A Fuzzy Nesterov-Accelerated Nonnegative Latent Factorization of Tensors Model for Efficient Representation to Dynamic Directed Graph Supplementary File

Anonymous

I. INTRODUCTION

This is the supplementary file for the paper entitled "A Fuzzy Nesterov-Accelerated Nonnegative Latent Factorization of tensors Model for Efficient Representation to Dynamic Directed Graph". We have put the proof of Lemma 1-2 and Theorem 1-2 in Section II, detailed discussion of related work in Section III and a table of experimental results in Section IV.

II. MODEL CONVERGENCE ANALYSIS

For the convenience of analysis, we alternatively fix the counterpart of the active parameter as constant, i.e., we treat $\boldsymbol{w}_j$ and $\boldsymbol{z}_k$ as constant when performing the analyses with $\boldsymbol{u}_i$. Note that the update of xi by NASGD is given as: For the t-th iteration counts.

$$\begin{aligned}
&t=1: u_i^{t+1}=u_i^t-\eta\nabla\varepsilon_{i,j,k}\left(u_i^t\right)-\gamma\eta\nabla\varepsilon_{i,j,k}\left(u_i^t\right)\\
&t\geq 2: u_i^{t+1}=u_i^t-\eta\nabla\varepsilon_{i,j,k}\left(u_i^t\right)\\
&\qquad+\gamma\left(\left(u_i^t-\eta\nabla\varepsilon_{i,j,k}\left(u_i^t\right)\right)-\left(u_i^{t-1}-\eta\nabla\varepsilon_{i,j,k}\left(u_i^{t-1}\right)\right)\right).
\end{aligned} \tag{S1}$$

We define $\boldsymbol{\omega}^t$ as:

$$\boldsymbol{\omega}^t=\begin{cases}\dfrac{\gamma}{1-\gamma}\left(\boldsymbol{u}_i^t-\boldsymbol{u}_i^{t-1}+\eta\nabla\varepsilon_{i,j,k}\left(\boldsymbol{u}_i^{t-1}\right)\right), t\geq 1\\ 0, \qquad\qquad , t=0\end{cases}. \tag{S2}$$

Thus, we present the following results.

Lemma 1: $\forall t\geq 0$,the following equation stands:

$$\begin{aligned}
\boldsymbol{u}_i^{t+1}+\boldsymbol{\omega}^{t+1}&=\boldsymbol{u}_i^{t+1}+\frac{\gamma}{1-\gamma}\left(\boldsymbol{u}_i^{t+1}-\boldsymbol{u}_i^t+\eta\nabla\varepsilon_{i,j,k}\left(\boldsymbol{u}_i^t\right)\right)=\frac{1}{1-\gamma}\boldsymbol{u}_i^{t+1}+\frac{\gamma}{1-\gamma}\left(\eta\nabla_{\boldsymbol{u}_i}\varepsilon_{i,j,k}\left(\boldsymbol{u}_i^t\right)-\boldsymbol{u}_i^t\right)\\
&=\frac{1}{1-\gamma}\left(\boldsymbol{u}_i^t-\eta\nabla_{\boldsymbol{u}_i}\varepsilon_{i,j,k}\left(\boldsymbol{u}_i^t\right)+\gamma\left[\left(\boldsymbol{u}_i^t-\eta\nabla_{\boldsymbol{u}_i}\varepsilon_{i,j,k}\left(\boldsymbol{u}_i^t\right)\right)-\left(\boldsymbol{u}_i^{t-1}-\eta\nabla_{\boldsymbol{u}_i}\varepsilon_{i,j,k}\left(\boldsymbol{u}_i^{t-1}\right)\right)\right]\right)-\frac{\gamma}{1-\gamma}\left(\boldsymbol{u}_i^t-\eta\nabla_{\boldsymbol{u}_i}\varepsilon_{i,j,k}\left(\boldsymbol{u}_i^t\right)\right)\\
&=\frac{1}{1-\gamma}\left(\boldsymbol{u}_i^t-\eta\nabla_{\boldsymbol{u}_i}\varepsilon_{i,j,k}\left(\boldsymbol{u}_i^t\right)\right)+\frac{\gamma}{1-\gamma}\left(\boldsymbol{u}_i^t-\eta\nabla_{\boldsymbol{u}_i}\varepsilon_{i,j,k}\left(\boldsymbol{u}_i^t\right)\right)-\frac{\gamma}{1-\gamma}\left(\boldsymbol{u}_i^{t-1}-\eta\nabla_{\boldsymbol{u}_i}\varepsilon_{i,j,k}\left(\boldsymbol{u}_i^{t-1}\right)\right)-\frac{\gamma}{1-\gamma}\left(\boldsymbol{u}_i^t-\eta\nabla_{\boldsymbol{u}_i}\varepsilon_{i,j,k}\left(\boldsymbol{u}_i^t\right)\right)\\
&=\frac{1}{1-\gamma}\left(\boldsymbol{u}_i^t-\eta\nabla_{\boldsymbol{u}_i}\varepsilon_{i,j,k}\left(\boldsymbol{u}_i^t\right)\right)-\frac{\gamma}{1-\gamma}\left(\boldsymbol{u}_i^{t-1}-\eta\nabla_{\boldsymbol{u}_i}\varepsilon_{i,j,k}\left(\boldsymbol{u}_i^{t-1}\right)\right)\\
&=\frac{1}{1-\gamma}\boldsymbol{u}_i^t-\frac{\gamma}{1-\gamma}\boldsymbol{u}_i^t-\frac{1}{1-\gamma}\eta\nabla_{\boldsymbol{u}_i}\varepsilon_{i,j,k}\left(\boldsymbol{u}_i^t\right)-\frac{\gamma}{1-\gamma}\boldsymbol{u}_i^{t-1}+\frac{\gamma}{1-\gamma}\eta\nabla_{\boldsymbol{u}_i}\varepsilon_{i,j,k}\left(\boldsymbol{u}_i^{t-1}\right)+\frac{\gamma}{1-\gamma}\boldsymbol{u}_i^t\\
&=\boldsymbol{u}_i^t+\frac{\gamma}{1-\gamma}\left(\boldsymbol{u}_i^t-\boldsymbol{u}_i^{t-1}+\eta\nabla_{\boldsymbol{u}_i}\varepsilon_{i,j,k}\left(\boldsymbol{u}_i^{t-1}\right)\right)-\frac{1}{1-\gamma}\eta\nabla_{\boldsymbol{u}_i}\varepsilon_{i,j,k}\left(\boldsymbol{u}_i^t\right)\\
&=\boldsymbol{u}_i^t+\boldsymbol{\omega}^t-\frac{\eta}{1-\gamma}\nabla_{\boldsymbol{u}_i}\varepsilon_{i,j.k}\left(\boldsymbol{u}_i^t\right).
\end{aligned} \tag{S3}$$

Lemma 2: Let $\boldsymbol{u}_i^{-1}=\boldsymbol{u}_i^0$, for any $t\geq 0$, we have:

$$\begin{aligned}\left\|\boldsymbol{u}_i^{t+1}+\boldsymbol{\omega}^{t+1}-\boldsymbol{u}_i\right\|^2&=\left\|\boldsymbol{u}_i^t+\boldsymbol{\omega}^t-\frac{\eta}{1-\gamma}\nabla_{\boldsymbol{u}_i}\varepsilon_{i,j,k}\left(\boldsymbol{u}_i^t\right)-\boldsymbol{u}_i\right\|^2\\&=\left\|\boldsymbol{u}_i^t+\boldsymbol{\omega}^t-\boldsymbol{u}_i\right\|^2-\frac{2\eta}{1-\gamma}\left\langle\boldsymbol{u}_i^t+\boldsymbol{\omega}^t-\boldsymbol{u}_i,\nabla_{\boldsymbol{u}_i}\varepsilon_{i,j,k}\left(\boldsymbol{u}_i^t\right)\right\rangle+\left(\frac{\eta}{1-\gamma}\right)^2\left\|\nabla_{\boldsymbol{u}_i}\varepsilon_{ijk}\left(\boldsymbol{u}_i^t\right)\right\|^2\\&=\left\|\boldsymbol{u}_i^t+\boldsymbol{\omega}^t-\boldsymbol{u}_i\right\|^2-\frac{2\eta}{1-\gamma}\left\langle\boldsymbol{u}_i^t+\frac{\gamma}{1-\gamma}\left(\boldsymbol{u}_i^t-\boldsymbol{u}_i^{t-1}+\eta\nabla_{\boldsymbol{u}_i}\varepsilon_{i,j,k}\left(\boldsymbol{u}_i^{t-1}\right)\right)-\boldsymbol{u}_i,\nabla_{\boldsymbol{u}_i}\varepsilon_{i,j,k}\left(\boldsymbol{u}_i^t\right)\right\rangle+\left(\frac{\eta}{1-\gamma}\right)^2\left\|\nabla_{\boldsymbol{u}_i}\varepsilon_{i,j,k}\left(\boldsymbol{u}_i^t\right)\right\|^2\\&=\left\|\boldsymbol{u}_i^t+\boldsymbol{\omega}^t-\boldsymbol{u}_i\right\|^2-\frac{2\eta}{1-\gamma}\left\langle\boldsymbol{u}_i^t-\boldsymbol{u}_i,\nabla_{\boldsymbol{u}_i}\varepsilon_{i,j,k}\left(\boldsymbol{u}_i^t\right)\right\rangle-\frac{2\eta\gamma}{(1-\gamma)^2}\left\langle\boldsymbol{u}_i^t-\boldsymbol{u}_i^{t-1},\nabla_{\boldsymbol{u}_i}\varepsilon_{i,j,k}\left(\boldsymbol{u}_i^t\right)\right\rangle\\&\quad-\frac{2\eta^2\gamma}{1-\gamma}\left\langle\nabla_{\boldsymbol{u}_i}\varepsilon_{ijk}\left(\boldsymbol{u}_i^{t-1}\right),\nabla_{\boldsymbol{u}_i}\varepsilon_{i,j,k}\left(\boldsymbol{u}_i^t\right)\right\rangle+\left(\frac{\eta}{1-\gamma}\right)^2\left\|\nabla_{\boldsymbol{u}_i}\varepsilon_{i,j,k}\left(\boldsymbol{u}_i^t\right)\right\|^2.\end{aligned}\tag{S4}$$

According to the properties of a convex function, the following inequality can be achieved.

$$\begin{aligned}\left\|\boldsymbol{u}_i^{t+1}+\boldsymbol{\omega}^{t+1}-\boldsymbol{u}_i\right\|^2&=\left\|\boldsymbol{u}_i^t+\boldsymbol{\omega}^t-\boldsymbol{u}_i\right\|^2-\frac{2\eta}{1-\gamma}\left\langle\boldsymbol{u}_i^t-\boldsymbol{u}_i,\nabla_{\boldsymbol{u}_i}\varepsilon_{i,j,k}\left(\boldsymbol{u}_i^t\right)\right\rangle-\frac{2\eta\gamma}{(1-\gamma)^2}\left\langle\boldsymbol{u}_i^t-\boldsymbol{u}_i^{t-1},\nabla_{\boldsymbol{u}_i}\varepsilon_{i,j,k}\left(\boldsymbol{u}_i^t\right)\right\rangle\\&\quad-\frac{2\eta^2\gamma}{1-\gamma}\left\langle\nabla_{\boldsymbol{u}_i}\varepsilon_{i,j,k}\left(\boldsymbol{u}_i^{t-1}\right),\nabla_{\boldsymbol{u}_i}\varepsilon_{i,j,k}\left(\boldsymbol{u}_i^t\right)\right\rangle+\left(\frac{\eta}{1-\gamma}\right)^2\left\|\nabla_{\boldsymbol{u}_i}\varepsilon_{i,j,k}\left(\boldsymbol{u}_i^t\right)\right\|^2\\&\le\left\|\boldsymbol{u}_i^t+\boldsymbol{\omega}^t-\boldsymbol{u}_i\right\|^2-\frac{2\eta}{1-\gamma}\left(\varepsilon_{i,j,k}\left(\boldsymbol{u}_i^t\right)-\varepsilon_{i,j,k}\left(\boldsymbol{u}_i\right)\right)-\frac{2\eta\gamma}{(1-\gamma)^2}\left(\varepsilon_{i,j,k}\left(\boldsymbol{u}_i^t\right)-\varepsilon_{i,j,k}\left(\boldsymbol{u}_i^{t-1}\right)\right)+\left(\frac{\eta}{1-\gamma}\right)^2(2\gamma+1).\end{aligned}\tag{S5}$$

Note that (S5) can yield the appearance of $\boldsymbol{u}_i$ when $t = 0$. By setting $\boldsymbol{u}_i^{-1}=\boldsymbol{u}_i^0$ the above inequality still holds. Hence, *Lemma2* stands.

Based on *Lemmas* 1-2, we present the following important result.

Theorem 1 (Convergence of NASGD): $\forall t\ge 0$, let $\left\|\nabla\varepsilon_{ijk}\left(\boldsymbol{u}_i\right)\right\|\le G$ as G be a positive constant, by setting $\eta=L/\sqrt{T+1}$ as L be a positive constant, When $t\in\{0,\cdots,T\}$, the formula (25) is cumulatively summed to get:

$$\frac{2\eta}{1-\gamma}\sum\nolimits_{t=0}^{T}\left(\varepsilon_{i,j,k}\left(\boldsymbol{u}_i^t\right)-\varepsilon_{i,j,k}\left(\boldsymbol{u}_i\right)\right)\le\frac{2\eta}{(1-\gamma)^2}\left(\varepsilon_{i,j,k}\left(\boldsymbol{u}_i^0\right)-\varepsilon_{i,j,k}\left(\boldsymbol{u}_i^T\right)\right)+\left\|\boldsymbol{u}_i^0-\boldsymbol{u}_i\right\|^2+\left(\frac{\eta}{1-\gamma}\right)^2(2\gamma+1)(T+1)G^2.\tag{S6}$$

Since $\boldsymbol{u}_i=\boldsymbol{u}_i^*$, and $\varepsilon_{i,j,k}(\boldsymbol{u}_i^t)\ge\varepsilon_{i,j,k}(\boldsymbol{u}_i^*)$,we get:

$$\sum\nolimits_{t=0}^{T}\left(\varepsilon_{i,j,k}\left(\boldsymbol{u}_i^t\right)-\varepsilon_{i,j,k}\left(\boldsymbol{u}_i\right)\right)\le\frac{\gamma}{(1-\gamma)^2}\left(\varepsilon_{i,j,k}\left(\boldsymbol{u}_i^0\right)-\varepsilon_{i,j,k}\left(\boldsymbol{u}_i^*\right)\right)+\frac{1-\gamma}{2\eta}\left\|\boldsymbol{u}_i^0-\boldsymbol{u}_i^*\right\|^2+\frac{\eta}{2(1-\gamma)}(2\gamma+1)(T+1)G^2.\tag{S7}$$

Let $\bar{\boldsymbol{u}}_i=\sum\nolimits_{t=0}^{T}\boldsymbol{u}_i^t/(T+1)$, for a convex function $\varepsilon_{i,j,k}(\boldsymbol{u}_i)$ have

$$\varepsilon_{i,j,k}\left(\bar{\boldsymbol{u}}_i\right)-\varepsilon_{i,j,k}\left(\boldsymbol{u}_i^*\right)\le\frac{\gamma}{(1-\gamma)(T+1)}\left(\varepsilon_{i,j,k}\left(\boldsymbol{u}_i^0\right)-\varepsilon_{i,j,k}\left(\boldsymbol{u}_i^*\right)\right)+\frac{1-\gamma}{2\eta}\left\|\boldsymbol{u}_i^0-\boldsymbol{u}_i^*\right\|^2+\frac{\eta}{2(1-\gamma)}(2\gamma+1)(T+1)G^2.\tag{S8}$$

The proof of Theorem 1 can be done by plugging in the value of the learning rate η. According to the same principle, $\varepsilon_{i,j,k}(\boldsymbol{w}_j)$ converges by training $\boldsymbol{w}_j$ by fixing $\boldsymbol{u}_i$ and $\boldsymbol{z}_k$ as a constant, and $\varepsilon_{ijk}(\boldsymbol{z}_k)$ converges by training $\boldsymbol{z}_k$ by fixing $\boldsymbol{u}_i$ and $\boldsymbol{w}_j$ as a constant. Therefore, Theorem 1 stands.

III. Related Work

Until now, various of graph representation learning models emerge, aiming at acquiring valuable knowledge by representing a given graph into low-dimensional space [1,2]. Considering a static undirected graph, Yang *et. al.* [3] introduce a node centrality-based representation method to address the binary link prediction problem, which combines various node centrality metrics to extract features related to nodes. Wang *et. al.* [4] introduce a heterogeneous network represent approach based on the concept of adversarial neural networks, where a generator and a discriminator are used to represent a heterogeneous social network in concert. Li *et. al.* [5] proposes an hTransM method based on knowledge graph embedding, which realizes the prediction of the missing parts of triples in the knowledge graph by defining a hierarchy-constrained margin. Fu *et. al.* [6] propose a heterogeneous attributed networks embedding approach, which leverages information from both structural space and content space to capture data from two different perspectives.

In many practical applications, a dynamic graph is frequently encountered. and corresponding representation learning method is designed. Liu *et. al.* [7] propose a method based on graph transformer for processing edge behaviors, which combines specific feature generation, decoding, and loss function settings to achieve the prediction of links in dynamic networks. Qin *et al*. [8] propose an inductive dynamic embedding aggregation model via combining multiple objectives regarding the scale difference minimization and error minimization for predicting dynamic weighted links. Li *et al*. [9] propose a type-aware anchor representation learning method that uses a two-layer attention architecture to combine type information and fusion information to extract the feature representation of user nodes. Li *et al*. [10] propose an dynamic graph neural networks method, which is based on reinforcement learning and uses a time-aware attentional aggregating module and a reinforced neighbor selection module to adaptively determine node updates.

Although the above representation learning models can represent a given graph, they are facing challenges in generalization and computation when representing a DDG. In contrast, the proposed FNL model focuses on the nonlinear modeling and high convergence rate, which fuses nonlinear activation function into learning objective to represent the nonlinearity hidden in DDG and designs a fuzzy Nesterov-accelerated parameter learning scheme, thereby achieving significantly higher generalization and efficiency.

IV. Experimental results

TABLE S1

Hyper-parameter Setting of M2-8 on D1-6

Datasets	**Hyper-parameter Setting**						
D1	M2: R=10, $\lambda=10^{-5}$, $\eta=0.001$	M3: R=10, $\lambda=10^{-5}$ $\eta=0.001$	M4: R=10, $\lambda=4\times10^{-8}$ $\gamma=480$	M5: R=10, $\lambda1=0.04$, $\lambda2=0.2$	M6: R=10, $\lambda1=0.5$, $\lambda2=0.0625$	M7: R=10, K=3 $\eta=0.1$, $\lambda=0.1$	M8: R=10, K=3, $\eta=0.1$, $\lambda=0.001$
D2	M2: R=10, $\lambda=10^{-5}$, $\eta=0.001$	M3: R=10, $\lambda=10^{-5}$, $\eta=0.002$	M4: R=10, $\lambda=10^{-8}$ $\gamma=500$	M5: R=10, $\lambda1=0.04$, $\lambda2=0.2$	M6: R=10, $\lambda1=0.125$, $\lambda2=0.25$	M7: R=10, K =3 $\eta=0.2$, $\lambda=0.1$	M8: R=10, K=3, $\eta=0.1$, $\lambda=0.0001$
D3	M2: R=10, $\lambda=10^{-5}$, $\eta=0.001$	M3: R=10, $\lambda=10^{-5}$, $\eta=0.004$	M4: R=10, $\lambda=10^{-8}$ $\gamma=450$	M5: R=10, $\lambda1=0.005$, $\lambda2=0.4$	M6: R=10, $\lambda1=0.25$, $\lambda2=0.0625$	M7: R=10, K =3 $\eta=0.1$, $\lambda=0.2$	M8: R=10, K=3, $\eta=0.2$, $\lambda=0.01$
D4	M2: R=10, $\lambda=10^{-5}$, $\eta=0.002$	M3: R=10, $\lambda=10^{-5}$, $\eta=0.005$	M4: R=10, $\lambda=4\times10^{-8}$ $\gamma=500$	M5: R=10, $\lambda1=0.007$, $\lambda2=0.4$	M6: R=10, $\lambda1=0.5$, $\lambda2=0.0625$	M7: R=10, K =3 $\eta=0.1$, $\lambda=0.1$	M8: R=10, K=3, $\eta=0.2$, $\lambda=0.001$
D5	M2: R=10, $\lambda=10^{-4}$, $\eta=0.001$	M3: R=10, $\lambda=10^{-4}$, $\eta=0.008$	M4: R=10, $\lambda=10^{-7}$ $\gamma=480$	M5: R=10, $\lambda1=0.07$, $\lambda2=0.3$	M6: R=10, $\lambda1=0.5$, $\lambda2=0.0625$	M7: R=10, K =3 $\eta=0.2$, $\lambda=0.2$	M8: R=10, K=3, $\eta=0.1$, $\lambda=0.002$
D6	M2: R=10, $\lambda=10^{-5}$, $\eta=0.001$	M3: R=10, $\lambda=10^{-5}$, $\eta=0.004$	M4: R=10, $\lambda=10^{-8}$ $\gamma=450$	M5: R=10, $\lambda1=0.006$, $\lambda2=0.5$	M6: R=10, $\lambda1=0.25$, $\lambda2=0.03125$	M7: R=10, K =3 $\eta=0.1$, $\lambda=0.1$	M8: R=10, K=3, $\eta=0.2$, $\lambda=0.01$

References

[1] X. Luo, M. Zhou, S. Li, and M. Shang, “An inherently nonnegative latent factor model for high-dimensional and sparse matrices from industrial applications,” *IEEE Transactions on Industrial Informatics*, vol. 14, no. 5, pp. 2011–2022, 2018.

[2] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge univer sity press, 2004.

[3] C. Yang, C. Wang, Y. Lu, X. Gong, C. Shi, W. Wang, and X. Zhang, “Few-shot link prediction in dynamic networks,” *in Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining*, p. 1245–1255, 2022.

[4] H. Wang, Z. Cui, R. Liu, L. Fang, and Y. Sha, “A multi-type transferable method for missing link prediction in heterogeneous social networks,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 11, pp. 10981–10991, 2023.

[5] M. Li, Y. Wang, D. Zhang, Y. Jia, and X. Cheng, “Link prediction in knowledge graphs: A hierarchy-constrained approach,” *IEEE Transactions on Big Data*, vol. 8, no. 3, pp. 630–643, 2022.

[6] Y. Fu, X. Yu, Y. Wu, X. Ding, and S. Zhao, “Robust representation learning for heterogeneous attributed networks,” *Information Sciences*, vol. 628, pp. 22–49, 2023.

[7] M. Liu, Z. Tu, T. Su, X. Wang, X. Xu, and Z. Wang, “Behaviornet: A f ine-grained behavior-aware network for dynamic link prediction,” *ACM Trans. Web*, vol. 18, 2024.

[8] M. Qin, C. Zhang, B. Bai, G. Zhang, and D.-Y. Yeung, “High-quality temporal link prediction for weighted dynamic graphs via inductive embedding aggregation,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 9, pp. 9378–9393, 2023.

[9] X. Li, Y. Shang, Y. Cao, Y. Li, J. Tan, and Y. Liu, “Type-aware anchor link prediction across heterogeneous networks based on graph attention network,” *in Proceedings of AAAI Conference on Artificial Intelligence* vol. 34, pp. 147–155, 2020.

[10] H. Li, C. Li, K. Feng, Y. Yuan, G. Wang, and H. Zha, “Robust knowledge adaptation for dynamic graph neural networks,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 36, no. 11, pp. 6920–6933, 2024.