

# 動的リンクフロー予測モデルと銀行送金ネットワーク

数理情報学専攻 48-226216

高橋 秀

指導教員

久野 遼平 講師

## 1 概要

送金データは口座を頂点、送金を辺とみなすことで動的ネットワークであるとみなせる。この送金ネットワークの将来予測は市場予測への応用が期待される [1]。本研究では大規模ネットワークにおける動的リンクフロー予測手法を構築し、またこれを実データである銀行送金データと仮想通貨送金データを用いて評価した。貿易ネットワークやインターバンク取引ネットワークなどの小さく密なネットワークと異なり、大規模ネットワークを対象とするため、提案モデルでは深層学習モデルである動的リンク予測モデルでの送金活動の集約と、Extreme Multi-Label 分類問題の手法である階層ソフトマックス木を使用した。また評価では、新規送金発生や既存送金消失の予測精度を確認するとともに、強力なベンチマークモデルである EdgeBank モデルに一部勝る性能を確認した。

## 2 背景

送金ネットワークから得られるネットワーク指標は経済の状況を反映していることが知られている。そのため、送金ネットワークの分析は実社会の経済状況のナウキャスト技術となりえる。ここでもし送金ネットワークの将来予測が可能であれば、現時点の経済状況にとどまらず、将来時点のネットワークを予測しそこから指標を生成することで、将来時点の経済状況の予測が可能になる。またウイルスの世界的な流行といったイベントがあったときに、イベントの情報を含まない範囲のデータから出力されるモデルの予測と実際のデータを比較することで、イベントの影響を評価することも可能である。このように送金ネットワークの将来予測は市場予測への応用が期待される (図 1)。

**動的リンクフロー予測**とは動的ネットワークにおいて辺の有無と辺重みの両方を予測することである。送金ネットワークのような大規模ネットワークにおいて動的リンクフロー予測を試みている先行研究は極めて少ない。貿易ネットワークやインターバンク取引ネットワークにおける先行研究では対象としているネット

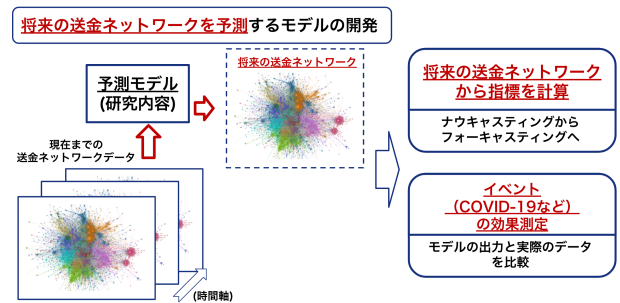


図 1. 提案手法の活用例。

ワークが小さく密であり、ベクトル自己回帰モデルや動的潜在空間モデルといったモデリング手法を用いている。また交通量予測の領域や既存の送金予測研究では辺の有無は既知とし、重みだけを予測している。

近年深層学習モデルが注目されており、ネットワーク領域においても深層学習モデルが存在する。しかし深層学習を用いた動的ネットワークモデルで主流となっている動的リンク予測モデルでは、辺の有無のみ予測しており、辺重みは辺特徴量の一種とされ予測対象とされない。

## 3 提案手法

本研究では大規模ネットワークの動的リンクフロー予測モデルを構築した。

**予測対象の分割** 動的リンクフロー予測とは、重み付き隣接行列の予測に他ならない。本研究では隣接行列の各行について、その行和と割合のそれぞれを予測する方法を取った。割合を予測するモデルを**送金割合予測モデル**、行和を予測するモデルを**総送金量予測モデル**と呼ぶ。

**動的リンク予測モデルの活用** 送金ネットワークには二次マルコフ性が見られることから、過去の送金・着金活動の情報を用いて将来の送金予測が可能であると推測できる。そこで本研究では動的リンク予測モデルである TGAT[2] を用いて、過去の送金・着金活動の情報を集約し、これを予測に用いることとした。

**階層ソフトマックス** 頂点数が  $O(10^4)$  程度の巨大な送金ネットワークへの応用を見据え、Extreme Multi-Label 分類問題の手法である階層ソフトマックス木を

導入し、計算量を削減するとともに予測精度向上を図った。

**損失関数の工夫** 送金割合予測モデルでは離散確率分布を予測する。この損失関数として、本研究ではクロスエントロピーではなく、ベクトルの各成分について二値クロスエントロピーを計算する方法を取った。これは真の送金割合が 0 の部分に関して陽に損失を計上することで、出力が疎に近づくことを目的としている。同様に出力を疎に近づけるため、正則化項としてエントロピーを加える工夫も行った。またインターバンク市場の先行研究に基づき、既存送金に関する損失に重みを付ける工夫も行った。

## 4 実験

評価には銀行送金データと、オープンデータである仮想通貨送金データを用いた。データの詳細は表 1 に示すとおりである。

表 1. データセットの概要

| データ      | ノード  | エッジ       | 直径 | 平均次数  | 密度    | 期間             | 頻度 |
|----------|------|-----------|----|-------|-------|----------------|----|
| bank100  | 90   | 100,522   | 3  | 39.7  | 32%   | 2019-04~2021-2 | 月  |
| bank500  | 607  | 949,562   | 4  | 120   | 12.3% | 2019-04~2021-2 | 月  |
| bank3000 | 2919 | 3,939,715 | 4  | 161.4 | 3.1%  | 2019-04~2021-2 | 月  |
| eth      | 500  | 20,612    | 7  | 7.38  | 0.8%  | 2018-04~2020-4 | 月  |

提案モデル (DLFlow) の他、ベンチマークモデルとして動的リンク予測モデルで提案されている EdgeBank モデル [3] を用意した。過去送金の平均を返すモデル (EdgeBank) と、過去 1 ヶ月の送金をそのまま返すモデル (EdgeBank<sub>tw</sub>) でも同様の評価を行った。また、提案モデルと EdgeBank モデルの結果の加重平均を取ったモデル (DLFlow<sub>comb</sub>) も評価対象とした。

## 5 結果・考察

表 2 に主結果を示す。割合予測においては混合モデルが EdgeBank モデルに優越するという結果になった。モデルの学習内容を調べるために、追加で評価を行った。まず、EdgeBank モデルと提案モデルの両方で予測値が非ゼロの部分に限定して割合予測のスコアを調べたところ、混合モデルのスコアはわずかに EdgeBank モデルに劣るといった結果になった。一方で EdgeBank モデルが送金無しと予測した部分に対する新規リンク予測精度や、EdgeBank モデルが送金ありと予測した部分に対するリンク消失予測精度を計測したところ、提案モデルは ROC-AUC において高いスコアを記録した (表 3)。このことから、提案モデルは既存送金の予測精

度は EdgeBank モデル程度であるものの、新規リンク予測やリンク切断予測において一定の性能を示しており、この点において混合モデルが EdgeBank モデルに優越していると考えられる。

表 2. 主結果.

| データ      | 指標  | EdgeBank <sub>tw</sub> | EdgeBank           | DLFlow             | DLFlow <sub>comb</sub>               |
|----------|-----|------------------------|--------------------|--------------------|--------------------------------------|
| bank100  | BCE | 4.38                   | 2.81               | 4.03               | <b>2.74</b>                          |
|          | CE  | 3.48                   | 2.07               | 3.09               | <b>2.01</b>                          |
|          | MAE | $8.02 \times 10^7$     | $13.6 \times 10^7$ | $9.18 \times 10^7$ | <b><math>7.59 \times 10^7</math></b> |
| bank500  | BCE | 5.52                   | 3.65               | 5.56               | <b>3.59</b>                          |
|          | CE  | 4.48                   | 2.83               | 4.61               | <b>2.78</b>                          |
|          | MAE | $4.75 \times 10^6$     | $12.0 \times 10^6$ | $6.19 \times 10^6$ | $4.85 \times 10^6$                   |
| bank3000 | BCE | 6.17                   | 3.99               | 7.13               | <b>3.97</b>                          |
|          | CE  | 5.04                   | 3.18               | 6.23               | <b>3.16</b>                          |
|          | MAE | $7.85 \times 10^5$     | $13.4 \times 10^5$ | $11.8 \times 10^5$ | <b><math>7.84 \times 10^5</math></b> |
| eth      | BCE | 13.79                  | 5.55               | 8.51               | <b>4.49</b>                          |
|          | CE  | 12.66                  | 4.31               | 6.85               | <b>3.65</b>                          |
|          | MAE | 0.57                   | <b>0.51</b>        | 0.75               | 0.62                                 |

表 3. 送金割合予測モデルの新規リンク予測及びリンク切断予測の精度.

| データ      | リンク予測 | リンク切断予測 |
|----------|-------|---------|
| bank100  | 0.778 | 0.929   |
| bank500  | 0.720 | 0.917   |
| bank3000 | 0.766 | 0.882   |
| eth      | 0.815 | 0.791   |

## 6 今後の課題

本研究では大規模な送金ネットワークに対し、深層学習を用いた動的リンクフロー予測モデルを提案した。このモデルは現実的な計算量で計算可能であり、また一部 EdgeBank モデルに優越している部分はあるものの、全体で見ても同等程度の性能しかない。EdgeBank モデルの出力をベースとして残渣を学習するようになるなど、実用化に向けてはさらなる工夫が必要であるといえる。

## 参考文献

- [1] 久野遼平, 長澤達也, 高橋秀, 近藤亮磨, 大西立顕. 銀行送金ネットワークの内在的構造と時間変化. 人工知能, Vol. 38, No. 2, pp. 131–138, 2023.
- [2] Da Xu, Chuanwei Ruan, Evren Körpeoglu, Sushant Kumar, and Kannan Achan. Inductive representation learning on temporal graphs. In *Proceedings of ICLR 2020*. OpenReview.net, 2020.
- [3] Farimah Poursafaei, Shenyang Huang, Kellin Pelrine, and Reihaneh Rabbany. Towards better evaluation for dynamic link prediction. In *Advances in Neural Information Processing Systems 35 (NeurIPS2022)*, pp. 32928–32941. Curran Associates, Inc., 2022.