Research on Technical and Tactical Action Mining of Basketball Based on Apriori Algorithm

Fang Liu

Chongqing Jiaotong University sports Chongqing 400074, China

Keywords: Apriori algorithm; Basketball skills and tactics; Data mining

Abstract: Nowadays, the general application and development of computer technology made the data mining technology play a crucial role in sports technology and tactics. This paper was based on the Apriori algorithm to explore the data mining of basketball skills and tactics. In this paper, the rules which could cause our own interest were deeply excavated by introducing the lift-measure interest independent method. And considering the mutex characteristic points included in the data mining, we optimized the classical Apriori algorithm. It effectively improved the efficiency of the Apriori algorithm for mining frequent itemsets. The optimized AD-apriori algorithm can reduce the complexity of the mining process in time and space.

1. Introduction

Data mining for sports training is to simulate and reproduce physical education teachers or coaches' teaching experience, training methods, managers' organization plan and training process by computer modeling technology. Therefore, we can explain, analyze, predict, organize and evaluate the experimental technical science of sports training (Hong-Xing W U et al.2015) [1]. As far as physical training is concerned, it can be used in the development of sports training simulator, human body information detection, visualization, tactical and executive choreography, optimization of movement technology and so on (Zhang X N et al.2016) [2]. It is necessary to realize that the research of physical training simulation is only in the initial stage of development. This technology has appeared soon, and there is not enough in-depth and systematic research in both theory and practice (Niu H L ET al.2016) [3]. Competitive sports have gradually turned to high speed and rapid development, which leads to sports training must rely on modern scientific and technological means (Zhang Y et al.2017) [4]. Based on the purpose of digging deep and thoroughly the potential of human beings it is necessary to systematically adopt the subject knowledge related to sports science. And it takes a systematic and scientific way to explore the inherent laws of sports. Physical training simulation belongs to an experimental technical science.

2. State of The Art

In recent years, the computer data mining auxiliary training in China has made a substantial breakthrough. The Sports Science Research Institute of the State Sports General Administration and the Computer Research Institute of Chinese Academy of Sciences jointly face the advantages of our country, diving, weightlifting and quasi advantage gymnastics. They are effective in developing "digital 3D human motion data mining system" in addition to the existence of intellectual property and computer virtual data mining technology. It has been effectively applied in the training process of the Athens Olympic Games. In addition to ensuring the absolute superiority of our athletes in the diving project, it provided the first generation of trampoline athlete Huang Shanshan for the first time in the Olympic Games to get the bronze medal results provided an effective help(Cheng X et al.2015)[5]. Zhuo Xianlin, the Chinese trampoline head coach said the bronze's secret weapon was the "digital trampoline training aid system". The project was jointly organized and funded by the Ministry of science and technology, the State Sports Administration, the Chinese Academy of Sciences and the Beijing Municipal Science and Technology Commission. Project undertaking unit: Institute of computing technology. The overall goal is to study the computer data mining technology of the digitized 3D human motion of competitive sports in China in the 2008 (Beijing) Olympic Games. They also take trampoline and diving as an example to develop a set of computer aided sports training system based on 3D human body motion data mining technology. The twenty-third world trampoline Championships (Olympic qualifiers) held in Hannover, Germany. After the fig officials watched the demo, they expressed the hope that the purchase of the software for trampoline training and international referee training. They give high attention to it.

3. Methodology

3.1 The Introduction of Lift-Measure's Interest Measurement

The data of the item set in the traditional association rules is static and does not change. With the introduction of the "state transition" model, the "state" changes of the association rule precursor and the post can be made (Dutta A ET al.2016) [6]. The following examples are used to explore how to introduce the concept of state transfer. The beginning is the addition of the lift-measure concept. The relational data set *D* is as table 1. It is assumed that the value of different attributes in the attribute *F* can cause a difference in the attribute *R* result. We can think of the attribute *F* as "cause" and the attribute *R* as a "result". That is the difference between the reasons and the results.

Table 1. Relational data sets

Id	F	R
1	a	x
2	a	x
3	b	x
4	c	x
5	a	y
6	b	y
7	c	y

We can understand that there are three classes of values in the property *F*, each of which is *a*, *b*, and *c* (only a certain value of three in a certain time attribute *F*). The property *R* has two types of cases, each of which is *x*, *y* (a certain moment can be only *x* or *y*). At the beginning, the rules associated with the *x* case of the attribute *R* are excavated, and Rule 1-3 is obtained. When a specified threshold is reached, rule 1 shows a strong association between the attribute value *a* of the attribute *F* and the attribute value *x* of the attribute *R*. If we want the attribute value of this database attribute *R* to be *x*, it can be reduced as much as possible. What is the process that can be taken? Because there are only three factors that can find the value of the attribute R is x. That is the *a*, *b*, and *c* of the attribute *F*. To reduce the incidence of *x*, it is necessary to connect the reasons from *a* to *b* or from *a* to *c*.

By comparing the above rules *1-3*, we can get the attribute *F* that is *b* to the attribute *R* is *x's* lifting degree is *1*; the attribute *F* is the *c* to the attribute *R* is the *x's* lifting degree is *2/3*. To understand the properties of the *F* is compared with *c* attribute *F* is *b* on the properties of *R* is relatively small increase degree of *x*, which is the emergence of *x* appears to improve the influence of the lowest. The ability to transform attribute *F* values from *a* to *c* can reduce the attribute *R* to the proportion of *x* in the overall database. Formalization shows this idea:

$$\text{F}\,(a \rightarrow c) \Rightarrow R(x \rightarrow \sim x) \qquad (1)$$

Rule r1:

$$[\sup = 2/8, lift\_measure = \frac{4}{3} - \frac{2}{3} = 2/3, conf = (\frac{4}{3} - \frac{2}{3}) / \frac{4}{3} = 1/2] \qquad (2)$$

Rule *r1* reflects that the attribute *F* is *a* and the attribute *R* is *x*, and the amount of data in the total amount is *2/8*. The *F* property is *a* compared with the ascent of the attribute *F* is *c* to the attribute *R* is *x* beyond 2/3. The proportion of the part beyond the attribute *F* is the *1/2* of the attribute *R* as the *x* lifting degree. The $\mathrm{F}(a \rightarrow c)$ meaning in rule *r1* is that "the value of the attribute *F* is transformed from *a* to *c*". With the above examples, we can learn that a change concept is added to the premise of association rules. We define the validity of the rules by the elevation of the height. The sup, Hft-measure, and conf in the rules bring the exact advice.

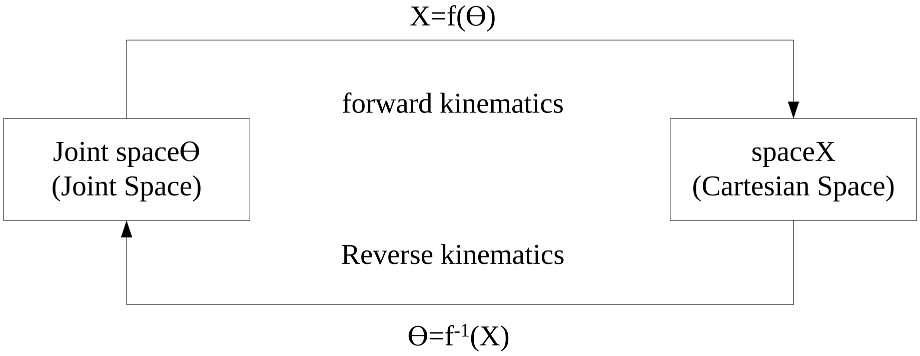


Figure 1. Forward kinematics and inverse kinematics

We define the lift-measure interest measurement. As far as relational databases are concerned, each attribute has a probability of having multiple attribute values. When a relational database is matched to a transaction database, each of the values of an attribute can be matched as an item. For a certain attribute, at a specific time, it can not take the value of two or two above the attribute values. Like the "gender" property, the value of this attribute is only male or female in a record. We define the various cases of attribute values in attributes as mutual exclusion. In the transaction database, the mutex is transformed into the same attribute. The size of the mutually exclusive set of items belongs to the attribute value of this attribute. Each item in the same mutex is the mutual exclusion of each other. The size of the mutex set belongs to the number of attribute values of this property. The set of relational data sets D, the user's clear reason attribute set and the result attribute set are respectively:

$$\mathrm{F} = \{F_1, F_2, \ldots, F_m\} \text{ and } R = \{R_1, R_2, \ldots, R_n\}, \quad F \cap R = \phi \qquad (3)$$

With the transformation of a relational database into a transaction database, a set of items from a variety of factors in a cause attribute concentration is recorded as a F. A set of items is set up from a result attribute to set up each of the results of a variety of results, and the following definition is made. One is to have a frequent itemset *L '*, which, the size of the frequent itemsets is |*L*|; $L' \subseteq F$ is present for all the item F_{i}' in the frequent itemsets. There is a F_{q}' in the result attribute set, and there is a $F_q' \in R$; The association rule $r': F_1' \wedge F_2' \wedge \ldots \wedge F_{\mathrm{L}'}' \Rightarrow \mathrm{F}_q'$ belongs to the strong association rule, and the support of rule R 'is *sup (r')*, and the degree of lifting is *lift（r'）*. The two is that the item $\mathrm{F}_i(a_i \rightarrow a_j)$ is expressed $\mathrm{F}_{\mathrm{i}}^{\mathrm{m}}$, and the item F_{q}' is set. The result attribute in the matching relational database is the case of R_k and the attribute value is *d*, and the item $\mathrm{R}_{\mathrm{k}}(d \rightarrow \sim d)$ is expressed as F_q^m, then $r^m : F_1^m \wedge F_2^m \wedge \ldots \wedge F_{\mathrm{L}'}^m \Rightarrow \mathrm{F}_{\mathrm{q}}^{\mathrm{m}}$. F_{i}'. It belongs to the state before the change, $\mathrm{F}_{\mathrm{i}}''$ belongs to the state after the change. In this case, the support degree *sup (r)*, lifting quantity *lift-measure (r)* and confidence *conf (r)* are as follows:

$$\sup(r) = \sup(r') \qquad (4)$$

$$lift\_measure(r) = lift(r') - lift(r'') \qquad (5)$$

$$con(r) = \frac{lift(r') - lift(r'')}{lift(r')} = \frac{lift\_measure(r)}{lift(r')} \qquad (6)$$

The degree of support in the rule represents the scope of influence, which shows the usefulness of the rules; the amount of promotion shows the extent of the change. After the upgrade of the higher representative will be the value of the F_i property from a_i to a_j, the R_k property value is *d* in reducing the more. Confidence is the certainty that shows the change. Then it is the nature of *lift_measure*. For association rules, the degree of lifting belongs to a class of relative measures with low complexity. The formula for calculating the lifting degree between *A* and *B* is:

$$lift(A,B) = \frac{P(A \cup B)}{P(A)P(B)} \qquad (7)$$

By virtue of the expression of lift in the association rules, the correlation metric formula based on the lifting quantity can be expressed as:

$$\mathrm{corr}_{(A1 \to A2)?(B \to \sim B)} = lift(A1,B) - lift(A2,B) \qquad (8)$$

It should be noted that *A1* in the form represents the set of initial itemsets. *A2* represents the set of item sets mutually exclusive by *A1*, and *B* represents the set of item sets of results. In this connection, we reduce the situation of *B*.

3.2 The Improvement of the Frequent Itemset Mining Algorithm Apriori

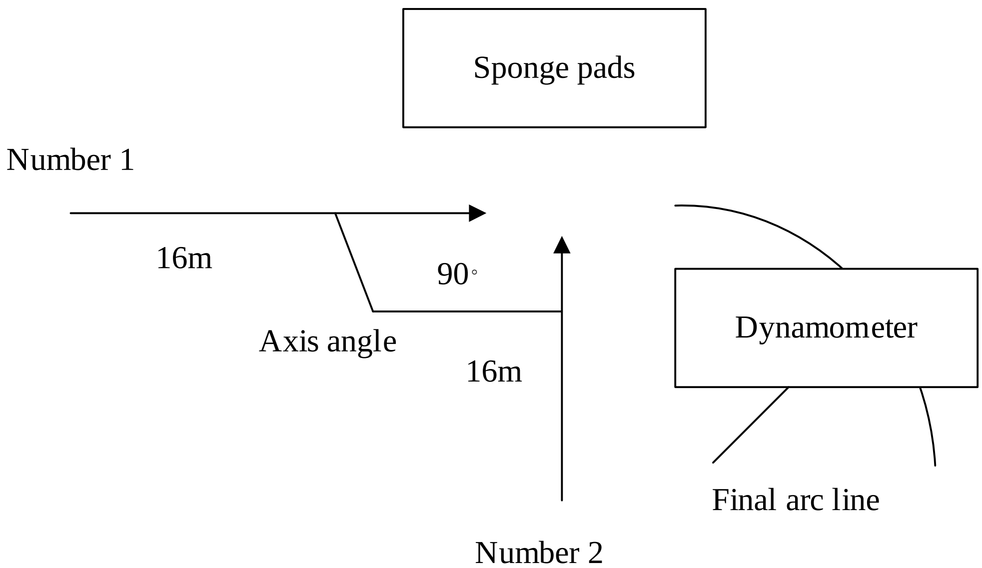
When we carry out frequent item set mining for students' physical health data, we can choose Apriori algorithm or other itemsets mining algorithm. The characteristics of the student's system health data are analyzed. If you have to implement frequent itemsets mining, this part of the algorithm has some limitations. It covers the validity of the frequent itemsets, the problem of the item set can not be combined, and so on. First, for the classical Apriori algorithm, the item in the initial item set covers only one item, which belongs to a single item set(Alekhya M et al.2015)[7]. For this algorithm, there are two items in the initial set of terms. By virtue of the relative attributes of the comparison database, a mutex Necklace table map-exctran is built. Each item of the mutex necklaces is constructed by using the different values of the related attributes. It also takes a map-assist that is applied to record the information of the first *k-1* item and the *k* items for the scanned items. The *KEY* of this map belongs to the first *k-1* item set of a particular item, and *VALUE* belongs to the *LIST* built by all types of item *k* that have been scanned and obtained.

In addition, there are differences in defining whether the two *k* item sets can be combined to get a *k+1* item set and the traditional approach. As for the classic Apriori algorithm, the way to define whether two *k* item sets can be combined into a *k+1* item set is: The first *k-1* item of the two item sets are consistent. The first set of item *k* in *ID* after a set of section *k* of the *ID* to be smaller. The criterion of the algorithm is that the first *k-1* item of the two item sets are consistent. The feature attributes of item *k* of the first set of items and the *k* items of the second item sets are not consistent. The first set of item *k* attributes is compared with second sets of *k* attributes of the *ED* are smaller. As a result of third conditions, second conditions must be reached. Therefore, the first two conditions are adopted in the algorithm to implement the specific decision. For the linked list matching the *k-1* item set before map-assist storage, we only need to merge the scanned items in the current itemset and the *k* items in the matching *k-1* itemset list into *k+1* itemsets. Furthermore, it prevents the occurrence of recurrent traversal to achieve the goal of reducing time and space complexity. We have optimized the Apriori algorithm and named it to AD-apriori, which is limited by the length of space, and we don't discuss the code. The improved algorithm input belongs to the transaction data obtained by the relational database to the transaction data and the minimum support technology min_sup_cnt. min_sup_cnt = min_sup*reord_cnt，It is the minimum expenses and the product of the total number of records.

Compared with the classic Apriori, the advantage of the AD-apriori algorithm is that the time complexity of the AD-apriori algorithm is more advantageous. It is assumed that a total of *m* items can be involved in the mining of an item set. Then the spatial complexity of the computed *k* set is reduced from the previous c_m^k to $c_m^{k-1}/n*m$, and the average time complexity also decreases to a certain extent. It must be emphasized that in the AD-apriori algorithm, the number of candidate sets is reduced by an order of magnitude, and it does not lose the information needed. The *k+1* item sets built using the *k* item set are relatively more than they were originally reduced. A collection of mutex that does not have the value of existence is reduced. And it reduces the user's less focused set, the efficiency can be significantly increased.

4. Result Analysis and Discussion

In this paper, the athletes were tested by a three dimensional dynamometer and high speed camera to obtain the kinematic and dynamic parameters of the supporting stage of the swinging leg before the take-off. The image data of the stage of the swing leg support stage to the landing stage of the take-off is obtained. The instrument used for the experiment is a Swiss KISTLER three-dimensional dynamometer（60cm×90 cm×lO cm). There are two Fastec500 convenient high speed cameras (120 frame /s). Synchronizer. Auxiliary testing tools: Aijie 3D DLT radiation frame, body weight, tape. In the specific test, the two lenses of the high speed camera are placed separately on the left and left side of the site. The main axis is perpendicular to the plane of motion. The distance between the camera and the center is 14.5 meters, and the machine is 1.2 meters high. The shooting frequency is 120 frames per second. The angle between the main optical axis of the two cameras is about 90 degrees (see figure



2).

Figure 2. Scene map of high jump shooting

After getting the video of the photographed motion process, we applied the algorithm. The video in sports is selected to test the algorithm. The experimental video is a video for the training and competition of the National Road skaters. The video format is AVI. The video resolution is 720x576 pixels. The frame rate is 29 frames per second. The video sampling size is 24 bits. The track of the speed skating of the road can be approximated as a local linear. The parameters are set: the number of initial particles is 300, the size of the target area is 10x10 (a part of the body of the athlete), and the variance of the weight value is a=0.20.

Figure.3 Images captured by the video screen

1200 frames were successfully tracked in this experiment. The above image is the image captured in the tracking process (limited to the number of frames in the length of the column). The first frames are manually calibrated initial frames. The white point in the figure is the location of the tracking target generated by the algorithm (the first frame is manually calibrated to be tracked), and the blue point is the position of the particles in the current frame. There is not much deviation from the actual position from the tracking results. And the target can still be tracked after the occlusion of the site staff continuously, and the same is true of the 848th and 992nd frames. It can be seen from the screenshot that the size of the athletes in different frames has changed greatly. Light and color also change to a certain extent. However, the HSV color histogram is adopted as the feature and the updating mechanism of the observation model. So it can also keep track of the target for a long time. And the robustness of the algorithm for continuous occlusion tracking is also good, and the parameters of the specific experimental results are as follows:

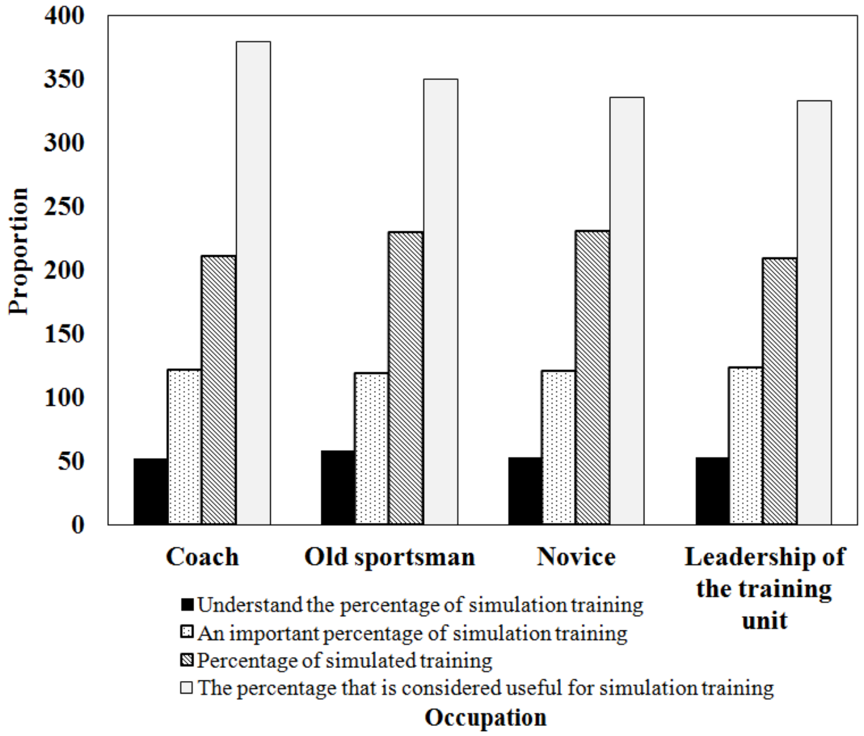


Figure 4. Questionnaire on the Cognition of simulation training

The smaller the noise, the higher the accuracy of the Apriori estimates. The number of particles is 200, the length of the sequence is 100, and the initial value of the system is 0.1. The results of the experiment are shown in the following drawings.

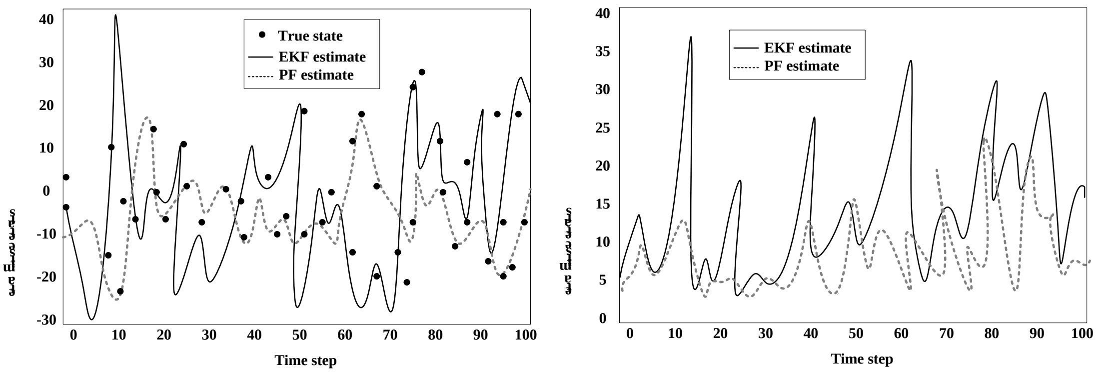


Figure 5. *a=0.3* State estimation and estimation Error

Figure 4 and figure 5 show the estimation and error of Apriori and EKF as the *a* becomes larger. It can be understood that in the case of relatively small *a*, the target technology is relatively strong. But as the *a* becomes larger, the nonlinearity weakens and the maneuverability of the target is weakened. As in Figure 3, the tracking performance of EKF is gradually approaching to Apriori. When *a=0.9*, the tracking performance of EKF is sometimes even better than that of Apriori. This experiment can be learned; Apriori in the nonlinear system of high intensity or relatively strong maneuvering conditions, it compared with the ordinary EKF performance is more excellent. The tracking performance of the two of the approximate linear systems is excellent and the difference is relatively small.

This part is based on the discussion of the design of Apriori and the key technology. We study the adaptive Apriori tracking system for the problem of target occlusion and the real time tracking loss and tracking in practical applications. Through the tracking experiment of linear moving target and nonlinear moving target in sports competition, it shows that the system has strong robustness to solve the problem of continuous occlusion and recovery after occlusion. Because of the adaptive mechanism of the number of particles, the tracking is quasi real time. In a word, the traditional sports teaching method is still our main training method, but it has some limitations. It is slow, costly and risky. Computer simulation can be better aided in many aspects. Especially, it is more practical for the action innovation with certain security risks. Computer technology can be used to simulate and simulate and assist in training. With the help of simulation system, athletes can receive more abstract coaches' oral training guidance from the past, and it can be changed into graphic simulation instruction and quantitative analysis of related movements displayed in graphic ways.

Conclusion

In the era of rapid development of computer technology, the information degree of basketball competition has been gradually improved. Whether the mass data accumulated in the basketball database hides more useful information or not, it needs to be solved with the help of data mining technology. This paper was based on the Apriori algorithm for the research of basketball skills and tactics mining. In this paper, the relevant concepts of the traditional association rules were analyzed in detail. This article was based on the support of the traditional association rules and the "mining model" of confidence. It optimized the association rules by referring to the "state transfer" pattern idea. And it relied on the introduction of lift-measure mining interest measure to us real interest rules. It also improved the shortcomings of the classical Aprirori algorithm, which was oriented to the mutual exclusiveness in the data of basketball technical and tactical data mining. The efficiency of the frequent itemset mining in the basketball technical and tactical data was low. We use this improved method, which belongs to the construction of exclusive Necklace table map-exctran and set mapping table map-assist. In this way, the efficiency of mining frequent itemsets by Apriori algorithm was improved. The adjusted AD-apriori algorithm can effectively reduce the time and the complexity of the mining process.

Reference

[1] Hong-Xing W U, Wang H. Research on Enterprise Web Log Mining Based on Improved Apriori Algorithm[J]. Computer Technology & Development, 2015, 1 (4): 88-91.

[2] Zhang X N, Na L I, Zhu F E, et al. Research and analysis of ship defect data using Apriori algorithm [J]. Ship Science & Technology, 2016, 8(4):99-102.

[3] Niu H L, Hui-Min L U, Liu Z J. The improvement and research of Apriori algorithm based on Spark[J]. Journal of Northeast Normal University, 2016, 9 (12): 99-100.

[4] Zhang Y, Hu C, Huang S, et al. Apriori Algorithm Based Data Mining and Analysis Method for Secondary Device Defect [J]. Dianli Xitong Zidonghua / automation of Electric Power Systems, 2017, 41(19):147-151.

[5] Cheng X, Su S, Xu S, et al. DP-Apriori: A differentially private frequent itemset mining algorithm based on transaction splitting [J]. Computers & Security, 2015, 50(1):74-90.

[6] Dutta A, Ghosh A. Improvement of Apriori Algorithm by Reducing Number of Database Scans and Generation of Candidate Keys[J]. International Journal of Advanced Research in Computer Science, 2016, 7(2):35-41.

[7] Alekhya M, Ram B R, Hanmanthu B. Apriori Algorithm base Model of Opinion Mining for Drug Review[J]. International Journal of Advanced Trends in Computer Science & Engineering, 2015, 9(10):55-59.