

Network Coordinate System using Non-negative Matrix Factorization based on KL Divergence

Lixiang Chai*, Xiangyang Luo*, Fan Zhao*, Mingyue Li*, Siqi Liu*

* The State Key Laboratory of Mathematical Engineering and Advanced Computing,
Zhengzhou Information Science and Technology Institute, Zhengzhou, China

clxwise@sina.com, luoxxy_ieu@sina.com, zhaofan_123@yeah.net, lmypretty@163.com, liusiqi_i@yeah.net

Abstract— Network coordinate is used to predict the network delay between the network nodes, for the problem that the accuracy of predicting the network delay of the network coordinate system based on non-negative matrix factorization is low, proposing a new network coordinate system using non-negative matrix factorization based on KL divergence (KL-NMF). First, the advantages and disadvantages of classical algorithm for delay prediction of network coordinate system based on matrix factorization are analysed. Then, according to the characteristics of network link delay distributions, proposed a new network coordinate system based on KL-NMF algorithm and the algorithm flow. Experiment based on the four commonly used network delay matrix data set, and the results show that compared to the existing delay prediction algorithm based on NMF network coordinate system, the delay prediction algorithm based on KL-NMF network coordinate system not only reduce the error to reproduce delay matrix, but also improve the network node delay prediction accuracy.

Keywords— Network Coordinate, Non-negative Matrix Factorization, KL Divergence, Link Delay Distribution, Node Delay Prediction

I. INTRODUCTION

Network coordinate system is a good method to predict the delay between network nodes. The main idea of network coordinate system is to use the existing mathematical theory and method to construct the network delay mapping space and in this space to solve practical network problem: each host in the actual network is mapped into a N dimension coordinate system, with a set of coordinates, such as a set of numbers, to describe the host node position in the IP network [1]. Thus the network delay (network distance) can be estimated by calculating the distance function of the host coordinate. In recent years, many the application involving the network link selection, such as P2P file sharing [2], application layer multicast [3], content distribution network [4], online game [5], this application need a large number of end-to-end network delay measurements, but the large end-to-end delay measurement requires high time complex degree and high measuring cost, and increase the load of the whole network, Therefore, the end-to-end delay measurement methods are not suitable for large scale network environment. Network coordinate system can predict the network delay, Therefore, the study of delay prediction method based on network coordinate system has important significance.

At present, the research of network coordinate system is mainly divided into two categories [6], network coordinate system based on Euclidean distance and network coordinate system based on matrix factorization.

The main idea of the network coordinate system based on Euclidean distance is network nodes in the actual network are embedded into a geometric space, the coordinates of network nodes in the geometric space as the network coordinates, using the coordinates of any two nodes calculate the distance in the geometric space, regarding the distance as the approximate network delay in the real network, where network delay can be called network distance. T.S.E.Ng et al. [1] first proposed actual network can be simulated as a geometric space, some network nodes as reference nodes are embedded into geometric space and calculate their coordinates, the remaining nodes are embedded into geometric space, and calculate their coordinates according to the coordinates of the reference nodes. This method avoids lots of end-to-end delay measurement, reducing the amount of delay measurement overhead, but the fixed reference node selection can affect the prediction accuracy. To solve this problem, Dabek et al. [7] first proposed a distributed network coordinate Vivaldi, the process of network nodes are embedded into geometric space can be seen as a process of minimizing prediction delay error, and also can be seen as a process of the spring elastic potential energy minimization, Vivaldi is a fully distributed network coordinate system, node delay prediction accuracy is not affected by reference node performance. However, the convergence speed is slow, the coordinates of the nodes are too concentrated. A layered architecture of the network coordinate system proposed by Ng et al [8], nodes in the network is divided into a plurality of layers, every node coordinates can be obtained in each layer, but makes the algorithm complexity increases. Yang Chen et al. [9] proposed the extended hierarchical network coordinate system based on Vivaldi, the same node give multiple sets of different network coordinate, One group is global scope (i.e. the entire Internet) network coordinate, and the other coordinates in a relatively small range. The network coordinate system based on Euclidean distance can effectively predict the network delay between any two nodes, and have a certain degree of scalability, due to the presence of TIV (triangle inequality

violation) [10] and network delay non symmetric [11] problem, the prediction accuracy still need to be improved.

In order to overcome the two main flaws result from the network coordinate system based on Euclidean distance, researchers proposed the network coordinate system based on the matrix factorization. The main idea is a delay matrix can be approximation decomposed into two small matrixes, the row vectors of two small matrix as the coordinates of the network nodes, the prediction delay between two nodes can be calculated according to the definition of vector multiplication rule. Yun Mao et al [12] first proposed a network coordinates based on matrix factorization (IDES), network nodes are divided into ordinary nodes and reference nodes, using reference nodes construct the initial network coordinate system, ordinary nodes calculate itself coordinates according to the coordinates of reference nodes, this method first brings matrix factorization into network coordinate system, and improve the network delay prediction accuracy, but this method sensitives to reference nodes. Liao [13] et al proposed a distributed network coordinate system based on matrix factorization, which not distinguish nodes attributes, each node can be randomly selected as reference nodes, to solve the flaw result from fixed reference nodes, but this method is shocked by the nodes coordinates. Chen et al [6] proposes a network coordinates system based on matrix factorization using weight (Phoenix), aiming at solving the problem of error propagation, the introduction of weighting schemes, the absolute error of the higher reference node given low weight, absolute error of the low reference nodes give high weight. Liao et al [14] introduce the machine learning into the network coordinate system, using stochastic gradient descent algorithm to update the nodes coordinates, proposed a network coordinates based on stochastic gradient decrease algorithm, the node coordinates can quickly converge. Rizk [15] et al. according to the network node in the real network topology delay similarity theory, the nodes of the network is divided into different classes, each class using matrix factorization to acquire the node coordinates, propose a network coordinate system based on matrix factorization using term clustering.

The above methods are based on the NMF algorithm to calculate the initial coordinates of the reference node, such as IDES using non-negative matrix factorization algorithm, and no analysis of network delay data distribution characteristics,. It is considered that the link delay distribution obeys a Gaussian distribution, and research indicates that in the actual Internet, network link delay approximation follows a Poisson distribution [16], [17]. Therefore, the NMF on delay matrix factorization in IDES, may not be a good fit to the actual Internet network delay. Aiming at this flaw, this paper presents a network coordinate system based on non-negative matrix factorization using KL divergence (KL-NMF). First, it analyses the network link delay distribution, then according to the link delay approximate obeys the Poisson distribution, using the non-negative matrix factorization algorithm based on KL divergence [18] into the network coordinate system. Finally, the node according to its own coordinate combined with definition of computational algorithms to predict the

delay between other nodes. Compared with the classical algorithm IDES, the results show that the proposed method (KL-NMF) improves the delay prediction accuracy.

II. THE BASIC PRINCIPLE OF IDES

network coordinate system based on Euclidean distance need to meet the condition that the sum of two sides is greater than third side in a triangle, but the literature [19] pointed out: the influence of actual network routing strategy results in the network nodes delay often appear TIV (Triangle Inequality Violation) phenomenon, that is, the network distance between them may violate the principle of triangle inequality, leading to predict delay error is too large. To solve this problem, Mao et al. first proposed that matrix factorization can be applied into the network coordinates system, they think that any two nodes are adjacent to each other in topological network, the delay from them to other nodes is similar, therefore, consists of a large number of delay of nodes network delay matrix, their row vector approximately equal or approximately equal to a linear combination of the other, and the delay matrix can also be expressed as the product of two lower dimensional delay matrix. According to this idea, we can find the approximate factorization of delay matrix:

$$\mathbf{D} = \mathbf{XY}^T \quad (1)$$

Where $\mathbf{D}$ is composed of $N \times N$ delay matrix, $\mathbf{X}$ and $\mathbf{Y}$ are $N \times d$ matrix ($d \ll N$), the network distance between any two nodes H_i and H_j can be estimated according to the:

$$D_{i,j} = \overrightarrow{X_i} \cdot \overrightarrow{Y_j} = \sum_{k=1}^d X_{i,k} Y_{j,k} \quad (2)$$

$D_{i,j}$ represents the network distance between two nodes H_i and H_j , $\overrightarrow{X_i}$ and $\overrightarrow{Y_j}$ represent the outgoing vector of H_i (delay from host to reference node) and incoming vector of H_j (delay from reference node to host), respectively. Therefore, it is a key problem to get the outgoing vector matrix and the incoming vector matrix of the host from the delay matrix, IDES applies the non-negative matrix factorization algorithm proposed by [18] to obtain the outgoing vector matrix and incoming vector matrix of hosts. The algorithm uses the typical square Euclidean distance as the objective function, that is:

$$D_E(\mathbf{D} \parallel \mathbf{XY}) = \frac{1}{2} \|\mathbf{D} - \mathbf{XY}\|^2 \quad (3)$$

For the formula (3), using the gradient descent algorithm to solve the minimum value, to minimizing the error of delay matrix $\mathbf{D}$ and $\mathbf{XY}$. The iterative rule of the algorithm is as follows:

$$X_{i,a} \leftarrow X_{i,a} \frac{(\mathbf{DY})_{i,a}}{(\mathbf{XY}^T \mathbf{Y})_{i,a}} \quad (4)$$

$$Y_{j,a} \leftarrow Y_{j,a} \frac{(\mathbf{X}^T \mathbf{D})_{a,j}}{(\mathbf{X}^T \mathbf{XY}^T)_{a,j}} \quad (5)$$

a is the number of columns in the matrix $\mathbf{X}$ and $\mathbf{Y}$, $X_{i,a}$ and $Y_{j,a}$ respectively represent the element of matrix $\mathbf{X}$ and $\mathbf{Y}$. The matrix $\mathbf{X}$ and $\mathbf{Y}$ can be obtained by the iterative formula (4), (5), then the row vector $\bar{X}_i$ and $\bar{Y}_i$ of the matrix are obtained, therefore, the predicted value of delay between any two nodes can be obtained by means of the formula (2). IDES can avoid the TIV problem, but owing to using formula (3) as objective function, network link delay distribution should be subject to Gauss distribution, however, research pointed out that network link delay approximation obeyed the Poisson distribution[16][17], therefore, according to formula (3) to decompose delay matrix will reduce the accuracy of network coordinate of reference nodes.

III. KL-NMF NETWORK COORDINATE SYSTEM ALGORITHM

Algorithm is divided into two parts, first part of the construction for the network coordinates, using the selected reference nodes, delay matrix is constituted by KL divergence based non-negative matrix factorization algorithm to get reference nodes of vector. The second part is the ordinary nodes delay prediction, the ordinary nodes measurement the delay from the reference nodes, then the coordinates in the network coordinate system are calculated, and the delay between the nodes is estimated according to their own coordinates. Figure 1 illustrates the general flow of the algorithm.

Input: network nodes measurement delay; reference nodes number; matrix dimension; iteration number.

Output: coordinates of reference nodes in the network coordinate system.

Step 1 Construction of reference nodes delay matrix.

In a given network environment, all the nodes measure the delay data which is used to construct the network nodes delay matrix. In this paper, we choose the network nodes delay data set which is widely used in the current literature, and adopt random selection algorithm to select a certain number of nodes as reference nodes from the given network environment. Take King data set as example, we use randperm() algorithm to randomly select N nodes as reference nodes from all the network nodes, and measure the delay by each pair of reference nodes to construct an $N \times N$ reference nodes delay matrix.

Step 2 Acquisition of reference nodes coordinates.

Use KL-NMF algorithm to decompose the reference node delay matrix as two low dimension matrix $\mathbf{X}$ and $\mathbf{Y}$. The specific process is as follows: firstly, set the dimension of

decomposed matrix $\mathbf{X}$ and $\mathbf{Y}$, and the iteration number. Then, initialize matrix $\mathbf{X}$ and $\mathbf{Y}$ with randomly number between 0-1; the next, obtain the non-negative matrix $\mathbf{X}$ and $\mathbf{Y}$ using iterative formula (6), (7) deduced by KL-NMF algorithm.

$$\mathbf{X} \leftarrow (\mathbf{X} / (\mathbf{I} \times (\mathbf{Y}^T * \mathbf{K}))) * ((\mathbf{d} / (\mathbf{X} \times \mathbf{Y})) \times \mathbf{Y}^T) \quad (6)$$

$$\mathbf{Y} \leftarrow (\mathbf{Y} / ((\mathbf{X}^T * \mathbf{I}) * \mathbf{K}^T)) * (\mathbf{X}^T \times (\mathbf{d} / (\mathbf{X} \times \mathbf{Y}))) \quad (7)$$

Where $\mathbf{I}$ denotes the i dimensional column vector, $\mathbf{K}$ denotes the k dimensional row vector, and $*$ denotes the dot production of matrix. The detailed deduction process of formula (6) and (7) will be listed in chapter 4. At last, algorithm ends with reaching the max iteration number. From the above process we can get two $N \times v$ matrixes named $\mathbf{X}$ and $\mathbf{Y}$, which N denotes reference nodes number.

Step 3 Construction of initial network coordinate system.

The two $N \times v$ matrixes in step 2 using KL-NMF algorithm are as follows:

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_N \end{pmatrix} = \begin{pmatrix} x_{11} & \cdots & x_{1v} \\ \vdots & \ddots & \vdots \\ x_{N1} & \cdots & x_{Nv} \end{pmatrix} \quad (8)$$

$$\mathbf{Y} = \begin{pmatrix} \mathbf{Y}_1 \\ \vdots \\ \mathbf{Y}_N \end{pmatrix} = \begin{pmatrix} y_{11} & \cdots & y_{1v} \\ \vdots & \ddots & \vdots \\ y_{N1} & \cdots & y_{Nv} \end{pmatrix} \quad (9)$$

Formula (8) denotes outgoing vector matrix of reference nodes, and (9) denotes ingoing vector matrix of reference nodes. The outgoing vector and incoming vector of any reference nodes are $\mathbf{X}_i = (x_{i1}, x_{i2}, \dots, x_{iv})$ and $\mathbf{Y}_i = (y_{i1}, y_{i2}, \dots, y_{iv})$ respectively, which denotes the reference nodes coordinates of network coordinate system.

Step 4. Ordinary nodes delay predict.

Similar to IDES, first, the step 1 get an ordinary node join into the network coordinate system; then, ordinary nodes measure the delay from the m reference nodes and to the m reference nodes, according to formula (10), (11) to calculate the ordinary nodes coordinates in the network coordinate system:

$$\mathbf{X} = \arg \min_{\mathbf{X} \in \mathbb{R}^d} \sum_{i=1}^m (D_{i,out} - \mathbf{X} \cdot \mathbf{Y}_i)^2 \quad (10)$$

$$\mathbf{Y} = \arg \min_{\mathbf{Y} \in \mathbb{R}^d} \sum_{i=1}^m (D_{i,in} - \mathbf{X}_i \cdot \mathbf{Y})^2 \quad (11)$$

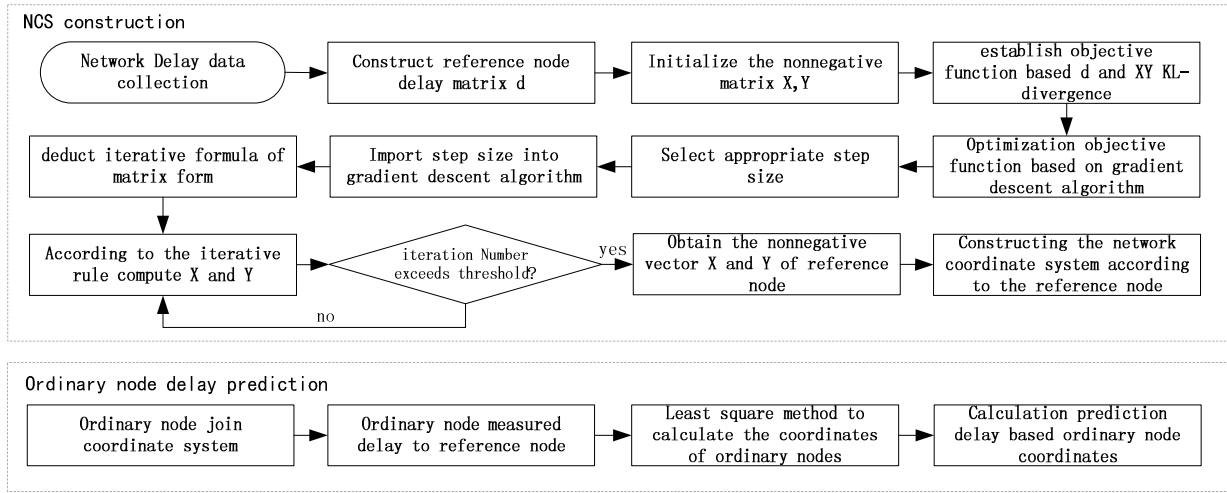


Figure 1. KL-NMF network coordinate system algorithm

Among them, $D_{i,out}$ is measured delay from ordinary nodes to the reference node i , $D_{i,in}$ is measured delay from the reference node i to ordinary nodes, $\mathbf{X}_i$ and $\mathbf{Y}_i$ respectively for the outgoing vector and incoming vector of the reference node i , $\mathbf{X}$ and $\mathbf{Y}$ respectively for the outgoing vector and incoming vector of ordinary nodes. When all ordinary nodes get their own coordinates, the network delay calculation method is defined by the formula (12), which can be used to calculate the delay between any two nodes:

$$D_{i,j} = \mathbf{X}_i \cdot \mathbf{Y}_j \quad (12)$$

Where i, j for any two ordinary nodes, $D_{i,j}$ for the ordinary node i and j prediction delay, $\mathbf{X}_i$ and $\mathbf{Y}_j$ are the outgoing vector and the incoming vector of ordinary node i .

The KL non-negative matrix factorization based network coordinate system uses minimized KL divergence multiplication and take the difference (i.e. KL divergence) of delay matrix $\mathbf{d}$ and the probability density distribution of matrix $\mathbf{XY}$ as target function:

$$D_{KL}(\mathbf{d} \parallel \mathbf{XY}) = \sum_{i=1}^I \sum_{j=1}^J (d_{i,j} \ln \frac{d_{i,j}}{[\mathbf{XY}]_{i,j}} - d_{i,j} + [\mathbf{XY}]_{i,j}) \quad (13)$$

where which $d_{i,j}$ denotes the i row and j column element of matrix $\mathbf{d}$. We apply unconstrained minimization technique to formula (13), because the partial derivative of formula (14)(15)

$$\frac{\partial D_{KL}(\mathbf{d} \parallel \mathbf{XY})}{\partial x_{i,k}} = -\sum_{j=1}^J (y_{k,j} d_{i,j} / [\mathbf{XY}]_{i,j} + y_{k,j}) \quad (14)$$

$$\frac{\partial D_{KL}(\mathbf{d} \parallel \mathbf{XY})}{\partial y_{k,j}} = -\sum_{i=1}^I (x_{i,k} d_{i,j} / [\mathbf{XY}]_{i,j} + x_{i,k}) \quad (15)$$

exists, the optimization problem of target function can be solved by gradient descent algorithm as formula (16), (17) shows:

$$x_{i,k} \leftarrow x_{i,k} - \mu_{i,k} \times \left[-\sum_{j=1}^J (y_{k,j} d_{i,j} / [\mathbf{XY}]_{i,j} + y_{k,j}) \right] \quad (16)$$

$$y_{k,j} \leftarrow y_{k,j} - \eta_{k,j} \times \left[-\sum_{i=1}^I (x_{i,k} d_{i,j} / [\mathbf{XY}]_{i,j} + x_{i,k}) \right] \quad (17)$$

where $\mu_{i,k}$ denotes the update stride of the i row and k column element of matrix $\mathbf{X}$, $\eta_{k,j}$ denotes the update stride of the k row and j column element of matrix $\mathbf{Y}$, $x_{i,k}$ denotes the i row and k column element of matrix $\mathbf{X}$, $y_{j,k}$ denotes the j row and k column element of matrix $\mathbf{Y}$. Making stride in formula (18), (19):

$$\mu_{i,k} = \frac{1}{\sum_{j=1}^J y_{k,j}} \quad (18)$$

$$\eta_{k,j} = \frac{1}{\sum_{i=1}^I x_{i,k}} \quad (19)$$

so that formula (18), (19) can be transformed as:

$$x_{i,k} \leftarrow x_{i,k} \frac{\sum_{j=1}^J y_{k,j} d_{i,j} / [\mathbf{XY}]_{i,j}}{\sum_{j=1}^J y_{k,j}} \quad (20)$$

$$y_{k,j} \leftarrow y_{k,j} \frac{\sum_{i=1}^I x_{i,k} d_{i,j} / [\mathbf{XY}]_{i,j}}{\sum_{i=1}^I x_{i,k}} \quad (21)$$

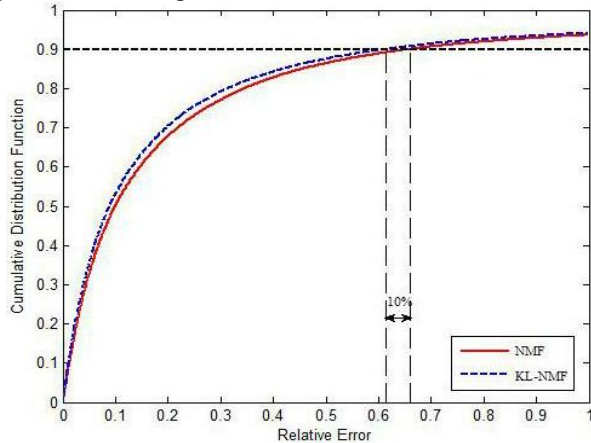
and formula (20), (21) can be transformed as formula (6),(7), so that we can get matrixes $\mathbf{X}$ and $\mathbf{Y}$ which are decomposed from delay matrix $\mathbf{d}$ with above two formula.

IV. EXPERIMENTAL RESULTS

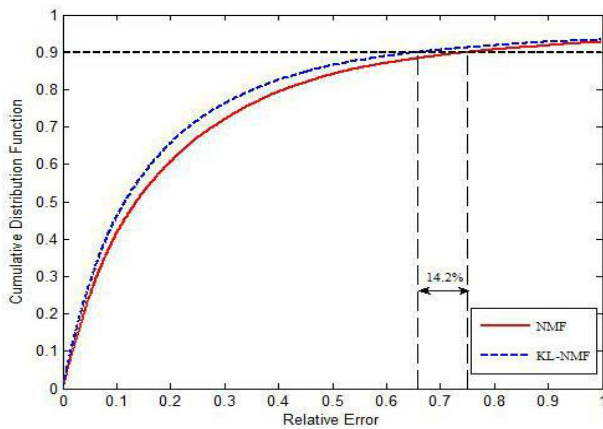
In this paper, the validity and rationality of the algorithm is verified in the Matlab experimental environment, and the experiment is based on four different delay matrices (King [20], PL-RTT [21], Toread [22], Meridian [23]). Among them, the King data set consists of 1740 DNS server delay

measurement. PL-RTT data set includes 169 nodes from the PlanetLab node in the Ping project in 3/23/2004 measurement. Toread data set includes 335 nodes obtained by PlanetLab data. Meridian data set from 2500 host Cornell Meridian project in each measurement delay, in order to approximate the steady state characterization of the network delay, delay in the data sets respectively at different time of repeated measurements of a minimum value, based on the experimental data can be more close to actual network performance.

In the delay prediction, select a small number of nodes as a reference nodes, the remaining nodes as ordinary nodes, using the classical algorithm IDES and the proposed algorithm to conduct the experimental, as shown in Figure 2, KL-NMF represents the algorithm proposed in this paper, NMF represents IDES algorithm.

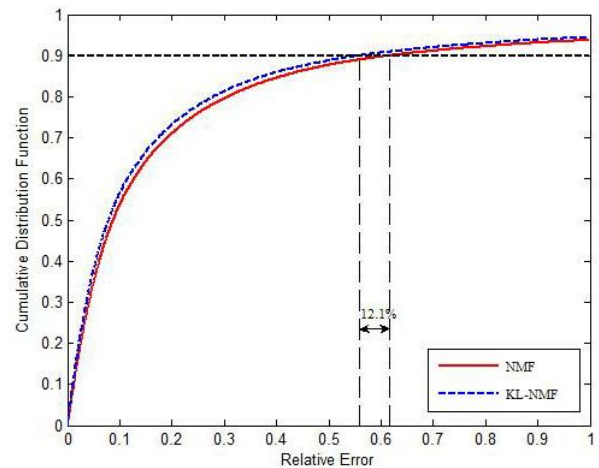


(a) King dataset

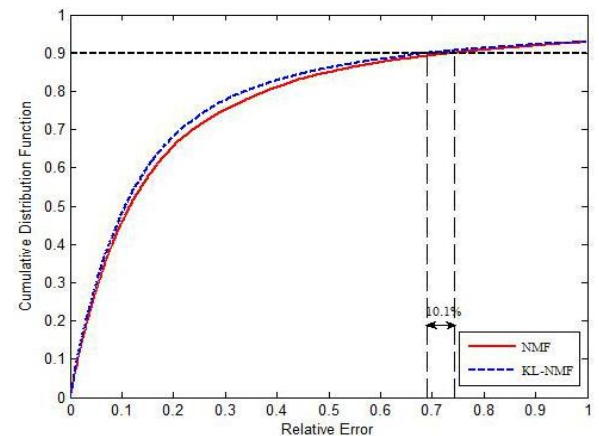


(a) PL-RTT dataset

Figure 2 shows that, in the condition of same dimension and the same number of references, in figure 2(a), the KL-NMF improve the accuracy of delay prediction by 10% than IDES algorithm, and in figure 2(b), the KL-NMF improve the accuracy of delay prediction by 14.2% than IDES algorithm, in figure 2(c), the KL-NMF improve the accuracy of delay prediction by 12.1% than IDES algorithm, in figure 2(d), the KL-NMF improve the accuracy of delay prediction by 10.1% than IDES algorithm. the reasons is the improved network



(c) Toread dataset



(d) Meridian dataset

Figure 2. Ordinary node delay prediction

coordinate system based on non-negative matrix factorization using KL divergence algorithm is more consistent with the characteristics of delay matrix data distribution, thus the delay prediction method is proposed in this paper is better than the classical IDES algorithm.

V.CONCLUSION

In order to improve the network coordinate system in network delay prediction accuracy, this paper proposed a new network coordinate system based on non-negative matrix factorization using KL divergence. The algorithm using the KL divergence as the objective function, using the gradient descent algorithm, delay matrix factorization iterative formula is obtained, combined with the row vectors of two matrices, coordinates of the nodes is calculated by using the formula proposed in this paper, According to the four different data sets, conducted a series of experiments, results show that the method proposed in this paper can improve the delay prediction accuracy. In the next step, we will study more efficient delay prediction algorithm based on network coordinate system to further improve the accuracy of network delay prediction.

ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation of China (Grant No. 61379151, 61272489, 61302159, 61401512, 61572052) and the Excellent Youth Foundation of Henan Province of China (Grant No. 144100510001) and the Foundation of Science and Technology on Information Assurance Laboratory (Grant No. KJ-14-108).

REFERENCES

- [1] Ng T S E, Zhang H, "Predicting Internet network distance with coordinates based approaches", in *Proc INFOCOM*, '02, 2002, paper 1.2, p. 170-179.
- [2] Pacifici, Valentino, Frank Lehrieder, and György Dán, "Cache bandwidth allocation for P2P file-sharing systems to minimize inter-ISP traffic," *IEEE/ACM Transactions on Networking*, vol. 24, pp. 437-448, 2016.
- [3] Le, Tien Anh, Hang Nguyen, and Manh Cuong Nguyen, "Application-network cross layer multi-variable cost function for application layer multicast of multimedia delivery over convergent networks," *Wireless Networks*, vol. 21, pp. 2677-2692, Aug. 2015.
- [4] Zhang, Li xia, et al, "Named data networking," *ACM SIGCOMM Computer Communication Review*, vol. 44, pp. 66-73, Mar. 2014.
- [5] Véron, Maxime, Olivier Marin, and Sébastien Monnet, "Matchmaking in multi-player on-line games: studying user traces to improve the user experience," in *Proc NOSSDAV*, '14, 2014, paper 3, p.7.
- [6] Chen, Yang, et al, "Phoenix: A weight-based network coordinate system using matrix factorization," *IEEE Transactions on Network and Service Management*, vol.8, pp. 334-347, 2011.
- [7] Dabek, Frank, et al, "Vivaldi: A decentralized network coordinate system," in *ACM SIGCOMM Computer Communication Review*, '04, 2004. Paper 34.4, p. 15-26.
- [8] Ng, TS Eugene, and Hui Zhang, "A Network Positioning System for the Internet," in *USENIX Annual Technical Conference*, '04, 2004, paper 7, p. 141-154.
- [9] Chen, Yang, et al, "Pharos: A decentralized and hierarchical network coordinate system for internet distance prediction," in *Proc GLOBECOM*, '07, 2007, paper 11, p. 421-426.
- [10] Mokhtarian, Kianoosh, and Hans-Arno Jacobsen, "Minimum-delay multicast algorithms for mesh overlays," in *Proc TON*, '15, 2015, paper 23, p. 973-986.
- [11] Cunha, Ítalo, et al, "DTRACK: a system to predict and track internet path changes," *IEEE/ACM Transactions on Networking*, vol. 22, pp. 1025-1038, 2014.
- [12] Mao, Yun, Lawrence K. Saul, and Jonathan M. Smith, "Ides: An internet distance estimation service for large networks," *IEEE Journal on Selected Areas in Communications*, vol. 24, pp. 2273-2284, 2006.
- [13] Liao, Yongjun, Pierre Geurts, and Guy Leduc, "Network distance prediction based on decentralized matrix factorization," *Springer Berlin Heidelberg*, Mar.2010.
- [14] Liao, Yongjun, et al, "DMFSGD: A decentralized matrix factorization algorithm for network distance prediction," *IEEE/ACM Transactions on Networking*, vol.21, pp. 1511-1524, 2013.
- [15] Rizk, Rawya, Walaa Saber, and Hany Harb, "Conditional clustered matrix factorization based network coordinate system," *Journal of Network and Computer Applications*, vol.45, pp. 191-202, 2014.
- [16] Liang, Gang, and Bin Yu, "Maximum pseudo likelihood estimation in network tomography," *IEEE Transactions on Signal Processing*, vol.51, pp. 2043-2053, 2003.
- [17] Kang M X, "Research and implementation of IP network delay measurement and estimation," *Beijing University of Posts and Telecommunications*, 2014
- [18] Lee, Daniel D., and H. Sebastian Seung, "Algorithms for non-negative matrix factorization," *Advances in neural information processing systems*, pp. 556-562, 2001.
- [19] Seibert, Jeffrey, et al, "Securing Virtual Coordinates by Enforcing Physical Laws," in *Proc ICDCS*, '12, paper 7, p.315-324, 2012.
- [20] Burnett, Sam, and Nick Feamster, "Encore: Lightweight measurement of web censorship with cross-origin requests," *ACM SIGCOMM Computer Communication Review*, vol. 4, pp. 653-667, 2015.
- [21] (2004)[Online]. Available:http://www.pdos.lcs.mit.edu/~strib/pl_app.
- [22] (2010)[Online]. Available:<http://code.google.com/p/toread/>.
- [23] Wong, Bernard, Aleksandrs Slivkins, and Emin Gün Sirer, "Meridian: A lightweight network location service without virtual coordinates," in *ACM SIGCOMM Computer Communication Review*, '05, 2005, paper 35.4, p. 85-96.



Measurement, IP Geolocation and Information Security.



"International Journal of Internet" and "Multimedia Tools and Applications". LUO is currently an associate professor of Zhengzhou Information Science and Technology Institute. His research interest is Networking and Information Security.



His primary interest is in Network Space Mapping and Network Entity Geolocation.



research interest includes landmark mining of network entities, IP geolocation and Information Security.



Technology Institute. His primary interest is in Network Entity Geolocation Technique.

CHAI Lixiang was born in Shaanxi Province, China, in 1991. Chai got the B.S. degree from Xian University of Science and Technology, Xi'an, China, in 2013. He is currently a M.S. candidate in The State Key Laboratory of Mathematical Engineering and Advanced Computing at Zhengzhou Information Science and Technology Institute. His research interest includes Network Delay

LUO Xiangyang was born in Hubei Province, China, 1978. Luo received the B.S. degree, the M.S. degree and the Ph.D. degree from Zhengzhou Information Science and Technology Institute, Zhengzhou, China, in 2001, 2004 and 2010, respectively. He is the author or co-author of more than 70 refereed international journal and conference papers. He is also a guest editor for

ZHAO Fan was born in Jiangsu Province, China, in 1989. He got the B.S. degree and the M.S. degree from Zhengzhou Information Science and Technology Institute, Zhengzhou, China, in 2011 and 2014, respectively. He is currently a Ph.D. candidate in The State Key Laboratory of Mathematical Engineering and Advanced Computing at Zhengzhou Information Science and Technology Institute.

LI Mingyue was born in Henan Province, China, in 1991. LI got the B.S. degree from Tianjin University of Science and Technology, China, in 2013. She is currently a M.S. candidate in The State Key Laboratory of Mathematical Engineering and Advanced Computing at Zhengzhou Information Science and Technology Institute. Her

LIU Siqi was born in Henan Province, China, in 1992. He got the B.S. degree from Zhengzhou Information Science and Technology Institute, Zhengzhou, China, in 2010 and He is currently a MS candidate in The State Key Laboratory of Mathematical Engineering and Advanced Computing at Zhengzhou Information Science and Technology Institute. His primary interest is in Network Entity Geolocation