

インターネット広告におけるスパースなユーザー行動ベクトルからのユーザー特徴抽出モデル

Users' Feature extraction model from their online behavior composed of sparse vectors for the use of internet advertising.

河本 哲 *1 秋光 淳生 *2
Satoshi Kawamoto Toshio Akimitsu

*1株式会社アイモバイル 技術本部, 放送大学大学院文化科学研究科文化科学専攻

Technical Department, I-mobile Inc. Division of Arts and Sciences, The School of Graduate Studies, The Open University of Japan

*2放送大学大学院文化科学研究科文化科学専攻

Division of Arts and Sciences, The School of Graduate Studies, The Open University of Japan

Finding out users who have high conversion probability is crucial task for internet advertising. Because advertisers' budgets are limited and they have to increase their profits efficiently. But users' behavior consists of a sparse vector and has high dimensionality. So general clustering or regression model doesn't work efficiently in many cases without any feature extraction. We propose a model that uses Autoencoder tolerating missing values and evaluated efficiency of extracted feature in this model.

1. はじめに

インターネット広告を限りのある予算内で配信することを考える際、配信効果が高いと見込まれるユーザーを配信対象とすることで広告費対効果が改善する。そのため、ユーザーの広告に対する成果要因となる特徴を獲得することは重要な課題となっている。

ユーザーの Web 上における行動特徴として考えられるものは、例えば「どの広告をクリックしたか」「どのサイトに訪問したか」などがある。しかしこれらの特徴は一般的に非常に高次元であり、かつスパース性の高いデータである。高次元で疎な Web 上のユーザー行動データを低次元化し、ユーザー特徴を獲得することが出来れば広告配信に効果的なユーザーを抽出することが可能になる。

本研究では、ユーザーの行動ベクトルのゼロ成分を欠損値とみなし、結合荷重の更新を一部スキップしたオートエンコーダを用いて、ユーザー特徴を効率的に獲得をするモデルを提案する。また獲得した特徴量をもとにユーザー群をクラスタリングし、分類されたユーザー群の特性についても評価を行う。

2. 関連研究

インターネット広告におけるユーザーの行動特性を推定する手法のひとつとして、ユーザーの Web 上の行動履歴から DBN によって抽出された特徴量を説明変数としてロジスティック回帰により CTR(クリック確率)を推定するモデルが提案されている [Jiang 16]。高次元の入力から特徴を獲得するモデルにはオートエンコーダや制限ボルツマンマシンなどがあり、また、ユニットをエージェントとみなしデノイジングオートエンコーダを再構成したモデルも提案されている [黒滝 16]。またショッピングサイトのユーザーの離脱防止の施策にフォーカスを当て、検索ワードや注文カテゴリ及び注文履歴からユーザーの特性把握を試みる研究も行われている [笹谷 16]。

3. ユーザーの特徴獲得モデル

3.1 提案手法

本章では Web 上のユーザーが保持しているデータと、ユーザーの特徴獲得に用いるモデルについて述べる。各ユーザーは広告のクリック回数およびコンバージョン回数の特徴量として保有している。コンバージョンとは商品購入やゲームのインストールなどの Web 上で測定可能な広告成果イベントのことであり、コンバージョン件数をクリック件数で割った値をコンバージョンレート (CVR) と呼ぶ。インターネット広告のクリックおよびコンバージョンの特性として、高次元でスパース性が高いことが挙げられる。すなわち、広告数は多いがユーザーがクリックおよびコンバージョンする広告は限定的であり、その回数も少ない。高次元なベクトルや行列をそのまま用いても有効な特徴量として機能せず、また計算時間上のデメリットも大きい。本稿ではユーザーがクリックおよびコンバージョンをしなかった広告を欠損値として扱い、欠損値に対して一定確率で明示的にゼロを与えたベクトルをオートエンコーダへの入力値として与え、有効なユーザー特徴を抽出するモデルを提案する。以降に具体的な手法を示す。

3.2 変数定義

本研究ではユーザー v が広告 a に対して行ったクリックおよびコンバージョン回数をデータとして与える。本稿で用いる変数は下記の通りである。

v : ユーザー番号

N : クリック実績のある広告数

M : コンバージョン実績のある広告数

a : クリックの発生した広告番号

b : コンバージョンの発生した広告番号

$x_{v,a}$: ユーザー v が広告 a をクリックした回数

$y_{v,b}$: ユーザー v が広告 b にてコンバージョンした回数

X_v : ユーザー v の特徴ベクトル

i : 3 層ニューラルネットワークの入力層のユニット番号

j : 3 層ニューラルネットワークの中間層のユニット番号

k : 3 層ニューラルネットワークの出力層のユニット番号

$w_{A,B}$: ユニット A, B 間の結合荷重 ($(A, B) = (i, j) \text{ or } (j, k)$)

z_A : ユニット A のバイアス項 ($A = i, j, k$)

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

h_A : ユニット A への入力 ($A = i, j, k$)

g_A : ユニット A の出力 ($A = i, j, k$)

$f_A(x)$: ユニット A の活性化関数 ($A = i, j, k$)

J : 誤差関数

α : 学習率の初期値

ここで、 X_v は下記の通りに定義される特徴ベクトルとする。

$$X_v = (x_{v,1}, x_{v,2}, \dots, x_{v,a}, \dots, x_{v,N}, y_{v,1}, y_{v,2}, \dots, y_{v,b}, \dots, y_{v,M})$$

3.3 スパースデータからの特徴獲得モデル

高次元なデータの次元を削減し、特徴を獲得するモデルのひとつにオートエンコーダが存在する。オートエンコーダは入力層に与えるデータと出力層の正解データに同じ値を与えたニューラルネットワークである。また中間層のユニット数を入力層のユニット数よりも少なくすることで、中間層に特徴量を蓄積することが可能になる。

ニューラルネットワークの課題のひとつとして、ユニット数が増えれば、ニューラルネットワークの学習処理に掛かる計算時間が大きくなる問題が挙げられる。一般的に Web 上の広告数およびユーザー数は非常に多いため (本研究では表 1 に示す数値となった)、効率的な学習モデルによる高速な計算が求められる。そこで X_v のスパース性を利用し、 X_v のゼロ成分を確率的に欠損値とみなして計算処理をスキップさせ、オートエンコーダの学習の高速化をさせるモデルを提案する。すなわち、入力値がゼロの成分のユニットの結合がランダムで存在しないものとみなす。具体的には図 1 に示すように「0(NA)」となっている箇所の結合が存在しないものとみなし、太線の箇所のみ結合を用いた学習処理を実行する。「0(NA)」の箇所は、学習ステップ毎にランダムで変更される。

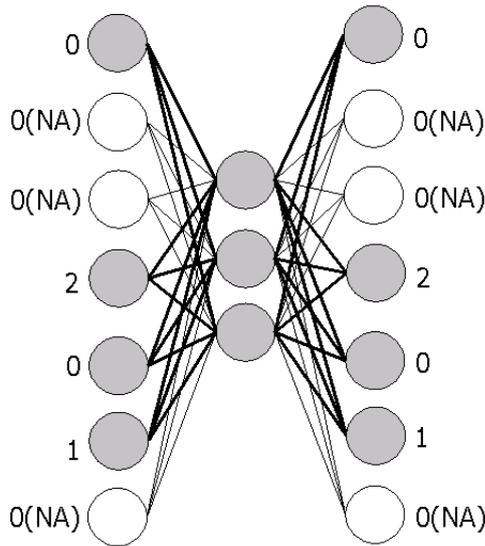


図 1: スパースデータ許容型オートエンコーダ

本研究では入力層と中間層の活性化関数はシグモイド関数を用い、出力層の活性化関数には恒等写像を用いる。すなわち、下式の通りに表現される。

$$f_i(x) = f_j(x) = \frac{1}{1 + \exp(-x)} \quad (1)$$

$$f_k(x) = x \quad (2)$$

3.4 スパースデータ許容型オートエンコーダの学習処理

本研究で用いるモデルも通常のニューラルネットワークと同じく、入力されたデータは、各層のユニットに伝播されており、最終的に出力層へと伝播する。入力層のユニットの出力 g_i は

$$g_i = \begin{cases} f_i(x_{v,i} + z_i) & (\text{when } i \leq N) \\ f_i(y_{v,i-N} + z_i) & (\text{otherwise}) \end{cases} \quad (3)$$

となる。また入力が欠損値であるノードと連結しているエッジとの結合荷重は一時的にゼロとみなす。結合荷重が一時的にゼロとみなされる箇所は計算をスキップ出来るため、学習の高速化が実現される。このモデルの仮定として、

$$h_j = \sum_i w_{ij} g_i = \sum_{i'} w_{i'j} g_{i'} \quad (4)$$

が成り立つとしている。ただし $i = 1, 2, \dots, N + M$ であり、 i' は i から欠損値の発生しているユニット番号を除外した番号である。すなわち入力層から中間層に入力が順伝播する際に、欠損値の情報が中間層に伝わる値は合計すると 0 になると仮定していることになる。

またオートエンコーダの二乗和誤差関数は

$$J = \frac{1}{2} \left\{ \sum_{k=1}^N (g_k - x_{v,k})^2 + \sum_{k=N+1}^{N+M} (g_k - y_{v,k-N})^2 \right\} \quad (5)$$

となるが、欠損値の発生しているユニット番号に関しては誤差が発生していないと仮定している。すなわち、

$$\begin{cases} g_k - x_{v,k} = 0 & (\text{when } x_{v,k} = \text{NA}, k \leq N) \\ g_k - y_{v,k-N} = 0 & (\text{when } y_{v,k-N} = \text{NA}, k > N) \end{cases} \quad (6)$$

とみなしている。結合荷重を一時的にゼロとみなしているので、誤差観測自体されないと言い換えることも出来る。

以上の仮定をもと、ニューラルネットワークの結合荷重を更新する。ニューラルネットワークの結合荷重の更新に用いる勾配は下記の式で表現される。

$$G_{ij}(t) = \sum_{k=1}^N \{(g_k - x_{v,k}) f'_k(h_k) w_{jk}\} f'_j(h_j) g_i + \sum_{k=N+1}^{N+M} \{(g_k - y_{v,k-N}) f'_k(h_k) w_{jk}\} f'_j(h_j) g_i \quad (7)$$

$$G_{jk}(t+1) = \begin{cases} (g_k - x_{v,k}) f'_k(h_k) g_j & (\text{when } k \leq N) \\ (g_k - y_{v,k-N}) f'_k(h_k) g_j & (\text{when } k > N) \end{cases} \quad (8)$$

よって、学習率が一定の場合は結合荷重の更新式は下記の通りとなる。

$$w_{ij}(t+1) = w_{ij}(t) - \alpha G_{ij}(t) \quad (9)$$

$$w_{jk}(t+1) = w_{jk}(t) - \alpha G_{jk}(t) \quad (10)$$

ここで一般的に学習率 α には小さな値を用いるが、固定の学習率を与えても学習の収束性は良くない。そこで本研究では、学習の進行とともに学習率を低減させる AdaGrad を用いて学習を行う。AdaGrad は、下記の式で表現される学習モデルである。

$$w(t+1) = w(t) - \frac{\alpha}{\sqrt{1 + \sum_{i=1}^t G(t)^2}} G(t) \quad (11)$$

ここで、 $w(t)$ は学習ステップ t における結合荷重であり、 $G(t)$ はステップ t における勾配である。

4. ユーザー特徴の獲得実験

4.1 シミュレーション概要

スパースな入力を許容したオートエンコーダにてユーザーベクトルの次元を圧縮し、結合荷重と誤差関数の変遷を概観する。それにより、ユーザー特徴を獲得しつつ次元削減されているかどうかを判断する。本研究では、株式会社アイモバイルの2016/6/1~2016/6/21の期間におけるクリックおよびコンバージョンのデータをスパースデータ許容型オートエンコーダの訓練データとして用いた。

4.2 シミュレーションの構成

シミュレーションに用いた訓練データの概要は表1の通りである。本研究では、集計期間中に何らかの広告を1回以上クリックしたユーザーを解析の対象としている。また、ユーザーベクトル X_u のゼロ成分の内ランダムで10成分のみ明示的にゼロ(非欠損)とみなし、残りのゼロ成分は欠損値とみなした。また、学習の繰り返し回数は10回とした。

表1: 表1:訓練データの統計量

項目	訓練データ
N	23339
M	5619
クリック総数	5644230
コンバージョン総数	32020
ユーザー数	3452491

活性化関数は入力層と隠れ層(中間層)にはシグモイド関数を用い、出力層には恒等写像 $y = x$ を用いた。

また学習率の初期値 α は0.01とし、結合荷重の初期値は一樣分布 $U(-0.01, 0.01)$ に従う乱数、バイアス項の初期値は一樣分布 $U(-0.51, -0.49)$ に従う乱数とした。

4.3 結合荷重の変遷

表1の訓練データでは $N = 23339, M = 5619$ であり、入力層のユニット数は28958となる。この訓練データをスパースデータ許容型オートエンコーダにより500次元にまで圧縮し(第1段階)、500次元のユーザー特徴をさらに通常のオートエンコーダにより146次元にまで圧縮(第2段階)した。

図2には28958次元から500次元に圧縮した際の、入力層と隠れ層間の結合荷重のヒストグラムを示した。図2にて示される通り、学習により一部の結合荷重が大きくなり、ユーザー特徴を獲得しながら次元の圧縮が行われていることを示している。図3には500次元から146次元に圧縮した際の、入力層と隠れ層の間の結合荷重のヒストグラムを示している。結合荷重の初期値が $U(-0.01, 0.01)$ に従っていたため、初期の結合荷重が並行移動しているような挙動を示していることが分かる。つまり、ユーザーの特徴は28958次元から500次元に圧縮されるときに獲得済みであり、500次元から146次元に圧縮されるときは情報圧縮のみが行われていることになる。

4.4 対数誤差関数の変遷

図4には28958次元から500次元に圧縮したときの誤差関数の変化を示した(グラフでは見やすさのため、二乗と誤差関数の対数を縦軸としている)。図から分かる通り、トリビアルな入力データに対して誤差関数が大きくなっている。また5,000,000データほど与えると、おおよそ学習が収束することが分かる。学習に用いたデータのユーザー数は3452491件であり、同じデータセットを繰り返し与えることで誤差関数が小

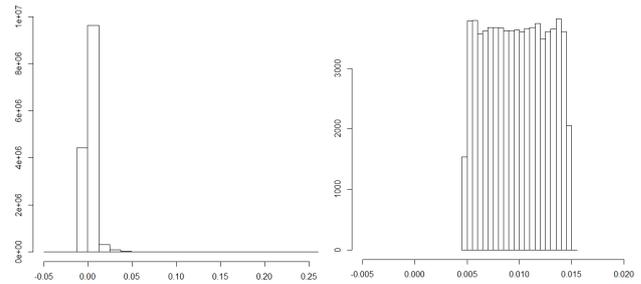


図2: 結合荷重(第1段階)

図3: 結合荷重(第2段階)

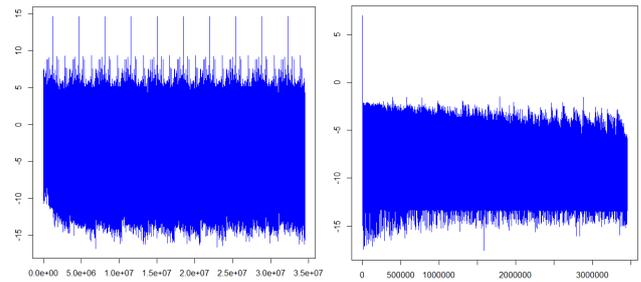


図4: 対数誤差(第1段階)

図5: 対数誤差(第2段階)

さくなっていくことが分かる。但し、(学習率などにも依存するが)3回以上同じデータセットを与えて繰り返し学習させることに大きな意味はなさそうだ。

図5には、500次元に圧縮されたユーザー特徴を146次元に圧縮する過程における誤差関数の変遷になる(こちら誤差関数に對数を掛けたものを縦軸にプロットしている)。学習が進むにつれ誤差関数は低減する傾向にあり、誤差の振れ幅も低下している。これは情報劣化を抑えながら500次元から146次元へとユーザー特徴が圧縮されていることを示している。

5. 獲得ユーザー特徴のCVRを用いた評価

5.1 クラスタリングによるユーザー群のCVRの評価

4.1~4.4で示したオートエンコーダによる学習により、ユーザー特徴を獲得出来た可能性が示唆された。獲得特徴の妥当性を評価するため、獲得特徴量を標準化したベクトルを用いてユーザー群をK-means法によりクラスタリングし、分類されたユーザー群のCVRを評価した。ユーザー群は50分類され、訓練データ(2016/6/1~6/21)におけるCVRとテストデータ(2016/6/22~2016/7/12)におけるCVRをプロットして獲得特徴の妥当性を評価した。

図6はクラスタリングされたユーザー群の訓練データでのCVR(横軸)及びテストデータでのCVR(縦軸)をプロットしている。横軸は「訓練データにおいて各ユーザー群が何らかの広告でコンバージョンをした確率」を示している。縦軸は「テストデータにおいて各ユーザー群が何らかの広告でコンバージョンした確率」を示している。

図6で分かる通り、訓練データでCVRが高かったユーザー群はテストデータにおいてもCVRが高かった。即ち、「過去にインターネット広告経由で成果を挙げたユーザー群はこれからも何らかのインターネット広告で成果を挙げやすい」といえる。つまり、広告効果の出やすいユーザー群と広告効果の出に

くユーザー群をクリックとコンバージョン特徴にて、分離可能であることが示される。また、図 6 の左上には訓練データにて CVR が低く、テストデータで CVR の高いユーザー群がいる。これは広告に対する認知が発生したユーザー群が存在している可能性を示唆している。但し、図 6 において示されるのは「何らかの広告でコンバージョンするユーザーであるかどうか」の情報を示しているに過ぎない。そのため、ある特定の広告キャンペーンで CVR の高いユーザー群が、今後も CVR の高い状態を維持するかどうかまでは分からない。

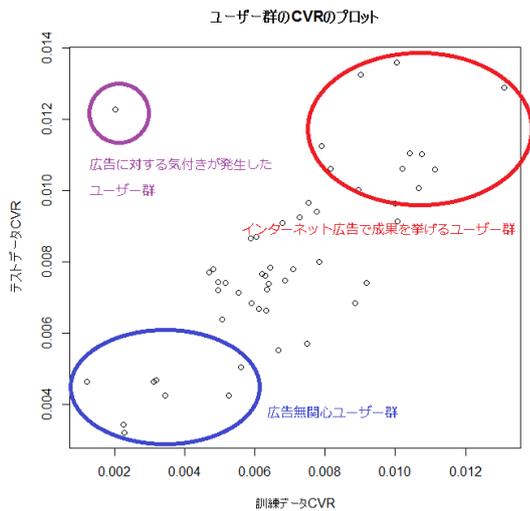


図 6: ユーザー群の CVR 特性

そこで、ある特定の広告キャンペーンについて各ユーザー群ごとの訓練データにおける CVR とテストデータにおける CVR をプロットした (図 7)。青の縦線は訓練データにおける全ユーザー群の CVR である。また、緑の横線は訓練データの集計期間内にクリック行動を示したユーザー群の、テストデータにおける CVR である。赤色の横線は訓練データの集計期間内にクリック行動を示さなかったユーザー群の、テストデータにおける CVR である。

図 7 から分かる通り、訓練データにおいて CVR の高かったユーザー群がテストデータでも CVR が高いとは限らないことが示される。それどころか、訓練データでクリック実績のあったユーザー群はテストデータにおいて、CVR が低くなる可能性があることが分かる。

この現象を説明する仮説は下記の 2 つがある。ひとつは、コンバージョンはゲームのインストールや商品購入などのイベントであり、一度コンバージョンしたユーザーが短い期間で再度コンバージョンする確率が低くなる可能性がある。即ちコンバージョンは非復元抽出を前提としたモデルを用いたほうが再現性が高くなる可能性がある。もうひとつは、訓練データの集計期間内で特定の広告をクリックしたユーザー群は当該キャンペーンを既に認知している状態になっているため、同じ広告を見せられてもコンバージョンに至る可能性が低くなっている可能性がある。

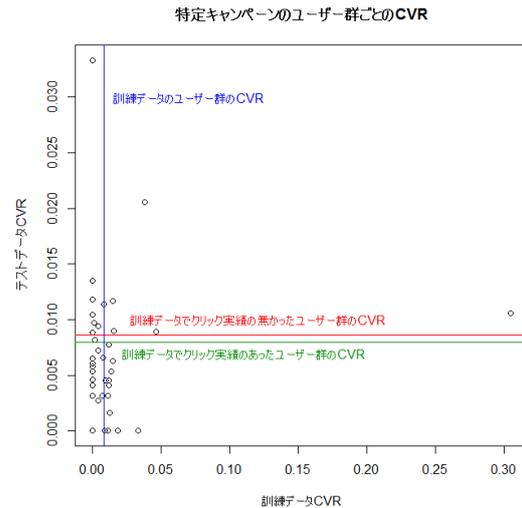


図 7. 特定のキャンペーンにおけるユーザー群の CVR 特性

6. 結論

本研究ではユーザーのクリックおよびコンバージョンデータからスパースなベクトルを生成して、スパースな入力を許容するオートエンコーダを用いてユーザー特徴の獲得を試みた。オートエンコーダに与えるベクトルに欠損値の存在を許すことで、高速にユーザーの特徴を獲得するモデルを実現出来た。また獲得特徴を用いたクラスター分析によりユーザーの過去のクリックおよびコンバージョンを用いて、CVR の高いユーザー群を分離抽出することが可能であることが分かった。

今後の課題として、図 6 の「広告に対する気付き」が発生した可能性があるユーザー群に対して、気付きを発生させる要因の発見と定量的な評価が挙げられる。また、クリックおよびコンバージョン以外のデータを広く用いたユーザーの特徴獲得も課題となる。また図 6 にて示された通りインターネット広告で成果を挙げやすいユーザー群をある程度特定すること自体は可能だが特定の広告キャンペーンにて成果を挙げやすいユーザー群も予測出来るモデルを構築することが今後の課題となる。

参考文献

- [Jiang 16] Zilong Jiang, Shu Gao, Wei Dai : Research on CTR Prediction for Contextual Advertising Based on Deep Architecture Model, in Journal of Control Engineering and Applied Informatics (Vol.18, No.1, pp.11-19, 2016)
- [田頭 16] 田頭 幸浩, 小林 隼人, 小野 真吾, 田島 玲 : オンライン広告におけるウェブ閲覧系列の分散表現の獲得, 人工知能学会全国大会 (第 30 回) 論文集 (2016)
- [黒滝 16] 黒滝 紘生, 中山 浩太郎, 松尾 豊 : マルチエージェントによる Deep Learning の提案, 人工知能学会全国大会 (第 30 回) 論文集 (2016)
- [笹谷 16] 笹谷 奈翁美, 坪内 孝太, 田代 昭悟, 鍛冶 伸裕, 清水 伸幸 : ショッピングサイトにおける優良顧客の離脱防止施策について, 人工知能学会全国大会 (第 30 回) 論文集 (2016)