

時系列グラフにおける着目ノードに特化したリンク予測

山口 寛人[†]大阪大学情報科学研究科[†]前川 政司[‡]大阪大学情報科学研究科[‡]佐々木 勇和[§]大阪大学情報科学研究科[§]鬼塚 真[¶]大阪大学情報科学研究科[¶]

1 はじめに

グラフのリンク予測は SNS 上の広告の推薦や将来の消費者の購買予測、道路の劣化予測などに利用可能であり、広く研究が行われている。リンク予測のユースケースは、道路の劣化予測のようにグラフ全体のリンク予測が必要なものと、ある顧客企業の SNS 上の企業アカウントのフォロー推薦のように特定の 1 ノードに対してリンクを予測するものの二つに分類することができる。

既存手法では、グラフ全体のリンク予測を目的とする深層学習モデルが提案されている [1, 2]。これらの手法はグラフ全体の構造情報や属性情報を学習することでグラフ中の任意の 2 ノード間のリンクを予測している。これらの深層学習モデルでは、既知のグラフ構造からノード間の類似度を示すような指標を計算し、それに基づいてリンクの存在を判断するヒューリスティックな手法 [3, 4] と比べて高精度なリンク予測が可能である。中でも、時系列グラフを活用したリンク予測手法 [2] はグラフ構造の時系列変化の傾向も学習することが可能であり、ある 1 時刻のグラフのみを利用してリンク予測を行う静的グラフのリンク予測手法と比べて高精度な予測が可能である。

一方で、1 ノードのリンク予測を目的とした深層学習モデルは存在しない。ただし、上記の既存手法を用いて特定の 1 ノードのリンクのみを予測対象とすることで、特定の 1 ノードのリンク予測を行うことは可能である。

そこで本研究では、静的グラフのリンク予測手法と時系列グラフのリンク予測手法の 1 ノードのリンク予測精度を比較することで、1 ノードのリンク予測における既存手法の予測精度を評価するとともに、1 ノードのリンク予測においても時系列グラフを活用することが精度向上に有用であるかを検証する。

2 問題定義

本研究では時系列グラフのリンク予測を行う。時系列グラフは $G = \{G^1, \dots, G^T\}$ で表され、 T はタイムステップ数とする。また、 $G^t = (V, E^t)$ はタイムステップ t のグラフを表す。 $V = \{v_1, \dots, v_n\}$ はノードセット、 $E^t \subseteq V \times V$ はタイムステップ t のリンクセットを表す。また、リンクを予測する特定の 1 ノードを着目ノードと呼ぶ。本研究では、タイムステップ $1 \sim t$ までの時系列グラフ $G = \{G^1, \dots, G^t\}$ とリンクを予測する特定の 1 ノードを入力として、タイムステップ $t+1$ のにおける着目ノードとその他の任意のノードとのリンクの存在を予測する。静的グラフのリンク予測手法を用いる際には、タイムステップ $1 \sim t$ までのリンクを 1 つのグラフに集約することで、時系列グラフを静的グラフに変換してモデルに入力する。

Dynamic Link prediction for a specified node

[†] Hiroto Yamaguchi, Osaka University[‡] Seiji Maekawa, Osaka University[§] Yuya Sasaki, Osaka University[¶] Makoto Onizuka, Osaka University

データセット	baseline	SEAL	DySAT
Enron	0.616	0.650	0.896

表1 着目ノードのリンク予測結果

3 比較手法

本研究では、静的グラフのリンク予測で最新技術である SEAL と時系列グラフのリンク予測で最新技術である DySAT の2つ手法を着目ノードのリンク予測に適用する。本章では、SEAL と DySAT の概要を紹介する。

3.1 SEAL

SEAL はグラフ中の任意の2ノード間のリンクを予測する静的グラフのリンク予測手法である。学習時には、任意の2ノードから h ホップのサブグラフを取得し、このサブグラフの構造情報や属性情報からその2ノード間にリンクが存在するかを予測するモデルを学習する。

3.2 DySAT

DySAT は自然言語処理で用いられている Transformer をベースにした時系列グラフのリンク予測手法である。構造と時系列の二次元で自己注意ネットワークを構成することで、構造的特徴と時系列変化の両方を考慮したリンク予測が可能である。

4 実験

本実験では、SEAL と DySAT の着目ノードのリンク予測精度を評価する。手法の評価のために、評価の基準となる baseline を設定する。これは、予測対象となるタイムステップのグラフ構造は直近のタイムステップのグラフ構造と類似している可能性が高いという観点から、直近のタイムステップの着目ノードのリンクを予測結果とする手法である。データセットは Enron^{*1} と呼ばれるヨーロッパの研究機関における電子メールの送受信履歴を表した時系列

グラフを用いる。評価指標にはリンク予測で一般的に用いられている AUC を用いる。

本実験の結果を表1に示す。静的グラフのリンク予測手法である SEAL では baseline と比較して 0.04 ほど AUC が高くなっていることが確認できる。このことから、深層学習を用いたリンク予測はヒューリスティックな手法に比べて高精度な予測が可能であることを示している。また、時系列グラフを活用したリンク予測手法である DySAT は baseline や SEAL と比較して高精度な予測を行うことができている。このことから、リンク予測において時系列グラフを活用することが精度向上に繋がっていることが確認できる。

謝辞

本研究は JSPS 科研費 JP20H00583 の助成を受けたものです。

参考文献

- [1] Palash Goyal, Sujit Rokka Chhetri, and Arquimedes Canedo. dyngraph2vec: Capturing network dynamics using dynamic graph representation learning. *Knowledge-Based Systems*, 187:104816, 2020.
- [2] Aravind Sankar et al. Dysat: Deep neural representation learning on dynamic graphs via self-attention networks. In *Proceedings of the WSDM*, page 519–527. Association for Computing Machinery, 2020.
- [3] Lada A Adamic et al. Friends and neighbors on the web. *Social Networks*, 25:211 – 230, 2003.
- [4] M. E. J. Newman. Clustering and preferential attachment in growing networks. *Phys. Rev. E*, 64:025102, 2001.

^{*1} <http://networkrepository.com/ia-enron-email-dynamic.php>