

1. 研究の背景

1994 年にオンラインマガジン HotWired にてバナー広告が誕生して以来、情報技術の発展とともにコンテンツマッチ配信や Web 訪問履歴などを利用したターゲティング配信といった様々な広告配信手法が発展してきた。

特にユーザーの趣味嗜好に合わせた広告配信が可能になるターゲティング広告は、広告効果の高い配信手法として注目されているが、その実現のためにはユーザー特徴を適切なベクトルで表現する必要がある。しかし、Web 上の端末数は億のオーダーになり、広告及びメディアの数は数十万の規模に上るため、一般的な手法ではベクトルの次元が非常に大きなものとなる。加えて、広告がクリックされる確率はおおよそ 0.1 パーセント程度であり、広告をクリックした人が商品購入などの成果地点に到達するのはさらにその数パーセント程度であるため、ベクトルのスパース性が高くなってしまいう問題がある。そのため効果の高い広告配信を実現するためには高次元でスパースなベクトルを効果的に低次元化するモデルが必要になる。

本研究ではユーザーの広告クリック情報および広告成果測定地点への訪問（コンバージョン）情報を用いて、精度と汎化性能の高い特徴量を獲得する手法を提案した。具体的には、TF-IDF（Term Frequency and Inverse Document Frequency）を用いてユーザーベクトルの重み付けを行い、後述のスパースベクトル許容型オートエンコーダを用いて、重み付けされたベクトルの次元を圧縮した。また、獲得特徴量を用いてユーザー分類を行い、広告効果の高いユーザー群が明確に検出されることを示した。

2. 提案モデル

(1)TF-IDF によるユーザーベクトルの重み付け

前処理として、Web 上のユーザーを文書とみなし、広告 ID を単語インデックスとみなした TF-IDF によるユーザーベクトルの重み付けを行った。例えば、あるユーザー v が広告 a を $x_{v,a}$ 回クリックして、広告 b にて $y_{v,b}$ 回コンバージョンしたとする。TF-IDF による変換を掛ける前のユーザーベクトル \mathbf{X}_v は次のように表現される。

$$\mathbf{X}_v = (x_{v,1}, \dots, x_{v,N}, y_{v,1}, \dots, y_{v,M})$$

ここに、 N をクリック実績のある広告数とし M をコンバージョン実績のある広告数とする。 \mathbf{X}_v に TF-IDF を掛けたベクトル \mathbf{Y}_v は、一般的にクリックされやすい広告の重みが小さくなり、ユーザーを特徴付けるような広告へのアクションの重みが大きくなる。 \mathbf{Y}_v を後述のスパースベクトル許容型オートエンコーダを用いて次元圧縮する。

(2)スパースベクトル許容型オートエンコーダ

オートエンコーダは、入力層と出力層に同じデータを与えることで中間層にデータの特徴を蓄積させるニューラルネットワークであるが、欠点として、ユニット数が多くなると計算時間が大きくなる問題がある。また \mathbf{Y}_v はその成分のほとんどがゼロであるため、「任意の入力に対してゼロを返す」ような特徴学習をしてしまう問題点も存在する。これらの 2 点の問題点を解決するため、図 1 に示すように \mathbf{Y}_v のゼロ成分を確率的に欠損値と見做してスキップさせるモデルを提案する。これをスパースベクトル許容型オートエンコーダと定義した。 \mathbf{Y}_v のゼロ成分を全て欠損値(NA)とせず、一部明示的にゼロを与えることで「広告 A をクリックしたユーザーは広告 B をクリックしにくい」といった非共起現象も学習させることが可能になる。

本研究では、第 1 段階としてユーザーベクトルを 28958 次元から 500 次元に圧縮し、第 2 段階でユーザーベクトルを 500 次元から 146 次元に圧縮するような積層型のスパースベクトル許容型オートエンコーダを用いた。また、入力層と中間層の活性化関数はシグモイド関数とし、出力層の活性化関数は恒等写像を用いた。

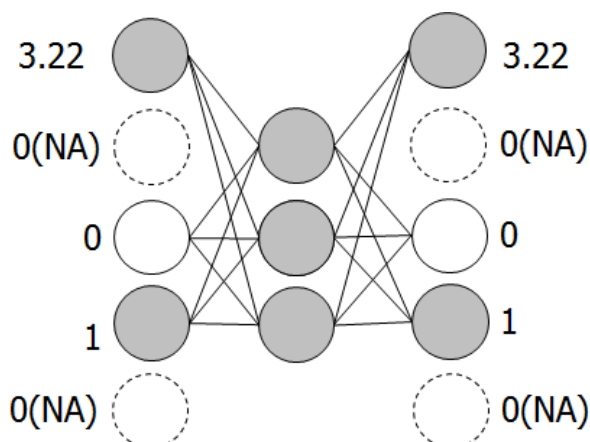


図1スパースベクトル許容型オートエンコーダ

70 スパースベクトル許容型オートエンコーダを用いるこ
71 とにより、Web 上の各ユーザーの特徴をベクトル表現でき
72 るようになった。このベクトルを用いて K-means により
73 ユーザー分類をし、広告効果の高いユーザー群が検出され
74 るかどうかを評価した。

77 (1)シミュレーションの概要

88

| 項目 | 訓練データ | テストデータ |
|-------------------|---------|---------|
| クリック実績のある広告数 N | 23339 | 25646 |
| コンバージョン実績のある広告数 M | 5619 | 5552 |
| クリック総数 | 5644230 | 5801623 |
| コンバージョン総数 | 32020 | 38547 |
| ユーザー数 | 3452491 | 3571259 |

92 図2 各ユーザー群のCVR

図 2 はユーザー群を 30 分類し、各ユーザー群の訓練データにおける CVR とテストデータにおける CVR を示している。すなわち CVR の高い群は、なんらかのジャンルの広告で、商品購入などの成果に到達しやすいユーザー群である。訓練データで高い CVR を示したユーザー群はテストデータでも CVR が高く、提案モデルによって広告効果の高いユーザー群が抽出出来ていることを示している。

本研究では、ユーザーがクリックおよびコンバージョン
した広告情報の次元を圧縮し、汎化性能の高い特徴を獲得
するモデルの考案と検証を行った。提案モデルによって、
広告効果の高いユーザー群と効果の低いユーザー群を明
確に分離することが可能になった。また、ニューラルネッ
トワークの計算量を軽減出来たことも意義が大きい。

109 今後は単純に広告効果の高いユーザー群を発見するだ
110 けではなく、各ユーザーの趣味嗜好の存在を検証すること
111 が重要な課題として挙げられる。また、ユーザーのサイト
112 訪問情報や滞在時間といった情報を加えてユーザー特徴
113 を獲得することも課題となる。

114