

## 伊呂原研究室

## Convolutional LSTM に基づくネット広告におけるコンバージョン数の 予測に関する研究

### Conversion Predicting System in Online Advertising based on Convolutional LSTM

CHEN LONGREN(ちん りゅうじん)

## 1 はじめに

近年、企業の広告の手段は多様化しており、インターネットを介した広告の重要性が高まりつつある。インターネット広告にはサイト広告や、リスティング広告などが存在するが、本研究ではユーザーがキーワードを検索した際に、キーワードに応じたサイトのリンクを表示させるリスティング広告に着目する。

実際に企業が広告を配信する場合には、企業と広告代理店の間で 1 カ月分の予算を決め、それを基に広告代理店が広告を企画する。広告会社は、検索キーワードに対応させて広告を配信する。配信の結果はクリック数、インプレッション数、コンバージョン数で評価され、その結果をもとに実際に広告会社から広告代理店が支払うコストを決定する。広告代理店および広告依頼企業にとって、予算によってどの程度の結果を得られるのかが分かれば、大いに役立つため、これを本研究の目的とする。

予算の設定による広告効果を予測する際、予算の情報が直接求められないことが大きな問題となっている。ここで、広告が問題なく配信された場合、予算とコストが一致するという固有技術的知見が存在する。コストに関する情報は得られていることから、予算をコストに置き換え、予測モデルを構築する。

本研究では、過去の複雑な情報を考慮するために、過去 1 か月分のカレンダー情報を用いて広告効果を予測するために、これまで主に画像認識において使用されていた CNN (畳み込みニューラルネットワーク) を LSTM モデルの中に組み込んだ、Convolutional LSTM のモデル[1]を援用する。また、このモデルを推定する際に、学習データの不足に対応するために、類似するデータも学習の中に取り入れる枠組みを提案する。さらに、重回帰分析、および LSTM によって分析した場合との違いについても考察する。

## 2 準備

### 2.1 ConvLSTM[2]

畳み込み LSTM (ConvLSTM) は、動画をはじめとするテンソルの時空間情報を考慮した、長期的な依存関係を持つ時系列データの高度な予測モデルを構築するための方法である[2]。モデルは LSTM を基に構

成されており LSTM におけるアダマール演算子が CNN で用いられる畳み込み演算子[2]に置き換えられる。この構造は LSTM に対して入力ゲート、出力ゲート、忘却ゲートにおいて時系列の前後の関係のみならず、周期的な前後関係も考慮することを可能としている。畳み込みは  $t-1$  時点で学習した潜在情報  $H_{t-1}$  とセル情報  $C_{t-1}$ 、 $t$  時点でのテンソルデータ  $X_t$  を畳み込み演算することで、求められる。(\*) は畳み込み演算子を示す。

$$C_t = f_t \circ C_{t-1} + i_t \circ \tanh(W_{xc} * X_t + W_{hc} * H_{t-1} + b_c) \quad (1)$$

$$H_t = o_t \circ \tanh(C_t) \quad (2)$$

ここでの畳み込みは下図における 3 次元テンソル  $X_t$  の 2 次元テンソルへの分割に相当する。

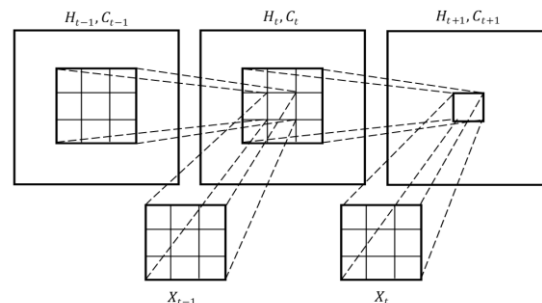


図 1 ConvLSTM の内部構造

## 3 分析データ

株式会社オプトにより提供された、インターネット広告会社で広告業務を行った結果データを扱う。また、今回の実験では、曜日情報、キャンペーン ID、Cost (広告費用)、Imp (インプレッション数)、Click (クリック数)、Cv (コンバージョン数)。

今回の実験で扱うデータは Google データセットにおける任意のキャンペーン ID を示す。キャンペーン ID のそれぞれの Imp、Click、Cv は予測対象とする。

集計対象日における日付情報は '2017/1/1' から '2019/9/30' まで総計 1003 日のデータを有している。

## 4 提案手法

先行研究 (Horita and Yamashita) により[3]、特に Cv データに関する予測精度が低いことが分かった。こ

のデータは時系列データにおける日の情報が複雑に関係していることが原因であると考えられる。また、少ないデータから推定している点にも問題がある。そこで日の情報をうまく使い切れるカレンダーの形式の情報に基づき分析していくことにする。

#### 4.1 カレンダーデータの活用

今回の実験は、カレンダー情報を考慮したモデルを提案すべく、ConvLSTMを適用する。これにより、時系列の情報のみならず、周期的な情報(曜日情報、週の情報、月の情報)を考慮することができる。

図3は、月ごと、週ごと、曜日ごとのデータ3次元テンソルデータを表す。一つずつのセルにはその時点での予測すべきデータが入るものとする。

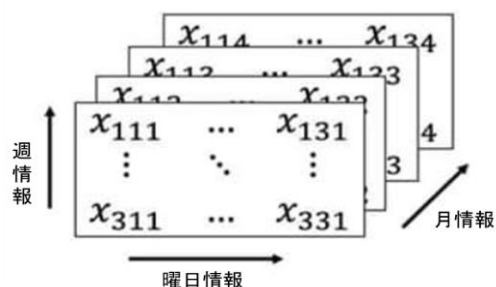


図2 カレンダー情報の設定

曜日、週、月の情報をモデルに反映させるため、5週分の2次元テンソルを5つ並べ、1週間は7日、データは5週間を1ユニットとする。それを5つまとめてタイムステップとすることで3次元テンソルデータを形成する。図2のようなカレンダーの3次元テンソルデータになる。このテンソルにImp、Click、Cvをそれぞれ代入し、それぞれを予測する。

#### 4.2 異なるデータセットのモデル学習の活用

深層学習におけるモデルは多くのデータを必要とする。本研究が与えられたデータでは十分な数を確保することができないため、似た値を持つデータを活用できるモデルの学習を行うことが望ましい。

そこで時系列データの順序性を維持しながら、N回のバッチ学習を行う。データの不足に対応する防止するため、同じデータセットの学習はM回繰り返す。分析対象となるデータに類似したK組の異なるデータセットも使って学習する。

### 5 実験

#### 5.1 実験の設定

ここではImp、Click、Cvのデータを予測する。ここでは二組のキャンペーンIDを実験データセットとする。データセット1(キャンペーンID:652057083)からの学習パラメータをデータセット2(キャンペーンID:648028230)の学習の初期値とし、テストデータはデータセット2の最後の35日のデータと設定する。評価指標はMSEと決定系数、重回帰分析およびLSTMを比較手法とする。

#### 5.2 分析結果

図3は予測結果と実データとの比較図である。時系列による変化傾向を予測できていることがわかる。

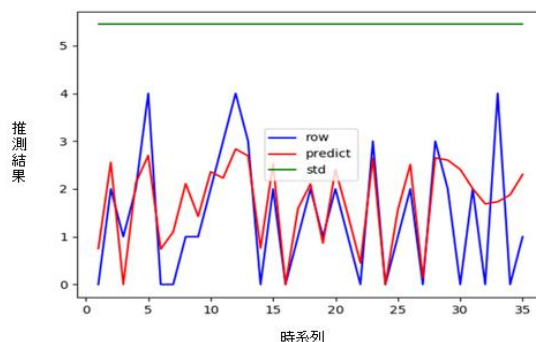


図3 予測値と実測値の比較

表1は3つの手法で、Imp、Click、Cvに対してそれぞれのMSEと決定系数を比較した結果である。

表1 3つの手法を用いた結果の比較

|          |       | MSE   | R2_score |
|----------|-------|-------|----------|
| REG      | Cv    | 3.476 | -0.243   |
|          | Imp   | 5.956 | 0.426    |
|          | Click | 2.087 | 0.201    |
| LSTM     | Cv    | 0.698 | 0.093    |
|          | Imp   | 6.698 | 0.179    |
|          | Click | 1.551 | 0.552    |
| ConvLSTM | Cv    | 0.881 | 0.483    |
|          | Imp   | 9.499 | -2.627   |
|          | Click | 2.060 | -0.013   |

回帰分析はImpの予測、LSTMはClickの予測、Cvの予測はConvLSTMの一番精度が良いと考えられる。

### 6 おわりに

本研究では、日による情報や閲覧情報を考慮し、カレンダー情報を用いたクリック数、インプレッション数、コンバージョン数の予測モデルを提案した。また、ConvLSTMに基づいてモデルを構築し、カレンダーデータという形式のデータを用いて研究を行って、モデルの問題点も明らかにした。3つの目標変数の予測が可能となるが、精度の改善および汎化性を高めるモデルの構築の必要がある。

### 7 参考文献

- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural computation, 9(8), 1735-1780.
- Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., ... & Chen, T. (2018). Recent advances in convolutional neural networks. Pattern Recognition, 77, 354-377.
- Yuki Horita & Haruka Yamashita, A Study on Prediction Model of Internet Advertisement Effect Depends on the Budget Based on Long Short Term Memory Model, Proceedings in 2nd International Conference on Economic Growth, Business Strategy and Social Sciences Reforms, 2019.