

反復機構による属性付き時系列グラフ予測の精度検証

山崎 翔平[†] 佐々木 勇和[†] Panagiotis Karras^{††} 鬼塚 真[†][†] 大阪大学大学院情報科学研究科 ^{††} Department of Computer Science, Aarhus University

1 はじめに

グラフ構造は現実世界の複雑な現象をモデル化する一般的なデータ表現である。実世界に存在するグラフ構造の多くはノードが属性を有し、また時間変化によってグラフ構造が変化する。このような時間変化を伴う属性付きのグラフは、属性付き時系列グラフと呼ばれる。将来の属性付き時系列グラフを予測する技術は多様な分野への応用が期待できるとされ大きな注目を集めている。特に、将来のグラフにおけるリンク関係を予測するリンク予測や、ノード属性を予測する属性予測は実世界での利用性の高さから多くの技術が提案されている。しかし、既存研究ではリンク予測と属性予測は独立に研究されており、属性付き時系列グラフに対して属性予測とリンク予測を同時に行った研究事例はない。また、これらのタスクでは共通して過去に観測されたノードのみを予測の対象としており、将来新たに出現するノードや消失するノードについては考慮していない。実世界の属性付き時系列グラフを予測する場合、リンクやノード属性の依存関係やノードの出現、あるいは消失する可能性も考慮しながらグラフの全体像を予測する必要がある。

そこで、本稿では属性付き時系列グラフに対してノード集合、リンク集合、ノード属性が互いに影響し合うことを仮定し、それら全てを同時に予測する問題に取り組む。この問題を解くための新たなグラフ予測技術として、個々の予測結果の相互作用を活用可能な反復機構を提案する。反復機構の初期値は任意の手法を用いて事前に実行されるノード予測、リンク予測、およびノード属性予測の結果である。反復機構では個々の予測結果を利用して、繰り返し予測することで精度の向上が可能である。実験では実世界のデータセットを用いて反復機構の有効性を示し、ノード予測、リンク予測、属性予測の全てにおいて高精度な予測結果を得られることを示す。

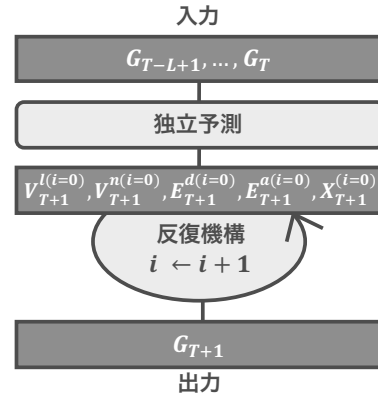


図1 提案手法概要

2 問題定義

グラフのノード集合 V 、リンク集合 E 、ノード属性 X が時間変化を伴う場合、タイムステップ t におけるグラフを $G_t := (V_t, E_t, X_t)$ と表記し、これを属性付き時系列グラフと呼ぶ。タイムステップ T から L ステップ前までの一連の属性付き時系列グラフ $\{G_{T-L+1}, \dots, G_T\}$ が与えられたとき、 $T+1$ におけるグラフ G_{T+1} を予測する。

予測対象は消失ノード $V_{T+1}^l (= V_T \setminus V_{T+1})$ 、出現ノード $V_{T+1}^n (= V_{T+1} \setminus V_T)$ 、消失リンク $E_{T+1}^d (= E_T \setminus E_{T+1} \setminus E_{T+1}^l)$ 、出現リンク $E_{T+1}^a (= E_{T+1} \setminus E_T \setminus E_{T+1}^n)$ 、およびノード属性とする。ただし、 E_{T+1}^l は消失ノードに接続されるリンク、 E_{T+1}^n は出現ノードに接続されるリンクである。

3 反復機構

本提案の反復機構ではノード集合、リンク集合、ノード属性が互いに影響し合うことを仮定し、個々の予測結果を活用して繰り返し予測を行うことで精度向上を狙う。ノード属性が隣接ノードから影響を受けるグラフの場合、事前にリンク集合を予測しておくことで属性予測の精度向上が期待できる。これは、リンク予測の結果を活用して属性予測を行うことで達成されるが、最初にリンク予測を独立に行っておく必要がある。このような、反復機構の初期値として事前に実行される個々の予測のことを独立予測と呼ぶ。反復機構ではそれぞれの独立予測の結果を活用して繰り返し予測を行い、グラフ全体の予測精度を向上させる。これらの予測の概要を図1に示す。観測された時系列長 L のグラフを入力とし将来1ステップのグラフを予測する。本提案では消失ノード、出現ノード、消失リンク、出

Does a repeat mechanism improve prediction accuracy on time-evolving attributed graphs?

[†] Shohei Yamasaki

[†] Yuya Sasaki

^{††} Panagiotis Karras

[†] Makoto Onizuka

Graduated School of Information Science and Technology, Osaka University ([†])

Department of Computer Science, Aarhus University (^{††})

表1 消失ノード予測, 消失リンク予測, および出現リンク予測タスクにおける反復機構の予測精度

Methods	消失ノード予測				出現リンク予測				消失リンク予測			
	ROC AUC		Average precision		ROC AUC		Average precision		ROC AUC		Average precision	
	NBA	Reddit	NBA	Reddit	NBA	Reddit	NBA	Reddit	NBA	Reddit	NBA	Reddit
Baseline	0.5000	0.5000	0.1871	0.5110	0.5000	0.5000	0.0137	0.0005	0.5000	0.5000	0.4136	0.3929
Random	0.5141	0.5138	0.1916	0.5181	0.4988	0.5116	0.0136	0.0005	0.5020	0.5027	0.4146	0.3942
FNN	0.8533	0.7750	0.5358	0.7456	0.4988	0.4838	0.0135	0.0005	0.4949	0.4876	0.4104	0.3777
LSTM	0.8547	0.8264	0.5394	0.8163	0.5003	0.8271	0.0135	0.0572	0.4887	0.7194	0.3999	0.5664
DynGEM	0.4731	0.6735	0.1824	0.6344	0.4945	0.4551	0.0136	0.0004	0.5106	0.4625	0.4253	0.3518
GCN	0.5743	0.7506	0.2453	0.7196	0.5014	0.8253	0.0136	0.0202	0.5062	0.6308	0.4093	0.4751
STGCN	0.6007	0.8222	0.2686	0.8072	0.4937	0.7359	0.0132	0.0011	0.4883	0.6530	0.4013	0.5039
EvolveGCN-H	0.5742	0.6131	0.2211	0.5791	0.4932	0.5652	0.0134	0.0007	0.4865	0.5109	0.4005	0.3982
EvolveGCN-O	0.5719	0.6460	0.2313	0.6052	0.4928	0.5110	0.0135	0.0005	0.4762	0.5252	0.3936	0.4116
TGGNN	0.8619	0.8284	0.5667	0.8195	0.5007	0.7664	0.0135	0.0014	0.5034	0.6218	0.4072	0.4741
Proposal	0.9520	0.8668	0.7889	0.8746	0.5114	0.8326	0.0140	0.0066	0.5135	0.7356	0.4663	0.5865

表2 データセット統計

	ノード数	リンク数	属性数	ステップ数
NBA	3,859	94,861	35	68
Reddit	8,077	22,375	300	16

現リンク, およびノード属性を同時に予測する汎用フレームワークである. 独立予測は他の予測結果に依存しない単体の予測タスクであり, GCN [1] 等の任意の既存手法を用いて予測する. 既存研究では, 時間的に変化するグラフ全体を予測するのではなく, 独立予測に焦点を当てているため, 個々の予測間の依存関係については研究していない. 本提案は反復機構を用いることで, 個々の予測間の依存関係をモデル化しているため, 予測タスクが別の予測タスクに与える影響を陽に知ることができるため, グラフ全体を高精度に予測可能である.

4 評価実験

4.1 実験設定

表2に示す NBA*と Reddit†を用いて実験を行い, 反復機構の有効性を検証する.

比較手法として, 観測時系列の中で最新のグラフを予測結果 (Baseline) とする手法およびランダム予測 (Random), FNN, LSTM, DynGEM [2], GCN [1], STGCN [3], EvolveGCN [4], TGGNN を用いる. TGGNN は STGCN の拡張手法であり, Spatial-Graph-Convolution ブロックとして GGNN [5] を採用している. 繰り返し予測を行う場合, 独立に予測した手法のうち最も精度が高かった手法を用いる.

4.2 結果

消失ノード予測, 消失リンク予測, および出現リンク予測タスクにおける反復機構の予測精度を表1に

示す. いずれの予測も二値分類タスクであることから, ROC 曲線の AUC 値および Precision-Recall 曲線の Average Precision を用いて各手法の性能を評価する. 提案手法は独立予測の中で最も良い精度の手法の結果を活用して繰り返し予測する. 消失ノード予測では NBA, Reddit ともに出現ノードの属性, およびそれに接続されるリンクの予測結果を活用して, TGGNN で予測したとき最も良い精度で予測しており, 新たに出現するノードを予測することで消失ノード予測の精度向上に影響を与えることが分かる. 出現・消失リンク予測についても反復機構を活用することで独立予測と同等以上の精度で予測している. しかし, Reddit における出現リンク予測では提案手法は独立予測の精度を向上させておらず, 反復機構が有効に機能していない. これは, Reddit データセットではノード属性が時間変化しないことや, 元々の独立予測の精度が低いことが原因として考えられる.

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] Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *ICLR*, 2017.
- [2] Palash Goyal, Nitin Kamra, Xinran He, and Yan Liu. DynGEM: Deep embedding method for dynamic graphs. In *Workshop on Representation Learning for Graphs.*, 2017.
- [3] Bing Yu, Haoteng Yin, and Zhanxing Zhu. Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. In *IJCAI*, 2018.
- [4] Pareja et al. EvolveGCN: Evolving graph convolutional networks for dynamic graphs. In *AAAI*, pages 5363–5370, 2020.
- [5] Yujia Li, Daniel Tarlow, Marc Brockschmidt, and Richard Zemel. Gated graph sequence neural networks. In *ICLR*, 2016.

* <https://www.basketball-reference.com/>

† <http://snap.stanford.edu/data/soc-RedditHyperlinks.html>