Network Big Data Knowledge Augmentation Algorithm Based on Variable-Granularity Opportunistic Scheduling (Postprint)

Authors: El Dorado, Liu Tao, Zhou Xianchun, Yan Xijun

Date: 2018-05-18T00:00:00+00:00

Abstract

To satisfy the high-precision requirements for data knowledge dissemination and eliminate interference from low-quality data in the context of network big data, a knowledge expansion algorithm for network big data based on opportunistic scheduling is proposed using a variable granularity adjustment scheme. Building upon the analysis of network big data characteristics, the algorithm captures the heterogeneous characteristics of network big data through adaptive vector encoding, employs multi-stage backpropagation to normalize heterogeneous network big data, and achieves real-time transmission of network big data via opportunistic scheduling. Simultaneously, based on the knowledge engineering system composed of network big data, the algorithm partitions fine-grained big data, performs dimensionality reduction on multi-dimensional features to transform knowledge granularity into known states, and subsequently adjusts the dynamic characteristics of granularity, thereby endowing the big data sets in knowledge engineering with linear characteristics and explicit geometric properties, and improving knowledge acquisition accuracy through knowledge expansion. Experimental results, through comparison with fine-grained knowledge acquisition algorithms, demonstrate that the proposed algorithm exhibits high reliability and real-time performance in network data transmission as well as high efficiency in knowledge acquisition.

Full Text

Preamble

Title: Network Big Data Knowledge Extension Algorithm Based on Variable Granularity and Opportunistic Scheduling

Authors: Huang Jinguo[1], Liu Tao[1], Zhou Xianchun[2], Yan Xijun[3]
[1]School of Information & Mechanical and Electrical Engineering, Jiangsu Open

University, Nanjing 210017, China;
[2]School of Electronic & Information Engineering, Nanjing University of Information Science & Technology, Nanjing 210017, China;
[3]College of Computer & Information, Hohai University, Nanjing 210098, China

Abstract: To meet the high-precision requirements of data knowledge transmission and eliminate interference from inferior data in the context of network big data, this paper proposes a network big data knowledge extension algorithm based on variable granularity adjustment schemes with opportunistic scheduling. By analyzing the characteristics of network big data, the algorithm employs adaptive vector encoding to capture the heterogeneous features of network big data and utilizes multi-order backpropagation for normalization of heterogeneous network big data, followed by real-time transmission through opportunistic scheduling. Simultaneously, the knowledge engineering system composed of network big data is used to partition fine-grained big data, enabling dimensionality reduction of multi-dimensional features to transform knowledge granularity into a known state. The dynamic characteristics of granularity are then adjusted to impart linear features and clear geometric properties to the big data set in knowledge engineering, thereby improving knowledge acquisition accuracy through knowledge extension. Experimental results comparing the proposed algorithm with a fine-grained knowledge acquisition algorithm demonstrate its high reliability and real-time performance in network data transmission, as well as high efficiency in knowledge acquisition.

Keywords: network big data; knowledge engineering; knowledge extension; variable granularity; opportunistic scheduling

0 Introduction

In the context of network big data, ensuring high-precision data knowledge transmission and eliminating interference from inferior data to obtain effective problem-solving knowledge has become a key issue in knowledge engineering. Reconstructing data knowledge bases from knowledge disseminated through network big data has garnered widespread attention. Network big data exhibits inherent characteristics such as type heterogeneity, data diversity, and distributed propagation, which pose significant challenges. In the domain of network big data dissemination, literature [7] proposed a spatial structure scheme that converts symbolic data into numerical values, effectively preserving original symbolic features while reconstructing sample similarity. Literature [8] presented an optimized prototype system that accelerates multi-batch data transmission server clusters while maximizing the utilization of bandwidth and dispersed random linear network coding for optimal information propagation. Literature [9] investigated optimal data distribution in resource-constrained mobile opportunistic networks, addressing the minimum-cost multicast problem with delay constraints.

In the area of network scheduling, literature [10] proposed a priority scheduling scheme based on random linear network coding, which leverages linear relationship feedback information such as packet reception status to determine the effective information scale of relay nodes. Literature [11] studied nodes and scheduling schemes across multi-channel wireless links. Literature [12] introduced a fully decentralized distributed scheduling strategy where each node determines the number of units to schedule based on its traffic requirements.

Regarding knowledge engineering, literature [13] examined geographic knowledge characteristics in virtual geographic environments, studying the classification and engineering architecture of geographic knowledge. Literature [14] proposed interdisciplinary and multicultural approaches to address issues and challenges in the knowledge society. Literature [15] investigated a knowledge engineering framework for fragmented knowledge modeling and online learning from multiple information sources, focusing on nonlinear fusion of fragmented knowledge and automated demand-driven knowledge navigation.

Building upon these studies in network big data dissemination and knowledge mining, this paper combines an opportunistic scheduling network big data model to investigate a knowledge extension algorithm that can improve both the efficiency and quality of network big data transmission.

1 Opportunistic Scheduling Network Big Data Model

Compared with traditional network data, network big data exhibits distinct characteristics such as complex and heterogeneous data types, structural differences, high data mining complexity, and significant network scheduling challenges. To address these features, this paper employs adaptive vector encoding to capture the heterogeneous characteristics and type features of network big data, utilizes multi-order backpropagation to unify heterogeneous network big data, and implements opportunistic scheduling for network big data transmission.

First, the network big data mining object is transformed from a one-dimensional vector group into a multi-dimensional vector encoding space, yielding an adaptive multi-dimensional vector encoding model driven by the heterogeneous features of network big data. Second, based on the scale and dimensionality of big data, the model stimulates the capture of heterogeneous characteristics and type features of network big data in the multi-dimensional vector encoding space. Third, a multi-order backpropagation algorithm is proposed based on network big data mining objects and opportunistic scheduling, organically integrating the multi-dimensional vector encoding space with the multi-order backpropagation space. Finally, feature capture and multi-order backpropagation are performed in the multi-dimensional vector encoding space to establish the opportunistic scheduling network big data model.

The definition of the vector encoding space is as follows: In an m-dimensional finite field Euclidean space G_m, for any one-dimensional vector space, the mapping relationship between the heterogeneous feature vector B of network big data and multi-dimensional vectors is given by:

$$
\begin{aligned}
&A_i \in \{A_{i1}, A_{i2}, \dots, A_{im}\} \\
&A_i \in \{A_{i1}, A_{i2}, \dots, A_{im}\} \\
&A_i \in \{A_{i1}, A_{i2}, \dots, A_{im}\} \\
&A_i = \begin{bmatrix} a_{i1} & a_{i2} & \cdots & a_{im} \end{bmatrix} \\
&A_i = \begin{bmatrix} a_{i1} \\ a_{i2} \\ \vdots \\ a_{im} \end{bmatrix} \\
&A_i = \begin{bmatrix} A_{i1} & A_{i2} & \cdots & A_{im} \end{bmatrix} \\
&A_i = \begin{bmatrix} a_{i1} & a_{i2} & \cdots & a_{im} \end{bmatrix} \\
&A_i = \sum_{j=1}^{m} a_{ij} \\
&A_i = \sum_{j=1}^{m} a_{ij}
\end{aligned}
$$

where b represents elements of vector B, j denotes the vector dimension in space, and c represents the vector after encoding vector A based on vector B drive.

The adaptive multi-dimensional vector encoding form is described by equation (3):

$$
\begin{aligned}
&C = \{C_1, C_2, \dots, C_m\} \\
&\delta = \frac{1}{m} \sum_{j=1}^{m} c_j \\
&\pi_i = \prod_{j=1}^{m} (a_{ij} - b_j)^2 \\
&d = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(c-\delta)^2}{2}\right)
\end{aligned}
$$

where the vector represents the encoding of network big data in multi-dimensional vector space, the variable denotes the vector dimension offset, and the parameter represents the difference between vectors A and B after multi-dimensional space encoding.

During network propagation, adaptively multi-dimensionally vector-encoded network big data undergoes iterative calculation as shown in equation (4):

$$
\begin{aligned}
c_i &= \int \prod_{j=1}^{m} A_{ij}\, dC \\
C &= \{C_1, C_2, \dots, C_m\} \\
B &= \{b_1, b_2, \dots, b_m\} \\
a_i &= \int \sum_{j=1}^{m} a_{ij}\, dC
\end{aligned}
$$

where A_{ij} and c_i represent the iterative deformation processing of multi-dimensional space encoding vectors and their parameters during network propagation.

To improve network big data propagation efficiency, opportunistic scheduling is employed, with opportunistic parameters calculable via equation (5):

$$
\begin{aligned}
\omega_i &= \frac{1}{m} \sum_{j=1}^{m} \omega_{ij} \\
t &= \int \sum_{j=1}^{m} a_{ij} b_j\, dt \\
o_m &= \int \sum_{i=1}^{m} c_i\, d\omega
\end{aligned}
$$

where ω_i represents the opportunistic scheduling weight on each of the m dimensions, and parameter t denotes the network big data propagation time. Throughout the network big data encoding propagation process, opportunistic scheduling is performed based on multi-dimensional space encoding differences.

The main steps of multi-order backpropagation for network big data are as follows:

a) Obtain the multi-dimensional space encoding of network big data and its parameters to get C and δ.

b) Perform iterative deformation processing on multi-dimensional space encoding vectors and their parameters during network propagation to obtain c_i and ω_i.

c) For encoding parameters c in each dimensional space, solve for opportunistic scheduling weights ω.

d) For each order of network big data propagation, solve for forward set F and backward data set R.

e) Calculate the difference between vectors A and B after multi-dimensional space encoding.

f) Compute residuals of F and R, and modify the backward network big data set.

The multi-order backpropagation algorithm for opportunistic scheduling network big data is described as:

```
**Input:** $m$, $A_i$ for $i = 1, i++, i \leq m$
**Process:**
For $j = 1, j++, j \leq m$
    $B_j = B_j \times B_{j-1}$; obtain $c_i$
For $i = 1, i++, i \leq m$
    $temp = \frac{temp}{m}$; computing the amending parameter
**Return:** $C$, $\omega$
```

2 Variable Granularity Knowledge Extension Algorithm

The knowledge engineering system composed of network big data is defined as a triple $K = \langle R, E, AR \rangle$, where R represents the network big data set, E denotes knowledge description objects, and AR represents the knowledge attribute set of all elements in the big data set. For any $r \in R$ and $e \in E$, a linear relationship attribute mapping $T_r : r \to ar$ is defined, where T_r represents the knowledge attribute mapping relationship for any element r in network big data set R, such that $ar \in AR$.

Thus, a rough knowledge engineering system can be defined as $KR = \langle R, E, AR \rangle_\lambda$, where λ denotes the granularity roughness weight. Assuming K is a multi-granularity rough knowledge engineering system, there exists a fuzzy rough mapping relationship among R, E, and AR, and the element mapping between R and E exhibits a many-to-one phenomenon, i.e., $|R| > |E|$. Consequently, the fine-grained precise knowledge set K_f in K and the fine-grained knowledge set KR_f in the rough knowledge engineering system KR satisfy the relationship shown in equation (6):

$$\begin{aligned} \lambda_1 &= \sum_{i=1}^{m} \lambda_i \\ \lambda_2 &= \sum_{i=1}^{m} \lambda_i^2 \\ \lambda &= \frac{\lambda_1}{\lambda_2} \end{aligned}$$

where λ represents the fine-grained threshold of rough sets.

In network big data knowledge engineering systems, knowledge granularity possesses multi-dimensional features in the multi-dimensional vector space of big data network transmission, rendering knowledge granularity unknown and dynamic. To seek fine-grained big data for dimensionality reduction of multi-dimensional features, knowledge granularity must be transformed into a known state while adjusting its dynamic characteristics to impart linear features and clear geometric properties to the knowledge engineering big data set, thereby improving knowledge mining precision and enabling unique definition of knowledge processing objectives. For coarse-grained multi-dimensional big data, adjusting granularity features and performing dimensionality reduction enables linear description of unknown properties and reveals the multi-dimensional spatial geometric characteristics of hidden unknown big data. The correspondence between multi-dimensional vector space and fine-grained knowledge geometric feature space is illustrated in [Figure 1: see original paper], where the knowledge engineering triple is reduced to a binary group, transforming unknown factors through deterministic conversion and geometric mapping.

Therefore, for knowledge engineering system K, the parameter and attribute descriptions of network big data knowledge based on variable granularity are given by:

$$\begin{aligned} &f : R^m \to V^m \\ &\rho = (r\sin\alpha + e\cos\beta) f(r, e, ar) \\ &\alpha = \arctan\left(\frac{r}{\rho}\right) \\ &\beta = \arctan\left(\frac{e}{\rho}\right) \end{aligned}$$

where f represents the dimensionality reduction mapping from the original multi-dimensional vector space R^m to the variable granularity feature space V^m.

Variable granularity conversion can be analytically completed through the equation $\rho = (r\sin\alpha + e\cos\beta) f(r, e, ar)$, where ρ denotes variable granularity, α represents the horizontal cross-radian of any big data element r in multi-dimensional vector space, and β represents the vertical cross-radian generated by any knowledge description object e during spatial dimensionality reduction. Consequently, the iterative relationship between variable granularity and network big data knowledge engineering is expressed as equation (8):

$$\rho_i = \sum_{i=1}^{m} \lambda_i \rho_i$$

$$\alpha_i = \sum_{i=1}^{m} \lambda_i \alpha_i$$

$$\beta_i = \sum_{i=1}^{m} \lambda_i \beta_i$$

After dimensionality reduction, data points on the knowledge plane undergo variable granularity conversion, transferring the entire multi-dimensional spatial knowledge set into the fine-grained geometric feature space. Data knowledge in this space possesses deterministic relationships and linear features. At this point, the network big data knowledge engineering system KR effectively resolves the interference of irregular geometric spaces caused by coarse granularity and the impact of dynamic granularity changes on the knowledge space. [Figure 2: see original paper] presents the granularity scheduling scheme in network big data knowledge acquisition, using λ as the threshold to partition coarse-grained and fine-grained sets. Fine-grained data directly enters the acquisition results, while coarse-grained data undergoes variable granularity scheduling to eliminate unknown properties and irregularities, converting into fine-grained data.

Following the variable granularity scheduling shown in [Figure 2: see original paper], the network big data knowledge engineering system performs knowledge extension through equation (9) after fine-grained segmentation at threshold λ:

$$K_{ext} = \sum_{i=1}^{m} \rho_i \cdot K_i$$

$$KR_{ext} = \int \sum_{i=1}^{m} \lambda_i \, dK_i$$

---

3 Experimental Results Analysis

The proposed algorithm based on variable granularity and opportunistic scheduling for network big data knowledge extension is denoted as NKE-VOS. Performance analysis and verification focus on data error after network scheduling, data transmission delay, and convergence times for knowledge acquisition. Under identical experimental conditions, the performance of NKE-VOS is compared with the fine-grained knowledge acquisition algorithm (FGKA). The experimental platform is described in .

** Experimental Platform** - Network terminals: 50 - Network servers: 5 - Server CPU: Intel Xeon E3 v2 - Server storage: 1 TB - Wireless communication protocol: IEEE 802.11g - Server OS: Ubuntu Server 16.04.2 LTS - Algorithm development language: Python - Network terminal storage: 50 min

[Figure 3: see original paper] illustrates the performance of both algorithms in terms of data precision as network big data volume increases with progressively activated network terminals. Over a 50-minute period, the big data transmitted by both algorithms is compared with original data to compute data error. The comparison reveals that FGKA' s static scheduling responds slowly to changes in big data scale, leading to data loss or errors that severely constrain data quality. Conversely, the proposed NKE-VOS algorithm employs opportunistic scheduling that transforms network big data mining objects from one-dimensional vector groups into multi-dimensional vector encoding space, capturing heterogeneous characteristics and type features based on big data scale and dimensionality, thereby achieving high-efficiency network scheduling and improved data precision.

[Figure 4: see original paper] shows the real-time performance of both algorithms as network big data volume increases. The average end-to-end delay of big data transmitted by both algorithms over 50 minutes is measured. The comparison demonstrates that NKE-VOS shortens network data transmission delay and ensures real-time performance by obtaining multi-dimensional space encoding and parameters of network big data, performing opportunistic scheduling for encoding parameters in each dimensional space, solving forward and backward data sets for each order of network big data propagation, and employing the multi-order backpropagation algorithm with opportunistic scheduling.

[Figure 5: see original paper] presents the iteration counts required for knowledge acquisition as network servers increase. NKE-VOS performs variable granularity scheduling on coarse-grained multi-dimensional big data while reducing dimensions, enabling linear description of unknown properties, eliminating unknown big data, and reconstructing multi-dimensional spatial geometric characteristics, thereby acquiring knowledge in fewer iterations and extending network big data knowledge.

---

4 Conclusion

Network big data applications impose higher requirements on data propagation real-time performance, data precision, and knowledge acquisition efficiency. To satisfy these demands, this paper proposes a network big data knowledge extension algorithm based on variable granularity and opportunistic scheduling. First, the algorithm establishes a multi-dimensional vector space from the heterogeneous features of network big data, captures heterogeneous characteristics in real-time through adaptive multi-dimensional vector encoding, and ensures real-time and reliable network big data transmission via multi-order backpropagation and opportunistic scheduling. Second, the knowledge engineering system of network big data is partitioned into fine-grained segments according to variable granularity thresholds, achieving dimensionality reduction of multi-dimensional features to transform knowledge granularity into a known state with explicit dynamic geometric characteristics, and employing a variable granularity-based knowledge extension algorithm. Comparative experiments with the fine-grained knowledge acquisition algorithm demonstrate that the proposed algorithm exhibits significant advantages in network big data transmission reliability, real-time performance, and knowledge acquisition efficiency.

References

[1] Gao J, Lu W F, Dai Z J, et al. A computational approach to characterizing the impact of social influence on individuals’ vaccination decision making [J]. Plos One, 2013, 8(4): 601-611.

[2] Przulj N, Malod-Dognin N. Network analytics in the age of big data [J]. Science, 2016, 353(6295): 123-124.

[3] Gao S, Pang H, Gallinari P, et al. A novel embedding method for information diffusion prediction in social network big data [J]. IEEE Trans on Industrial Informatics, 2017, 3(99): 1-11.

[4] Kim R, Lim H, Krishnamachari B. Prefetching-based data dissemination in vehicular cloud systems [J]. IEEE Trans on Vehicular Technology, 2016, 65(1): 292-306.

[5] 康文文, 李浩敏, 汤超, 等. 面向复杂工程系统设计的云知识平台 [J]. 系统工程与电子技术, 2017, 39(5): 1078-1084.

[6] Zheng X, Wang J, Dong W, et al. Bulk Data dissemination in wireless sensor networks: analysis, implications and improvement [J]. IEEE Trans on Computers, 2016, 65(5): 1428-1439.

[7] 王齐, 钱宇华, 李飞江. 基于空间结构的符号数据仿射传播算法 [J]. 模式识别与人工智能, 2016, 29(12): 1132-1139.

[8] Liu Y, Niu D, Khabbazian M. Cooper: Expedite Batch Data Dissemination in Computer Clusters with Coded Gossips [J]. IEEE Trans on Parallel & Distributed Systems, 2017, 28(8): 2204-2217.

[9] Liu Y, Wu H, Xia Y, et al. Optimal Online Data Dissemination for Resource Constrained Mobile Opportunistic Networks [J]. IEEE Trans on Vehicular Technology, 2017, 66(6): 5301-5315.

[10] 王练, 梁申虎, 彭代渊. 多源多中继无线网络中基于随机线性网络编码的调度方案 [J]. 电子与信息学报, 2017, 39(3): 532-538.

[11] Moharir S, Krishnasamy S, Shakkottai S. Scheduling in densified networks: algorithms and performance [J]. IEEE/ACM Trans on Networking, 2017, 25(1): 164-178.

[12] Domingo-Prieto M, Chang T, Vilajosana X, et al. Distributed PID-based scheduling for 6tisch networks [J]. IEEE Communications Letters, 2016, 20(5): 1006-1009.

[13] 林珲, 游兰. 虚拟地理环境知识工程初探 [J]. 地球信息科学学报, 2015, 17(12): 1423-1430.

[14] García-Peñalvo F J. Engineering contributions to a multicultural perspective of the knowledge society [J]. IEEE Revista Iberoamericana De Tecnologias Del Aprendizaje, 2015, 10(1): 17-18.

[15] Wu X, Chen H, Zhang Q, et al. Knowledge Engineering with Big Data [J]. IEEE Intelligent Systems, 2015, 30(5): 46-55.

Note: Figure translations are in progress. See original paper for figures.

Source: ChinaXiv —Machine translation. Verify with original.