Chaotic Association Feature Extraction of Big Data Clustering Based on the Internet of Things

Xiaoming Liu[1*], Thipendra Pal Singh[2], Rajeev Kumar Gupta[3], Edeh Michael Onyema[4]
[1]JingZhou Vocational College of Technology, Software Engineering Institute, Jingzhou, Hubei, 434000, China
[2]School of Computer Science, University of Petroleum and Energy Studies, Dehradun, India
[3]Pandit Deendayal Energy University, Gandhinagar, India
[4]Department of Mathematics and Computer Science, Coal City University, Enugu, Nigeria
Emails: xiaomingliu7@126.com, thipendra@gmail.com, rajeev.gupta@sot.pdpu.ac.in, michael.edeh@ccu.edu.ng

This article addresses the stabilization of chaotic characteristics in abnormal data by proposing chaotic correlation feature extraction of big data clustering based on the Internet of things. The chaotic features in big data usually show complex folding and distortion without obvious rules and order and non-synchronization. In this article, the dimension of extracted correlation is utilized as the chaotic feature for the clustering of big data. The one-dimensional time series that can be extended in multi-dimensional space is analysed based on phase space reconstruction, to extract the chaotic correlation dimension (CCD) features. After the relevant experimental analysis, this paper mainly compares the energy consumption and processing time of the two respective algorithms. In the simulation parameter design, the time interval of big data packet generation is 0.1s, and the data is generated from the simulation time of 300s. The results obtained show that when dealing with the same amount of data, the energy consumption of this algorithm is significantly lower than that of the traditional algorithm. When dealing with the same amount of data, the time required by this algorithm is significantly lower than that of the traditional algorithm. This is because this algorithm is easy to implement and has good clustering efficiency for data, so the clustering time is short. With the gradual increase in the amount of data, the correlation dimension of this algorithm tends to be stable. While the correlation dimension of the traditional algorithm fluctuates greatly, it is revealed that the proposed approach has high data clustering efficiency and verifies the effectiveness of this algorithm.

Povzetek: Za internet stvari je analizirana možnost stabilizacije nenavadnih podatkov znotraj velikih podatkov.

1   Introduction

With the rapid expansion of network technology, the network crime activities in the big data environment are gradually increasing, increasing the amount of abnormal data in the environment of huge data [1]. Therefore, seeking effective big data mining methods is of great consequence to ensure the security of related systems in a big data environment [2]. Most of the current big data mining methods carry out big data mining according to the known abnormal characteristics, which reduces the reliability and efficiency of big data mining, increases the overhead of processing big data, and reduces the overall availability and performance of big data. As revealed in Figure 1, the framework of big data mining and analysis platform [3]. Therefore, how to analyse the failure rate, probability analysis, and adjustment scheme of big data in different regions without interfering with the performance of huge data has an emphasis on the analysis of data mining [4]. In large-scale data mining, massive data brings great difficulties to the existing abnormal data mining

efficiency [5]. How to design sub-region mining algorithms for massive data has gained attention and becomes a research hotspot. Due to the huge amount of data, to reduce the pressure of hardware, when the data scale exceeds the upper limit, it is necessary to partition big data [6]. In the distributed cluster environment without fault tolerance, the efficiency of big data partitioning is inversely proportional to the hardware involved in mining [7]. Therefore, anomaly data mining of massive data is a challenging task. The traditional partition mining algorithm based on mean clustering is affected by data similarity. This kind of partition mining algorithm will produce a high communication load in the parallel process, which is difficult to achieve a high degree of parallelism [8].

There are certain research gaps in the traditional work like the problem of stabilization of chaotic characteristics in abnormal data by proposing the chaotic correlation feature extraction of huge data clustering based on the Internet of things. Also, the chaotic features in big data usually show complex folding and distortion

without obvious rules and order and non-synchronization. The chaotic features are very complex, which are described by the correlation dimension.

Thus, this article contributes to the extraction of the correlation dimension as the chaotic feature of huge data clustering. Based on the reconstruction of phase space, the 1D (one dimensional) time series can be extended in multi-dimensional space, to extract the chaotic correlation dimension features. Cluster analysis of big data is carried out according to the extracted chaotic correlation dimension (CCD). Relevant experimental analysis is carried out in this article and the traditional neural network algorithm is compared in terms of the energy consumption and processing time of the two algorithms. In the simulation parameter design, the time interval of big data packet generation is and the data is

generated at varying simulation times. In the experiment, the amount of data varied from 100MB to 1GB. The correlation dimension of this algorithm is observed to be stable, while the correlation dimension of the traditional algorithm fluctuates greatly, verifying the effectiveness of the proposed algorithm for high data clustering efficiency.

The structure of this paper is arranged as: the literature review is provided in section 2 and the huge data clustering process based on chaotic correlation dimension (CCD) feature extraction is depicted in section 3. The experimental outcomes are presented in section 4 while the conclusion is presented in section 5 of this article.
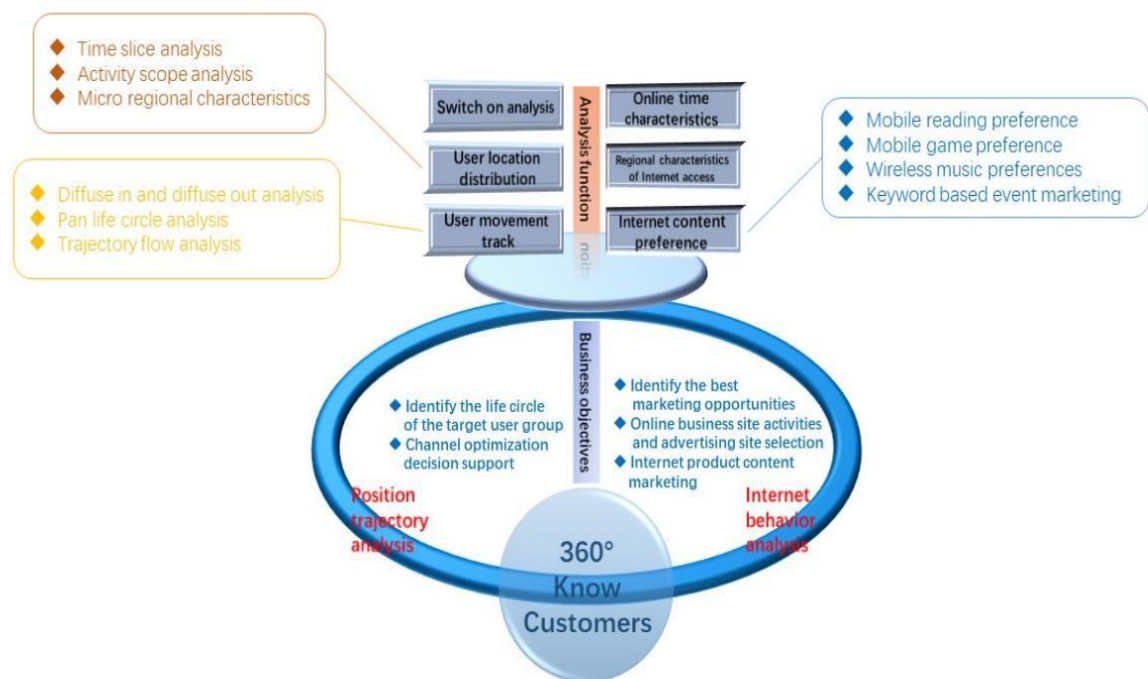

Figure 1: Big data mining and analysis platform

2   Related work

In this section, various state-of-the-art works in the field of feature extraction of big data clustering based on the Internet of Things are discussed.

For this research problem, there are many research methods related to big data clustering of the Internet of things. For example, the cluster analysis method of big data of the Internet of things proposed by Liu *et al.* [9]. Boushaki *et al.* proposed a multi-view fuzzy clustering algorithm based on the condensed information bottleneck audio event clustering method and representing point consistency constraints [10]. Single pass Bayesian fuzzy clustering algorithm and dynamic optimization cellular genetic fuzzy clustering method proposed by Yang *et al.* [11]. RNA SEQ data clustering method proposed by Park

and Lee [12]. Grid coupled data stream clustering method proposed by Cui [13].

Roy *et al.* proposed an uncertain data clustering algorithm based on Voronoi diagram in obstacle space

[14]. Fast density clustering algorithm for location big data proposed by Li *et al.* [15]. Mdfuzzyk modes clustering algorithm based on classification matrix object data proposed by Yan *et al.* [16]. Fast adaptive clustering algorithm based on representative comment scoring strategy and geographic spatiotemporal big data clustering method proposed by Chen *et al.* [17]. The clustering method of Internet of things data in the cloud proposed by song, t, and others has the ability to classify the event big data with chaotic correlation characteristics into their respective clustering centres, and can obtain

satisfactory clustering results. However, from the actual clustering effect, the above traditional methods have some key problems to be solved, such as large time consumption, slow speed, low agility, low data access load, slow convergence, large error, low efficiency of load balanced collaborative filtering. Research on more effective Internet of things big data clustering algorithm based on cloud mode event chaotic correlation feature extraction is rare [18].

Based on the current research, this paper presents the chaotic correlation feature extraction of huge data clustering based on the Internet of things. The chaotic features in big data usually show complex folding and distortion without obvious rules and order and non-synchronization. The chaotic features are very complex, which are described through the correlation dimension. In this article, the dimension of extracted correlation is used as the chaotic characteristic of huge data clustering. Based on the reconstruction of phase space, time series of one-dimensional space can be extended in multi-dimensional space, so as to extract the chaotic features of correlation dimension. Cluster analysis of big data is presented according to the extracted chaotic correlation dimension. The relevant experimental analysis depicts some simulation outcomes which show that the proposed method can accurately mine abnormal data for different large data sets, and has high feasibility and efficiency.

3 Huge data clustering algorithm depending on CCD feature extraction

This section includes the description of clustering algorithm based on chaotic correlation dimension along with the big data clustering implementation.

3.1 Feature extraction and analysis of CCD

The chaotic characteristics in big data are usually complex folding and distortion without obvious rules and order and non-synchronization. The chaotic characteristics are very complex and need to be described by correlation dimension [19].

A. Reconstruction of phase space
The data sequence belongs to nonlinear time series to a great extent, and the key of nonlinear time series is phase space reconstruction. Phase space reconstruction can keep many geometric features in the original system unchanged, establish a bridge between the original time series and multi-dimensional space analysis, and effectively extract the chaotic correlation dimension (CCD) features of data in multi bit phase space. The phase space reconstruction method is as follows: assuming that the time series is $\{x_1, x_2, \dots, x_N\}$, the phase space reconstruction result can be described as:

$$X = \left[X_1, X_2, \dots, X_K\right]$$
$$= \begin{bmatrix} x_1 x_2 \dots x_K \\ x_{1+\tau}, x_{2+\tau}, \dots, x_K \\ x_{1+(m-1)\tau}, x_{1+(m-1)\tau} \dots x_{M+(m-1)\tau} \end{bmatrix} \quad (1)$$

Wherein, $K = N - (m-1)\tau$, τ is used to describe the time delay; M is used to describe the embedding dimension. If $m \geq 2d + 1$ the geometric structure of the dynamic system will be completely opened, and d is used to describe the dimension of the chaotic attractor of the system. The selection of embedding dimension m and time delay is the key to phase space reconstruction. Only by selecting reasonable m and τ can we accurately reconstruct the phase space reflecting the characteristics of the original system. The detailed selection method is given below. For the selection of time delay τ. This study considers time delay τ denoted by the abscissa when the mutual data of delay time takes the first minimum value as the finest time delay for recreating phase space [20]. In the interval of data distribution, the probability distribution curve of data is established. p_i is used to describe the probability that $x(t)$ appears in the interval I of the data distribution curve; $p_{ij}(\tau)$ is used to describe the joint probability that $x(t)$ appears in i and delay $x(t + \tau)$ after a certain amount of delay τ appears in region j. Then the delay time mutual information can be described as:

$$I(\tau) = -\sum_{ij} p_{ij}(\tau) \ln \frac{p_{ij}(\tau)}{p_i p_j} \quad (2)$$

If $I(\tau) = 0$, $x(t + \tau)$ cannot be predicted, that is, $x(t)$ and $x(t + \tau)$ are independent of each other, and the smaller $I(\tau)$ is more independent $x(t)$ and $x(t + \tau)$. Therefore, when $I(\tau)$ reaches the minimum, the time delay τ corresponding to the abscissa can be utilized as the finest time delay for recreating the phase space. For the selection of embedding dimension m, this paper uses the virtual nearest neighbor algorithm for the estimation [21]. According to Takens theorem, the $m - dimensional$ vector formed in the $m - dimensional$ phase space can be described as:

$$X(n) = \{x(n), x(n + \tau), \dots, x(n + (m-1)\tau)\} \quad (3)$$

Obtaining the minimum embedding dimension of phase space reconstruction needs to meet the conditions described in equation (4). If yes, $X_{\eta(n)}$ is called the false nearest neighbor of X_n.

$$\frac{\left| x_{\eta(n)+m\tau} - x_{n+m\tau} \right|}{\left\| X_{\eta(n)} - X_n \right\|_2^{(m+1)}} \geq R_{tol} \quad (4)$$

Where R_{tol} is used to describe the threshold, usually R_{tol} takes 15. At this time, the proportion curve of false nearest neighbour points is required. If the proportion of false nearest neighbour points is less than 5%, it is considered that the obtained m is the minimum embedding dimension of phase space reconstruction [22].

B. Feature extraction of chaotic correlation dimension

In this paper, the extracted CCD is utilized as the chaotic element of huge data clustering. Based on phase space rebuilding, 1-D time series can be stretched out in multi-layered space to separate chaotic element aspect highlights [23]. As per the procedure analyzed in earlier section, the recreated time series can be acquired:

$$X_i = \left(x_i, x_{i+\tau}, ..., x_{i+(m-1)\tau}\right)^T \qquad (5)$$

In the $m - dimensional$ phase space recreated by the above-mentioned procedure, the focuses whose separation from phase point x_j to x_i additional x_i itself doesn't surpass r can be portrayed as:

$$Q = \sum_{j \neq i} H\left(r - \|x_i - x_j\|\right) \qquad (6)$$

Where H (*) is utilized to portray the Heaviside work. The idea of connection work is given here. All focuses that might be more modest than the given distance r are comparative with one another The extent of the complete point logarithm is known as the connection work, and the equation is portrayed as follows:

$$C_N(r) = \frac{2}{Q(Q-1)} \sum_{i=1}^{N} \sum_{j=i+1}^{N} H\left(r - \|x_i - x_j\|\right) \qquad (7)$$

In the equation, the numerator is 2 to wipe out continued counting. The distance between two-stage focuses can be acquired by depicting the distance between two-stage focuses with standard, or at least, the greatest contrast among two vectors:

$$\|x_i - x_j\| = \max_{1 \leq k \leq m} |x_{i-(k-1)\tau} - x_{j-(k-1)\tau}| \qquad (8)$$

For a vector whose distance doesn't surpass r, it tends to be called a cooperative vector [24]. Expecting that there is n $1D$ estimated succession information, the quantity of vector focuses in stage space remaking is $N = m - (m-1)\tau$. Compute the extent of the stage point logarithm with connection in all conceivable $N(N-1)/2$ sets, which is known as the relationship aspect. The recipe is depicted as follows:

$$C_m(r) = \frac{2}{N(N-1)} \sum_{i=1}^{N} \sum_{j=i+1}^{N} H\left(r - \|x_i - x_j\|\right) \qquad (9)$$

Then the relationship aspect got above is the tumultuous trademark amount of large information grouping, and the bunching of huge information is acknowledged by the connection aspect.

3.2 Big data clustering implementation

The cluster analysis is to divide different samples into several categories, and make the samples of an aggregate class more similar than those of different aggregate classes [25]. In this paper, huge data is clustered and analysed as per the extracted CCD [26-28]. The flowchart of big data clustering implementation in this article is depicted in Figure 2 and the detailed implementation is provided in this section.

A. Input samples and parameters

Enter n data samples $\{x_1, x_2, ..., x_n\}$, According to the characteristics of chaotic correlation dimension, n cluster centers are selected from the above samples and described by $\{Z_1, Z_2, ..., Z_n\}$.

B. Divide n samples into the nearest cluster according to the following principles ω_j

$$\|x - z_j\| = \min\left(\|x - z_j\|\right) \qquad (10)$$

Where $\|x - Z_j\|$ is used to describe the distance between x and Z_j. At the same time, it is assumed that there are N_i samples in ω_j.
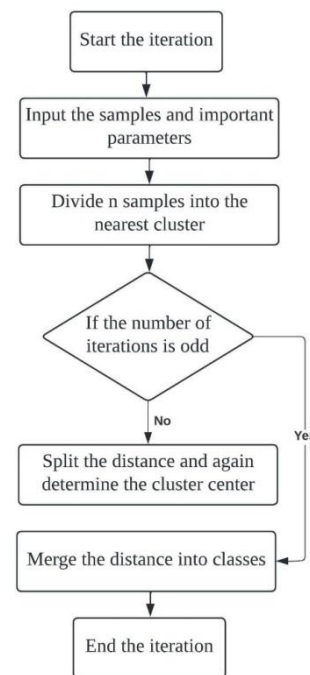

Figure 2: Flowchart of huge data clustering implementation

C. The cluster centre value is obtained by the following formula

$$z_j = \frac{1}{N_j} \sum_{x \in \omega_j} x C_m(r) \tag{11}$$

If the number of iterations is odd, proceed directly to step (5); Otherwise, follow next step.

D. Split

Assuming $L = \max(x - Z_i)$, $x \in \omega_j$, d_1 is used to describe the splitting distance. If $L > d_1$, ω_j is divided into two categories. At this time, the cluster center can be described as:

$$\begin{cases} z_{i1} = z_i + \lambda L \\ z_{i2} = z_i - \lambda L \end{cases} \tag{12}$$

Where λ is used to describe a constant greater than 0. If $L < d_1$ and the last merge operation was not performed, proceed to step (6).

E. Merge

Assuming $l = \|Z_I - Z_J\| = \|Z_i - Z_j\|$, use d_2 to describe the merge distance. If $l < d_2$, then, ω_I, ω_J are merged into one class, and the merging center can be described as:

$$z_{IJ} = \frac{1}{N_I + N_J} [N_I z_I + N_J z_J] \tag{13}$$

If $l < d_2$, and not classified last time, proceed to step (6), otherwise proceed to step (3).

F. End iteration

In this paper, the data with the same chaotic correlation characteristics are divided into one class through the above clustering analysis process, so as to realize the effective clustering of big data [29-31]. This work is also considered for the industrial applications and contributing towards social life with the integration of the Internet of Things, AI, and robotics [32-35].

4   Results and Analysis

This section presents the result analysis obtained for from the proposed big data clustering algorithm and finally presents its discussion and summary in conclusion section.

In order to validate the efficiency of the huge data clustering algorithm based on chaotic correlation feature extraction proposed in this paper, relevant experimental analysis is needed [36-38]. Taking the traditional neural network algorithm as a comparison, the energy consumption and processing time of the two algorithms are mainly compared [39-42]. In this paper, the algorithm is verified by simulation data. All the experimental programs are written in C++, which is in Ubuntu 12.04 operating system. The experimental hardware platform is

LenovoM4390 (i3-2100 CPU, 4UB memory, 2TB disk), processor Intel (R) core (TM) 2duocpu2 94GHz, memory: 8.00GB. In the simulation parameter design, the time interval of big data packet generation is 0.1s, and the data is generated from the simulation time of 300s. In the experiment, the amount of data is from 100MB to 1GB, with 100MB as the unit, the data increases nonlinearly, discrete scheduling and interval boundary approximation are carried out for big data, the time interval of big data feature acquisition is 0.1s, and the parameter configuration is listed in Table 1.

Parameter	Value (Mbps)
Data quantity	1000
Number of big data distribution Characteristics	5
Load per data access system	16
Data complexity size (GB)	2
Data execution time delay (MS)	2400
Maximum queue size	2200

Table 9: Parameter configuration

The algorithm in this paper and the traditional algorithm are used to cluster different amounts of data, and the clustering efficiency of the two algorithms is counted. The outcomes are listed in Table 2 and graphical represented is provided in Figure 3.

Data volume	Time required for the proposed algorithm (s)	Time required for traditional algorithm (s)
200	1925	5998
400	4433	12769
800	8343	29151
1024	10151	35832

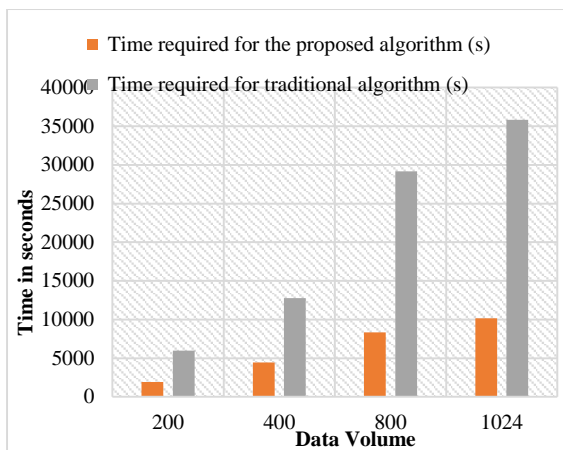Table 10: Comparison results of clustering efficiency of two algorithms

Figure 3: Graphical comparison of clustering efficiency of two algorithms

It can be observed from the analysis of Table 1 and Figure 3 that with the gradual increase of the amount of data, the time for data clustering of this algorithm and the traditional algorithm gradually increases. While this improvement occurs, the processing time required by this algorithm has been potentially lower than that of the traditional algorithm, which shows that this algorithm has high data clustering efficiency and verifies the effectiveness of this algorithm.

In order to further validate the effectiveness of this algorithm, this paper compares the energy consumed by the two algorithms to process the same amount of data. The results are shown in Figure 4.

By analysing Figure 4, it can be seen that when processing the same amount of data, the energy consumption of this algorithm is significantly lower than that of the traditional algorithm. This is because this algorithm is easy to implement and has high clustering efficiency for data, so it consumes less energy, which verifies the effectiveness of this algorithm.
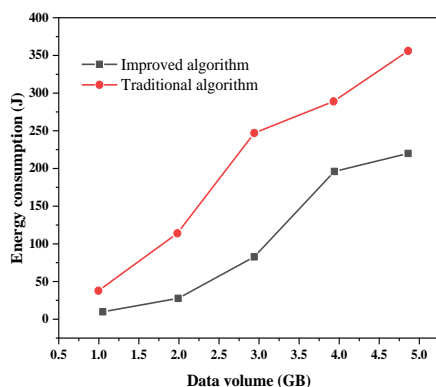

Figure 4: Comparison results of energy consumption of two algorithms

By analysing Figures 5, 6 and 7, it can be seen that when processing the same amount of data, the time required by the algorithm in this paper is significantly lower than that of the traditional algorithm. This is

because the algorithm in this paper is easy to implement and has good clustering efficiency for data, so the clustering time is short, which further verifies the effectiveness of the algorithm in this paper.
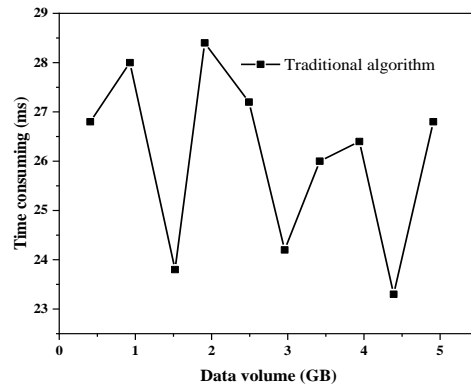

Figure 5: Time consuming of traditional algorithm
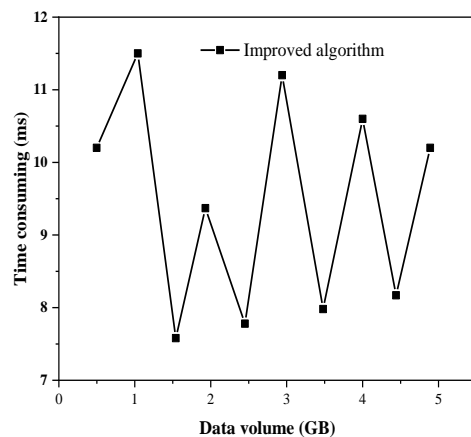

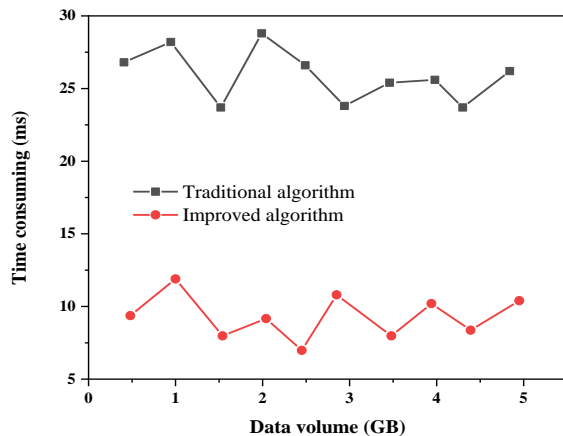Figure 6: Time consuming of improved algorithm

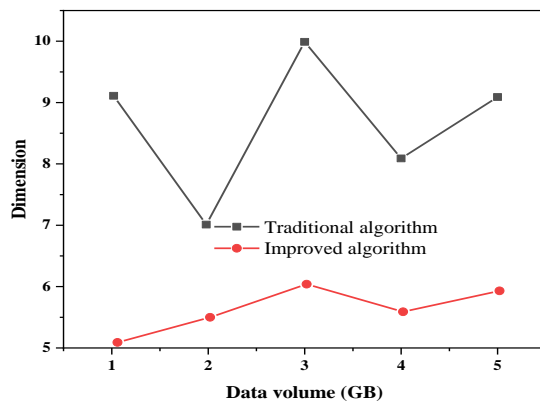Figure 7: Time consuming comparison results of two algorithms

Figure 8: Comparison results of correlation dimensions of two algorithms

It can be seen from the analysis of Figure 8 that with the gradual increase of the amount of data, the correlation dimension of the algorithm in this paper tends to be stable, while the correlation dimension of the traditional algorithm fluctuates greatly. This fluctuation shows that the algorithm in this paper has high data clustering efficiency and verifies the effectiveness of the algorithm in this paper [43-44].

5   Conclusions

This article presents the CCD feature extraction of huge data clustering based on the Internet of things is proposed. By reconstructing the phase space, a multi-dimensional state space vector and chaotic trajectory are established. It was revealed that many geometric features in the creative scheme remain unchanged, which provides an effective basis for analysing the chaotic characteristics of the original system. The false adjacent neighbour procedure is used to select the finest embedding dimension. The extracted CD is used as the chaotic feature of huge data clustering, and the big data is clustered according to the extracted chaotic correlation dimension. Simulation outcomes show that the proposed method can accurately mine abnormal data for different large data sets, and has high feasibility and efficiency. At present, the composition structure, operation mechanism and relevant standards of the Internet of things in cloud mode have not been completely unified. This can act as the future research scope of this article and therefore, the research on big data clustering of the Internet of things needs to be further discussed in many aspects in the future part of this research work.

References

[1] Bu, F. (2018). An efficient fuzzy c-means approach based on canonical polyadic decomposition for clustering big data in IoT. *Future Generation Computer Systems*, *88*, 675-682. *https://doi.org/10.1016/j.future.2018.04.045*

[2] Bu, F., Hu, C., Zhang, Q., Bai, C., Yang, L. T., & Baker, T. (2020). A Cloud-Edge-aided Incremental High-order Possibilistic c-Means Algorithm for Medical Data Clustering. *IEEE Transactions on Fuzzy Systems*, *29*(1), 148-155. *10.1109/TFUZZ.2020.3022080*

[3] Liu, Y., Zhang, J., & Zhan, J. (2021). Privacy protection for fog computing and the internet of things data based on blockchain. *Cluster Computing*, *24*(2), 1331-1345. *https://doi.org/10.1007/s10586-020-03190-3*

[4] Lye, G. X., Cheng, W. K., Tan, T. B., Hung, C. W., & Chen, Y. L. (2020). Creating personalized recommendations in a smart community by performing user trajectory analysis through social internet of things deployment. *Sensors*, *20*(7), 2098. *https://doi.org/10.3390/s20072098*

[5] Cai, G., Fang, Y., Chen, P., Han, G., Cai, G., & Song, Y. (2020). Design of an MISO-SWIPT-aided code-index modulated multi-carrier M-DCSK system for e-health IoT. *IEEE Journal on Selected Areas in Communications*, *39*(2), 311-324. *10.1109/JSAC.2020.3020603*

[6] Xia, H., Huang, W., Li, N., Zhou, J., & Zhang, D. (2019). PARSUC: A parallel subsampling-based method for clustering remote sensing big data. *Sensors*, *19*(15), 3438. *https://doi.org/10.3390/s19153438*

[7] Arora, S., Sharma, M., & Anand, P. (2020). A novel chaotic interior search algorithm for global optimization and feature selection. *Applied Artificial Intelligence*, *34*(4), 292-328. *https://doi.org/10.1080/08839514.2020.1712788*

[8] Maddumala, V. R. (2020). Big Data-Driven Feature Extraction and Clustering Based on Statistical Methods. *Traitement du Signal*, *37*(3). *10.18280/ts.370305*

[9] Liu, W., Wang, X., & Peng, W. (2019). Secure remote multi-factor authentication scheme based on chaotic map zero-knowledge proof for crowdsourcing internet of things. *IEEE Access*, *8*, 8754-8767.

10.1109/ACCESS.2019.2962912

[10] Boushaki, S. I., Kamel, N., & Bendjeghaba, O. (2018). A new quantum chaotic cuckoo search algorithm for data clustering. *Expert Systems with Applications*, *96*, 358-372. *https://doi.org/10.1016/j.eswa.2017.12.001*

[11] Yang, Q., Ruan, J., Zhuang, Z., & Huang, D. (2019). Chaotic analysis and feature extraction of vibration signals from power circuit breakers. *IEEE Transactions on Power Delivery*, *35*(3), 1124-1135. *10.1109/TPWRD.2019.2934123*

[12] Park, S. W., & Lee, I. Y. (2019). Enhanced signature RTD transaction scheme based on Chebyshev polynomial for mobile payments service in IoT device environment. *The Journal of Supercomputing*, *75*(8), 4617-4637. *https://doi.org/10.1007/s11227-018-2546-8*

[13] Cui, Y. (2018). Application of the improved chaotic self-adapting monkey algorithm into radar systems of internet of things. *IEEE Access*, *6*, 54270-54281. *10.1109/ACCESS.2018.2869632*

[14] Roy, S., Chatterjee, S., Das, A. K., Chattopadhyay, S., Kumari, S., & Jo, M. (2017). Chaotic map-based anonymous user authentication scheme with user biometrics and fuzzy extractor for crowdsourcing Internet of Things. *IEEE Internet of Things Journal*, *5*(4), 2884-2895. *10.1109/JIOT.2017.2714179*

[15] Li, L., Wen, G., Wang, Z., & Yang, Y. (2019). Efficient and secure image communication system based on compressed sensing for IoT monitoring applications. *IEEE Transactions on Multimedia*, *22*(1), 82-95. *10.1109/TMM.2019.2923111*

[16] Yan, Z., Liu, J., Vasilakos, A. V., & Yang, L. T. (2015). Trustworthy data fusion and mining in Internet of Things. *Future Generation Computer Systems*, *49*(C), 45-46. *https://doi.org/10.1016/j.future.2015.04.001*

[17] Chen, F., Li, Q., Li, M., Huang, F., Zhang, H., Kang, J., & Wang, P. (2021). Unclonable fluorescence behaviors of perovskite quantum dots/chaotic metasurfaces hybrid nanostructures for versatile security primitive. *Chemical Engineering Journal*, *411*, 128350. *https://doi.org/10.1016/j.cej.2020.128350*

[18] Song, T., Li, R., Mei, B., Yu, J., Xing, X., & Cheng, X. (2017). A privacy preserving communication protocol for IoT applications in smart homes. *IEEE Internet of Things Journal*, *4*(6), 1844-1852. *10.1109/JIOT.2017.2707489*

[19] Alarifi, A., Sankar, S., Altameem, T., Jithin, K. C., Amoon, M., & El-Shafai, W. (2020). A novel hybrid cryptosystem for secure streaming of high efficiency H. 265 compressed videos in IoT multimedia applications. *IEEE Access*, *8*, 128548-128573. *10.1109/ACCESS.2020.3008644*

[20] Niu, Z., Zheng, M., Zhang, Y., & Wang, T. (2019).

A new asymmetrical encryption algorithm based on semitensor compressed sensing in WBANs. *IEEE Internet of Things Journal*, *7*(1), 734-750. *10.1109/JIOT.2019.2953519*

[21] Li, L., Liu, L., Peng, H., Yang, Y., & Cheng, S. (2018). Flexible and secure data transmission system based on semitensor compressive sensing in wireless body area networks. *IEEE Internet of Things Journal*, *6*(2), 3212-3227. *10.1109/JIOT.2018.2881129*

[22] Yung, C., Chen, C. C., Yuan, Y. L., & Li, C. (2019). A Systematic Model of Big Data Analytics for Clustering Browsing Records into Sessions Based on Web Log Data. *J. Comput.*, *14*(2), 125-133. *10.17706/jcp.14.2.125-133*

[23] Jang, S. W., & Kim, G. Y. (2017). A monitoring method of semiconductor manufacturing processes using Internet of Things–based big data analysis. *International Journal of Distributed Sensor Networks*, *13*(7), 1550147717721810. *https://doi.org/10.1177/1550147717721810*

[24] Gong, X., Liu, L., Fong, S., Xu, Q., Wen, T., & Liu, Z. (2019). Comparative research of swarm intelligence clustering algorithms for analyzing medical data. *IEEE Access*, *7*, 137560-137569. *10.1109/ACCESS.2018.2881020*

[25] Lee, Y. C., Huang, S. C., Huang, C. H., & Wu, H. H. (2016). A new approach to identify high burnout medical staffs by kernel k-means cluster analysis in a regional teaching hospital in Taiwan. *Inquiry: The Journal of Health Care Organization, Provision, and Financing*, *53*, 0046958016679306. *https://doi.org/10.1177/0046958016679306*

[26] Shabaz, M., Sharma, A., Al Ajrawi, S., & Estrela, V. V. (2022). Multimedia-based emerging technologies and data analytics for Neuroscience as a Service (NaaS). *Neuroscience Informatics*, *2*(3), 100067. *https://doi.org/10.1016/j.neuri.2022.100067*

[27] Poongodi, M., Hamdi, M., Malviya, M., Sharma, A., Dhiman, G., & Vimal, S. (2022). Diagnosis and combating COVID-19 using wearable Oura smart ring with deep learning methods. *Personal and ubiquitous computing*, *26*(1), 25-35. *https://doi.org/10.1007/s00779-021-01541-4*

[28] Kumbinarasaiah, S., & Raghunatha, K. R. (2021). A novel approach on micropolar fluid flow in a porous channel with high mass transfer via wavelet frames. *Nonlinear Engineering*, *10*(1), 39-45. *https://doi.org/10.1515/nleng-2021-0004*

[29] Wang, H., Sharma, A., & Shabaz, M. (2022). Research on digital media animation control technology based on recurrent neural network using speech technology. *International Journal of System Assurance Engineering and Management*, *13*(1), 564-575. *https://doi.org/10.1007/s13198-021-01540-x*

[30] Ting, L., Khan, M., Sharma, A., & Ansari, M. D. (2022). A secure framework for IoT-based smart climate agriculture system: Toward blockchain and

edge computing. *Journal of Intelligent Systems*, *31*(1), 221-236. *https://doi.org/10.1515/jisys-2022-0012*

[31] Sharma, D., Kaur, R., Sandhir, M., & Sharma, H. (2021). Finite element method for stress and strain analysis of FGM hollow cylinder under effect of temperature profiles and inhomogeneity parameter. *Nonlinear Engineering*, *10*(1), 477-487. *https://doi.org/10.1515/nleng-2021-0039*

[32] Ren, Y., Rubaiee, S., Ahmed, A., Othman, A. M., & Arora, S. K. (2022). Multi-objective optimization design of steel structure building energy consumption simulation based on genetic algorithm. *Nonlinear Engineering*, 11(1), 20-28. *https://doi.org/10.1515/nleng-2022-0012*

[33] Singh, P. K., & Sharma, A. (2022). An intelligent WSN-UAV-based IoT framework for precision agriculture application. *Computers and Electrical Engineering*, *100*, 107912. *https://doi.org/10.1016/j.compeleceng.2022.107912*

[34] Zeng, H., Dhiman, G., Sharma, A., Sharma, A., & Tselykh, A. (2021). An IoT and Blockchain-based approach for the smart water management system in agriculture. *Expert Systems*, e12892. *https://doi.org/10.1111/exsy.12892*

[35] Sharma, A., & Singh, P. K. (2021). UAV-based framework for effective data analysis of forest fire detection using 5G networks: An effective approach towards smart cities solutions. *International Journal of Communication Systems*, e4826. *https://doi.org/10.1002/dac.4826*

[36] Sharma, A., Singh, P. K., & Kumar, Y. (2020). An integrated fire detection system using IoT and image processing technique for smart cities. *Sustainable Cities and Society*, *61*, 102332. *https://doi.org/10.1016/j.scs.2020.102332*

[37] Gunupudi, R. K., Nimmala, M., Gugulothu, N., & Gali, S. R. (2017). CLAPP: A self constructing feature clustering approach for anomaly detection. *Future Generation Computer Systems*, *74*, 417-429. *https://doi.org/10.1016/j.future.2016.12.040*

[38] Sharma, A. (2021). Integrity and Multimedia Data Management using Emerging Technologies in the Healthcare Applications-Part II. Recent Advances in Electrical & Electronic Engineering (Formerly Recent Patents on Electrical & Electronic Engineering), 14(7), 698-699. *https://doi.org/10.2174/235209651407211104091930*

[39] Wang, Y., Chen, Q., Kang, C., & Xia, Q. (2016). Clustering of electricity consumption behavior dynamics toward big data applications. *IEEE transactions on smart grid*, *7*(5), 2437-2447. *https://doi.org/10.1109/TSG.2016.2548565*

[40] Guo, Z., & Xiao, Z. (2021). Research on online calibration of lidar and camera for intelligent connected vehicles based on depth-edge matching. *Nonlinear Engineering*, 10(1), 469-476. *https://doi.org/10.1515/nleng-2021-0038*

[41] Deng, Z., Hu, Y., Zhu, M., Huang, X., & Du, B. (2015). A scalable and fast OPTICS for clustering trajectory big data. *Cluster Computing*, *18*(2), 549-562. *https://doi.org/10.1007/s10586-014-0413-9*

[42] Chen, Y., Zhang, W., Dong, L., Cengiz, K., & Sharma, A. (2021). Study on vibration and noise influence for optimization of garden mower. *Nonlinear Engineering*, 10(1), 428-435. *https://doi.org/10.1515/nleng-2021-0034*

[43] Sharma, A., Singh, P. K., Hong, W. C., Dhiman, G., & Slowik, A. (2021). Introduction to the Special Issue on Artificial Intelligence for Smart Cities and Industries. Scalable Computing: Practice and Experience, 22(2), 89-91. *https://doi.org/10.12694/scpe.v22i2.1939*

[44] Luna-Romera, J. M., García-Gutiérrez, J., Martínez-Ballesteros, M., & Riquelme Santos, J. C. (2018). An approach to validity indices for clustering techniques in big data. *Progress in Artificial Intelligence*, *7*(2), 81-94. *https://doi.org/10.1007/s13748-017-0135-3*