$
$\theta_4 = E(\zeta_4)\cdot\frac{E(\zeta_1)}{E(\zeta_5)}$

但し、

$$
E(\zeta_i) = \pi_0 + \frac{(1-\pi_0)y_i - \pi_0(1-\pi_0)x_i}{B_0-A_0} \times\log\left(\frac{B_0+(1-\pi_0)x_i}{A_0+(1-\pi_0)x_i}\right) \quad (3)
$$

とする。ここに、一様分布のパラメータとして $A_0 = 1.0, B_0 = 10.0$ を与える。また、確率命題は $\theta_0,\theta_1,\theta_2,\theta_4$ を周辺化すると $E(\zeta_1)$ となるように設計されており、s,c の CVR の弱い推定値となっている。

周辺化すると $E(\zeta_1)$ を満たすような確率命題変数を作成することにより、枠 s に特徴量がほとんど無い状態であっても、ロバストな CVR 推定がなされるようになる。即ち特徴量が希薄な枠 s については $E(\zeta_1)$ の状態に応じて $\pi(s,c)$ が調整され、キャンペーン全体の CVR 推定の整合性が保たれる。但し、$0 < \theta_i \leq 0.5$ となるように制限を掛けている。

3.4 混合分布による群 C の s,c における CVR 推定

3.3 で定義した確率命題 θ_i が事前分布 $w(\theta_i)$ に従うとき、事後分布は次のような混合ポアソンモデルにて求められるとする。

$$
\begin{aligned}
w(\theta_i|x_0,y_0) &= \frac{w(\theta_i)\cdot w(x_0,y_0|\theta_i)}{\sum_{j=0}^{4} w(\theta_j)\cdot w(x_0,y_0|\theta_j)} \\
&= \frac{w(\theta_i)\cdot\frac{\lambda_i^{y_0}}{y_0!}\cdot\exp(-\lambda_i^{y_0})}{\sum_{j=0}^{4} w(\theta_j)\cdot\frac{\lambda_j^{y_0}}{y_0!}\cdot\exp(-\lambda_j^{y_0})} \quad (4)
\end{aligned}
$$

ただしポアソンパラメータ λ_i は $\lambda_i = x_0\theta_i$ とし、事前分布には下記の値を与える。

$$
w(\theta_i) = \frac{\exp\left(-\frac{(\theta_i-\hat{\mu})^2}{2\hat{\sigma}^2}\right)}{\sum_{j=0}^{4}\exp\left(-\frac{(\theta_j-\hat{\mu})^2}{2\hat{\sigma}^2}\right)} \quad (5)
$$

ここに、$\hat{\mu} = \bar{\theta}, \hat{\sigma}^2 = \frac{1}{4}\sum_{j=0}^{4}(\theta_i - \hat{\mu})^2$ とする。事後分布 $w(\theta_i|x_0,y_0)$ を用いると、群 C が広告枠 s、キャンペーン c にてコンバージョンする確率 $\pi(s,c)$ は

$$
\pi(s,c) = E(\boldsymbol{\theta}|x_0,y_0) = \sum_{j=0}^{4} w(\theta_i|x_0,y_0)\cdot\theta_i \quad (6)
$$

となる。

3.5 群 C の s,a における CVR の推定

群 C が s,a でコンバージョンする確率 $\pi(s,a)$ は $\pi(s,c)$ を用いて下式で推定される。

$$
\pi(s,a) = \pi(s,c) + \frac{(1-\pi(s,c))y_6 - \pi(x,c)(1-\pi(s,c))x_6}{B_1 - A_1} \times\log\left(\frac{B_1+(1-\pi(s,c))x_6}{A_1+(1-\pi(s,c))x_6}\right) \quad (7)
$$

パラメータ A_1, B_1 にはそれぞれ $A_1 = 1.0, B_1 = 100.0$ を与える。このモデルを下記のロジスティック回帰モデルと比較し、CVR の推定精度比較を実施する。

$$
\log\left(\frac{\pi(s,a)}{1-\pi(s,a)}\right) = \boldsymbol{\gamma^T Z} \quad (8)
$$

この式を $\pi(s,a)$ について解くと

$$
\pi(s,a) = \frac{1}{1+\exp(-\boldsymbol{\gamma^T Z})} \quad (9)
$$

となる。ここに、$\boldsymbol{Z}$=$100\times((E(\zeta_0), E(\zeta_1), ..., E(\zeta_6)))$ なる特徴量とし、係数 $\boldsymbol{\gamma}$ は確率的勾配法で学習させる。また、ロジスティック回帰による CVR の推定値 $\pi(s,a)$ は $0.001 \leq \pi(s,a) \leq 0.5$ となるように制限を加えた。

4. 評価実験

4.1 評価実験用データ条件

本研究における評価実験には、株式会社アイモバイルにおける 2015 年 11 月 1 日〜2015 年 11 月 8 日の期間における各広告のクリックおよびコンバージョンのデータを使用した。また、訓練データとして 11 月 1 日〜11 月 7 日のデータを用い、テストデータには 11 月 8 日のデータを用いた。

4.2 広告キャンペーン全体の CVR 予測性能の評価

3.5 にて求められた $\pi(s,a)$ および、ロジスティック回帰モデルで求められた CVR の推定値の予測精度を比較する。予測精度の評価には LogLoss を用いる。LogLoss は下記の式で与えられる。

$$
\mathrm{LogLoss} = -\frac{1}{N}\sum_{i=1}^{N}(y_i\log(\hat{y}_i) + (1-y_i)\log(1-\hat{y}_i)) \quad (10)
$$

ここに N はテストデータにおける全キャンペーンの、全広告枠における総クリック数であり、y_i はコンバージョンに至ったクリックの場合 1, コンバージョンに至らないクリックの場合は 0 とする。また、$\hat{y}_i$ は CVR の推定値である。LogLoss の差は以下の通りとなった。

表 1: 表 1:LogLoss の性能比較

評価モデル	LogLoss	N
提案モデル C_0	0.0394597128565462	3986217
提案モデル C_1	0.0617590371161231	142020
ロジスティック回帰 C_0	0.0469903961224792	3986217
ロジスティック回帰 C_1	0.0893563755247816	142020

ここに C_0 を広告主サイト未訪問のユーザー群とし、C_1 を広告主サイト既訪問ユーザー群とする。表 1 に示されている通り、提案モデルにおける LogLoss は C_0, C_1 いずれのユーザー群に対してもロジスティック回帰モデルによる推定値を用いた LogLoss よりも小さく、全体的な CVR の推定精度が高いことを示している。

4.3 異常値に関するロバスト性の評価

4.2 にて提案モデルを用いた際、全体的な CVR の推定精度の向上が得られることを示した。しかし全体的な推定精度を向上させるとともに、$\pi(s,a)$ が異常値となってしまうことを防止することも重要な課題である。ここで、異常値とはテストデータ (例えば翌日のデータなど) の枠 s におけるクリック数が十分あるにも関わらずテストデータにおける CVR の実値と推定された CVR が大幅に異なる現象が発生してしまい、広告主あるいはメディアの運営者に裁定機会が生じていることと定義する。つまり (推定 CVR)≫(実 CVR) となるような推定をした場合は、広告主側は過剰な単価での入札をして損失が発生してしまい、(推定 CVR)≪(実 CVR) となる推定をした場合は、メディア側には適正な報酬が支払われなくなってしまう。

本研究では、前述の提案モデルとロジスティック回帰モデルの異常値の発生確率の比較評価も実施した。異常値の判断基準として、テストデータにおける有意水準 5%の信頼区間を算出し、CVR の推定値が信頼区間外の値であった場合は異常値であるとみなした。

ここで信頼区間の算出には Clopper&Pearson の信頼区間を用いた。Clopper&Pearson の信頼区間とは、下記の式で定義される二項分布の精密な信頼区間である。

$$\left(\frac{n_2}{n_1 F^{n_1}_{n_2}\left(\frac{\epsilon}{2}\right) + n_2}, \frac{n'_1 F^{n'_1}_{n'_2}\left(\frac{\epsilon}{2}\right)}{n'_1 F^{n'_1}_{n'_2}\left(\frac{\epsilon}{2}\right) + n'_2} \right) \tag{11}$$

$n_1 = 2(n-k+1), n_2 = k, n'_1 = 2(k+1), n'_2 = 2(n-k)$ とする。但し、n をテストデータでのクリック回数、k をテストデータでのコンバージョン件数とする。表 2 に、提案モデルとロジスティック回帰モデルでの検定の結果を示す。ただし、M はクリックの発生した枠と広告の組合せの数とする。

表 2 に示す通りユーザー群 C_0, C_1 のいずれにおいても、提案モデルを用いた異常値の発生確率はロジスティック回帰モデルよりも低減されている。即ち全体的な CVR の推定精度を維持しつつ、異常値が発生する確率の低減も実現できていることが示された。

表 2: 表 2:異常値発生確率の比較

評価モデル	異常値	M
提案モデル C_0	215	852121
提案モデル C_1	2	90423
ロジスティック回帰 C_0	1860	852121
ロジスティック回帰 C_1	104	90423

5. 結論

本研究では、ユーザー群 C が s, c でコンバージョンする確率 $\pi(s,c)$ を確率命題の混合ポアソン分布で表現し、$\pi(s,c)$ の期待値を制約条件としたベータ分布に従う事前分布から $\pi(s,a)$ を推定するモデルを提案した。その際、ベータ分布のパラメータ β を一様分布に従うハイパーパラメータと設定することで、高いロバスト性と $\pi(s,a)$ を陽に表現することを両立させた。

また、ロジスティック回帰モデルと比較して異常値の発生確率が低減されていることも示した。今後は、ハイパーパラメータの適正化、CVR の非斉次な現象のモデル化および、ユーザー群分類を精緻化した提案モデルの有効性の検証などを課題として考えていきたい。また、3.3 にて示した θ_i は手動で作成したモデルであるが、確率命題の自動抽出なども検討すべき課題である。

参考文献

[原 15] 原 淳史, 高野 雅典,Roman Shtykh, 川端 貴幸: インターネット広告におけるコンバージョンに近いユーザの抽出方法の検討, 人工知能学会全国大会（第 29 回）論文集 (2015)

[宮西 15] 宮西　一徳, 高野 雅典, 吉田　岳彦: 大規模リワード広告システムにおける行動履歴と広告属性を利用したコンバージョン予測モデルの構築, 人工知能学会全国大会（第 29 回）論文集 (2015)

[山口 14] 山口 景子：頻度の時間変化を考慮した階層ベイズモデルによるウェブサイト訪問行動の分析, マーケティング・サイエンス, Vol. 22, No. 1, pp. 13-29 (2014)

[本橋 12] 本橋　永至, 磯崎　直樹, 長尾　大道, 樋口　知之:状態空間モデルによるインターネット広告のクリック率予測, オペレーションズ・リサーチ : 経営の科学 57(10) (2012)

[Zhang 15] Weinan Zhang, Jun Wang: Statistical Arbitrage Mining for Display Advertising,in *Proc. of KDD* 2015, pp.1465-1474(2015)

[Lee 12] Kuang-chih Lee, Burkay Orten, Ali Dasdan, Wentong Li: Estimating Conversion Rate in Display Advertising from Past Performance Data ,in *Proc. of KDD* 2012, pp.768-776(2012)

[Chappele 14] Olivier Chappele, Eren Manavoglu, Romer Rosales: Simple and scalable response prediction for display advertising, WSDM(2014)

[Zhang 14] Weinan Zhang, Shuai Yuan, Jun Wang: Optimal Real-Time Bidding for Display Advertising,in *Proc. of KDD* 2014 pp.1077-1086(2014)