 Vol.2019-AVM-107 No.21
2019/12/6

ネットワーク・ワードエンベディングのための負値残差低減および半直交制約付き非負値行列分解

橋本 陸[1]　笠井 裕之[2]

Abstract：Network embedding is intended to produce low-dimensional vector representations of nodes in a network to preserve and extract the latent network structure, which has higher robustness to noise, outliers, and redundant data. Although a recently proposed multi-level nonnegative matrix factorization (NMF)-based approach has exhibited superior performance on network analysis, it is adversely affected by performance degradation because of discarded negative residual and redundant base selection throughout sequential multiple factorization processes. To alleviate this shortcoming, this paper presents a proposal of a sequential semi-orthogonal NMF with negative residual reduction for the network embedding (SSO-NRR-NMF). The proposed approach reduces the negative residuals to be discarded, and avoids redundant bases with a semi-orthogonal constraint.

Sequential semi-orthogonal multi-level NMF with negative residual reduction for network embedding

1. Introduction

Many practical data systems use *network structured data*, which can include web page networks, social networks, road traffic infrastructure, biological networks, and information networks. It is challenging to handle such network data effectively for analytical tasks such as link prediction, node classification, node recommendation, and network visualization. Classical topology-based network representation techniques are hindered by onerous bottlenecks encountered in handling large-scale and high-dimensional network data because they handle an input *adjacency matrix* directly and because they are adversely affected by noise, outliers, and redundant data. By contrast, *network embedding* (NE) has come to be a promising approach that seeks *low-dimensional* vector representations of nodes in a network to preserve and extract its latent network structure efficiently [1], [2], [3], [4], [5], [6], [7]. Actually, NE alleviates such problems through low-dimensional representation while preserving the original intrinsic structure. Hence, it allows various *off-the-shelf* machine learning tools to apply this representation directly.

One recent advanced NE has been achieved in approaches attempting to *factorize* a given designed matrix to obtain low-dimensional representation assuming that the node connectivity matrix is globally *low-rank* [7]. It is, however, not always true when the matrix consists of a complex structure. This structure hinders ineffective representations from capturing all the observed connectivity patterns. In this regard, a novel multi-level NE framework (BoostNE) using *nonnegative matrix factorizations* (NMFs) has been proposed. The framework learns multiple embedding representations with different granularities, i.e., *globally and locally* low-rankness [8]. However, its *sequential* multi-level embedding discards *negative residuals* to enforce residual matrices that are nonnegative at the succeeding level. It also does not consider mutually redundant bases across multiple levels. To alleviate these problems, building on BoostNE, this paper presents a proposal of a sequential semi-orthogonal NMF with negative

[1] 電気通信大学情報理工学域
[2] 早稲田大学基幹理工学部情報通信学科

residual reduction, designated as SSO-NRR-NMF. Convergence analysis of SSO-NRR-NMF is also given. Numerical evaluations illustrate the effectiveness of SSO-NRR-NMF when used with several real-world datasets.

Throughout the paper, we represent scalars as lower-case letters $(a, b, \ldots)$, vectors as bold typeface lower-case letters $(\boldsymbol{a}, \boldsymbol{b}, \ldots)$, and matrices as bold typeface capitals $(\mathbf{A}, \mathbf{B}, \ldots)$. An element at (i, j) of a matrix $\mathbf{A}$ is represented as $[\mathbf{A}]_{i,j}$. Herein, $\boldsymbol{a}^l$ and $\boldsymbol{a}_k$ represent the l-th *row* vector and the k-th *column* vector of $\mathbf{A}$. The Frobenius norm of a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ is defined as $\|\mathbf{A}\|_F = \sqrt{\sum_{m,n}[\mathbf{A}]_{m,n}^2}$. Operators $\mathrm{Tr}(\cdot)$ and $(\cdot)^T$ respectively stand for the matrix trace and transpose. Operator $\max(a, b)$ outputs a when $a \geq b$, and b otherwise. Operator $\min(a, b)$ is opposite. $\odot$ represents the elemental-wise product. $\mathbb{R}_+^d$ represents d-dimensional nonnegative subspace.

2. NMF and BoostNE

2.1 Nonnegative matrix factorization (NMF)

Nonnegative matrix factorization (NMF) approximates a nonnegative matrix $\mathbf{X} \in \mathbb{R}_+^{m \times n}$ with a product of two nonnegative *factor matrices* $\mathbf{U} \in \mathbb{R}_+^{m \times r}$ and $\mathbf{V} \in \mathbb{R}_+^{n \times r}$ as $\mathbf{X} \approx \mathbf{U}\mathbf{V}^T$. Actually, r is usually chosen such that $r \ll \min\{m, n\}$, i.e., $\mathbf{X}$ is approximated in the two *low-rank* matrices. Consequently, this problem is formulated as a constrained minimization problem in terms of the Euclidean distance as

$$\min_{\mathbf{U},\mathbf{V}} \frac{1}{2}\|\mathbf{X} - \mathbf{U}\mathbf{V}^T\|_F^2, \quad \text{s.t.} \quad \mathbf{U} \geq 0, \mathbf{V} \geq 0. \tag{1}$$

Because Problem (1) is a *non-convex* optimization problem, finding its global minimum is NP-hard. For this problem, *multiplicative update* (MU) provides a simple but effective calculation algorithm [9], which is formulated as $\mathbf{V} \leftarrow \mathbf{V} \odot \frac{\mathbf{X}^T\mathbf{U}}{\mathbf{V}\mathbf{U}^T\mathbf{U}}$ and $\mathbf{U} \leftarrow \mathbf{U} \odot \frac{\mathbf{X}\mathbf{V}}{\mathbf{U}\mathbf{V}^T\mathbf{V}}$, where $\odot$ (resp. $\frac{\cdot}{\cdot}$) denotes the component-wise product (resp. division) of matrices.

2.2 Boosted network embedding (BoostNE) [8]

BoostNE addresses the fact that *matrices representing many real-world network data do not always have a low-rank structure*. Hence, it approximates an input network matrix by *multiple* matrices. Concretely, the approximation error generated by one NMF process is approximated recursively by another subsequent NMF. This leverages a power of *gradient boosting* [10]. Finally, we obtain multiple bases from a coarser characteristic to a finer characteristic, resulting in higher classification accuracy than that of others [8]. More specifically, denoting the index of the *boosting level* as $k (\in [K])$, where K stands for the total number of the levels, BoostNE seeks multiple nonnegative bases that can express $\mathbf{X}$ via multiple NMF processes. Given an input network matrix $\mathbf{X} \in \mathbb{R}_+^{m \times n}$, the mathematical definition is formulated using $\mathbf{U}^{(k)} \in \mathbb{R}_+^{m \times r_k}$ and $\mathbf{V}^{(k)} \in \mathbb{R}_+^{n \times r_k} (r_k \ll \min\{m, n\})$ as

$$\min_{\{\mathbf{U}^{(k)},\mathbf{V}^{(k)}\}_{k=1}^K} \sum_{k=1}^{K} \frac{1}{2}\|\mathbf{R}^{(k)} - \mathbf{U}^{(k)}\mathbf{V}^{(k)^T}\|_F^2, \tag{2}$$

where $\mathbf{R}^{(k)} \in \mathbb{R}_+^{m \times n}$ represents *residual elements* after the $(k-1)$-th level of the NMF process, which is defined as

$$\mathbf{R}^{(k)} = \begin{cases} \mathbf{X} & \text{if } k = 1 \\ \max(\mathbf{R}^{(k-1)} - \mathbf{U}^{(k-1)}\mathbf{V}^{(k-1)^T}, 0) & \text{if } k \geq 2. \end{cases} \tag{3}$$

Without loss of generality, we set r_k as $r_k = r$ for simplicity in the following discussion.

3. Sequential semi-orthogonal NMF with negative residual reduction

3.1 Negative residual reduction NMF

As the definition of $\mathbf{R}^{(k)}$ in (3) clearly represents, the (i, j)-th element which satisfies $[\mathbf{R}^{(k-1)}]_{ij} < [\mathbf{U}^{(k-1)}\mathbf{V}^{(k-1)^T}]_{ij}$ is *discarded* to keep $\mathbf{R}^{(k)}$ nonnegative. This step exacerbates the approximation capability and degrades the NE quality. To alleviate this shortcoming, we propose a new NMF with negative residual reduction, designated as NRR-NMF. NRR-NMF attempts, at the k-th level NMF, to force the value of the approximated element, i.e., $[\mathbf{U}^{(k)}\mathbf{V}^{(k)^T}]_{ij}$, lower than the corresponding target element $[\mathbf{R}^{(k)}]_{ij}$ to reduce the discarded elements. This feature reduces the approximation errors and improves the quality of subsequent analytics tasks. Specifically, we consider a penalty of such negative elements , which will be discarded at the subsequent $(k+1)$-th level at the original BoostNE. Consequently, the problem is formulated as

$$\begin{aligned} &\min_{\mathbf{U}^{(k)},\mathbf{V}^{(k)}} f_{\mathrm{NRR}} \\ &= \frac{1}{2}\|\mathbf{R}^{(k)} - \mathbf{U}^{(k)}\mathbf{V}^{(k)^T}\|_F^2 + \frac{\lambda_1}{2}\|\max(\mathbf{U}^{(k)}\mathbf{V}^{(k)^T} - \mathbf{R}^{(k)}, 0)\|_F^2 \\ &= \frac{1}{2}\|\mathbf{R}^{(k)} - \mathbf{U}^{(k)}\mathbf{V}^{(k)^T}\|_F^2 + \frac{\lambda_1}{2}\|\mathbf{M}^{(k)} \odot (\mathbf{U}^{(k)}\mathbf{V}^{(k)^T} - \mathbf{R}^{(k)})\|_F^2, \end{aligned} \tag{4}$$

where the second term is the regularizer with $\lambda_1 (> 0)$. The *mask* matrix $\mathbf{M}^{(k)}$ is introduced for efficient calculation as

$$[\mathbf{M}^{(k)}]_{ij} = \begin{cases} 1 & \text{if } [\mathbf{R}^{(k)} - \mathbf{U}^{(k)}\mathbf{V}^{(k)^T}]_{ij} < 0, \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

Hence, the update rules of $\mathbf{U}$ and $\mathbf{V}$ are derived as

$$\mathbf{V} \leftarrow \mathbf{V} \odot \frac{\mathbf{R}^T\mathbf{U} + \lambda_1(\mathbf{M} \odot (\mathbf{M} \odot \mathbf{R}))^T\mathbf{U}}{\mathbf{V}\mathbf{U}^T\mathbf{U} + \lambda_1(\mathbf{M}^T \odot (\mathbf{M}^T \odot \mathbf{V}\mathbf{U}^T))\mathbf{U}},$$
$$\mathbf{U} \leftarrow \mathbf{U} \odot \frac{\mathbf{R}\mathbf{V} + \lambda_1(\mathbf{M} \odot (\mathbf{M} \odot \mathbf{R}))\mathbf{V}}{\mathbf{U}\mathbf{V}^T\mathbf{V} + \lambda_1(\mathbf{M} \odot (\mathbf{M} \odot \mathbf{U}\mathbf{V}^T))\mathbf{V}}, \quad (6)$$

from which superscript (k) is omitted for notational simplicity. Also, although this is of major practical importance, we consider control of λ_1 in NRR-NMF to maintain better approximation at the end of each level as $\lambda_1(t) = (T_{\max} - t)/T_{\max}$, where $T_{\max}$ is the maximum number of iterations in each NMF process, and t is the index of the inner iterations.

3.2 Sequential semi-orthogonal NMF

BoostNE performs multiple NMFs sequentially, where these NMFs are processed entirely independently. Therefore, some column bases within $\mathbf{U}^{(k)}$ might be similar to those in $\{\mathbf{U}^{(l)}\}_{l<k}$. To avoid such redundant columns, we consider an orthogonal constraint on $\mathbf{U}^{(k)}$. It should be emphasized that this differentiates itself from conventional orthogonal NMFs [11], [12], [13] such that every column vector in $\mathbf{U}^{(k)}$ should be orthogonal to every column vector in $\{\mathbf{U}^{(l)}\}_{l<k}$. It should also be emphasized that the orthogonal constraint is posed in an *approximated* manner to avoid degradation of its convergence speed and the subsequent analytic tasks because of overly tight restriction. Hence, we define the problem as

$$\begin{aligned}
&\min_{\mathbf{U}^{(k)},\mathbf{V}^{(k)}} f_{\mathrm{SSO}} \\
&:= \frac{1}{2}\|\mathbf{R}^{(k)} - \mathbf{U}^{(k)}\mathbf{V}^{(k)^T}\|_F^2 + \lambda_2 \sum_{i=1}^{r}\sum_{l=1}^{k-1}\sum_{j=1}^{r} \boldsymbol{u}_j^{(l)}\boldsymbol{u}_i^{(k)} \\
&= \frac{1}{2}\|\mathbf{R}^{(k)} - \mathbf{U}^{(k)}\mathbf{V}^{(k)^T}\|_F^2 + \lambda_2 \sum_{q=1}^{r(k-1)} \mathrm{Tr}(\mathbf{Q}_q^T\mathbf{U}^{(k)}). \quad (7)
\end{aligned}$$

Therein, $\boldsymbol{u}_i^{(k)}$ is the i-th column vector of the k-th level basis $\mathbf{U}^{(k)}$. Furthermore, $\mathbf{Q}_q$ is defined as $\mathbf{Q}_q = [\ \boldsymbol{u}_q : \boldsymbol{u}_q : \cdots : \boldsymbol{u}_q\ ] \in \mathbb{R}_+^{m\times r}$, where $\boldsymbol{u}_q \in \mathbb{R}_+^m$ is the q-th ($q \in [r(k-1)]$) column vector of $\mathbf{U}$, which is the *concatenated matrix* of $\{\mathbf{U}^{(l)}\}_{l<k}$ as $\mathbf{U} = [\mathbf{U}^{(1)} : \mathbf{U}^{(2)} : \cdots : \mathbf{U}^{(k-1)}] \in \mathbb{R}_+^{m\times r(k-1)}$. In the sequel, noting that $\frac{\partial}{\partial \mathbf{Y}}\mathrm{Tr}(\mathbf{YB}) = \mathbf{B}^T$ and that the rule of $\mathbf{V}^{(k)}$ is identical to MU, the update rule of $\mathbf{U}^{(k)}$ is defined as

$$\mathbf{U} \leftarrow \mathbf{U} \odot \frac{\mathbf{RV}}{\mathbf{U}\mathbf{V}^T\mathbf{V} + \lambda_2 \sum_{q=1}^{R(k-1)} \mathbf{Q}_q}. \quad (8)$$

3.3 Convergence analysis

The convergence analysis is given as below.

Theorem 3.1. *The objective function f_{NRR} in (4) is non-increasing under the update rules in (6).*

4. Conclusion

This paper presented a proposal of a sequential semi-orthogonal NMF with negative residual reduction for boosted network embedding: SSO-NRR-NMF. The presentation will show some numerical evaluations using several real-world datasets which demonstrate the effectiveness of the proposed SSO-NRR-NMF.

References

[1] P. Cui, X. Wang, J. Pei, and W. Zhu, “A survey on network embedding,” *IEEE Transactions on Knowledge and Data Engineering*, 2018.

[2] B. Perozzi, R. Al-Rfou, and S. Skiena, “Deepwalk: Online learning of social representations,” in *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2014, pp. 701–710.

[3] S. Cao, W. Lu, and Q. Xu, “Grarep: Learning graph representations with global structural information,” in *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management (CIKM’15)*, 2015, pp. 891–900.

[4] J. Tang, M. Qu, M. Wang, M. Zhang, J. Yan, and Q. Mei, “LINE: Large-scale information network embedding,” in *Proceedings of the 24th international conference on world wide web*, 2015, pp. 1067–1077.

[5] S. Cao, W. Lu, and Q. Xu, “Deep neural networks for learning graph representations,” in *Thirtieth AAAI Conference on Artificial Intelligence*, 2016.

[6] R. Feng, Y. Yang, W. Hu, F. Wu, and Y. Zhang, “Representation learning for scale-free networks,” in *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[7] J. Qiu, Y. Dong, H. Ma, J. Li, K. Wang, and J. Tang, “Network embedding as matrix factorization: Unifying DeepWalk, LINE, PTE, and node2vec,” in *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining (WSDM)*, 2018, pp. 459–467.

[8] J. Li, L. Wu, and H. Liu, “Multi-level network embedding with boosted low-rank matrix approximation,” in *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 2019.

[9] D. D. Lee and H. S. Seung, “Algorithms for non-negative matrix factorization,” in *Annual Conference on Neural Information Processing Systems (NIPS)*, 2001.

[10] J. H. Friedman, “Greedy function approximation: a gradient boosting machine,” *The Annals of Statistics*, vol. 29, no. 5, pp. 1189–1232, 2001.

[11] D. Chris, L. Tao, P. Wei, and P. Haesun, “Orthogonal nonnegative matrix tri-factorizations for clustering,” in *12th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD)*, 2006.

[12] Y. Jiho and C. Seungjin, “Nonnegative matrix factorization with orthogonality constraints,” *Journal of Computing Science and Engineering*, vol. 4, no. 2, pp. 97–109, 2010.

[13] F. Pompili, N. Gillis, P.-A. Absil, and F. Glineur, “ONP-MF: An orthogonal nonnegative matrix factrization algorithm with application to clustering,” *Neurocomputing*, vol. 141, no. 2, pp. 15–25, 2013.