

Graph Attention Networks を用いた CtoC サービスにおけるユーザ定着・離脱予測

User Settlement/Churn Prediction on a CtoC service

工藤航 鳥海 不二夫
Wataru Kudo Fujio Toriumi

東京大学
The University of Tokyo

Recently, the form of transaction between consumers called CtoC (Consumer to Consumer) service has expanded the market scale and has attracted great interest. In CtoC services, where the decrease in users directly causes the decline of services, it is necessary to take measures to prevent existing users from leaving. In order to support such measures, the problem of predicting user departure with high accuracy and interpretability has been actively studied as "User Churn Prediction". The purpose of this research is to construct a high-performing and interpretable framework to predict user settlement and churn that incorporates not only users' own characteristics but also the effects of contact between users. We applied Graph Neural Networks(GNNs), which have been attracting attention in recent years, to the task of predicting user settlement and churn in the CtoC service.

1. 序論

近年のインターネットの普及に伴い、e コマース (電子商取引) が盛んに行われるようになった。それに伴い、消費者の消費行動も多様化している。特に、スマートフォンやそれを用いた決済システムの普及に伴い、CtoC (Consumer to Consumer) と呼ばれる一般消費者間での取引形態が市場規模を拡大し、大きな関心を集めている。

CtoC プラットフォームを含む e コマースプラットフォームでは、ユーザ数を増加・維持させて利益を最大化するために、離脱しそうなユーザにアプローチすることで離脱を防ぐ施策が積極的に行われている。学術界ではその施策を支援すべく、高い精度と解釈性をもってユーザの離脱を予測する研究が盛んに行われている。CtoC サービスを利用するユーザに対する定着・離脱予測を考えると、予測対象とするユーザ自身の特徴だけでなく、ユーザ間の接触の影響を詳細に考慮することは精度と解釈性の両面で非常に重要であると考えられる。特に、予測対象となるユーザの 1) 接触相手の性質、2) 接触の方法、そして 3) どの接触が重要であったかをモデリングすることで、高い精度と解釈性を期待できる。

このような背景から、本研究の目的は、CtoC サービスを利用するユーザの行動傾向を明らかにするとともに、ユーザの定着・離脱予測について、予測対象のユーザの特徴だけでなくユーザ間の接触がもたらす影響も組み込んだ定着・離脱予測フレームワークを構築し、高精度で解釈可能な予測を行うことである。結果の解釈を通して定着・離脱に関わるユーザの傾向を明らかにし、ユーザ数を増加させるための施策への活用を目指す。

本研究ではハンドメイド作品を売買できる CtoC プラットフォーム minne^{*1} の実データを対象とする。ユーザの特徴と接触の影響を考慮した高精度で解釈可能な定着・離脱予測を実現するため、近年注目を集めている Graph Neural Networks をベースとした予測モデルを実データ上の定着・離脱予測に適

用し、ベースライン手法と比較実験を行う。また、モデルの解釈性を活用し、定着・離脱を予測する際に重要であった他ユーザとの接触の種類や接触相手を抽出できることを示し、ユーザの定着を促す施策への活用可能性について考察する。

2. 関連研究

2.1 ユーザ離脱予測

サービスのユーザ数を増加させていくための戦略として、新たにユーザを獲得することは現存するユーザを維持することよりもコストが高くなりやすいことが知られている [Daly 02, Gillen 05]。そのため、離脱しそうなユーザにアプローチして離脱を防ぐ施策を講じることはビジネス上の観点から非常に重要である。学術界においても、現存ユーザの特徴から高精度で解釈可能なユーザ離脱予測モデルを構築する研究が盛んに行われてきた。本研究に関連する例として、ユーザ同士の接触があるサービスを対象としたユーザ離脱予測について紹介する。

Kwon ら [Kwon 19] はビジネスサイトに対するレビュープラットフォームを利用するユーザを対象として、ユーザ・地域などの特徴量に加え、ユーザ同士のフレンドネットワークにおける回数や近傍の離脱率も特徴として組み込み離脱予測を行うモデルを提案し、モデルの解釈により離脱と関係が強い特徴について考察した。Drachen ら [Drachen 18] はソーシャルゲームのユーザに対し、ユーザのプレイ・購入履歴とフレンドネットワークにおける回数、ページランク、各種中心性などから離脱を予測する手法を提案した。Yang ら [Yang 18] はソーシャルコミュニケーションアプリの Snapchat を対象に、個々のユーザの行動特徴に加えフレンド関係におけるエゴネットワークも特徴に組み込み、ユーザクラスタリングと離脱予測を行う解釈可能なフレームワークを提案した。

これらはユーザ同士の接触を考慮しているが、いずれも特徴量エンジニアリングによりネットワーク上のノードが持つトポロジカルな性質 (回数、ページランクなど) を抽出したり、近傍の持つ特徴を単純平均したりするアプローチである。そのため、ユーザが接触した個々の接触相手の性質や接触方法について個別に組み込むこと、および解釈することはできない。したがって、本研究で扱うような多種多様なユーザが複雑な接触を

連絡先: 工藤航, 東京大学大学院システム創成学専攻 鳥海研究室, 東京都文京区根津 1-24-5, 09030938349, kudo@torilab.net

*1 <https://minne.com>

行っているデータに適用する場合、モデルの表現力・解釈性の双方に改善の余地があると考えられる。

2.2 複雑ネットワークに関する手法

本研究で扱うデータは商品の提供者と購入者という異なる性質を持つユーザがあり、それらが購買関係やフォロー関係といった複数の複雑な関係で結ばれているネットワークデータであるとみなすことができる。近年、このように複雑なネットワークデータをモデリングする手法の研究が盛んに行われている。

近年、ノード分類やリンク予測へのアプローチとして大きな注目を集めている手法が Graph Neural Networks(GNNs)である。GNNsの分類や応用例については複数のサーベイ論文が見られる [Zhou 18, Wu 19]。GNNsは Network Embedding とは異なり、通常は教師あり学習として扱われる手法である。ノード特徴量の変換と隣接関係による集約という一連の流れを教師情報を用いて End-to-End で学習することにより、ネットワーク関連の種々のタスクで高いパフォーマンスを発揮してきた [Kipf 16, Hamilton 17, Ying 18]。Schlichtkrullら [Schlichtkrull 18]はエッジの種類が複数あるようなネットワークに対して効果的な Relational Graph Convolutional Networks(R-GCN)を提案し、ノード分類とリンク予測で高い性能を発揮することを示した。Veličkovićら [Veličković 17]が提案した Graph Attention Networks (GAT)は、隣接ノードの集約を行う際の重み付けを同時に学習を行うことで、エッジのなかでも本質的なものだけを使ってノード表現の更新を行うことができる。また、学習によって取得された重みを使ってモデルの解釈を行うことができる。

本研究では、R-GCN[Schlichtkrull 18]をベースとしてモデルと、GAT[Veličković 17]をベースとしたモデルをユーザ定着・離脱予測に適用し、その他のベースライン手法との比較を行うとともに、モデルの解釈性を活用したユーザ傾向の分析を行う。

3. CtoC サービスデータ概要

本研究では、GMO ペパボ株式会社*2より、同社が2012年1月から運営しているサービスである minne についてのデータ提供を受けた。

minne はハンドメイド作品を売買することができる CtoC サービスである。ユーザ間で取引される商品はユーザによって創作されたハンドメイド作品 (以下、作品) である。minne のユーザは利用方法から“作家”、“購入者”の2種類に分類することができる。作家とは、作品を創作し出品するユーザを指し、購入者は出品された作品を購入するユーザを指す。また、作家・購入者間の接触として、作家と購入者間の作品の売買、購入者が気に入った作家の登録する“フォロー”機能や購入者が購入した作品についての感想を投稿する“レビュー”機能、投稿されたレビューに対して作家が返信する“リプライ”機能などがある。

本研究に用いるデータは2018年12月1日から2019年12月1日の間 (以下、対象期間) に行われた作品の売買取引履歴、対象期間に取引を行なったユーザの全フォロー関係、レビュー・リプライ履歴、全出品履歴である。なお、minne のデータを研究目的で利用する際には、実際のユーザと紐づかないようにユーザ ID を振り直して利用している。

また本研究では、サービスの利用方法 (購入、販売) に従ってユーザを作家と購入者に分類する。そこでまず、対象期間中の取引履歴から作品の販売を行ったユーザ数 (=作家数)、および作品の購入を行ったユーザ数 (=購入者数) を算出したところ、作家数が141,906人、購入者数が2,438,592人であった。このことから、購入者は作家の約17.2倍存在し、少数の作家と多数の購入者間で作品の取引が行われていることがわかった。

また、期間中の月ごとの取引数の平均は約410618回、月ごとのレビュー数の平均は約155630回、リプライ数の平均は約81617回、月ごとのフォロー増加数の平均は約550053回であった。これらから、作家・購入者間で作品の取引、レビュー・リプライ、フォローといった接触が盛んに行われていることがわかる。

4. ユーザ定着/離脱予測

4.1 実験概要

4.1.1 目的

本研究では、1) 高い精度での定着・離脱ユーザの予測と、2) ユーザの定着や離脱と関連が強い要素を明らかにすることの2点を目的とし、minne ユーザの定着・離脱を予測するフレームワークを検討する。

4.1.2 実験設定

本実験では「2019年3月に出品・販売・被レビュー・被フォローを行なった作家」を対象とする。また、傾向をより明確に把握するため、作家を新規作家 (登録期間: 2019年1月~3月) と古参作家 (登録期間: 2018年1月~3月) に分けてそれぞれで学習と評価を行う。

特徴量としては、各作家の2019年3月における販売数、出品数、被フォロー数、被レビュー数、リプライ数、作品紹介文の平均長、取引の平均金額、購入者一人当たりの取引数、一取引あたりの作品の数、レビューの平均長、リプライの平均長、リピート購入者の数*3、の合計12個の特徴量を作成した。

目的変数は2つのパターンで定義した。1つは、対象のユーザが定着するかそうでないかを判別することに主眼を置いた、“定着/非定着”目的変数である。そして、もう一つは、対象のユーザが離脱するかそうでないかを判別することに主眼をおいた、“離脱/非離脱”目的変数である。“定着/非定着”は、「2019年9月以降に出品または販売を行った/行わなかった」と定義し、“離脱/非離脱”は「2019年4月以降に出品または販売を行わなかった/行なった」と定義した。

4.2 手法

本実験では GNNs をベースとした2つのモデル (R-GCN; Relational Graph Convolutional Network[Schlichtkrull 18], GAT; Graph Attention Networks[Veličković 17]) を定着・離脱予測に適用し、3つのベースライン手法 (決定木分析 [Breiman 84], ロジスティック回帰分析 (LR)[Cox 58], 多層ニューラルネットワーク (多層 NN)) との比較を行う。

4.3 結果

表1に、新規作家の定着予測、新規作家の離脱予測、古参作家の定着予測、古参作家の離脱予測についての結果を示す。新規作家に対する分類タスクで、GNNs をベースとしたモデルは有効であるということが明らかになった。

*2 <https://pepabo.com>

*3 期間中に2回以上の取引があった購入者のユニーク数

表 1: 作家についての分類タスクの AUC

手法	新規定着	新規離脱	古参定着	古参離脱
決定木	0.684	0.701	0.767	0.804
LR	0.666	0.696	0.778	0.804
多層 NN	0.675	0.700	0.784	0.809
R-GCN	0.721	0.729	0.778	0.808
GAT	0.722	0.728	0.783	0.811

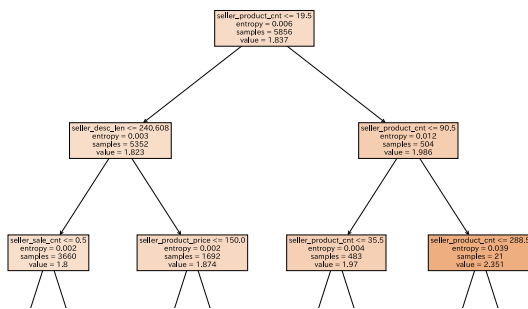


図 1: GAT の新規作家定着予測モデルにおける Self-loop への重要度の回帰木

4.4 考察

4.3 節の結果から、特に GNNs ベースの手法での性能改善が著しかった新規作家定着予測について、GAT のモデルの解釈を通して考察する。

図 1 に、GAT によって学習された Self-loop^{*4} の重要度算出について、作家の特徴量から重要度を説明する回帰木を示す。これより、作家が活発な場合は Self-loop の重要度が高くなっていることがわかる。図 2 に、Self-loop 以外のエッジについて、作家・購入者の特徴量およびエッジのタイプから重要度を説明する回帰木を示す。これより、購入者のフォロー数やレビュー数が多い場合に強い重要度が算出されていることがわかる。以上の重要度の解析を要約すると、以下のことが言える。

- 新規作家定着予測では、作家自身が活発な場合は作家の特徴量で定着かどうか主に判断する、
- 新規作家定着予測において、作家自身が活発ではない場合は相対的に接触した購入者の特徴を考慮する比率が高まる。特に、接触した購入者のフォローやレビューの回数が多い場合は重要度が高くなる傾向がある。

接触相手の特徴を適切に捉えることによる性能改善に加え、解釈可能性も GAT の利点である。GAT で学習された attention の値を見ることによって、個々のユーザについて予測タスクに重要だった接触と接触相手を明らかにすることができる。以下では例として、新規作家定着予測の対象とした 2 人の作家について各接触に算出された attention の値を可視化する。

図 3 は、特徴量作成期間中に 3 人の購入者からフォローされたある作家に対する attention の値を示したものである。ここでは Self-loop よりも 2 人の購入者 (buyer_239687, buyer_400135)

*4 Self-loop は、作家-購入者間のネットワークにおける自己ループを示す。Self-loop の重要度が高いことは、定着・離脱予測モデルがその作家自身の特徴量をもとに判断する比率が高かったことを意味する

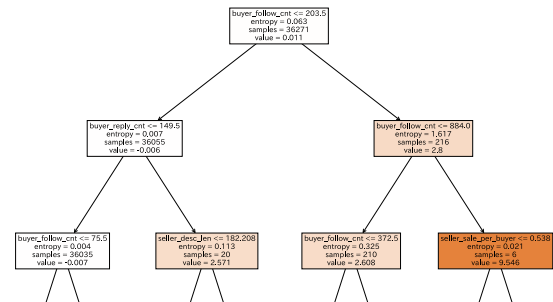


図 2: GAT の新規作家定着予測モデルにおける Self-loop 以外への重要度の回帰木

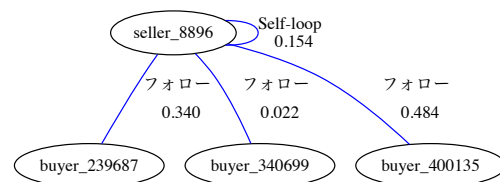


図 3: 新規作家定着予測における attention 値の例 1: Self-loop よりもフォローに大きな attention 値が算出された作家。接頭の “seller” は作家を表し, “buyer” は購入者を表す。なお, “seller” および “buyer” に付記された ID は研究での利用に際して振り直したもので、実際のユーザとは対応しない。

との “フォロー” エッジに大きな attention が算出されていることがわかる。このことから、GAT は、この作家の定着を予測するにあたっては作家自身の特徴量よりもフォローで繋がった 2 人の購入者の特徴量を重視したと言える。なお、この作家に付与された目的変数ラベルは “定着” であり、GAT が定着確率 0.926 と適切に予測を行えたのに対し、多層ニューラルネットワークは 0.463 と低い確率を予測してしまっている。この結果から、この作家は作家自身の特徴としては離脱しそうだが、フォローされた購入者の特徴量から定着と判断できる作家であったと考えられる。また、もう一つ着目すべき点として、接触した 3 人の購入者は全てフォローで繋がっているが、算出された attention 値には差があることが挙げられる。この状況下では、高い attention 値が算出された buyer_239687, buyer_400135 が作家の定着に寄与した購入者であると判断できる。

以上より、GAT には実運用上で 2 点の利点があると言える。一つは、予測対象ユーザの行動履歴を集約して作った特徴量のみを考慮するモデルでは捕捉できない個々の接触相手の特徴まで考慮したモデルであるため高い性能で予測を行えるという点である。もう一つは、図 3 で示したように個々のユーザとの接触単位で予測への寄与を解釈可能であるという点である。ユーザの定着に寄与したユーザを洗い出したり、シミュレーションのアプローチで離脱しそうなユーザに対するアイテムや他ユーザの推薦を行うときに参考にしたりすることもできる。

5. 結論と今後の展望

本研究では、ハンドメイド作品を売買できる CtoC プラットフォームを対象に、ユーザの時系列的な傾向の分析とユーザ定着・離脱予測を行った。

Graph Neural Networks をベースとした 2 つのモデルをユーザの定着・離脱予測に適用し、ベースライン手法との比較を行った。実験結果および構築したモデルの解釈を通して、予測対象となるユーザの特徴だけではなく接触相手の特徴にも有益な情報が含まれるとき、Graph Neural Network をベースとした手法が有効であることが示された。さらに、GAT を用いれば個々の定着・離脱予測に対して重要な意味を持った接触や接触相手を可視化できることも示した。

今後の展望としては、予測結果および解釈を利用し、実際に離脱を防止する施策を行い、その効果を検証することが主要な方向性である。解析の結果重要であった接触についての機能を充実させることや、個々のユーザに対してのリコメンデーションに離脱予測モデルを活用することでパーソナライズされた離脱防止施策を行い、実際の効果を評価することが必要であると考えられる。そのためには、モデルの解釈性を用いた分析を高度化し、定着・離脱と接触の因果を解明するアプローチが必要であると考えられる。

参考文献

- [Breiman 84] Breiman, L., Friedman, J., Olshen, R., and Stone, C.: Classification and regression trees. Wadsworth Int, Group, Vol. 37, No. 15, pp. 237–251 (1984)
- [Cox 58] Cox, D. R.: The regression analysis of binary sequences, *Journal of the Royal Statistical Society: Series B (Methodological)*, Vol. 20, No. 2, pp. 215–232 (1958)
- [Daly 02] Daly, J. L.: *Pricing for profitability: Activity-based pricing for competitive advantage*, Vol. 11, John Wiley & Sons (2002)
- [Drachen 18] Drachen, A., Pastor, M., Liu, A., Fontaine, D. J., Chang, Y., Runge, J., Sifa, R., and Klabjan, D.: To be or not to be... social: Incorporating simple social features in mobile game customer lifetime value predictions, in *Proceedings of the Australasian Computer Science Week Multiconference*, p. 40ACM (2018)
- [Gillen 05] Gillen, T.: *Winning new business in construction*, Gower Publishing, Ltd. (2005)
- [Hamilton 17] Hamilton, W., Ying, Z., and Leskovec, J.: Inductive representation learning on large graphs, in *Advances in Neural Information Processing Systems*, pp. 1024–1034 (2017)
- [Kipf 16] Kipf, T. N. and Welling, M.: Semi-supervised classification with graph convolutional networks, *arXiv preprint arXiv:1609.02907* (2016)
- [Kwon 19] Kwon, Y. D., Chatzopoulos, D., Haq, ul E., Wong, R. C.-W., and Hui, P.: GeoLifecycle: User Engagement of Geographical Exploration and Churn Prediction in LBSNs, *Proceedings of the ACM on Interactive,*

Mobile, Wearable and Ubiquitous Technologies, Vol. 3, No. 3, pp. 1–29 (2019)

- [Schlichtkrull 18] Schlichtkrull, M., Kipf, T. N., Bloem, P., Van Den Berg, R., Titov, I., and Welling, M.: Modeling relational data with graph convolutional networks, in *European Semantic Web Conference*, pp. 593–607Springer (2018)
- [Veličković 17] Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., and Bengio, Y.: Graph attention networks, *arXiv preprint arXiv:1710.10903* (2017)
- [Wu 19] Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., and Yu, P. S.: A comprehensive survey on graph neural networks, *arXiv preprint arXiv:1901.00596* (2019)
- [Yang 18] Yang, C., Shi, X., Jie, L., and Han, J.: I Know You'll Be Back: Interpretable New User Clustering and Churn Prediction on a Mobile Social Application, in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 914–922ACM (2018)
- [Ying 18] Ying, R., He, R., Chen, K., Eksombatchai, P., Hamilton, W. L., and Leskovec, J.: Graph convolutional neural networks for web-scale recommender systems, in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 974–983ACM (2018)
- [Zhou 18] Zhou, J., Cui, G., Zhang, Z., Yang, C., Liu, Z., Wang, L., Li, C., and Sun, M.: Graph neural networks: A review of methods and applications, *arXiv preprint arXiv:1812.08434* (2018)