July - August 2020
ISSN: 0193-4120 Page No. 5423 - 5432

Data Analysis and Evaluation Based on RFID Considering K-Means Algorithm of GPGPU

Juan Li[1,*]
[1]Jinling Institute of Technology, Nanjing, China, 211169

Article Info
Volume 83
Page Number: 5423 - 5432
Publication Issue:
July - August 2020

Article History
Article Received: *25 April 2020*
Revised: *29 May 2020*
Accepted: *20 June 2020*
Publication: *28 August 2020*

Abstract

As for the missing data in the material data collected by radio frequency identification (RFID), this phenomenon greatly reduces the correctness of the results used in RFID. In view of this phenomenon of missing reading data, the solution is to take the original RFID data reading as a unit, and expand the window to smooth according to the previous reading data of the tag itself. In this behavior, a lot of redundant data that has nothing to do with the query is parsed and perspective. In addition, it involves multiple RFID applications. Therefore, the correctness of analysis and perspective is lower. To solve the above problems, the first is to transfer the RFID data from the data layer to the RFID data layer, and take it as the processing granularity. Three kinds of computing data and evaluation analysis of GPU k-means algorithm are launched. By taking the predicted GPGPU sequence correlation, the future events are analyzed, evaluated and judged.

Keywords: *RFID Technology, Data Analysis and Evaluation, GPGPU, Data of Missed Reading, Redundant Data;*

1. Introduction

Radio frequency identification (RFID) technology is a non-contact automatic identification and data acquisition technology. The basic working principle of RFID technology is as the following[1-2]: The reader broadcasts energy to its surroundings, and the tag that senses the energy sends the data it carries back to the reader immediately. The reader decodes the received data and then transmits the data to the host for processing. RFID technology was first used to track and detect objects in the radar monitoring system. However, with the development of wireless communication technology and data management technology[3-4], RFID technology has been extensively used in supply chain management, traffic monitoring, intelligent exhibition halls, hospitals, and other fields. The broad application prospects of RFID technology have attracted more and more attention[5-6].

In RFID technology, wireless radio frequency signals are used for data communication. Since wireless radio frequency signals are highly susceptible to environmental influences and interference with each other, especially when the number of tags and readers is increasing[7-8], signal interference is enhanced, which leads to high unreliability of the RFID data. The unreliability of RFID data is mainly reflected in the phenomenon of missed readings of data. When the missed reading phenomenon is severe, the accuracy of the query results is decreased sharply, which has significantly hindered the extensive promotion of RFID technology[9-10]. Hence, in this paper, solving the problem of missed readings of data based on RFID is selected as the main research content. The path and time are combined to analyze and evaluate the data that are useful for the query. Different from the processing of the missed readings of data in the existing RFID and sensor data cleansing work, we put forward a brand new data analysis and evaluation model based on RFID in this paper. The existing data cleansing strategies are carried out based on RFID readings as the granularity. However, for most applications, they are not concerned about whether the underlying readings are lost. The raw data required for the query are abstracted RFID data information. Hence, we put forward a strategy of analysis and evaluation based on the logic GPGPU as the granularity in this paper to avoid analyzing and evaluating redundant data information. In the existing RFID data cleansing model, mainly the window smoothing method based on historical readings and the spatial and temporal association strategy are taken into consideration, which is not suitable for application scenarios based on path information with multiple RFID data involved. In particular, when all RFID data readings are missed, window smoothing will not be able to patch the RFID data. In this paper, a new Means algorithm architecture is adopted. At the same time, the important analysis and evaluation factors such as the path, the regional missed reading rate, and the residence time are taken into account. In the complex application with multiple RFID data involved, the accuracy of analysis and evaluation is higher than that of the existing strategies.

2. Problem Description

In normal cases, RFID data are expressed in the form of the triplet $o\left(T_{epc}, R_{epc}, t\right)$. In order to unify the model, we refer to it as a reading event in this paper, which is equivalent to a basic event. Among them, T_{epc} and R_{epc} stand for the EPC code of the tag and reader, respectively, and they are the unique identifiers; t stands for the time stamp, which represents the moment when the reader detects the tag. Further, the RFID data are extracted to the RFID data layer in this paper and labeled as $o\left(T_{epc}, L, t_{start}, t_{end}\right)$, which is referred to as GPGPU. Among them, L stands for the RFID data detected by the tag, t_{start} and t_{end} stand for the start time and end time of the tag when L is detected, respectively. On the basis of the data abstraction, three model definitions related to event analysis and evaluation algorithms are described in detail in this section, that is, the architecture of the Means algorithm, the similar Means algorithm model and the evaluation model.

2.1. Means algorithm

Definition 1 (Means algorithm). A tag is processed through one or more RFID data in turn, which is referred to as the Means algorithm and is denoted as E. It is expressed as $l_\alpha L_\beta l_\gamma$, in which, $l_\alpha \in L_{start}$, $L_\beta \in 2^{L_{all}}$, $l_\gamma \in L_{end}$, L_{start} stands for the set of all starting RFID data, L_{all} stands for the set of all the end RFID data, L_{all} stands for the set of all RFID data, including L_{start} and L_{end} , and the symbol 2L stands for the set of the power of L_{all} . The elements in the set are referred to as GPGPUs. For example, the number of GPGPUs included in the Means algorithm E is referred to as the length of the Means algorithm, which is denoted as Len(E), and the Means algorithms with the length of n are collectively denoted as S_n .

Definition 2 (occurrence rate). Given a certain period of time T, the occurrence rate of a Means algorithm E within T is defined as the ratio of the total number of occurrences of this Means algorithm to the total number of occurrences of all Means algorithms, which is denoted as PO(T, E), as shown in the equation (1) below:

$$P_O\left(T, E\right) = Count\left(T, E\right) / Count\left(T, U\right) \tag{1}$$

In the above equation, Count (T, E) stands for the sum of the number of occurrences of Means algorithm E in the time period T, and U stands for all Means algorithms. Hence, it can be easily obtained that within a certain period of time T, the sum of the incidence rates of all Means algorithms is 1, as shown in the equation (2) below:

$$\sum_{E_i \in U} P_O\left(T, E_i\right) = 1 \tag{2}$$

Definition 3 (Missed reading rate of data based

July - August 2020
ISSN: 0193-4120 Page No. 5423 - 5432

on RFID). The missed reading rate of certain data L based on RFID refers to the ratio of the sum of the number of tags that have passed through the RFID data but have not been detected to the sum of the number of all tags that have passed the RFID data within a given time T, which is denoted as $P_{ML}(T,L)$.

In addition, RFID streaming data continues to arrive and are constantly changing. In order to ensure that the Means algorithm can predict the subsequent events of arrival accurately, we need to update the Means algorithm and maintain it in real time. In this paper, the sliding window technology is used to carry out incremental and decremented maintenance of the Means algorithm. The sliding window model always maintains a fixed-length data set in the latest time period. As the window is sliding, new data are continuously added, and expired data are deleted at the same time. In order to facilitate the incremental and decremented maintenance of the data set, the sliding window can be further broken down into several hop windows, that is, the hop window is the basic unit of the sliding window, and it is also the basic unit of the sliding range of each window.

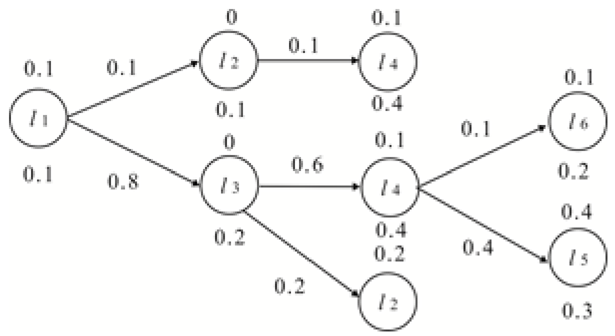

Figure 1. Architecture of the Means algorithm.

When the sliding window technology is used to maintain the Means algorithm, the size of the sliding window and the sliding amplitude of each window has to be determined according to the specific application so as to ensure that the Means algorithm obtained through statistical learning can predict the subsequent events that have arrived as accurately as possible. In this paper, the appropriate size of the sliding window and the window sliding amplitude will be obtained through the experiments under the given application conditions.

Definition 5 (Similar Means algorithm). Given two Means algorithms E_1 and E_2, they are denoted as $E_1 = l_{i_1}l_{i_2}\cdots l_{i_m}, E_2 = l_{j_1}l_{j_2}\cdots l_{j_n}$, in which l_{i_k}, l_{j_k} stand for a certain RFID data. If Means algorithms E_1 and E_2 meet the following two conditions at the same time: (1) m = n; (2) $l_{i_k} = l_{j_k}, k \in \{1,2,\cdots,m\}$, they are referred to as the same Means algorithm. If the Means algorithm E_1 only requires that k RFID data should be analyzed and evaluated, it becomes the same Means algorithm as E_2, then E_2 is referred to as the k-similar Means algorithm to E_1, which is denoted as $E_1 \xrightarrow{k} E_2$. Hence, the same Means algorithm can also be referred to as the 0-similar Means algorithm. The symbol $S_i(E,K)$ stands for the k-similar Means algorithm of the Means algorithm E.

Definition 6 (Most similar Means algorithm). Given the observation value $o(t) = l_{i_1}l_{i_2}\cdots l_{i_m}$ at the time t, all its similar Means algorithms are collectively expressed as $S_{i\_all} = S_i(o(t),k), k \in 0,1,2,\cdots n$, and its most similar Means algorithm is defined as shown in the equation (3) below:

$$s_{i\_mostly}(o(t)) = \left\{ r'(t) \middle| P(r'(t)|o(t)) = \max\left(P\left(r(t) = l_{j_1}l_{j_2}\cdots l_{j_n} \middle| o(t) = l_{i_1}l_{i_2}\cdots l_{i_m}\right), r(t) \in S_{i\_all}\right)\right\} \tag{3}$$

Definition 7 (Common sub-event and the longest common sub-event). Given two Means algorithms E_1 and E_2, if $\exists E\left(E \xrightarrow{k_1} E_1 \wedge E \xrightarrow{k_2} E_2\right)$ is true (k_1, k_2 are natural numbers), then E is referred to as the common sub-event E_1 of E_2 and are denoted as $E = C(E_1, E_2)$. If $\exists E'\left(E' \xrightarrow{k_1'} E_1 \wedge E' \xrightarrow{k_2'} E_2\right)$ is also true (k_1' is a natural number), then E is referred to the longest common sub-event of E_1 and E_2, which is denoted as $E = C'(E_1, E_2)$.

2.2. Evaluation model

July - August 2020
ISSN: 0193-4120 Page No. 5423 - 5432

Definition 7 (Reduction rate). Within a given time period T, the reduction rate is defined as the ratio of the volume of data reduced by the raw data after the data are abstracted to the volume of data not abstracted, which is denoted as $P_D(T)$, and the calculation expression is as the following

$$P_D(T) = \left(Account(T, R_a) - Account(T, R_c)\right) / Account(T, R_a) \tag{4}$$

In the above equation, $Account(T, R_a)$ and $Account(T, R_c)$ stand for the volume of source data that arrive within the time T and the volume of data that remain after the abstraction.

Definition 8 (Accuracy rate). Given two data sets, that is, the real data set R_e and the cleansed data set R_c, within a certain period of time T, the accuracy rate $P_A(T)$ is defined, as shown in the equation (5) below:

$$P_A(T) = R_e(T) \cap R_c(T) / R_e(T) \tag{5}$$

3. Data Analysis and Evaluation Model

In this paper, a data analysis and evaluation model framework based on RFID is put forward, as shown in Figure 2. The data analysis and evaluation model based on RFID is mainly composed of data abstraction, detection of missed readings, data analysis and evaluation mechanism. In the data abstraction mechanism, the RFID data are remodeled, that is, the triplet data $o(T_{epc}, R_{epc}, t)$ is modeled as $o(T_{epc}, L, t_{start}, t_{end})$. The event table is used to store all the real Means algorithms that may occur, and it is predictable. If a Means algorithm that occurs does not exist in the event table, it means that a missed reading phenomenon of GPGPU occurred when the Means algorithm was detected. The missed reading detection mechanism is to monitor the ongoing Means algorithm in real time combined with the event table. If the detected Means algorithm shows any missed reading phenomenon, it is transmitted to the data analysis and evaluation mechanism. Otherwise, the event is directly output for the query layer. It should be emphasized that the data analysis and evaluation mechanism is to analyze and evaluate the GPGPU where a missed reading has occurred. It is mainly composed of three parts, that is, the Means algorithm architecture, the matching engine, and the analysis and evaluation strategy mechanism. The Means algorithm architecture is established based on the statistical learning of the historical event sets, and it is continuously updated and maintained according to a set period. The matching engine searches for the similar Means algorithm of the current event based on the specific analysis and evaluation algorithm. Subsequently, it is transmitted to the analysis and evaluation strategy mechanism. Based on the Means algorithm architecture and the similar Means algorithm, the missed reading events at present are analyzed and evaluated based on the analysis and evaluation strategy mechanism.

In the data abstraction strategy put forward in this paper, a certain tolerance for the missed reading of data is considered while the data are abstracted, that is, setting a threshold t_smooth, which indicates the size of the time smoothing window. When the interval of the two subsequent data time is less than t_smooth time units, the data between them are smoothed. Figure 3 describes the data abstraction process through examples. Through the extraction of data, the GPGPUs that make sense for queries are analyzed and evaluated to avoid analysis and evaluation of redundant data.

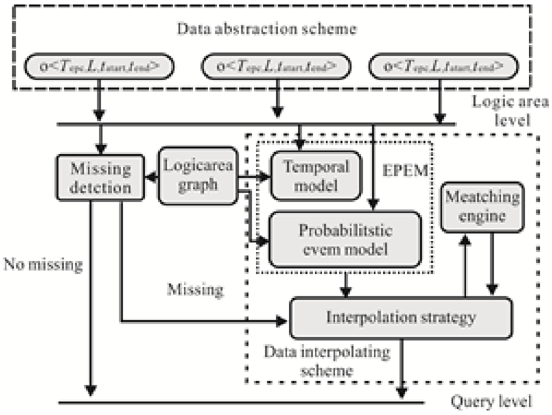

Figure 2. Framework of the data analysis and

Published by: The Mattingley Publishing Co., Inc.

July - August 2020
ISSN: 0193-4120 Page No. 5423 - 5432

evaluation model.

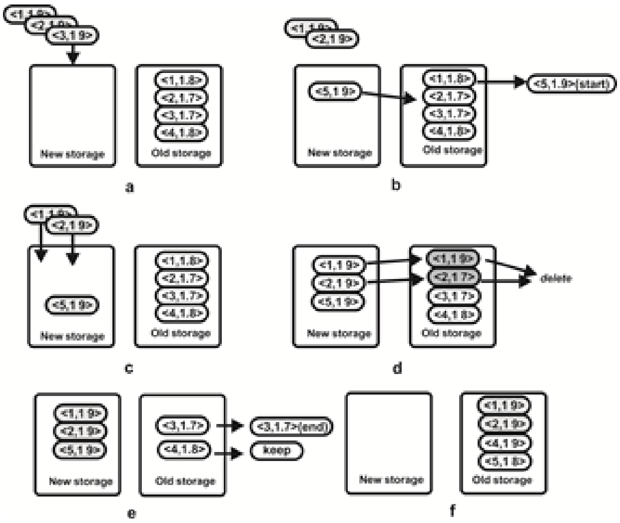

Figure 3. Data abstraction process.

4. Data Analysis and Evaluation Based on RFID

In the RFID application, when the data are acquired, a relatively important attribute is obtained at the same time, that is, the time stamp. In normal cases, when an object with a tag passes through each RFID data, there is a certain rule not only when different RFID data are selected, there is also a certain regularity in the time interval that different RFID data are retained. For example, in the park system mentioned in the previous section, the time of stay of the tourists in the small theater is probably the performance time of the drama. In places with a lot of fun, such as playgrounds, most tourists will stay longer, while in some places with a few landscapes, most tourists will not stay too long. Hence, the interval of the residence time can be added to the architecture of the Means algorithm. When the missed reading of a certain RFID data occurs in the tag, not only is the correlation of the RFID data sequence taken into consideration, but the consideration of the missed reading interval of the tag is increased, and the missed readings of data are analyzed and evaluated from these two aspects.

Minkowski distance is a generalization of Euclidean distance, also known as the distance L_p. Given the time series $X=\{x_1,x_2,\cdots,x_n\}$ and $Y=\{y_1,y_2,\cdots,y_n\}$, the distance L_p distance is defined, as shown in the equation (6) below:

$$L_p(X,Y)=\left[\sum_{i=1}^{n}|x_i-y_i|^p\right]^{\frac{1}{p}},\ p\geq 1 \tag{6}$$

In the above equation, when p = 2, it is Euclidean distance.

When the residence time distribution of the tags in the RFID data is relatively concentrated, the time is estimated by using the histogram method, and the accuracy rate will decrease accordingly. At this point, if the European distance is used to perform evaluation, a better result will be obtained, and this method will be adopted by the K-Means algorithm, which is described as the following.

The K-Means algorithm is also put forward on the basis of the extended Means algorithm. However, the algorithm is established based on an accurate time model. While the analysis and evaluation algorithm are improved based on Theorem 4, the impacts on the two aspects of event sequence and label residence time can also be considered in parallel when the data are analyzed and evaluated at the same time. In addition, the weight of these two impacts can be adjusted through a parameter according to this concept. In this paper, an extended definition of the most similar event is provided, as shown in equation (7) below:

$$S''_{i\_mostly}(o(t))=\alpha\cdot S_{i\_mostly}(o(t))+(1-\alpha)\cdot\beta'(o(t),r(t)) \tag{7}$$

In the above equation, the definition of $S_{i\_mostly}(o(t))$ is given in the equation (3), the definition of β'(o(t), r(t)) is given in the equation (8), α is an adjustment factor, and the specific value can be obtained through the experiment.

$$\beta'(o(t),r(t))=\sum_{l\in o\wedge l\notin r}g\left(L_2\left(t_l,D(t)\right)\right) \tag{8}$$

In the above equation, g(x) is a monotonically

Published by: The Mattingley Publishing Co., Inc.

July - August 2020
ISSN: 0193-4120 Page No. 5423 - 5432

decreasing function, and D(t) is the t_{end}-t_{start} set of each GPGPU maintained by the precise time model at the time t.

The K-Means algorithm is proposed on the basis of equation (9).When the missed reading Means algorithm is analyzed and evaluated, it not only considers the two factors of the occurrence rate of the Means algorithm and the missed reading rate of RFID data but also considers the factor of the tag retention time in the missed reading area. In addition, their respective influence on the missed reading area the tag belongs to is adjusted accordingly. In different application cases, the value taken for α will be different. As a result, three improved data analysis and evaluation algorithms based on β+ can be obtained, that is, the β+greedy algorithm, the β+minimum k-similarity algorithm, and the β+full similarity algorithm. The three improved algorithms need to solve the most similar Means algorithm to the missed reading Means algorithm and make modifications to the basic strategy accordingly. It can be seen that a key step of the K-Means algorithm is to solve β'(o(t) and r(t)). In the algorithm 1, the Euclidean distance algorithm of r(t)) is described for solving the β'(o(t).

5. Experimental evaluation

5.1. Experimental settings

In general, due to the limitation of the site and application, the existing RFID data query and cleansing literature use simulation data to carry out experiments. In order to ensure that the simulation effect is as close to the real situation as possible, we adopt the famous NetLogo in this paper, which is extensively used in the simulation of sensor equipment. The simulation system is configured based on the features of the RFID device. To consider the complexity of the application and the diversity of data, we use NetLogo to simulate a play scene in a park in this paper: It is assumed that the park has a total of 30 scenic spots (that is, RFID data), and each scenic spot is equipped with 2 readers (The reading rate of the readers is 5KBPS, and each reader is equipped with two antennas).

5.2. Testing and evaluation of the data analysis and evaluation algorithm

In order to illustrate the data analysis and evaluation strategy, we have tested the appropriate window size and sliding amplitude of Data1, Data2, and Data3 (the experimental figure is omitted). In the following experiment, for different simulation data, the window size and the sliding amplitude are set according to the most appropriate window size WS and the sliding amplitude SR measured in the experiment. Figure 8 shows the result measured by using the simulation data Data1, the degree of entry and exit of nodes in the RFID data graph = 3, and the missed reading rate PML of the RFID data is 0.1 ~ 0.7. From the figure, it can be seen that when PML is < 0.3, the four analysis and evaluation algorithms can achieve a very high accuracy rate, which is almost close to 100%. However, as the missed reading rate of the RFID data continues to increase, the accuracy rates of the three analysis and evaluation algorithms proposed in this paper are significantly higher than those of the PLA algorithm. The reason is that in the RA, MKA and AKA algorithms, the data are analyzed and evaluated based on the RFID data layer, that is, when a tag passes through an RFID data and is not detected even once, it can be analyzed and evaluated by these three algorithms according to the architecture of the Means algorithm. However, the PLA algorithm is based on the data layer. When the tag is not detected at all in certain RFID data, the tag cannot be evaluated by the PLA algorithm. Hence, when the missed reading rate of the RFID data increases, the accuracy rate of the PLA algorithm is decreased the fastest. On the other hand, among the three algorithms proposed in this paper, the AKA algorithm has the highest accuracy rate. When PML is < 0.7, it can ensure that the accuracy rate of the data after analysis and evaluation is as high as more than 90%. The accuracy rate of the MKA algorithm is slightly lower, and the accuracy rate of the RA algorithm is the lowest, which is consistent with the results of theoretical analysis. At the same time, as

Published by: The Mattingley Publishing Co., Inc.

July - August 2020
ISSN: 0193-4120 Page No. 5423 - 5432

the missed reading rate increases, all the accuracy rates of the RA, MKA, and AKA algorithms tend to decline. The reason is that all the three algorithms are put forward on the basis of the Means algorithm architecture when the missed read rate is high, the unreliability of the RFID data increases, and the accuracy of the Means algorithm architecture based on statistical learning source data decreases, which leads to a decrease in the accuracy of the algorithm.

Figure 5 is measured by using the simulated data Data3, θ = 2, PML value is 0.1 ~ 0.8. Compared with the test results of Data1 in Figure 8, the accuracy rates of the four algorithms are the same. However, as the missed reading rate increases, the accuracy rate of the PLA is decreased at a faster rate. On the other hand, the accuracy rates of the other three algorithms are decreased slowly, with very high accuracy rates. The reason is that 30 RFID data are introduced into Data3, and the length of the Means algorithm is more than 20; while only 10 RFID data are introduced into Data1, and the length of the Means algorithm is less than 10. As the missed reading rate increases, the Means algorithm with a relatively large length of Data3 is more likely to have the case when the tag is completely missed by a certain GPGPU, which will lead to a decrease in PLA accuracy. When there are many RFID data, and the length of the Means algorithm is long, the randomness of the RFID data selected by the tag becomes large, so that the length of the longest common sub-event between the two events will not be very long. From Theorem 3, it can be known that in this case, the accuracy of the minimum k-similarity algorithm and the full similarity algorithm will be very high. Hence, it can be seen that, in the context of larger-scale applications, the analysis and evaluation algorithm put forward in this paper is superior to PLA in terms of accuracy.

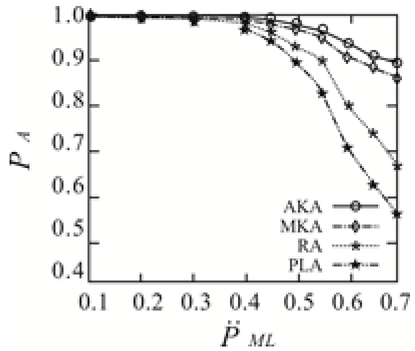

Figure 4. Comparison of the accuracy-1.

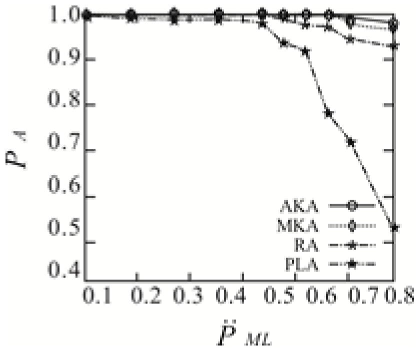

Figure 5. Comparison of the accuracy-2.

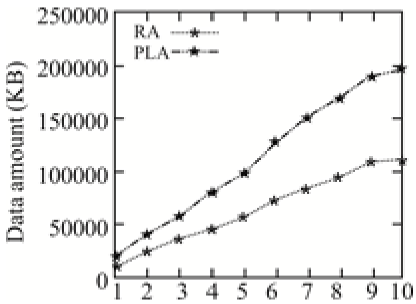

Figure 6. Comparison of the data volume-2.

Subsequently, the algorithm PLA and algorithm RA are compared in the aspect of data redundancy, and the results are shown in Figure 6. The figure is measured by using the simulated data Data1, and the number of tags is 300. It can be seen that the volume of RFID data analyzed and evaluated based on the RA algorithm is about 60% of the data volume analyzed and evaluated based on the PLA algorithm, where the data volume has been reduced by about 40%. The reason is that in the RA algorithm, analysis and evaluation are carried out on the basis of remodeling the RFID data, that is, there is a data abstraction process to delete the redundant data and ensure that the data redundancy after the RA algorithm analysis and evaluation is very small. On the other hand, the PLA algorithm only analyzes and

Published by: The Mattingley Publishing Co., Inc.

TEST Engineering & Management

July - August 2020
ISSN: 0193-4120 Page No. 5423 - 5432

evaluates the missed readings of data and fails to process the redundant data. Hence, the data after the analysis and evaluation based on the PLA algorithm show a high degree of data redundancy. In the data after the analysis and evaluation based on the algorithms MKA and AKA, the data volume and the RA curve are approximately the same as well. Hence, it can be seen that the three algorithms put forward in this paper are superior to the PLA algorithm in the aspect of processing redundant data.

As the PLA algorithm performs data analysis and evaluation based on the data layer, it does not involve the concept of events, which is not on the same level as the real-time performance put forward in this paper, the real-time performance of only the three algorithms RA, MKA and AKA are compared in the following section. The comparison result is shown in Figure 7, which is measured by using the simulated data Data1. The horizontal coordinates stand for the system running time, and the vertical coordinates stand for the cumulative delay time from the start time to the current time. From the figure, it can be seen that the greedy algorithm has the shortest delay time and the highest real time performance. Within a fixed time period, the delay time is almost 0. The second is the minimum k-similarity algorithm, which indicates that sometimes its curve is smooth, and sometimes the rising amplitude is relatively large. It suggests that the delay time of the algorithm fluctuates significantly in each unit time period. The reason is that it is related to the specific event of arrival. When the same event of arrival is identified in the set of the correct events, it is not necessary to traverse other events, and it is considered that no missed reading has occurred in the event. At this point, the delay time is the shortest. When all events are traversed, the event that is most similar to the event of arrival is identified, and the delay time is the longest. For the full-similar algorithm, for any event of arrival, it is necessary to traverse the correct set of events to identify the most similar event. Hence, the delay time of the full-similar algorithm is the largest, and the real time performance is the poorest. The curve rises almost linearly, which is consistent with the theoretical analysis results.

All the above experiments are measured under the condition that the equation (7) is applied. In this case, it can be seen that the performance of the MKA algorithm is very excellent. The following changes in the application conditions make it fail to meet the equation (7). Under this condition, the accuracy of the algorithm MKA is evaluated. The experimental results are shown in Figure 8 as the following. The figure is measured by using Data3, and the value taken for PML is 0.1 ~ 0.8. It can be seen that when the application conditions do not meet the equation (7), the accuracy rates of the algorithms PLA, RA and AKA are almost unaffected, and only the accuracy of the MKA algorithm is declined significantly, which is even slightly lower than RA. The reason is that when the application conditions do not meet the equation (7), the missed reading Means algorithm is more inclined to read the missed reading GPGPUs more frequently. At this point, if the analysis and evaluation are carried out according to the principle of least analysis and evaluation, it will increase the probability of error analysis and evaluation. Hence, the accuracy rate of the MKA algorithm is decreased. Hence, it can be seen that equation (7) is a prerequisite for the application of the algorithm MKA. When the application conditions do not meet the equation, the RA and AKA algorithms can be selected to analyze and evaluate the data.

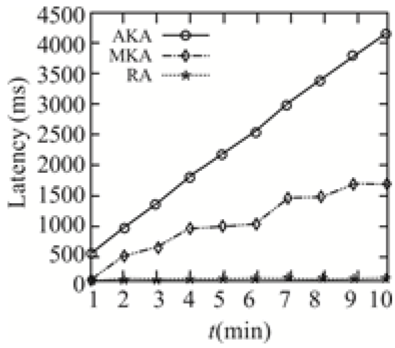

Figure 7. Comparison of the real time performance.

Published by: The Mattingley Publishing Co., Inc.

July - August 2020
ISSN: 0193-4120 Page No. 5423 - 5432

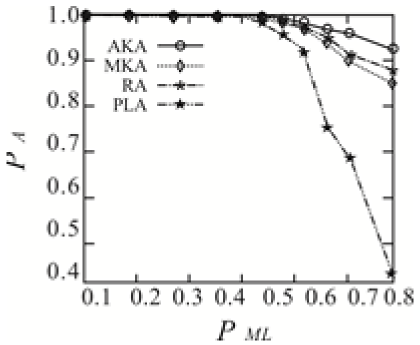

Figure 8. Comparison of the accuracy -3.

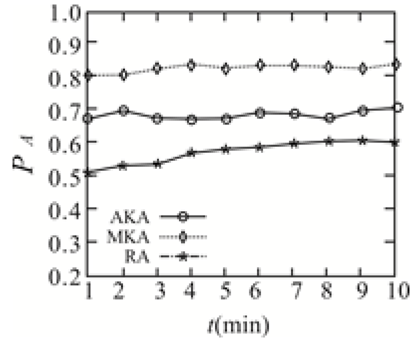

Figure 9. Influence of the missed reading rate factors.

Figure 9 shows the analysis of the influence of the missed reading rate of RFID data on the accuracy of the three analysis and evaluation algorithms put forward in this paper. It is measured without considering the missed reading rate of RFID data, that is, when missed reading of GPGPU occurs, the data are analyzed and evaluated directly based on the incidence rate of the Means algorithm. The experimental data used include Data1, PML = 0.6, and the other parameters are the same as those in Figure 9. Compared with the three algorithms when PML = 0.6 in Figure 9, all the accuracy rates of the corresponding algorithms in Figure 8 have shown a slight decrease, in which the accuracy rate of the AKA algorithm shows the fastest decline. Hence, it can be seen that the missed read rate is a crucial factor affecting the accuracy of an algorithm, especially the AKA algorithm. The reason is that when the missed read rate is not taken into account, the full-similar algorithm will follow all the Means algorithms that have arrived in real time based on the Means algorithm with the highest occurrence rate to carry out analysis and evaluation.

6. Conclusions

On the one hand, RFID technology is extensively used in more and more fields due to its automatic and fast features. On the other hand, the data collected by RFID readers have high unreliability and redundancy, which has restricted the application occasions of RFID and presented new challenges for data management. To address these problems, we carried out in-depth research on RFID data analysis and evaluation algorithms in this paper and put forward an analysis and evaluation framework based on the dynamic Means algorithm model with the logical GPGPU as the analysis and evaluation granularity for the first time. Based on the triplet model of the RFID data, a data abstraction algorithm is proposed to abstract RFID data from the data layer to the RFID data layer. Subsequently, for the primary type of data unreliability in the applications of RFID – missed readings of data, three data analysis and evaluation algorithms are put forward on the basis of data abstraction, that is, the greedy algorithm, the minimum k-similarity algorithm, and the full similarity algorithm, to balance the real time performance, accuracy, and maintenance costs. Finally, the consideration of the time factor is added to improve the proposed analysis and evaluation algorithm, which has improved the accuracy of the analysis and evaluation results. Under different application conditions, the two improved algorithms have their respective strengths. A large number of experiments have verified that the data abstraction and various data analysis and evaluation strategies proposed in this paper have different performance strengths in different situations. For the application scenarios that involve multiple RFID data and have obvious path information, the framework put forward in this paper has more superior performance in the simplicity and accuracy of analysis and evaluation to the existing data analysis and evaluation strategies.

Acknowledgemens

Ministry of Education's first co-education project in 2019（201901051025）, Computer Basic Education

Published by: The Mattingley Publishing Co., Inc.

Research Project of the National Institute of Higher Learning（2019-AFCEC-044）.

References

[1] Mezzah I, Chemali H, Kermia O. Emulation-based fault analysis on RFID tags for robustness and security evaluation[J]. Microelectronics Reliability, 2016, 69:págs. 115-125.

[2] Fischer, J. R., Neebling, T. E., & Quist, M. C. (2012). Development and evaluation of a boat-mounted rfid antenna for monitoring freshwater mussels. Freshwater Science, 31(1), 148-153.

[3] Yang, J. (2010). Chip impedance evaluation method for uhf rfid transponder ics over absorbed input power. Etri Journal, 32(6), 969-971.

[4] Nayyar, S, Gavrilova, M, Greenough, A, & Ghali, W. (2010). Exploration of spatial-temporal patterns in healthcare implementation of rfid indoor navigation system. Current Problems in Surgery, 6(2), 3-52.

[5] Harnik, P. G., Fitzgerald, P. C., Payne, J. L., & Carlson, S. J. (2010). Terebratulide data for extinction selectivity analysesdata from: phylogenetic signal in extinction selectivity in devonian terebratulide brachiopods. current problems in surgery, 6(2), 3-52.

[6] Papapostolou, A., & Chaouchi, H. (2012). Integrating rfid and wlan for indoor positioning and ip movement detection. Wireless Networks, 18(7), p.861-879.

[7] Luo, H., Wen, G., Su, J., & Huang, Z.. (2018). Slap: succinct and lightweight authentication protocol for low-cost rfid system. Wireless Networks, 24(1), 69-78.

[8] Zhou, W., & Piramuthu, S.. (2014). Consumer preference and service quality management with rfid. Annals of Operations Research, 216(may), 35-51.

[9] Oztekin, A., Mahdavi, F., Erande, K., Kong, Z. J., Swim, L. K., & Bukkapatnam, S. T. S. . (2010). Criticality index analysis based optimal rfid reader placement models for asset tracking. International Journal of Production Research, 48(9), 2679-2698.

[10] Zhi, Z., Zhonghai, L. U., Qiang, C., & Xiaolang, Y. (2012). Design and optimization of a cdma-based multi-reader passive uhf rfid system for dense scenarios. Ieice Transactions on Communications, E95-B(1), 206-216.