Interpolation Analysis of Industrial Big Data Based on KDR Knowledge Recognition Algorithm Considering Singular Value Decomposition Theory

Cenglin Yao[1,2], Yongzhou Li[1*]
[1]Evergrande School of Management, Wuhan University of Science and Technology, Wuhan, Hubei,430081, China
[2]College of Mechanical and Electrical Engineering, Wuhan Business University, Wuhan, Hubei 430056, China
E-mail: yl200808@126.com, 20150511@wbu.edu.cn
[*]Corresponding author

Although various algorithms have made some progress in the current research of industrial big data interpolation, most of them are only suitable for static KDR operation methods. Most of the data is not achieved overnight but in an incremental manner. For example, the data will increase with time. In the process of data collection, to ensure the consistency of KDR calculation results under dynamic conditions, the same and different information in the old and new data must be merged, to disperse the dynamic data. According to the increasing properties of data in Industrial big data analysis, a dynamic KDR operation model is established by considering singular value decomposition (SVD) theory. To ensure consistency before and after static separation, a rough set method based on Manilkara is used. Under the influence of Yalo' s singular value decomposition (SVD) theory, the conventional interval is divided into two parts: the core and the blank to express the unstable interval. This method uses the method based on the middle interval. By dividing the middle interval again, the interval between the old and the new data is combined.

Povzetek: V raziskave industrijskih velikih podatkov so vpeljali dinamični model KDR, ki uporablja teorijo singularne razčlenitve (SVD) in grobo množico Manilkara za zagotavljanje konsistentnosti med dinamičnimi in statičnimi podatki.

1 Introduction

In 1987, KDR was first formally introduced. To solve the problem of incomplete numerical properties, Wang and Chiu2 gave EFKDR algorithm with equal frequency respectively, and the EWKDR algorithm with equal width was introduced in the same year. Then, the KDR operation method has many directions according to the development of the problem. For example, in 1991, Huang and Chiu3' proposed the KDR computation of the basic maximum direct coefficient, so that the corresponding interval number can be automatically obtained according to the characteristics of the data in the KDR computation processing. In 2018, "Hacibeyoglu and Ibrahim" proposed a KDR algorithm for Euonymus according to the EF method.

In maximum likelihood estimation, the distribution of boundaries is used to estimate an uncertain parameter, usually using the expected maximum (EM).EM method requires a large amount of data in theory to ensure the asymptotic and normality of the estimation. However, the EM algorithm's local optimal solution can be found quickly, its convergence is sluggish, and its operation is quite intricate.

The study assumed that the missing data also existed and that the missing data were the most valuable, rather than simply missing data, so the data were analyzed in a non-processing manner. For example, if a user uses the attribute "obesity," the item "weight" will not be added, but will appear directly on the "empty number". The method analyzes the original data directly, without any preprocessing. The most typical ones are Bayesian networks and neural networks. Bayesian network is used to

describe the probability of association between features, which is used to reveal the association between data and provide a natural expression for it. In a Bayesian network, features are represented as features, which correlate with attributes represented laterally. The necessity of a Bayesian network for data collection, and also is to have the understanding of the data collection and the correlation between each attribute is clear. Based on this, we must first analyze these features, or all of these features are added to the model, the cause of the complexity of the Bayesian network, but also as the number of properties and geometric ratio increase. Although the artificial neural network is also a very popular machine learning technique in the present situation, it needs more work to interpolate the missing data [1-6].

In the traditional KDR operation method, the solution of interval cut-off is a game method. In the actual KDR operation, a point cannot be used to describe the boundary point of an interval. In the same period, the boundary of an interval will also be different; At the same time, an interval can have multiple uncertain tangents. Based on this situation, we give an interval that takes into account the theory of singular value decomposition. Using the idea of a boundary interval in three intervals, the uncertain region is extracted to form a boundary interval. When performing dynamic fusion, the intervals in the image need to be re-segmented to achieve the purpose of image delay, to achieve the purpose of KDR operation on the image. The three intervals can not only express the uncertainty of the intervals but also delay the boundary intervals in the dynamic KDR operation [10-14].

2 Related work

In machine learning, clustering is one of the main methods used to analyze non-monitored data, but it still has many problems to be solved. When clustering, objects close to the edge of the cluster cannot be efficiently classified. Yul3l gave three clustering classification methods under the influence of singular value decomposition theory.

Compared with the traditional cluster represented by a single set, the 3D group is a new representation of clusters. In three subgroups, the objects of a cluster are divided into two groups, and the universe of a cluster is divided into core domain, edge domain, and irrelevant domain.

Taking into account the singular value decomposition theory, the central and boundary objects of different categories can be compared respectively. Based on considering the singular value decomposition theory, this topic has been deeply discussed and some results have been obtained. Given the uncertainty of objects and clusters in multi-view clustering, Yu et al. L6l proposed a low-order ternary principal component classification method, which can not only reflect the correlation between objects and clusters but also effectively raise the multidimensional clustering's accuracy. To solve the large-scale clustering problem, Yu et al. introduced the fusion architecture of three clusters in 2019, which combined the clustering method with the clustering method, which could not only ensure the quality of the cluster but also reduce the time cost of calculation. In 2016, Yu et al. proposed singular value decomposition (SVD) theory based on different vertex distribution characteristics to achieve different vertex representations. Under the influence of the singular value decomposition theory, Liu et al. proposed an approach based on centroids to address the dynamic problem of overlapping communities [7-9]. Table 1 shows the summary table of the previous literatures.

Table 1: Summary of the literature review

Author	Objective	Findings
[18]	The manuscript explores these concepts and provides a case study that demonstrates the implementation of new intelligent hybrid algorithms for Industry 4.0 applications with limited data.	An extremely accurate deterministic model that fits the real data was generated by applying the suggested approach. We also employed the UKF technique to strengthen the model's resistance to uncertainty.
[19]	Big data analytics and related applications in smart grids are	The result focuses on the complex applications of different data analytics in

	introduced in this study. The initial section covers the features of big data, smart grids, and huge data collection to illustrate the goal and potential advantages of incorporating advanced data analytics in smart grids.	smart grids. The current power system may gain a lot from handling massive amounts of data from geographical information systems, meteorological information systems, and energy networks, among other sources. In the big data era, this will also enhance customer service and societal welfare.
[20]	The work presented a novel method known as the interpretable kernel DR algorithm (I-KDR), which converts Information from the characteristics space into a lower-dimensional space where classes are closer together and there is less overlap.	Furthermore, the dimensions are created by the algorithm based on the local contributions of the data samples, which facilitates their interpretation by class labels. Furthermore, we effectively combine the feature selection task with the DR to identify the original space's most pertinent characteristics for the discriminative purpose.
[21]	This paper presents a new technique that transfers The feature space corresponds to data to a lower dimensional space where classes are closer together and there is less overlap: the interpretable kernel DR algorithm (I-KDR).	Additionally, the approach produces the dimensions based on the data samples' local contributions, which make it easier to comprehend the data by class labels.
[22]	For an improved intrusion detection system, the suggested system uses a Hybrid Deep Learning (HDL) network made up of Long Short-Term Memory (LSTM) and Convolutional Neural Network (CNN).	PySpark, which offers Python support for the study made use of the Apache Spark technology in the Google Colab environment. The CIDDS-001 data set was used to evaluate the model in multiclass mode, and the UNS-NB15 data set was used to evaluate the model in binary mode.

| [23] | The fusion of conventional machine learning algorithms with big data technology has created novel and captivating obstacles in domains such as social media and social networks. Data processing, data storage, data representation, and the use of data for pattern mining, user behavior analysis, data visualization, and data tracking are among the primary issues addressed by these new challenges. | Big data has emerged as a significant topic for several study fields, including social networks, data mining, machine learning, computational intelligence, information fusion, and the Semantic Web. The development of several big data frameworks for huge data processing based on the Map Reduce paradigm, like Apache Hadoop and, more recently, Spark, has made it possible to use data mining techniques and machine learning algorithms effectively across a variety of domains. |

3   Research methods

In general, statistical principles and machine learning techniques are used to interpolate the defaults to obtain a complete data set. At present, the following interpolation methods are usually used in data processing:

3.1 Manual interpolation

This method is done manually based on human experience. That is, over the years, they have relied on their years of experience and interpolating this type of data, so the manual interpolation method is usually more accurate. However, with a large number of missing values, this algorithm can take a lot of time and effort.

$$\bar{R}(X) = \cup \quad \{Y_i \mid Y_i \in U/IND(R) \wedge Y_i \cap X \neq \emptyset\} \tag{1}$$
$$\underline{R}(X) = \cup \quad \{Y_i \mid Y_i \in U/IND(R) \wedge Y_i \subseteq X\}$$

Where, represents the partition of ambiguous relation on. $U/IND(R) = \{X \mid (X \subseteq U \wedge \forall_{x \in X, y \in X, \in \in R}(a(x) = a(y)))\}RUY_i$ Represents a set of objects, which can be regarded as an equivalence class.$Y_i U/IND(R)Y_i$ When the intersection with the objective function is not an empty set, the upper approximation set requirement is satisfied.$Y_i X Y_i$ According to the theory of singular value decomposition (SVD), two classes are divided into three classes. In the case of ambiguity, the ambiguous
information is expressed, and the processing of the ambiguous information is realized. This strategy is based on human experience and is frequently regarded as correct. But, especially when dealing with big datasets, it can be labor-time-intensive. The volume of the data may make manual interpolation impractical in an industrial big data scenario where efficiency and accuracy are critical.

3.2 Specific value imputation

This algorithm uses a special value to interpolate the missing data, which is independent of other eigenvalues. If the missing data were all labeled "empty," a completely different set of data would be produced, but this would be highly biased and generally not beneficial.
Although the sequential clustering method is used to obtain the boundary of the original adjacent interval, there is still some uncertainty. There are two types of merging and disjunction between two adjacent regions. If the center distance between the two spaced groups is small, and the sample groups in the interval have high similarity, if the stable value between the two intervals is still high, then the combination operation is performed. On the contrary, a new segment can be formed by splitting the intersecting part of the two intervals.

$$\lambda_a = \frac{|V_a|}{F_a} \tag{2}$$

It represents a stable region, and the higher its value, the less stable it will be. A parameter represents a range. The total number of samples in the data represents the width of the total sample value of the individual number. Since the distribution density of each data is different, a parameter is adopted to balance the relationship between the sampling frequency and the width of the interval between each attribute. The relationship is as follows:

$$\varphi_a(d_i) = \lambda_a \frac{F_{d_1}}{F} + (1 - \lambda_a)\frac{W_{d_1}}{W}, \lambda_a \in [0,1] \tag{3}$$

The stability of the ordered cluster can be found by using this equation. The stability increases with the decrease of

the stable value. The formula method is used to describe the attribute KDR operation degree of numerical values. The closer the value is to 1, the smaller the attribute KDR operation is, and the higher the attribute KDR operation is. Regardless of other eigenvalues, this technique gives missing data a unique value. Even if it can be simple, if the assigned value is not indicative of the missing data, bias might be introduced. Furthermore, it might not fully represent the intricacy of industrial data, where missing values could have a big influence on studies performed later on.

3.3 Average imputation

In the data set, the eigenvalues are divided into continuous and discontinuous values, and the average interpolation is carried out according to the size of their eigenvalues. If the omission is of continuous type, the missing data are interpolated according to the average value of the eigenvalue. If the blank is of continuous type, the most (that is, the most frequent value) in the eigenvalue of the missing data is interpolated according to the method of the statistical method. A similar idea is used in the imputation algorithm of the conditional mean, that is, the average of the missing data is carried out by this algorithm, but it is not selected from all the objects in the data set, but obtained from the target which is consistent with the target's judged eigenvalue. In contrast, the basic idea of both algorithms is to use the maximum possible value to insert the missing data, but the specific implementation will be different depending on the difference in the specific data. Using the average eigenvalue as a guide, this method interpolates missing data. It is easy to use and straightforward, however, it assumes that the dataset is homogeneous, which may not be the case in industrial settings where data can be varied and heterogeneous.

3.4 Hot card insertion

Hot card interpolation is looking for and missing from the original data collection value closest to the object, the object's characteristics and use of interpolation, and through the study of the interpolation of the different characteristics of the object, and through the data between the interconnected to estimate the missing data, its deficiency is unable to define similarity and there is much subjective influence.

In this way, a single numerical feature can be assigned to the evaluation object. The maximum deviation value

represents a numerical value as an attribute, and the magnitude of the value determines the importance of the element. Represents the number of data samples. Is for the number characteristic in a pair. This is a special ability. Compared with the principal component analysis (PCA) method, the variance-based maximum shift method is more convenient, but it is not universal and does not involve the interaction between attributes.

MIC, the maximum information, is used to determine the correlation of each attribute and is the largest parameter less information-based inquiry can not only detect different attributes but also find different attributes, to reflect their importance. Conventional MIC only studies a single data type,

$$\mathrm{IMIC}(a, A) = \sum_{b \in A-a} \mathrm{MIC}(a, b) \qquad (4)$$

Based on MIC, an IMIC algorithm for determining the importance of attributes based on MIC is proposed. The MIC of each attribute is superimposed with the MIC of other attributes to reflect the correlation between each attribute, to determine the importance of each attribute without supervision.

In the practical case, we have assumed a normal prosperous present value. On this basis, the data sampling values of each interval meet the normal allocation, and it is regarded as the center point of the interval. Based on the sample size represented by the characteristic values of statistical data, the interval with the largest sample size in a certain region is regarded as the center of the cluster.

$$\begin{aligned} \mathrm{KL}_i &= \big(f(d_i) - f(d_{i-1})\big)/(d_i - d_{i-1}) \\ \mathrm{KR}_i &= \big(f(d_i) - f(d_{i+1})\big)/(d_i - d_{i+1}) \\ K_i &= \mathrm{KL}_i \times \mathrm{KR}_i \end{aligned} \qquad (5)$$

This technique uses the properties of neighboring values in the dataset to approximate missing data. However, defining similarity and figuring out which values are closest might be subjective, which could produce biased findings, particularly in intricate industrial datasets with a variety of features.

3.5 K-adjacent interpolation

This algorithm is based on the nearest K sampling values of the object to the missing data and uses the weighted average of the K sampling values to estimate [3]. K-nearest neighbor interpolation uses the hierarchical clustering

pattern to estimate the missing data class and interpolates the average of this class.

The idea is to insert all the values in the missing data and see if the inserted values achieve the best interpolation result. As you can imagine, such an algorithm is indeed a good solution in the case of a small amount of data, but if the number of data is too large, the data is lost too much, and a lot of data is needed, so a lot of data is needed to test. To prove that the data has the characteristics of regularity, that is data clustering. By the degree of concentration of these data, we can find the average value in these data. In this paper, the statistics and analysis of Sat image in UCI are presented.

On this basis, the maximum shift method is used to solve the attribute importance. The larger the value of variance is, the existence of the item is indicated.

The higher the uncertainty, the higher the importance. The specific algorithm is as follows:

$$Y_{a_i} = \sum_{j=1}^{n} \sum_{k=1}^{n} \sqrt{\left(\frac{v^j - v_i^k}{w_{a_i}}\right)^2}, i = 1, 2, \ldots, k \quad (6)$$

Using the closest K sampling values, this technique predicts missing data. Despite providing a mathematical methodology and its efficacy may differ based on the distribution and features of industrial big data.

3.6 A comprehensive approach

This method also allows for a test on the missing data, with the difference that the inserted data is the best of the final attribute reductions to be used as an interpolation for the missing data. This algorithm can improve the accuracy of the algorithm on the premise of increasing the complexity of the operation. Obviously, in the case of big data, there will be a lot of loss of these data, so this algorithm can obtain good interpolation accuracy, but it costs a lot of time. Most of the commonly used methods for missing information interpolation are regression-based methods. However, all regression methods are built based on complete data, so the data should be pre-processed. In the case of missing data, the known feature quantities are used to predict, and the method is used to predict the missing data. This paper also focuses on the interpolation method of the missing data. The solution of MIC value is the largest way of parameter-less search based on data. According to the experimental results of reference [66], the larger the MIC value, the greater the correlation between the two

properties. If the MIC of both personalities is 1, it indicates that there is a linear relationship between the two personalities' divides the two-dimensional data into grids, and accumulates the mutual information in the grids to obtain the initial mutual information system. Finally, the cumulative value obtained from the different grid assignments is the MIC value, square formula (2.7). MIC can be used to find out the direct relationship of each attribute, as well as their internal relationship, to reflect the importance of each attribute.

$$MIC(a, b) =$$

$$\max_{n \times m < B} \left(\frac{\sum_{x \in X, y \in Y} P(x,y) \log \frac{P(x,y)}{\sum_{x \in X} P(x,y) \sum_{y \in Y} P(x,y)}}{\log m \quad \{n,m\}} \right) \quad (7)$$

Each interval in the core interval is represented by the traditional interval representation, and the cut points in each interval are obtained by the static singular value decomposition (SVD) algorithm based on the data samples. These cut points are supported by the data, and the corresponding attribute values can be found in the original data samples [15-17].

$$ED_a = \{[p_{min}, p_{1,3}), [p_{1,3}, p_{2,3}), \ldots, [p_{k-1,3}, p_{max}]\} \quad (8)$$

$$BD_a = \{(p_{1,2}, p_{2,1}), (p_{2,2}, p_{3,1}), \ldots, (p_{k-1,2}, p_{k,1})\} \quad (9)$$

Each interval in the blank interval is represented by the traditional interval representation, and each interval is based on the interval segments not included in the corresponding core interval set, as shown in Table 2 below:

Table 2: Case table of singular value decomposition theory

u_1	b	c	d
u_2	0.8	2	1
u_3	1	0.5	0
u_4	1.3	3	0
u_5	1.4	1	1
u_6	1.4	2	1
u_7	1.6	3	1

Data are assumed to be incremented over time, and the nature of numerical data follows a normal assignment. If a data set containing numeric characteristic data is entered at a certain time, the continuity between all characteristics of the data is taken into account, so that the resulting data intervals have a certain order, that is, a priority in the case of sequence clusters. On this basis, firstly, the feature set of numbers is extracted to determine the importance of the attributes of the numbers, and then the corresponding attribute set is obtained by sorting the numbers in order of size. The expected KDR digit feature is regarded as the net ranking of each data in the sequence cluster, and the sequence clustering method is used to sort the unlabeled information to obtain the three interval sets of the initial KDR operation. Then, the uncertain regions in the middle are re-evaluated and segmented, and then the original data is KDR operation. Finally, the data set after the unlabeled KDR operation is obtained by successive KDR operations on the numeric attribute group. These features include the features of the KDR operation and the features of the KDR operation, and then these features are introduced into the new separated features. In this method, the dynamic KDR operation method is adopted, and the combination process is carried out to obtain a new KDR operation information system. Although it requires more computer power and complexity, this method combines several approaches to increase interpolation accuracy. In industrial settings where precision is crucial, this method might make up for the extra computing burden.

4   Result analysis

In Table 3 below, statistics of the NB algorithm are performed on the data after the KDR operation. In general, the NB algorithm has the same good performance as the Euonymus algorithm. From a personal point of view, this method is better than EF1l, EW9, and EF_Unique7.In general, the results of 27 out of 33 trials were better than those of the control method, reaching 81.82%. The average accuracy of the proposed method was 82.53%, ranking first. Compared with 81.07, which ranked second, it increased by 1.46 percentage points. The statistical analysis of the suggested approach combining KDR operation with NB, KNN, and C4.5 algorithms will yield encouraging results. With a noteworthy gain of 1.46 percentage points over the second-ranking approach, the average accuracy of 82.53% beats alternative methods. Furthermore, with 81.82% of trials producing better outcomes than control approaches, the suggested strategy shows a significant improvement over other method. The suggested approach specifically

shows superiority in 78.79% of cases with respect to recall rates and precision, indicating its efficacy in classification tasks. The advantages of the suggested technique are particularly evident when compared to rule-based systems, as demonstrated by the C4.5 process's lower production of leaf nodes. These results highlight how important it is to combine different methods to improve classification efficiency and accuracy. Moreover, the technique's enhanced performance in comparison to well-known algorithms like Euonymus, EF1l, EW9, and EF_Unique7 highlights its potential for real-world uses. The statistical tests validate the importance of these results, indicating the robustness and dependability of the suggested approach in successfully handling classification problems.

In this method, KNN and NB methods were used to calculate KDR operation data, and C4.5 was used to analyze KDR operation data. In a single aspect, this method is 6 times better than the EF algorithm, 8 times better than the EW method, and 6 times better than the Euonymus method. In general, the total number of times was 20, and the proportion was 60.61%. As a rule-based method, C4.5 will produce incomplete rules when the real rules are generated, which makes some problems difficult to classify accurately when conducting experiments. Therefore, we also make a statistic for the average recall rate of each data set. Comparison results show that the effect of this method is better than that of the control method in 26 cases, reaching 78.79%.On this basis, the average precision and average recall of the proposed algorithm have reached a good level. The number of leaf nodes produced by the C4.5 process is illustrated. As can be seen in Table 3 and Figure 1, after adopting the method of the invention, the number of leaf nodes produced is less than the others.

Table 3: Schedule of dynamic Singular value Decomposition theory

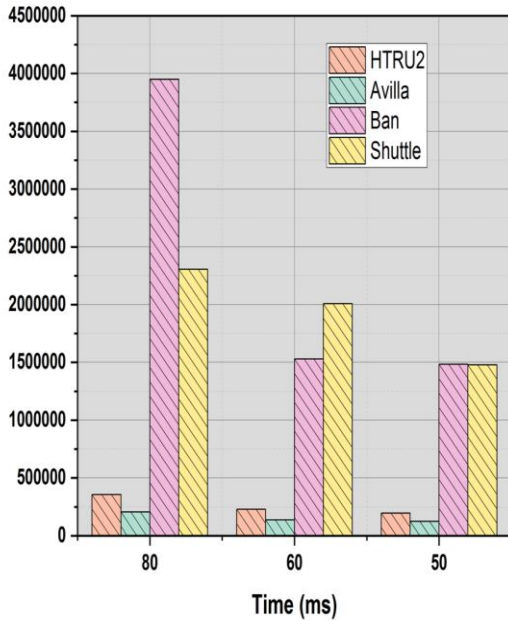| Time(ms)	80	60	50
HTRU2	354743	228578	195755
A villa	205466	135768	124841
B an	3950123	1528104	1481980
Shuttle	2306469	2007489	1477111

Figure 1: Scheduling performance of dynamic SVD theory

The performance of several algorithms, including KNN, KDR, PLS, and SVD-KDR, over various time intervals is shown in the Table 4 and Figure 2. KNN reaches 0.1 accuracy at 5%-time allocation, while KDR reaches the highest at 0.8. With 0.7 accuracy, KDR holds its lead when time rises to 10%. Fascinatingly, PLS and SVD-KDR both exhibit robustness and efficiency in time-sensitive tasks, consistently obtaining 0.9 accuracy or above across all time intervals.

Table 4: Outcomes of interpolation method

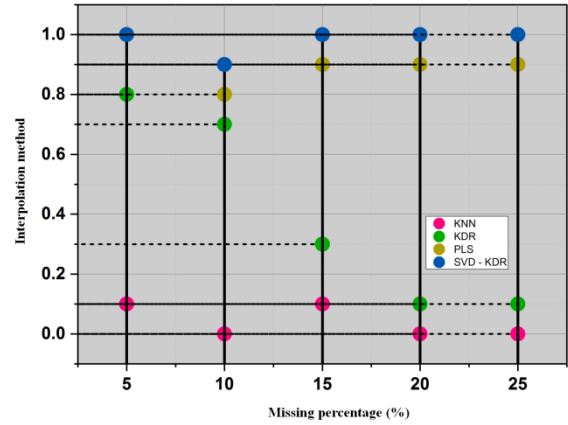| Missing percentage	Interpolation method			
	KNN	KDR	PLS	SVD-KDR
5	0.1	0.8	0.9	1.0
10	0	0.7	0.8	0.9
15	0.1	0.3	0.9	1.0
20	0.0	0.1	0.9	1.0
25	0	0.1	0.9	1.0

Figure 2: Comparison of interpolation method

Three algorithms, KNN, NB, and C45 are used in the experiments. The above studies show that the algorithm is effective for unsupervised data processing. In the NB and C4.5 methods, better results are obtained.

In the industrial big data environment, based on the characteristics of industrial data, this paper proposed a decentralized problem method based on industrial big data. In this paper, the definition of KDR computation and the research status at home and abroad are described in detail. This paper focuses on the basic principle of bidirectional selection and related theories at home and abroad. At present, most of the KDR calculation methods are only used to deal with static data and do not pay attention to the dynamic change of data. Therefore, this paper will further discuss the kinetic characteristics of KDR and Industrial big data analysis from the perspective of kinetics. There are three main aspects of this study:

4.1 A dynamic model of KDR operation is established by using a three-branch decision

In the case of Industrial imputation data, due to the dynamic nature of the original data, the relationship between the original data and the new time data cannot be guaranteed when KDR operation is carried out. When there is a big difference between the new and old-time data, it will cause an unsatisfactory separation. However, most of the current KDR computational theories ignore the dynamic properties of the data. To solve this problem, a dynamic data analysis method based on KDR operation is established by taking into account the singular value decomposition theory. This model can be divided into two categories: static and dynamic. Firstly, the static KDR operation method is used to perform the initial KDR operational processing at each

time point. On this basis, the interval method is used to fuse the KDR operation information at each time point. In consideration of singular value decomposition (SVD) theory, the delay decision method considering SVD theory is introduced, and the original interval is replaced by the form of a three-branch interval. In dynamic fusion, only the core region is fused, the existing blank region is discarded, and the delay method is used to segment the blank region. Lastly, the accuracy of the suggested analysis algorithm is confirmed using the UCI test results. The technique has a promising future for KDR computational processing of dynamic data, according to the experimental results.

4.2 Three-branch interval method taking into account singular value decomposition theory

According to the kinetic characteristics of Industrial imputation data, the three-branch interval analysis method is given. To solve the problem that the boundary of the KDR operation interval is uncertain in dynamic data, this paper uses the method of considering singular value decomposition (SVD) to redefine it, so that it cannot only express the boundary of the region but also update the edge of the dynamic region in real time. After describing the space interval, the space interval is used to distribute the space appropriately, to achieve dynamic adjustment of the KDR operation interval for massive information, and to solve the attribute KDR operation problem of incremental big data. The gap between the three root intervals can not only show the change in the region but also dynamically adjust the region according to the size of the space.

5   Discussion

The time measurements in milliseconds (ms) for the various tasks completed by HTRU2, A villa, B an, and Shuttle are shown in this table. A separate entity is represented by each row, and a different time condition 80, 60, and 50 milliseconds is represented by each column. For example, HTRU2 took 354,743 ms to finish its assignment under the 80 ms condition, while A villa took 205,466 ms, B and took 3,950,123 ms, and Shuttle took 2,306,469 ms. under a similar vein, HTRU2 took 228,578 ms, A villa took 135,768 ms, B and took 1,528,104 ms, and Shuttle took 2,007,489 ms under the 60 ms condition. In the end, HTRU2 took 195,755 ms, A villa took 124,841 ms, B and took 1,481,980 ms, and Shuttle took 1,477,111 ms under the 50 ms constraint. Singular value decomposition (SVD) theory is incorporated into our study's dynamic KDR

operating model for industrial big data interpolation, which takes into account the incremental nature of data accumulation. Notably, our strategy deviates from static techniques that are common in the literature and frequently ignore the dynamic change of data over time. We guarantee the consistency of KDR calculation results under dynamic situations by combining SVD theory with a rough set approach based on Manilkara. By contrasting our findings with previous research, we find that most algorithms are not built to take into consideration the incremental nature of industrial data because they are intended for static KDR processes. By overcoming this gap and offering a thorough foundation for dynamic interpolation, our methodology makes a novel contribution. Moreover, our methodology enables a more precise depiction of changing data trends, augmenting the resilience and dependability of industrial big data examination. We show the unique qualities and benefits of our suggested model through this comparative analysis, opening the door for more developments in the area.

5.1 Limitations

There are drawbacks to the suggested dynamic KDR operation model that integrates the rough set method with SVD theory. Its computational complexity might make it difficult to use with very large datasets. Additionally, problems with data quality or outliers may make it less effective to combine old and new data. Dependence on theoretical frameworks such as SVD may restrict the application to a variety of commercial datasets, necessitating rigorous cross-domain validation.

6   Conclusion

Although there has been research on the non-monitoring nature of interpolation in industrial big data, these problems caused by interpolation have not been completely solved. In the processing of industrial big data, there is often a lack of data. For such incomplete dynamic KDR operational methods, there is still a lack of relevant theories and methods at home and abroad. The processing of missing data often leads to the loss of data and thus affects the analysis of data. Therefore, considering the theory of singular value analysis, it is a very promising solution to explore the KDR operation. The study of industrial big data interpolation is still beset with numerous issues. According to the statistics of the industrial data, it is found that the data is extremely uneven. The root cause of the problem is that

the consistency between the pre-KDR and the interval is guaranteed by using a wide range of cell merging. Therefore, the correctness of the algorithm can be tested by using UCI data. Experiments show that the algorithm can not only solve the non-monitored static data effectively but also separate the non-monitored dynamic data effectively, which provides a useful reference for the subsequent KDR computation of dynamic data.

References

[1] Esmaeilbeigi M, Chatrabgoun O, Hosseinian-Far A, et al. 2020. A low cost and highly accurate technique for big data spatial-temporal interpolation [J]. *Applied Numerical Mathematics*, 153.

[2] Zhu Q X, Liu D P, Xu Y, et al, 2021. Novel space projection interpolation-based virtual sample generation for solving the small data problem in developing soft sensor [J]. *Chemometrics and Intelligent Laboratory Systems,* 217: 104425-.

[3] Luthra H, Nihith T, Pravallika V, et al, 2021. New Paradigm in Healthcare Industry Using Big Data Analytics [J]. *IOP Conference Series: Materials Science and Engineering*, 1099(1): 012054 (14pp).

[4] Yu F, Zhou Y, 2021. Development Planning and Path Analysis of Intelligent Logistics Industry in Big Data Age [J]. *Journal of Physics: Conference Series,* 1852(4): 042064 (8pp).

[5] Zhao L, Tao W, Wang G, et al, 2021. Intelligent anti-corrosion expert system based on big data analysis [J]. *Anti-Corrosion Methods and Materials*, ahead-of-print(ahead-of-print).

[6] Udugama I A, Gargalo C L, Yamashita Y, et al, 2020. The Role of Big Data in Industrial (Bio)Chemical Process Operations [J]. *Industrial & Engineering Chemistry Research*.

[7] Cheng C, Huang H, 2021. Big data and industrial innovation progress in Jiangxi Province incremental effect highlights enabling digital economy cultivation [J]. *Journal of Physics Conference Series,* 1852(2): 022005.

[8] Ram J, Zhang Z, 2021. Examining the needs to adopt big data analytics in B2B organizations: development of propositions and model of needs [J]. *Journal of Business & Industrial Marketing*, ahead-of-print(ahead-of-print).

[9] Yang K, 2020.The construction of sports culture industry growth forecast model based on big data [J]. *Personal and Ubiquitous Computing*, 24(1): 5-17.

[10] Chi J, Li Y, Huang J, et al, 2020. A secure and efficient data sharing scheme based on blockchain in industrial Internet of Things [J]. *Journal of Network and Computer Applications*, 167: 102710.

[11] Wei, HT, YY, et al, 2015. A k-d tree-based algorithm to parallelize Kriging interpolation of big spatial data [J]. *GISCI REMOTE SENS*, 2018,52(1)(-): 40-57.

[12] Wu W, Ahmad M O, Samadi S, 2019. Discriminant analysis based on modified generalised singular value decomposition and its numerical error analysis [J]. *Iet Computer Vision,* 3(3): 159-173.

[13] Luo L, Wang L, Hu J, 2019. On the Modeling and Analysis of an Improved CNC Interpolation Algorithm [J]. *Materials Science Forum*, 626-627: 459-464.

[14] Tao R, Kang X, Wen S, et al, 2017. Study of Dynamometer Cards Identification Based on Root-Mean-Square Error Algorithm [J]. *International Journal of Pattern Recognition & Artificial Intelligence*, 32(2).

[15] Gao Y, 2019. Constructing the social network prediction model based on data mining and link prediction analysis [J*]. Library Hi Tech*, ahead-of-print(ahead-of-print).

[16] Szczepanik M, Jozwiak I, 2019. Data management for fingerprint recognition algorithm based on characteristic points' group [J]. *Foundations of Computing & Decision Sciences*, 38(2): 123-130.

[17] Guo Y, Zhang B, Sun Y, et al, 2020. Machine learning based feature selection and knowledge reasoning for CBR system under big data [J]. *Pattern Recognition*, 112(6): 107805.

[18] Khayyam, H., Jamali, A., Bab-Hadiashar, A, Esch, T., Ramakrishna, S., Jalili, M. and Naebe, M, 2020. A novel hybrid machine learning algorithm for limited and big data modeling with application in industry 4.0. *IEEE Access*, 8, 111381-111393.

[19] Rani, R., Khurana, M., Kumar, A. and Kumar, N, 2022. Big data dimensionality reduction techniques in IoT: Review, applications and open research challenges. *Cluster Computing*, 25(6), pp.4027-4049.

[20] Hosseini, B. and Hammer, B. September 16–20, 2019, Interpretable discriminative dimensionality reduction and feature selection on the manifold. In Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2019, Würzburg, Germany, Proceedings, 310-326. Springer International Publishing.

[21] Ngiam, K.Y. and Khor, W, 2019. Big data and machine learning algorithms for health-care delivery. *The Lancet Oncology*, 20(5), pp.e262-e273.

[22] Bello-Orgaz, G., Jung, J.J. and Camacho, D, 2016. Social big data: Recent achievements and new challenges. *Information Fusion*, 28, pp.45-59.

[23] Islam, M.R., Liu, S., Wang, X. and Xu, G, , 2020. Deep learning for misinformation detection on online social networks: a survey and new perspectives. *Social Network Analysis and Mining* 10(1), p.82.