

深層学習を用いたユーザ離脱予測

Churn Prediction using Deep Learning

宮崎 邦洋^{*1} 村山 菜月^{*2} 山本 裕樹^{*1} 牛山 史朗^{*3} 大澤 昇平^{*1} 松尾 豊^{*1}
 Kunihiro Miyazaki Natsuki Murayama Yuki Yamamoto Fumiaki Ushiyama Shohei Ohsawa Yutaka Matsuo

^{*1}東京大学大学院工学系研究科 ^{*2}東京大学教養学部
 School of Engineering, The University of Tokyo Dept. of Arts and Science, The University of Tokyo

^{*3}ウェルスナビ株式会社
 WealthNavi Inc.

The number of companies using subscription business model is increasing, and churn prediction is getting a more important task. In existing research, various type of machine learning models have already been used, but churn prediction has to be trained by combining various data such as time series data and non-time series data, which has not been fully studied. On the other hand, the technique of deep learning is still being developed, and one of its characteristics is that it can learn various data and models from end-to-end. In this research, we propose a churn prediction model with deep learning using data of WealthNavi inc. which manages the service of Robo-adviser. Specifically, we propose a method to learn time series data and non-time series data with one model. In the experiment, the effectiveness of this method was demonstrated by obtaining the result exceeding the accuracy of the classifier of the existing research.

1. はじめに

ユーザ離脱予測はビジネスにおいて重要な課題である。企業にとってユーザの新規獲得コストは大きく、マーケティング領域の研究においては新規顧客に製品を販売するコストは既存顧客に販売するコストの5倍程度かかることが明らかになっている [Verbeke 12]。そのため新規顧客の獲得と同様以上のコストを既存顧客のリテンションに割くことは企業にとって経済合理性があり、より効果的に既存顧客を維持するためにも、近い将来での離脱が見込まれるユーザを予測することは重要である。特に昨今では Web 上の音楽や本などのコンテンツ分野においてサブスクリプション方式（製品を購入せずに、ある期間における製品の使用料をユーザが支払う形式）のサービス形態が台頭している。そのようなサブスクリプション方式においては、新規ユーザの獲得とともに既存ユーザにどれだけ継続的に利用されるかがサービスの成否を左右する。ユーザの定着は企業のビジネスを安定させ、企業がサービスの改善や使用料引き下げという形でユーザに還元することでさらにユーザを惹きつけるという好循環が生まれる可能性も見込める。Web サービスを中心にユーザの行動や状態のデータが蓄積されるなど分析の環境も整いつつあり、今後ユーザの離脱予測は企業の重点課題としてより取り上げられていくと思われる。

ユーザ離脱予測は主に通信の分野で研究されている技術であり、その基本的な手法は、ユーザの属性情報や行動データなどの特徴量を機械学習のモデルに読み込ませ、将来の一定期間における離脱をユーザ毎に予測するものである。既存研究では Random Forest や SVM などの手法が使われているが、深層学習を用いた研究は多く行われていない。本研究においては、ユーザの離脱にあたり、深層学習における以下2点の特徴を活かしたモデルを作成する。

- 再帰型のニューラルネットワークを使用することで、過去の情報を保持したまま時系列の学習ができる点

連絡先: 宮崎邦洋, 東京大学工学系研究科技術経営戦略学専攻, miyazaki@weblab.t.u-tokyo.ac.jp

- 様々なデータを一つのモデルで End-to-end のアンサンブル学習を行わせることにより、効率的な学習ができる点

実験では、サブスクリプション方式である資産運用の Web サービスにおいて実験を行った。具体的にはユーザの非時系列の情報と時系列の情報を組み合わせたデータを使用し、ランダムフォレストや多層パーセプトロンなど時系列性を考慮しないモデルに対し、再帰型ニューラルネットワークを組み合わせた提案モデルの予測精度を評価した。検証の結果、

- 時系列性を持つと思われるデータは時系列モデルを通して学習をさせた場合に予測精度が高いことがわかった
- 様々なデータをインプットとする上で、深層学習のようなモデル構造の自由度が高い手法の有用性が示された

2. 関連研究

ユーザ離脱の研究は多く行われている。主な手法としては、(1) ユーザ行動, (2) 顧客のサービスへの態度, (3) 顧客基本情報, (4) 外部要因の4カテゴリの特徴量を素性とし、それらを機械学習モデルに読み込ませて予測しているケースが多い [Kamalraj 13]。使用される機械学習は様々であるが [Ali 14], 深層学習を用いたユーザ離脱予測は多く行われていない。特に、時系列モデルやアンサンブル学習を使用している研究はあるが、それらを組み合わせて予測を行う研究は筆者の知る限り未だかつてない。

3. 提案手法

3.1 Long Short-Term Memory

本研究では時系列データを扱うモデルに Long Short-Term Memory (LSTM) を採用している。LSTM は再帰型ニューラルネットワークの一種であり、学習において過去の時系列相関を取得することが期待される。再帰型ニューラルネットワーク

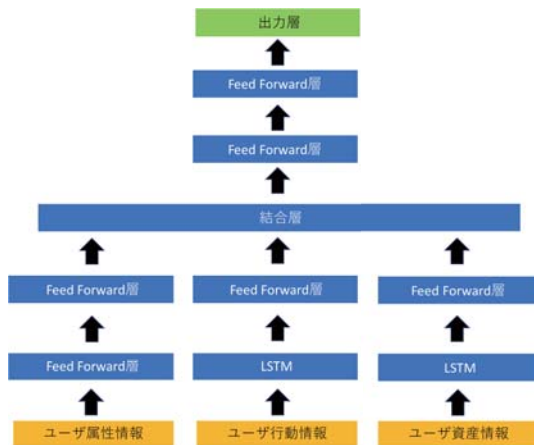


図 1: 提案手法の概要図

は時系列情報を記憶しながら学習できるニューラルネットワークである。再帰的な過程は以下のように表される。

$$z^t = f(W_{in}X^t + W_{z^{t-1}}) \quad (1)$$

ここで、 f は活性化関数、 W_{in} は入力層から隠れ層への重み行列、 W は隠れ層の重み行列、 X^t は時刻 t における入力データ、 z^t は時刻 t における隠れ層の出力を示す。特に今回の手法においては、過去のユーザーのアクセスパターンや資産価格の変動が、ユーザーの離脱の意思決定にどれだけ影響を与えるかを捉えることができる。LSTM は再帰型ニューラルネットワークの拡張版であり、より長期的な時間依存を扱える。

3.2 提案モデル

今回提案するモデルでは、複数のニューラルネットワークを組み合わせたモデルを提案している。深層学習の特徴として、ネットワーク構造の自由度が高く、扱うデータに合わせてモデルを改変し易いことがあげられる。特に今回のような時系列データと非時系列データを同時に扱う場合等で利便性が高い。また複数のデータとモデルを一つのネットワークに組み込むことで、人手を介さずに End-to-end でアンサンブル学習することも特徴としてあげられる。アンサンブル学習では通常低層の出力をモデルの作成者が人手で設計する必要があり、その場合作成者のバイアスがかかってしまう可能性と設計自体に手間がかかるという問題があるが、複数の入力から出力まで一気通貫して学習できる深層学習はその問題を解決できる。提案モデルの概要図は図 1 である。今回は入力を (1) ユーザーの属性情報、(2) ユーザーの日次のアクセス回数、(3) ユーザーの日次のポートフォリオ資産価格変動とした。それらをまずローカルなネットワークに情報として与え、それぞれの出力を途中で結合し、一つのアウトプット層につなげている。提案手法ではアクセス情報と資産情報は第一層を LSTM とした。

4. 実験

4.1 対象データ

今回、ロボアドバイザーを運営するウェルナビ株式会社のデータを用いて研究を行った。ロボアドバイザーとは、ユーザーの長期にわたる安定的な資産運用をサポートする Web サービスの通称であるが、大きく 2 つのタイプに分けられる。1 つはユーザーのリスク許容度などを診断してポートフォリオ（資産の組み合わせ）を提案し、さらには具体的な金融商品の購入やそ

の後のメンテナンスなども全て自動で行うもので「投資一任型」と呼ばれ、もう 1 つはポートフォリオの提案までは行うものの金融商品の購入からはユーザーが自ら行う必要があるもので「アドバイス型」と呼ばれる。ウェルナビ株式会社が提供するの「投資一任型」のサービスであり、ユーザーは自身のリスク許容度などを診断した上で入金すれば、ポートフォリオの構築からその後のリバランスなどの適切なメンテナンスまで自動で行われる。毎月一定額を自動的に投資にまわす自動積立機能もあり、ユーザーは必要に応じて運用状況を確認する以外はほとんど手間をかけることなく、長期にわたる資産運用を行うことができる。ユーザーが長期間利用を続けることでメリットを最大限に享受できることを狙ったサービス設計となっており、短期での離脱はユーザー自身の利益を損なうケースも少なくないと考えられる。

4.2 実験概要

今回使用したデータは 2017 年 8 月から 12 月のものである。実験では、過去 30 日のデータを用い、将来の 2 週間における離脱を予測する分類問題とした。検証したユーザーにおいて、実際に 2 週間で離脱するユーザーは約 0.5% である。偏ったデータに対応するため、学習時点では正例と負例が同じ数量になるようアンダーサンプリングによりバランスをさせた。また、学習データ自体を増やすため、過去のデータを 2 週間毎 10 ステップのデータを取得し、それらをつなぎ合わせて学習データとした。提案手法と比較する手法としては、(1) 既存手法で頻繁に使用されている Random Forest、(2) 時系列性を考慮しない深層学習の手法である中間層が 4 層の Multi Layer Perceptron (MLP) とした。

4.3 結果

実験の結果は 1 のようになっている。

表 1: 各手法の比較

手法	Recall	Precision	F 値
Random Forest	1.000	0.090	0.165
MLP	0.994	0.099	0.179
提案手法	0.994	0.104	0.188

提案手法が他 2 つの手法を精度で上回っていることがわかる。本研究ではもともとのデータの特性から、ほとんど Recall が 1 に近い精度となっている。離脱が見込まれないユーザーに対してメールを打つなどの離脱防止策を施すことはユーザーを不必要に煩わせ逆効果になることもありうるので、Recall Rate が高いことは実務上望ましいことである。その中で如何に False Positive を少なくするかが F 値にそのまま効いてきていることがわかるが、提案手法が最も Precision を高く学習できており、的中率が高いことを示している。

5. 結論と展望

今回の結果から、アクセス情報やポートフォリオ情報を時系列モデルで学習させた方が良かった。また、それら様々なデータをインプットとする上で、深層学習のようなモデル構造の自由度が高い手法の有用性が示された。

今後としては、この度作成したユーザー離脱モデルを基に、より詳細にユーザーの行動をモデリングする手法の研究を進めていく。一つは、今回アクセス回数をユーザーの行動情報として用いているが、より End-to-end に学習が可能にするため、ページ遷移を含めたアクセス情報を入力として精度向上を目指す。

また、今回は事前実験においてポートフォリオ情報が精度向上をもたらすことが判明しており、ポートフォリオの変動とユーザ行動の関係性の分析を進めることで、よりユーザにとって利便性の高いサービスとなる助けとなるような研究を目指す。

6. 謝辞

本研究は、ウェルスナビ株式会社の支援を受けて行われた、

参考文献

- [Ali 14] Ali, Ö. G. and Artürk, U.: Dynamic churn prediction framework with more effective use of rare event data: The case of private banking, *Expert Systems with Applications*, Vol. 41, No. 17, pp. 7889–7903 (2014)
- [Kamalraj 13] Kamalraj, N. and Malathi, A.: A survey on churn prediction techniques in communication sector, *International Journal of Computer Applications*, Vol. 64, No. 5 (2013)
- [Verbeke 12] Verbeke, W., Dejaeger, K., Martens, D., Hur, J., and Baesens, B.: New insights into churn prediction in the telecommunication sector: A profit driven data mining approach, *European Journal of Operational Research*, Vol. 218, No. 1, pp. 211–229 (2012)