

消費者の購買行動に対する ディープラーニングの適用可能性の検討

若林 憲人^{†1} 生田目 崇^{†2}

概要：本論文では、近年注目されているディープラーニングをマーケティングデータ、特に顧客の購買行動に着目して適用し、評価を行う。ディープラーニングは画像や音声の分野で高い判別能力を示しており、他の分野での適用も期待されている。本論文では、購買及びその関連データに関して、複数のディープラーニングによる予測分析を行う。分析結果から、モデルの特徴を考察し、今後の可能性について論じる。

キーワード：ディープラーニング, CNN, DNN, 購買生起, 予測精度

1. はじめに

近年、インターネットの普及により、消費者の商品・サービスの購入に対する意思決定タイミングの変化が起こった。その変化に対応するマーケティング手法として、コンテンツマーケティングが重要視されている。コンテンツマーケティングでは、マーケティング実施主体とその訴求先とが「コンテンツ」を媒介にコミュニケーションすることを前提としている。訴求先としては、消費者だけでなく組織や団体なども含まれる。また、コンテンツの作りや内容により、ブランディング、認知拡大、利用促進など多くの目的で活用できる。このように、コンテンツマーケティングは、ターゲットとなる顧客を定義・把握し、その顧客を引き寄せ、獲得し、関わり合い、購買に結び付く行動を促すことを目的としているため、消費者行動を理解する上で重要な役割を担っていると考えられる [7, 12]。

特に、EC サイトは購買チャネルとしてだけでなく、各種のプロモーションや情報提供なども担うチャネルである。したがって、顧客の購買までのプロセス、すなわち検索や絞り込みなどの情報をアクセスログから収集できることから、購買前後の顧客の購買関連行動といった、より広い範囲で顧客行動に関するデータ項目を得ることができる。こうしたデータから購買行動を探ることで、顧客理解やアップセルの促進といったマーケティング活動に寄与できることが期待されている。しかし、消費者の多様化と共に、顧客の行動や購買決定もより複雑なものになっており、これらを共通した単一のモデルで測ることは簡単なことではない。そこでより複雑な因果関係を表現するようなモデル分析が必要とされつつある。

本論文では、EC サイト上のコンテンツにおける顧客のアクセス行動の複雑性を、近年注目されているディープ

ラーニング（深層学習）を用いて学習し、顧客の購買の予測を試みる。そして分析結果をもとに、マーケティングデータの分析におけるディープラーニングの適用可能性について論じる。

2. ディープラーニングと適用分野

ディープラーニングは、多層構造を持つニューラルネットワークを用いた機械学習アルゴリズムである。多層構造モデルを学習させることで、変数間の複雑な関連性を抽出できるため、モデルは非常に高い柔軟性をもつ学習モデルになる。近年、画像認識や音声認識の精度を競う数々のコンペティションで過去の記録を大幅に塗り替える高い性能を発揮し、ここ数年で一気に広まった [6, 9]。

ディープラーニングの考え方自体はニューラルネットワークの分野で古くから認知されていたが、十分な学習データを集めることが困難であった。また計算機も大規模なデータに耐えうる能力が不足していた。しかし、近年の計算機の計算能力の飛躍的向上やウェブの整備が状況を大きく変化させ、複雑なモデルに耐えうる環境に整ってきた。

本論文では、大きくは2種類のディープラーニングモデルを用いる。

一つは畳み込みニューラルネットワーク (Convolutional Neural Network: CNN) [9, 10, 11]である。CNN は主に画像認識の分野で用いられるモデルである。CNN では、画像の近接する画素が強い相関を持つという特徴を用いている。従来のニューラルネットワークが、入力と隠れ層が全結合している構造であるのに対して、CNN では入力を近接する画素にまとめ、小さい領域に分割することで、分割した領域とだけ隠れ層が結合するような構造を作る。さらに畳み込んだ層の小さい領域の一部分をプーリングする。プーリン

^{†1} 中央大学大学院

^{†2} 中央大学 (連絡先: nama@indsys.chuo-u.ac.jp)

投稿: 2016年11月30日

採択: 2017年2月10日

グを行うことで画像の特徴位置の移動に対してロバストとなる。

もう一つが、多層ニューラルネットワーク (Deep Neural Network: DNN) [8] である。DNN は、入力層、隠れ層、出力層がそれぞれ一つのニューラルネットワークモデルについて、隠れ層を複数の層に拡張している。隣りあう隠れ層は全要素で結合している。CNN が各層での働きを定義しなければならないのに対して、DNN では各層に含まれる要素数のみ指定すればよい。

このように、ディープラーニングの各モデルは複雑な関係を表現することができる。マーケティングにおいても市場における消費者の行動は多様化しており、消費者の購買行動などにディープラーニングが応用できることが期待できるが、現状では研究事例は多くない。

3. データの概要

本論文では、あるゴルフ専門 EC サイトを分析の対象とする。利用するデータは、4つのカテゴリ (アクセスログデータ、購買データ、予約データ、デモグラフィックデータ) に大別できる。

検証実験として、購買観測期間を設定し、その観測期間内にドライバーを買うか買わないかについての購買予測モデルの構築を行う。ドライバーは比較的高額な購買になることから突発的な購買行動は起こりにくいと考えられるため、直近1ヶ月程度の検討期間を考えることで消費者の迷いの行動を定量化できると考える。また、会員登録から対象期間までに累積の行動をサイトへのロイヤルティを捉え、これらをモデルに含める。

3.1 購買観測期間と予測期間について

購買観測期間と予測期間については図1のように定めた。すなわち、アクセスログデータについては、2015年7月20日からの1カ月間、また、直近の購買と予約は8月13日からの1週間とした。そして、8月20日から8月27日までの1週間にドライバーを購入するかどうかについて分析を行う。なお、この予測期間において1842人の購買が観測された。このうち、モデル検証用データとして300人をランダム抽出し、残りの1542人を学習用データとしている。また、購買がない顧客をランダムに同数、すなわち1842人抽出した。

以下で、4カテゴリの変数についてまとめる。

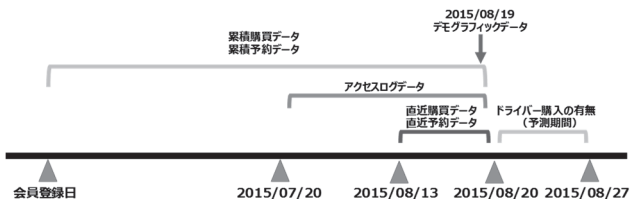


図1 データ集計期間

3.2 アクセスログデータ

アクセスログデータは、2015/07/20~2015/08/19 を集計期間とする。変数については、

当該 EC サイトで提供しているコンテンツページの閲覧回数とサイト訪問の回数を対象とする。

表1にアクセスログデータの変数名とその概要、表2にデータの様子を示す。表2の表頭の比率はそれぞれの項目のパーセンタイルを表す。すなわち、例えばショッピングページのページビュー (PV) である shop の下位 25% に該当する顧客は 190PV であるであることが分かる。

表1 アクセスログデータの変数一覧

変数名	変数の意味
purchase_count	累積購買回数(回)
sales_amt_Sum	累積購買金額の合計(円)
sales_amt_Mean	累積購買金額の平均(円)
sales_qty_Sum	累積購買個数の合計(個)
sales_qty_Mean	累積購買個数の平均(個)
Re_purchase_count	直近一週間の累積購買回数(回)
Re_sales_amt_Sum	直近一週間の購買金額の合計(円)
Re_sales_amt_Mean	直近一週間の購買金額の平均(円)
Re_sales_qty_Sum	直近一週間の購買個数の合計(個)
Re_sales_qty_Mean	直近一週間の購買個数の平均(個)

表2 アクセスログデータの概要

変数名	25%	50%	75%	100%
shop	190	580	1,510	76,115
news	0	5	61	19,640
campaign	0	0	0	100
lesson	0	10	120	11,840
reserve	0	5	105	25,815
beginner	0	0	0	85
dr	0	0	0	590
session_count	5	12	23	175

表1、表2より shop ページが最も良く見られていること、加えて、news、lesson、reserve ページといったメインコンテンツが閲覧されやすいことが分かる。

3.3 購買データ

購買データは、2015/08/13~2015/08/19 までの直近一週間の購買の金額の平均と合計、数量の平均と合計、購買回数と、各顧客の会員登録日からの累積の購買の金額の平均と合計、数量の平均と合計、購買回数を対象とする。表3に購買データの変数名とその意味、表4にデータの概要を示す。なお、表4の値は顧客ごとに集計したものであり、表頭の比率は表2と同様パーセンタイル点を表す。

表3 購買データの変数

変数名	変数の意味
purchase_count	累積購買回数(回)
sales_amt_Sum	累積購買金額の合計(円)
sales_amt_Mean	累積購買金額の平均(円)
sales_qty_Sum	累積購買個数の合計(個)
sales_qty_Mean	累積購買個数の平均(個)
Re_purchase_count	直近一週間の累積購買回数(回)
Re_sales_amt_Sum	直近一週間の購買金額の合計(円)
Re_sales_amt_Mean	直近一週間の購買金額の平均(円)
Re_sales_qty_Sum	直近一週間の購買個数の合計(個)
Re_sales_qty_Mean	直近一週間の購買個数の平均(個)

表4 購買データの概要

	25%	50%	75%	100%
purchase_count	4.0	12.0	26.0	405.0
sales_amt_Sum	39,002.5	123,096.0	311,887.3	19,890,966.0
sales_amt_Mean	3,983.0	6,492.7	10,394.0	157,500.0
sales_qty_Sum	6.0	20.0	46.0	2,049.0
sales_qty_Mean	1.0	1.0	1.1	7.2
Re_purchase_count	0.0	0.0	1.0	36.0
Re_sales_amt_Sum	0.0	0.0	7,313.0	608,955.0
Re_sales_amt_Mean	0.0	0.0	3,680.0	114,400.0
Re_sales_qty_Sum	0.0	0.0	1.0	46.0
Re_sales_qty_Mean	0.0	0.0	1.0	5.0

表3, 表4より, 累積で相当数購入している顧客がいることが分かる. また, 直近一週間のみの集計でも購買金額・個数が高い.

3.4 予約データ

予約データは, 2015/08/13~2015/08/19 までの直近一週間の予約の金額の平均と合計, プレー人数の平均と合計, 予約回数と各顧客の会員登録日時から 2015/08/19 までの累積の予約の金額の平均と合計, プレー人数の平均と合計, 予約回数を対象とする.

表5に予約データの変数名とその意味, 表6にデータ概要を示す. 表6の表頭の比率は表2と同様パーセンタイルを表す.

表5, 表6より, `reserve_count` の項目を見ると 75%タイル点で 17 回の予約があることから予約を相当数行っている顧客がいることが分かる. また, `Re_reserve_count` の値を見ると, 全体としては直近での予約はないものの, 逆に 6 回も予約をしている顧客も存在する.

3.5 デモグラフィックデータ

デモグラフィックデータは, 2015/08/19 時点の年代, 性別を対象とする. 表7にデモグラフィックデータの変数名とその意味と集計値を示す.

表5 予約データの変数と意味

変数名	変数の意味
reserve_count	累積予約プレー回数(回)
play_fee_amt_Sum	累積予約プレー金額の合計(円)
play_fee_amt_Mean	累積予約プレー金額の平均(円)
result_player_qty_Sum	累積予約プレー人数の合計(人)
result_player_qty_Mean	累積予約プレー人数の平均(人)
Re_reserve_count	直近一週間の予約プレー回数(回)
Re_play_fee_amt_Sum	直近一週間の予約プレー金額の合計(円)
Re_play_fee_amt_Mean	直近一週間の予約プレー金額の平均(円)
Re_result_player_qty_Sum	直近一週間の予約プレー人数の合計(人)
Re_result_player_qty_Mean	直近一週間の予約プレー人数の平均(人)

表6 予約データの概要

	25%	50%	75%	100%
reserve_count	0.0	4.0	17.0	271.0
play_fee_amt_Sum	0.0	151,752.0	749,991.8	17,645,492.0
play_fee_amt_Mean	0.0	31,477.4	49,146.8	470,208.0
result_player_qty_Sum	0.0	14.0	65.0	1,160.0
result_player_qty_Mean	0.0	3.0	4.0	32.0
Re_reserve_count	0.0	0.0	0.0	6.0
Re_play_fee_amt_Sum	0.0	0.0	0.0	525,600.0
Re_play_fee_amt_Mean	0.0	0.0	0.0	525,600.0
Re_result_player_qty_Sum	0.0	0.0	0.0	40.0
Re_result_player_qty_Mean	0.0	0.0	0.0	40.0

表7 デモグラフィックデータ

変数名	変数の意味	人数
age10	10代	1
age20	20代	54
age30	30代	280
age40	40代	1,142
age50	50代	1,411
age60	60代以上	796
men	男性	3,526
women	女性	158

表7より, 人数では 30 代以下は少なく 40 代以降の顧客が多い. 性別は男性が多いことが特徴である.

なお, 分析に用いたすべてのデータはそれぞれの項目の最小値と最大値をそれぞれ 0 と 1 とし, [0,1] の区間に正規化してから用いている.

4. 購買予測の比較

本論文では, 複数のディープラーニングモデルについての評価を行った. なお, 分析は Python 言語のディープラーニング用のライブラリ Lasagne [13] を用いている. 以下で, モデル構築及び結果について述べる.

表10 CNNによる正答率

	第1主成分			第2主成分			第3主成分			第1~第3主成分		
	最小値	平均値	最大値	最小値	平均値	最大値	最小値	平均値	最大値	最小値	平均値	最大値
絶対値CNN-SGD	62.00%	63.77%	65.83%	63.00%	64.43%	65.67%	64.00%	65.03%	67.50%	66.33%	67.57%	68.83%
絶対値CNN-Adam	62.17%	64.43%	66.00%	61.67%	63.87%	66.50%	61.83%	63.73%	65.50%	64.00%	65.16%	66.33%
正負CNN-SGD	60.17%	64.57%	67.50%	62.33%	63.73%	64.83%	63.67%	64.57%	65.83%	65.17%	66.07%	67.33%
正負CNN-Adam	61.50%	63.27%	65.50%	63.00%	64.97%	66.17%	64.67%	65.87%	67.33%	66.00%	66.40%	67.17%
両端CNN-SGD	62.17%	64.20%	65.50%	64.50%	65.13%	66.00%	61.33%	62.30%	63.50%	64.67%	65.87%	67.33%
両端CNN-Adam	57.83%	65.50%	64.33%	61.83%	64.00%	66.00%	59.83%	63.50%	64.67%	65.50%	67.33%	67.67%

4.1 データの並べ方に関する工夫

入力変数としては前述の4カテゴリのデータを用いるが、ディープラーニングで画像などの識別能力が高いのは、画像のなかで共通する部分が存在し、そうした特徴部分を畳み込みやプーリングによって抽出できるためと考えられる。しかし、本論文で用いるマーケティングデータの場合、データの並びに関連性はなく、ディープラーニングにおける特徴抽出がうまくいかない可能性がある。

そこで、本研究では、項目間の相関に着目し、相関が強いと考えられる順にデータを並び替えて用いることを行った。そのために説明変数について主成分分析を行い、各カテゴリに関して主成分負荷量をもとにデータを並び替える。ただし、第1主成分だけでは主成分負荷量の絶対値が小さい変数についてはうまく並び替えの効果が働かないことも考え、寄与率をもとに第3主成分までを用いる。

なお、入力変数を並べる際に、データのカテゴリによって次元が異なるが、不足する次元については共通して0を代入している。

4.2 比較モデル

表8と表9に本論文で比較したモデル一覧を示す。なお、各モデルの詳細や求解方法の詳細は付録にまとめる。また、クラス判別で広く用いられているロジスティック回帰分析をModel-31として比較に含める。これらの表内の「SGD」、「Adam」はそれぞれ最適化手法であり、詳細は付録にまとめる。また、「絶対値」は、得られた主成分負荷量について絶対値の大きい順に並び替えている。「正負」は、符号を含めて大きい順に並び替える。「両端」は、符号の順繰りにその絶対値の大きい順に端から内側に向かって埋めていくように並び替えている。例えば、4つの変数 x_1, x_2, x_3, x_4 の主成分負荷量が順に0.6, -0.4, 0.9, -0.1の場合は、まず最も絶対値の大きい x_3 を左端に、次に大きい x_1 を右端に、次に x_2 を x_3 の右に、最後に x_4 を x_1 の左側に配置する。

表9は、主成分分析を通さず、ランダムに変数の順を入れ替えた場合のモデル番号である。

表8 主成分分析の結果を用いるモデル一覧

	第1主成分	第2主成分	第3主成分	第1~第3主成分
絶対値CNN-SGD	Model-1	Model-2	Model-3	Model-4
絶対値CNN-Adam	Model-5	Model-6	Model-7	Model-8
正負CNN-SGD	Model-9	Model-10	Model-11	Model-12
正負CNN-Adam	Model-13	Model-14	Model-15	Model-16
両端CNN-SGD	Model-17	Model-18	Model-19	Model-20
両端CNN-Adam	Model-21	Model-22	Model-23	Model-24

表9 主成分分析の結果を用いないモデル一覧

	各主成分と同じサイズ	第1~第3主成分と同じサイズ
ランダムCNN-SGD	Model-25	Model-26
ランダムCNN-Adam	Model-27	Model-28
ランダムDNN-SGD	Model-29	—
ランダムDNN-Adam	Model-30	—

4.3 学習結果

本論文における計算実験は、モデルごとに5回実験を行った。5回の計算機実験の正答率の平均値、最大値、最小値を表10と表11にまとめる。ただし、正答率は出力について0.5を超えれば購入、越えなければ購入しないと予測した場合に正しく予測されたかどうかの比率である。なお、ロジスティック回帰モデル (Model-31) は解が一意に定まりその正答率は68.33%である。

表11 ランダムに変数をならべた場合の正答率

	各成分と同じサイズ			第1~第3主成分と同じサイズ		
	最小値	平均値	最大値	最小値	平均値	最大値
ランダムCNN-SGD	65.33%	65.87%	66.33%	65.00%	66.40%	68.33%
ランダムCNN-Adam	62.33%	64.50%	68.00%	65.83%	66.90%	68.33%
ランダムDNN-SGD	47.50%	55.63%	66.17%	—	—	—
ランダムDNN-Adam	57.17%	60.00%	62.17%	—	—	—

5. 考察

前節の結果から、CNN, DNNとも、ロジスティック回帰分析を大きく上回る判別能力はないと言える。CNNとDNNを比較すると、DNNよりCNNの方が正答率は高い。特にCNNのModel-4 (第1~第3主成分までを用い、それらについて主成分負荷量の絶対値の大きい順に並び替える。また、求解方法はSGDを用いている) においては、ほんの少しであるが、ロジスティック回帰分析を上回る正答率が得られた。

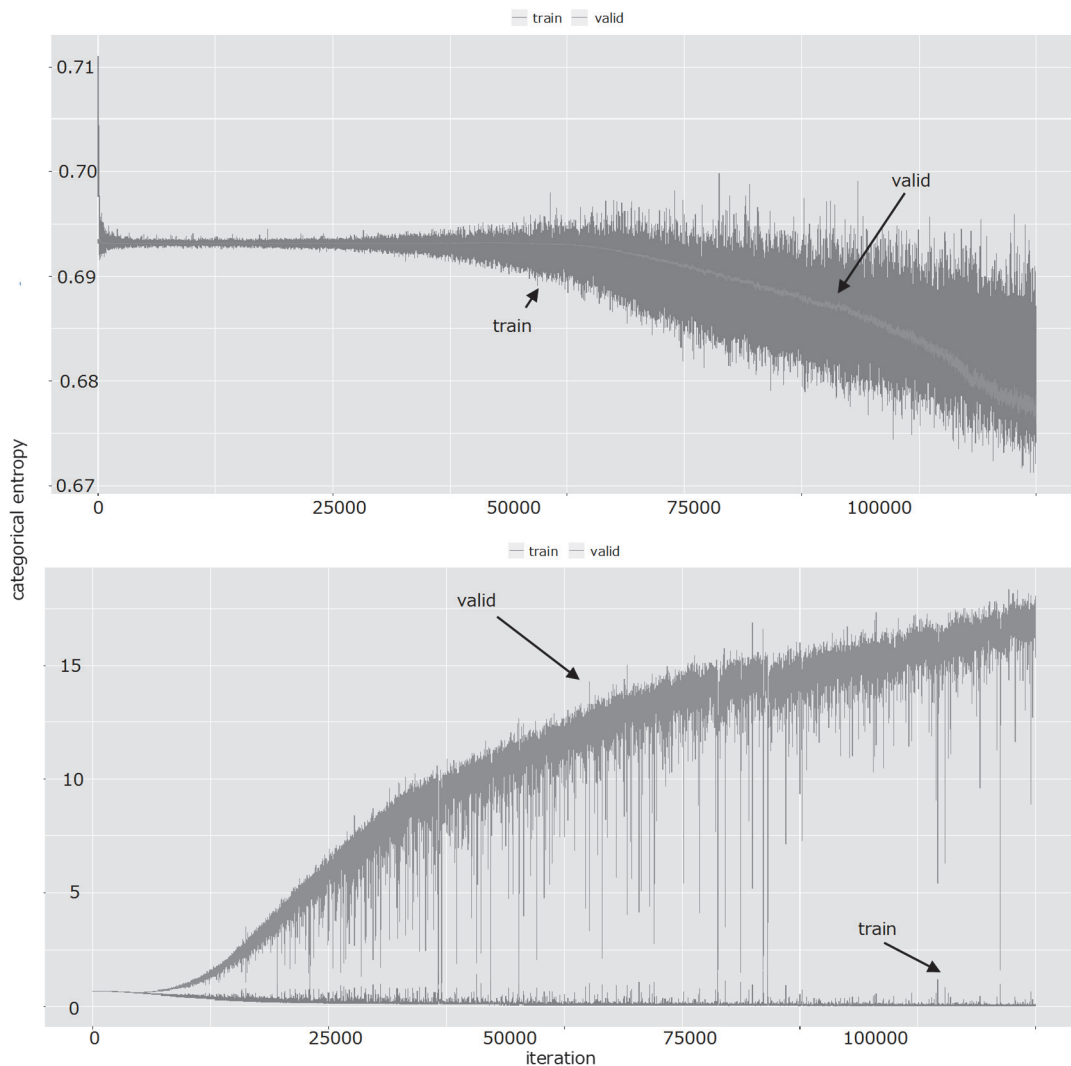


図2 エントロピー誤差
(上図：DNN，下図：CNN，train：学習データ，valid：検証データ)

全体的に DNN に比べて CNN の方が、また多くの主成分を用いる場合の判別能力が高いことから、入力の一部に特徴的なパターンを持ったデータについて、畳み込みやプーリングといった特徴抽出がパターン認識においては有効であることが示唆される。

以下、CNN と DNN それぞれで最も正答率の高い Model-4 と Model-29 についてさらに考察する。

図2は横軸が学習回数、縦軸がエントロピー誤差のグラフである。上のグラフが DNN (Model-29)、下のグラフが CNN (Model-4) であるが、DNN を見ると、学習回数が増えると学習データ、検証データとも誤差が小さくなる様子がわかる。これに比べて CNN の場合は、早い段階で収束するが、その後は過学習状態になっている。したがって、CNN についてはある程度の早い段階でも高い識別能力が期待できるが、その後のモデルの挙動については注意が必要であ

る。

また、正答、不正答についてケースごとにロジスティック回帰分析 (Model-31) と CNN (Model-4)、DNN (Model-29) を比較した。

ロジスティック回帰モデルは分散の大きな変数において極端に大きいもしくは小さい値を持つケースにおいて比較的正確しやすい。これは、分散が大きい場合にオッズ比が大きくなるためと思われるので、極端に大きいもしくは小さい説明変数にパラメータが引っ張られやすいためと考えられる。

また、DNN は変数内で値の分散が大きい場合に突出した値をもつユーザには正答しづらかった。DNN では変数間の関係性を抽出して学習を行っているため、ある値のみ突出した値をもっている場合でも変数全体から判別しようとしているためと思われる。

CNN についてはロジスティック回帰モデルの場合と正答・不正答のパターンが DNN の場合と比べて似ていた。これは、突出した値をもつ特殊なケースについて、各変数の特徴を抽出して学習が行えていると言える。畳み込みの強みである特徴抽出が上手く表現できた結果と言える。

6. まとめと今後の課題

本論文では、ディープラーニングの中でも特に画像認識の分野で用いられている CNN を中心に購買に着目したマーケティングデータに対して適用を検討した。本論文では特に、購買予測のためのモデル構築を目的とした。その中でモデル精度、データセットの構築、最適化手法、ユーザの予測、モデルの使い分けについて評価を行った。

モデルの予測精度に関しては、わずかではあるものの、既存のモデルのロジスティック回帰モデル(Model-31)や DNN (Model-29) より CNN を適用した Model-4 が高い精度を得られる場合があることが分かった。これは、CNN の特徴抽出の強みがチャンネル数を増やすことで活かせ、多数の異なるデータを分析に用いることで、精度向上の可能性があることが言える。ユーザごとの予測を見てみると、一般的な変数値をもつユーザに対しては軒並みどのモデルでも正答できている。しかし、突出した変数値を持つユーザに関しては DNN に比べ CNN とロジスティック回帰モデルの方が適切な予測がされやすい傾向があった。

データセットに関しては、主成分分析を用いてデータセットの順番を行う工夫を行った。大きな精度の改善は得られなかったが、入力データのチャンネル数を増やすことで精度向上が見込めることから、さらに多くの異なるデータを用いることで CNN の特徴抽出の強みを生かせ、精度向上の可能性について言及できよう。

本論文はまだ試行段階といえるため、いくつかの課題が残されている。まず、CNN の良さである特徴抽出の部分で過学習が起きているため、データセットに各ユーザの時系列の行動情報を複数入れる、もしくは変数やチャンネル数を増やすことで解決できる可能性があると考えられる。また、入力データについては主成分分析以外にも特異値分解や非負値行列因子分解など他の方法を考慮することも考えられる。これらについて評価することも今後の課題である。

参考文献

- [1] Ba, J. and Kingma, D., "Adam: A Method for Stochastic Optimization," *Proceedings of International Conference on Learning Representations*, arXiv:1412.6980v8 [cs.LG] 23 Jul 2015 (2015).
- [2] Duchi, J., Hazan, E. and Singer, Y., "Adaptive Subgradient Methods for Online Learning and Stochastic Optimization," *Journal of Machine Learning Research*, Vol. 12, pp. 2121-2159 (2011).
- [3] Boureau, Y., Ponce, J., and LeCun, Y., "A Theoretical Analysis of Feature Pooling in Visual Recognition," *Proceedings of the 27th*

- International Conference on Machine Learning* (2010).
- [4] LeCun, Y., Bottou, L., Bengio, Y. and Haffner, P., "Gradient-based Learning Applied to Document Recognition," *Proceedings of the IEEE 1998* (1998).
- [5] LeCun, Y., Bottou, L., Orr, R., and Muller, K., "Efficient Backprop" in *Neural Networks: Tricks of the Trade*, Springer (1998).
- [6] 川上和也, 松尾豊, "Deep Collaborative Filtering Deep Learning 技術の推薦システムへの応用", 第 28 回人工知能学会全国大会論文集, 3H3-OS-24a-1 (2014).
- [7] 近藤知幸, 井上佑介, "コンテンツマーケティング最前線 全国 28,000 店舗のコンビニを活用したサービス提供", *UNISYS TECHNOLOGY REVIEW EXTRA EDITION*, 第 124 号(2015).
- [8] 岡谷貴之, 「深層学習」, 講談社, (2009).
- [9] 岡谷貴之, 齋藤真樹, "ディープラーニング," 情報処理学会研究報告, Vol. 2013-CVIM-185, No. 19 (2013).
- [10] Simonyan, K. and Zisserman, A., "Very Deep Convolutional Networks for Large-scale Image Recognition," *Proceedings of ICLR 2015* (2014).
- [11] Hinton, G., Srivastava, N. and Swersky K., "Lecture6e - rmsProp: Divide the Gradient by a Running Average of its Recent Magnitude," *Coursera: Neural Networks for Machine Learning* (2012). http://www.cs.toronto.edu/~tijmen/csc321/slides/lecture_slides_lec6.pdf (2017/01/14 閲覧) .
- [12] ContentMarketingInstitute, "What Is Content Marketing?" <http://contentmarketinginstitute.com/what-is-content-marketing/> (2016/11/07 閲覧) .
- [13] Lasagne ウェブサイト, <https://github.com/Lasagne/Lasagne> (2016/11/20 閲覧) .

付録

A. 多層ニューラルネットワークと畳み込みニューラルネットワーク

A.1 多層ニューラルネットワーク

単層ニューラルネットワークでは、入力層と出力層の間に一層の隠れ層を挟み込み入力層から隠れ層、隠れ層から出力層にウェイトを乗じて足し合わせた上でバイアスを加えて入力値を伝播させていく。複数のユニットからなる隠れ層で、入力と出力の関係を複数の面で評価することができるので、複雑な入出力関係を評価することができる。ただし、データの次元が大きくなると、ユニットを増やすだけでは対応できなくなるほど複雑な関係が出てくる。そこで、隠れ層そのものを増やすことで、データに潜む特徴や関連性を表現しようというのが多層ニューラルネットワーク (Deep Neural Network, DNN) である [8]。本論文で DNN という場合は隣り合う層のユニット同士はすべてのペアで連結している場合を指す。

A.2 畳み込みニューラルネットワーク

DNN では隣り合う層のすべてのユニット同士での連結を考えるが、入力データのある領域に共通の特徴が出現しそれを抽出したい場合などにおいては、DNN ではうまくその特徴がうまく抽出できない場合がある。そこで、下記に示すような工夫が行われる。こうしたモデルは畳み込みニューラルネットワーク (Convolutional Neural Network,

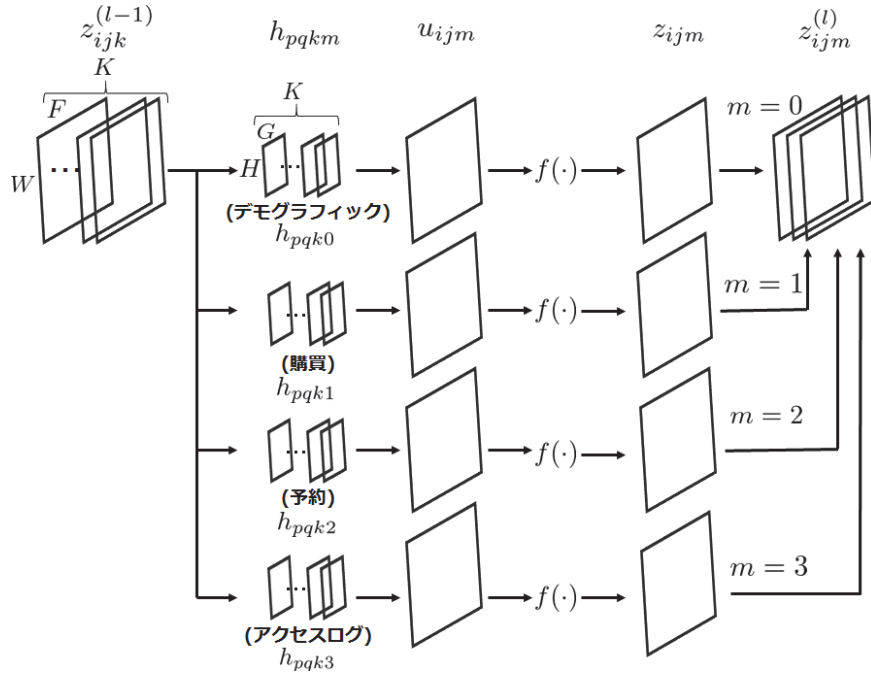


図 A.1 畳み込みの構造

CNN) と呼ばれる [9, 10, 11]. CNN の代表的なモデル上の工夫として、畳み込みとプーリングがある. これらを以下で説明する.

A.2.1 畳み込みについて

入力は、データのサイズが $W \times F$ の K 枚 (K : チャンネル), 変数のインデックスを (i, j) ($i = 0, \dots, W - 1; j = 0, \dots, F - 1$) とする. インデックス (i, j) の変数値を x_{ij} とする. チャンネル数は、データの特性に応じて作成する. 本研究では、4 カテゴリーのデータなので $K = 4$ である. そして、フィルタと呼ばれるサイズの小さいデータを $H \times G$ とし、フィルタのインデックスを (p, q) ($p = 0, \dots, H - 1; q = 0, \dots, G - 1$) で表し、変数値を h_{pq} とする.

データの畳み込みは、式 (A.1) に示すようなデータとフィルタ間での積和計算を行う.

$$u_{ij} = \sum_{p=0}^{H-1} \sum_{q=0}^{G-1} x_{i+p, j+q} h_{pq} \quad (\text{A.1})$$

データの畳み込みにおいて得られる働きは、フィルタの特徴的な変数や類似した変数値のパターンが入力データのどこにあるのかを検出する役割がある. つまり、データの特徴的な構造を抽出することが可能である.

畳み込み層は、畳み込みの演算を行う単層ネットワークである. 実際には、1 つのチャンネルに 1 つのフィルタを畳み込むより、複数のデータを用意して使い、多チャンネルに

複数のフィルタを並行して畳み込む方が実用的である.

図 A.1 では、第 l 層に位置する畳み込み層が直前の $l - 1$ 層から K チャンネルのデータ

z_{ijk}^{l-1} ($k = 0, \dots, K - 1$) を受け、これに M 種類のフィルタ h_{pqkm} ($m = 0, \dots, M - 1$) を適用している. 各フィルタは入力と同じチャンネル数 K を保有し、サイズは $H \times G \times K$ となる. 図 A.1 において $m = 0, 1, 2$ の 3 つの各フィルタを並行して計算が行われ、それぞれ 1 チャンネルずつで u_{ijm} の出力が得られる. この計算の終了後、結果を変数ごとに全チャンネルにわたっておこない、これは式 (A.2) のように表すことができる.

$$u_{ijm} = \sum_{k=0}^{K-1} \sum_{p=0}^{H-1} \sum_{q=0}^{G-1} z_{i+p, j+q, k}^{(l-1)} h_{pqkm} + b_{ijm} \quad (\text{A.2})$$

ただし、 $b_{ijm} = b_m$ はバイアスであり、フィルタごとに全ユニット共通としている. 最後に、得られた u_{ijm} に活性化関数を適用し、式 (A.3) の値が最終的な出力となり、次の層へと値が伝播する.

$$z_{ijm}^l = f(u_{ijm}) \quad (\text{A.3})$$

このように、入力のサイズが $W \times F \times K$ であったのが、畳み込み後の出力は $W \times F \times M$ になる.

A.2.2 プーリング層について

プーリング層は畳み込み層の後に置かれ対で使われ、畳み込み層からの出力がこの層への入力となる。データのどの位置でフィルタの応答が強かったかという情報を一部捨て、データ内に現れる微小な位置変化に対する応答の普遍性を実現するものである。

プーリング層の入力は直前の畳み込み層の出力であり、その次元は $W \times F \times M$ となる。その入力データ上の変数 (i, j) を中心とする $H \times F$ の領域をとり、この領域内の中に含まれるデータの集合を P_{ij} とする。この P_{ij} 内の変数について、チャンネル k ごとに H^2 個ある変数値を使って一つの変数値 u_{ijk} を算出する。方法の一つとして最大プーリング法があり、式 (A.4) のように最大値となるものが選択される。

$$u_{ijk} = \max_{(p,q) \in P_{ij}} z_{pqk} \quad (\text{A.4})$$

プーリング層においては学習によって変化するパラメータはなく、層間の結合の重みは固定されている。

A.3 確率的勾配降下法

確率的勾配降下法の元になる勾配降下法は、ニューラルネットワークモデルにおいて各層での入力変数ベクトルを $\mathbf{x}^{(i)}$ 、ウェイトとバイアスをそれぞれ $W^{(i)}$ 、 $\mathbf{b}^{(i)}$ して、出力関数を、

$$\mathbf{u}^{(i+1)} = W^{(i)}\mathbf{x}^{(i)} + \mathbf{b}^{(i)} \quad (\text{A.5})$$

とするとき、最終な出力層（第 L 層）の出力関数を、

$$y_k = \frac{\exp\{u_k^{(L)}\}}{\sum_{j=1}^K \exp\{u_j^{(L)}\}} \quad (\text{A.6})$$

として出力値を得る。このとき誤差関数は次の交差エントロピーを用いる。

$$E(\mathbf{w}) = - \sum_{n=1}^N \sum_{k=1}^K d_{nk} \log y_k(\mathbf{x}_n, \mathbf{w}) \quad (\text{A.7})$$

ただし、 d_{nk} は出力の第 k 要素の正答クラスであり、 d と \mathbf{y} が一致すれば 1、そうでなければ 0 となる。そしてこの誤差関数が小さくなるように学習する。しかし、式 (A.7) は一般には凸関数でないため、大域的最適解を求めるのは困難である。そこで、局所的な最適解をくり返し更新しながら

ら学習を進める。勾配降下法はこうした問題を解くための方法であり、 ε を学習係数として、式 (A.8) により解を更新する。

$$\mathbf{w}^{(t+1)} = \mathbf{w}^{(t)} - \varepsilon \frac{\partial E}{\partial \mathbf{w}} \quad (\text{A.8})$$

なお、 W の各要素の初期値については、下記の一様分布からランダムにサンプリングした値を用いることで広範囲の探索を行うようにしている。

$$W \sim U[-a, a], \quad a = \sqrt{\frac{12}{H_{in} + H_{out}}} \quad (\text{A.9})$$

H_{in}, H_{out} はそれぞれ、各層の前後の層の個数である。バイアス \mathbf{b} の初期値は広く行われているように 0 としている。

ただし全データを用いてこの計算を行うことは計算負荷が高いため、確率的勾配降下法 (Stochastic Gradient Descent Method, SGD) が用いられる。SGD では、ランダムに少数の学習データを選んで誤差関数を計算し、その勾配方向にパラメータを修正する操作を反復する手法である。パラメータが最適解から遠く離れている場合に、どのような訓練データを用いても停留点に近い位置に到達することが期待できる。

A.4 Adam

勾配降下法は、パラメータの更新量を学習係数によって変化させる。この学習係数は学習の良し悪しを大きく左右する非常に重要な役割を担っている。学習係数の決め方は分析者の裁量により、試行錯誤して値を決めることが一般的である。ネットワークの層毎に同じ学習係数を使用せず、層ごとに異なる学習係数を使うことで、各層に対する学習速度が均一になるように設定することが望ましいとされている [5]。

一方で、学習係数を自動的に定める方法で近年注目されているのが Adam (Adaptive Moment Estimation) 法である [1]。Adam 法は、先行手法の AdaGrad [2] の勾配が疎になる場合に強い性質と RMSProp [11] の非定常な問題に強い性質の 2 つの長所をもっている。

B. 本論文で用いたモデル

以下に Model-1~30 のモデルについて詳細をまとめる。表 B.1 は各要素に関する単語の一覧であり、表 B.2 および表 B.3 が CNN および DNN の構造である。

表 B.1 モデル層に関する説明

名称	意味
Input	入力層
Conv	畳み込み層
Pool	プーリング層
Full	全結合層
Output	出力層
ReLU	正規化線形関数
Sigmoid	シグモイド関数
Filter	畳み込みのフィルター数
Filter_Size	フィルター後の大きさ
Pool_Size	プールサイズ
Unit	ユニット数

表 B.2 CNN の層の概要

関数	Filter	Size	Pool	Size	Unit
Input	—	—	—	—	—
Conv	ReLU	4 or 16	(1,2)	—	—
Pool	—	—	—	(1,1)	—
Conv	ReLU	4 or 16	(1,2)	—	—
Pool	—	—	—	(1,1)	—
Conv	ReLU	4 or 16	(1,3)	—	—
Conv	ReLU	4 or 16	(1,3)	—	—
Conv	ReLU	4 or 16	(1,3)	—	—
Full	ReLU	—	—	—	250
Full	ReLU	—	—	—	100
Output	Sigmoid	—	—	—	2

Size が 4 の場合は Model-(1~3, 5~7, 9~11, 13~15, 17~19, 21~23, 25, 27) , 16 の場合は Model-(4, 8, 12, 16, 20, 24, 26, 28) で採用した。

表 B.3 DNN の層の概要

関数	Filter	Unit
Input	—	—
Full	ReLU	1024
Full	ReLU	512
Full	ReLU	256
Full	ReLU	128
Full	ReLU	64
Full	ReLU	32
Full	ReLU	16
Full	ReLU	8
Full	ReLU	4
Output	Sigmoid	2

本モデルは Model-(29, 30) で用いた。

A study of Possibility of Application of Deep Learning to Consumer Purchase Behavior

Kento WAKABAYASHI^{†1} Takashi NAMATAME^{†2}

Abstract: In this study, we focus on consumer purchase behavior, and apply deep learning which is a machine learning method to have a high profile in recent years to marketing data. Deep learning has high distinct ability in the field of image or document analyses, then deep learning is expected to apply the other research areas. In this study we predict consumers' purchase activity from purchase or related data by using plural deep learning models, i.e. deep neural network and convolutional neural network. From our result of analyses, we discuss the characteristics of our models and possibility of applying model.

Keywords: Deep Learning, Purchase Occurring, Accuracy of Prediction

^{†1} Graduate School, Chuo University
^{†2} Chuo University (Corresponding Author: nama@indsys.chuo-u.ac.jp)
Submitted: 30/11/2016
Accepted: 10/02/2017