Application research on long jump training and guidance based on data mining

Tao Lu[1]*, Xiaojun Zhang[2]

[1]*Qingdao Vocational and Technichal College of Hotel Management, Qingdao, Shandong, postcode 266100*

[2]*Henan Institute of Education, Henan Zhengzhou 450000, China*

**Corresponding author's e-mail:22480538@qq.com*

Received 01 October 2013, www.cmnt.lv

Abstract

This paper adopts the data ware house technology and data mining technology to establish a sports training assistant decision support system for long jump athletes. To make organic integration of the training factors for athletes, it applies scientific training theory and advanced training methods to the sports training management. We focus on the improvement of two classic data mining algorithms: association rules of Apriori and decision tree classification ID3. For Apriori algorithm, we improve the connections and pruning strategy when creating (k+1)-order frequent item set by k-order frequent item set, and the process pattern of transactions. For the defects of ID3 algorithm, we propose to reduce the computation of attribute gain selection when establishing the tree, and provide corresponding scheme to set the attribute importance. Then the actual examples are used to apply the improved models to the sports training assistant decision support system. The results show our algorithm improve the mining efficiency actually. The generated strong association rules which have higher association can be imported into knowledge library as important base for the sports training schemes.

Keywords: converter training samples, apriori, ID3, data warehouse, data mining

1 Introduction

In recent 20 years, China has accumulated large quantity of data related to sports training in sports field. However, at present, the training data management of national sports training is basically in disorder state and it cannot make use of various data which has been accumulated for dozens of years in national sports development. The laws outside experience and patterns outside prediction cannot be discovered effectively.

Data mining is a rapid rising subject in recent years. It extracts hidden, unknown, potential and useful information and knowledge from massive, incomplete, noisy, fuzzy and random data [1, 2]. Many scholars studied related mining technology for massive data in sports training and the application field of data mining technology is extended widely. Zhang in literature [3] studies two kinds of mining analysis methods: principal component analysis and Bayesian K-nearest neighbor algorithm. The actual implementation cases of these two methods are also provided. Principal component analysis is mainly used in comprehensive results evaluation. It reduces the information quantity by reducing correlative relationship between analysis factors. These two methods are used together to offer direction and evaluation in working direction for athletes [4]. Wang puts forward the data mining method of decision tree in literature [5]. His algorithm combines with SLIQ algorithm to be used in athletes' achievement database so as to analyze various data. Meanwhile, by analysis and establishment of decision tree model, the decision basis is provided for users. Jiang [6] further studies the mining technology and proposed the scheme based on ID3 algorithm. He provides the mining analysis on related management data in educational administration department of university to discover the association in curriculum design, offering data reference for university the decision. Dalal put forward multi-strategy idea in his study [7]. He comprehensively uses previously various mining discovery technology and applies it in grade database, to offer better decision program for universities students. In addition, he uses statistical analysis to query and analyze the operations on each subject scores under different conditions. Meanwhile, related score analysis report and tables are generated for teachers to make teaching decisions. Tang introduces Apriori association rule algorithm and decision tree ID3 algorithm in literature [8] and he respectively points out the application range of these two algorithms.

This paper adopts training assistant decision support system of long jump athletes which is constructed by database technology and data mining technology. It organically integrates long-jump athletes' different aspects, applying scientific training theory and advanced training method to the management of athletes' sports training, which integrates users' input to produce a set of reasonable sports training project. The system is based on the data including sports scores of college athletes, body checklist, etc, in one famous university to establish multi-dimensional dataset. It analyzes and studies the association rule mining and classification algorithm of decision tree in data mining in detail. At first, it appropriately improves the classical Apriori algorithm of association rule and improves efficient data rate in reading database. Meanwhile, it reduces some unnecessary data scanning which can quickly generate frequent item set. These two points obviously improve the mining efficiency in massive data of algorithm. Next, decision tree ID3 algorithm is improved and corresponding mining model is established which reduces the computation complexity of selection attribute gain value in tree-building, to improve the operation speed. The problem that the root node tends to the attributes with more values is solved. Meanwhile, the pruning closed value is set to limit the complete growing of tree, and it helps to produce decision tree of rational structure. This paper is organized as follows. In section 2 we introduce the decision support system for training and guidance, and discuss the function of mining methods adopted in this paper. In section 3 we mainly improve two data mining algorithm: apriori and ID3 decision tree, followed by the experimental evaluations in section 4. The conclusions are given in section 5.

2 Training guidance decision support system based on data mining

This system adopts intelligent decision model and it adds data mining and training program generation modules. The system structure and system functions are shown as Figure 1. The HCI system provides a convenient and friendly interaction environment for the decision maker. According to the suggestion of interface, professional athelets' physical quality indicator data is gradually input and submit to server for processing. Then, it provides the running state of system for users so they can fully understand the operation situation, running results and inference results. According to users' demand, corresponding training projects are output including individual training suggestion and grouping training project. The system has function to correct input errors and determination of the input data.

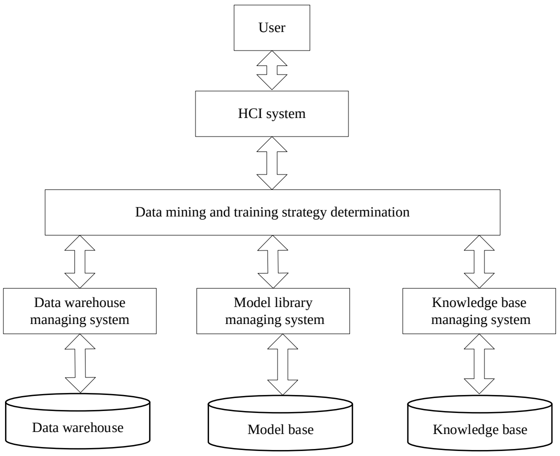

User
HCI system
Data mining and training strategy determination
Data warehouse managing system
Model library managing system
Knowledge base managing system
Data warehouse
Model base
Knowledge base

FIGURE 1 Functional architecture of the system

Data mining and problem solving system contains two modules: training project making module and data mining module for athletes. According to models from athletes system in model base, single table in data warehouse or the established multi-dimension dataset are performed data mining of athletes so as to obtain knowledge principle. Data mining needs some special algorithms such as association rule, clustering algorithm, decision tree, etc. The data mining results can be further used to enrich knowledge base and model base as new knowledge and model. Newly generated knowledge principle can be taken as decision support for sports teaching reform. Training project decision module is the function to process the data which is input to the form by user, in the module of servers. During the data processing, JavaBean is based on users' input and antecedent match of knowledge rule in knowledge base to obtain training items of long-jump athletes, so it provides individual training item and plan. With users' input continuance, all users' training projects will be grouped. Besides, when final input completes, all generated results will be summed up to produce a set of collective training projects to return to users.

The following steps describe system data mining methods in detail.

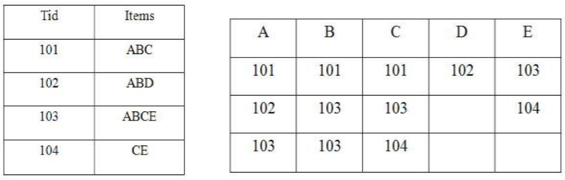
Association rule mining. In this system, data mining is performed on data warehouse by Apriori algorithm. The algorithm is realized by using association rule mining model in model base to investigate association among each factor that influencing athletes' physical quality. That is, it finds the factors which determine athletes' physical qualities and their weight together. At the same time, it will also produce new influencing factors related to different sports items and take it as new rules to be added to the knowledge base.

According to Decision tree ID3 algorithm this system integrates improved ID3 algorithm to classify examples ranking from root node to leaf node. The leaf is the classification belonging to the example. Each node in the tree indicates the test of one attribute on case. Then, each subsequent branch corresponds to one possible value of this attribute. According to training samples, one decision tree is finally produced in order to produce classification rule, which is used to group the athletes.

3 **Improvement of** mining algorit**hms**

3.1 IMPROVED APRIORI ALGORITHM

The process on database in Apriori is looked as horizontal structure, as is depicted in Figure 2(a). This paper adopts the idea of literature [9] and uses new data structure. It invents the corresponding vertical structure of project transactions as shown in Figure 2(b).



Tid	Items
101	ABC
102	ABD
103	ABCE
104	CE

(a) Original transaction database

A	B	C	D	E
101	101	101	102	103
102	103	103		104
103	103	104		

(b) New project transaction database

FIGURE 2 The transformation of data structure in this paper

Such transformation is easy for the computation of the amount of transactions supporting k-project: The transaction set of known support project A and B is T_{AB}, and the transaction set of support project C is T_C. Then the transaction supporting A, B and C is the intersection of two sets: $T_{ABC} = T_{AB} \cap T_C$. The detailed idea of our improved strategy is described as below:

1) Scan the source database, record the transaction code supporting each item during the scanning process. Then count the number of transactions supporting each item. Delete the item whose support item is less than the minimum support transaction, to acquire the frequent 1-item set.

2) We get the candidate 2-item set by connection one to one of 1-item set. Then the transactions set supporting each item set can be acquired by computing the intersection of transactions of two items in candidate 2-item. Delete the item whose support item is less than the minimum support transaction, to acquire the frequent 2-item set.

3) In this way, we continue to delete the item set that can not be frequent k-item set. The candidate k-item set is acquired by connection on to one. Then compute the transaction intersection of each (k-1)-item set in support k-item. Let $x \in L_{k-1}$, $y \in L_{k-1}$, then $x \infty y \in C_k$. The transaction set supporting each k-item set is acquired. Delete the item whose support item is less than the minimum support transaction to acquire frequent k-item set.

4) Repeat the above procedure until there is no frequent item.

The improvement is that it first filters the items in L_{k-1} that can not create frequent k-item, before the candidate k-item is created. It saves more time consumption than Apriori in filtering and greatly reduces the time for frequent k-item creation. It also records that the support transaction of each frequent set when creating the frequent k-item set. If

$$x \in L_k \ , \ y \in L_k \ ,$$

$$x_1 = y_1 \wedge x_2 = y_2 \wedge ..x_{k-1} = y_{k-1} \wedge x_k < y_k \ ,$$

we need not rescan the database when determine whether $x \infty y$ is frequent (k+1)-item set, and we just need to find the intersection of T_x and T_y . The transactions in the intersection in T_x and T_y all support $x \infty y$. If the amount of transactions in the section is greater than or equal to the number of the minimum support transactions, it is frequent item set; otherwise it is not. So it avoids the overhead caused by pattern match in Apriori algorithm.

The improved Apriori algorithm is described as follows:

```
Input: Transaction database $D$ ; the minimum support threshold $\min\_\sup$ .
Output: frequent item set $L$ in $D$ .

$L_1$ =find_ $L_1$ ( $D$ , $\min\_\sup$ )’
For (k=2; $L_{k-1} \neq \not\subset$ ; k++) {
$L'_{k-1} = L_{k-1}$ ;
// count the times of each item emerging in $L_{k-1}$ and record the frequent item set supporting this item
For each item $x \in L_{k-1}$
  For each field $f \in x$
    { $f.count$ ++;
      $I_f[f.count] = x$ ;}
    // Delete the items that can not create frequent k-item in $L'_{k-1}$
    Delete ( $L'_{k-1}$ , $I_f \mid f.count] < k-1$ );
    For each item $x \in L'_{k-1}$
       For each item $y \in L'_{k-1}$
    // Find the frequent item that can be connected
       If ( $x_1 = y_1 \wedge x_2 = y_2 \wedge .. \wedge x_{k-2} = y_{k-2} \wedge x_{k-1} < y_{k-1}$ )
       { n=0;
    // The character of transaction sequence is convenient to find the amount of the same transactions in two sets
       For (i=1, j=1; $i \leq |T_x|$ , $j \leq |T_y|$ ;)
       If ( $T_x[i] < T_y[j]$ )
       i++;
       else if ( $T_x[i] > T_y[j]$ )
         j++;
     else
      {n++;
      $T_{x\infty y} = T_{x\infty y} \cup \{T_x[i]\}$
       i++;
       j++; }
    if ( $n \geq N \times \min\_\sup$ ) // find one frequent set
    { $T^l_{x\infty y} = T_{x\infty y}$ ;
     $L_k = L_k \cup \{x \infty y\}$
    }
    return L= $\bigcup_k L_k$ ;
    // Create 1-item frequent item set
       Procedure find_ $L_1$ ( $D$ , $\min\_\sup$ );
     $C_1 = get\_item(D)$    // get the items of transactions database
    For each transaction $t \in D$
    { N++;
     For each item $x \in t$
     { $x.count$ ++;        // sequence of transaction in database
       $T_x[x.count] = t_{id}$;    // The sets are arranged in order
     }
    }
     $L_1 = \{x \mid x \in C_1 \wedge x.count \geq N \times \min\_\sup\}$
  Return $L_1$
```

3.2 IMPROVEMENT OF DECISION TREE ID3 ALGORITHM

ID3 algorithm proposed by Quinlan [10] is the most influential decision generation algorithm. It adopts the dividing and conquering strategy to create a decision tree by the learning of a training set. It uses the information gain to test each feature attribute in the dataset. Then the nodes for decision tree establishment which have the biggest gain of feature attribute are selected. The different value of this feature of attribute is adopted to establish branches. This method is iterated by the example set of each branch, to establish the next class of node and branch of the decision tree, until the examples in one subset belong to the same class. But there exists some deficiency in ID3 algorithm [11]:

1) The computation of mutual information depends on the features with more values.

2) When the maximum gain attribute is adopted to establish the tree, the operation involved is too much and it affects the speed of establishment.

3) When the training set is enlarging, it may cause complete growing of the decision tree and slow the speed of establishment.

So we make the following improvement for the above defects:

(1) The attribute importance is introduced in this paper. We can make advanced assumption according to assistant knowledge. It will influence the attribute gain value and avoid that the algorithm abandons the data topple with small amount of data. If we set the i_{th} attribute of domain

knowledge has importance as S_i, $gain(\alpha) = H(X) - S_i H(X \mid \alpha)$;

(2) Simplify the value of $gain(\alpha)$ to reduce the operation of determination on gain, so the algorithm speed can be improved. For simplification, we assume there are two classed of training cases set, that is, positive and negative example. Let the amount of positive set be p and the amount of negative set be n. The attribute α has value { $\alpha_1, \alpha_2, ..., \alpha_t$ }. When $\alpha = \alpha_i$, the elements of its subordinate branches will be $p_i + n_i$. There are n_i positive examples and p_i negative examples. Then we get the degree of uncertainty on the separation by the decision tree:

$$H(X) = -n(n+p)\log_2(n/(n+p)) - p/(n+p)\log_2(p/(n+p)) \tag{1}$$

When choosing testing attribute α, we separate the entropy of classified information of node X_j at each $\alpha = \alpha_i$:

$$H(X \mid \alpha) = 2\sum_{j=1}^{t} p_j n_j / (p_j + n_j) \tag{2}$$

Set Y_j as the examples set of $\alpha = \alpha_i$, then the uncertainty degree is the conditional entropy of training examples set to attribute α:

$$H(Y_j) = -n_j(n_j + p_j)\log_2(n_j/(n_j + p_j)) + p_j/(n_j + p_j)\log_2(p_j/(n_j + p_j)) \tag{3}$$

Substitute above equations to the equation of $gain(\alpha)$, we get $gain(\alpha) = H(X) - S_i H(X \mid \alpha)$. Since the value $H(X)$ is fixed for each attribute, we only need to consider $S_i H(X \mid \alpha)$. By simplification we get

$$H(X \mid \alpha) = S_i/(n+p)\ln^2 \sum(-n_j/(p_j + n_j) \cdot \ln(n_j/(p_j + n_j)) - p_j/(p_j + n_j)\ln(p_j p_j/(p_j + n_j) \tag{4}$$

Because $1/(n+p)\ln^2$ is fixed, so

$$H_1(X \mid \alpha) = S_i \sum_{j-1}^{t}(-n_j(n_j + p_j)\ln(n_j/(n_j + p_j)) + p_j/(n_j + p_j)\ln(p_j/(n_j + p_j))) \tag{5}$$

When $x \to 0$, $\ln(1+x) \approx x$. Because $n_j(n_j + p_j)$ and $p_j/(n_j + p_j)$ is far less than 1, then

$$H_1(X \mid \alpha) = S_i 2\sum_{j=1}^{t} n_j p_j / (n_j + p_j) \tag{6}$$

The attribute with minimum value is taken as the root node. So the attribute selection only involves plus, multiplication and division, which are easy to be implemented on computer. Its operation speed is much faster than original algorithm.

(3) In the subset corresponding to each attribute during the branching process, the proportion of each class is compared to predetermined threshold. If the value is larger the growing of decision tree is terminated.

The improved algorithm procedures can be summarized as follows:

Step 1: Read the training samples;

Step 2: Input the pruning threshold;

Step 3: Establish the tree downward;

Step 4: Select the attribute according $gain(\alpha)$

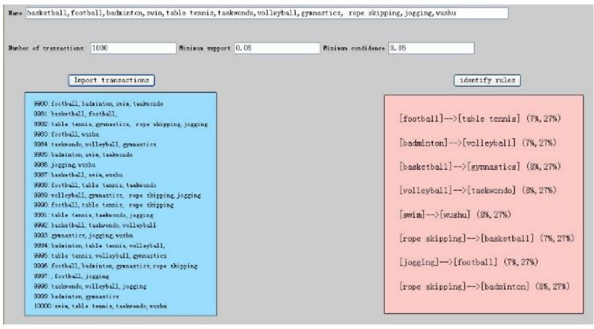
Step 5: Determine whether the attribute selection is over. If so then the process comes to an end; otherwise, compare the proportion of each class to predetermined threshold: if it is bigger, go to the end; else turn to step 3.

4 System realization and implementation of data mining model

We implement these two improved model in sports training decision support system of long-jump athletes. There are two aspects application of association rule mining. One of them is Apriori-based data mining model established in the model base. It is used to mine single transaction database and find the relationship among internal attributes in each table of database. The other one is applying data mining tool in SQLServer to establish the association rules model. It can find the attribute relationship among the tables and the factors affecting sports training. The improved ID3 algorithm is also used to generate classification rules to instruct the training plan.

4.1 IMPLEMENTATION OF APRIORI ALGORITHM MODEL

The mining is first used for the single table to find the relationship between each attribute in the table. The generated rule will be stored in knowledge base for decision support. The next one is applying this algorithm to mine table *avorate--sports* in literature [12]. The table lists the sort of athletes' favourite sports items, involving sports such as basketball, football, badminton, swimming, gymnastics, taekwondo, table-tennis, track and field, martial arts and volleyball. The athletes can fill the form according to their hobbies. However, at least, they are demand for filling in the first hobby. There are 1000 athletes' sports hobby items records which are excavated. The selected minimum support *min_support*=5% and the minimum confidence *min_conf*=30%. The operating results are shown as below:



Name basketball, football, badminton, swin, table tennis, taekwondo, volleyball, gymnastics, rope skipping, jogging, wushu

Number of transactions: 1000 Minimum support 0.05 Minimum condidence 0.05

```
9980: football, badminton, swim, taekwondo
9981: basketball, football,
9982: table tennis, gymnastics, rope skipping, jogging
9983: football, wushu
9984: taekwondo, volleyball, gymnastics
9985: badminton, swim, taekwondo
9986: jogging, wushu
9987: basketball, swim, wushu
9988: football, table tennis, taekwondo
9989: volleyball, gymnastics, rope skipping, jogging
9990: football, table tennis, rope skipping
9991: table tennis, taekwondo, jogging
9992: basketball, taekwondo, volleyball
9993: gymnastics, jogging, wushu
9994: badminton, table tennis, volleyball,
9995: table tennis, volleyball, gymnastics
9996: football, badminton, gymnastics, rope skipping
9997: , football, jogging
9998: taekwondo, volleyball, jogging
9999: badminton, gymnastics
10000: swim, table tennis, taekwondo, wushu
```

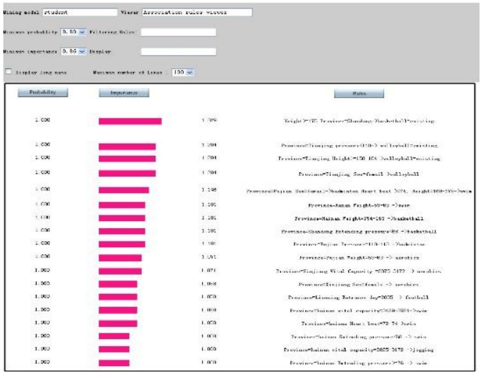
```
[football]-->[table tennis] (7%,27%)
[badminton]-->[volleyball] (7%,27%)
[basketball]-->[gymnastics] (8%,27%)
[volleyball]-->[taekwondo] (8%,27%)
[swim]-->[wushu] (8%,27%)
[rope skipping]-->[basketball] (7%,27%)
[jogging]-->[football] (7%,27%)
[rope skipping]-->[badminton] (8%,27%)
```

FIGURE 3 Mining results of the records

From the above result and the running result of classic algorithm we can see that the improved Apriori algorithm has advantage in rules generation efficiency. The effect is more obvious when the transaction database is larger.

The attribute column in student dimension is determined and some attribute columns of body status are taken as input. The sport-evaluation column is taken as the predictable column. The part of generating rule conclusion is athletes are appropriate for their joining sports. Left condition part is athletes' basic personal information or physically basic condition description. The generated rules can be taken as the next basis to formulate athletes' sports training plan and they are also used to instruct the sports teaching reform.

Previously determined minimum support and minimum confidence are input in mining model viewer to generate sports training rule for decision support. Then, the effective association rule is input in knowledge base to be used for later sports training program. The next one is the generated rules meeting the conditions after data mining process. If the minimum support *min_sup*=20, the minimum confidence, that is, the minimum probability is *min_con*=0.8. While the importance, that is, the correlation degree is *min_imp*=0.85. Parts of the generated rules are shown as Figure 4.



Mining model: student Viewer: Association rules viewer
Minimum probability: 0.80 Filtering Rules:
Minimum importance: 0.86 Display:
□ Display long name Maximum number of lines: 100

```
Probability   Importance   Rules
1.000         1.399        Height>=175 Province=Shandong->basketball=existing
1.000         1.204        Province=Tianjing pressure<110-> volleyball=existing
1.000         1.204        Province=Tianjing Height>=150 164->volleyball=existing
1.000         1.204        Province=Tianjing Sex=femail ->volleyball
1.000         1.148        Province=Fujian Sex=femail->badminton Heart beat >74, Height=168-176->swim
1.000         1.101        Province=Hanan Feight=58-63 ->swim
1.000         1.101        Province=Hainan Feight=154-163 ->basketball
1.000         1.101        Province=Shandong Extending pressure<88 ->basketball
1.000         1.101        Province=Fujian Pressure=110-113 ->badminton
1.000         1.091        Province=Fujian Feight=58-63 -> aerobics
1.000         1.071        Province=Xinjiang Vital Capacity =2072 3172 -> aerobics
1.000         1.068        Province=Xinjiang Sex=female -> aerobics
1.000         1.050        Province=Liaoning Entrance day=2005 -> football
1.000         1.058        Province=Hainan vital capacity=3630-3581->swim
1.000         1.050        Province=Hainan Heart beat=72 74 ->swim
1.000         1.000        Province=Hainan Extending pressure<56 -> swim
1.000         1.000        Province=Hainan vital capacity=2025 3172 ->jogging
1.000         1.000        Province=Hainan Extending pressure>=76 -> swim
```

FIGURE 4 The order of importance of rules

4.2 IMPLEMENTATION OF IMPROVED ID3 ALGORITHM MODEL

There are four attributes given in the example: height, flexibility, strength and speed. According to the experience of sports experts the importance of them is: 0.25, 0.27, 0.31 and 0.33. The rune off value is 0.8.

TABLE 1 An example of collection for long jump training

No.	Attributes				Categories
	Height	Flexibility	Strength	Speed	
1	Tall	good	strong	slow	Y
2	Tall	good	strong	fast	Y
3	short	good	strong	slow	Y
4	middle	common	strong	slow	N
5	middle	bad	weak	slow	N
6	middle	bad	weak	fast	N
7	short	bad	weak	fast	N
8	Tall	common	strong	slow	Y
9	Tall	bad	weak	slow	N
10	middle	common	weak	slow	N
11	Tall	common	weak	fast	N
12	short	common	strong	fast	Y
13	short	good	weak	slow	Y

According to improved ID3 algorithm, we use $H_2(X \mid \alpha) = S_i 2\sum_{j=1}^{t} n_j p_j / (n_j + p_j)$ to compute the gain of each attribute and the results are shown in Table 2.The decision trees are established respectively based on ID3 algorithm and the improved algorithm. The final result verifies they are the same one.

The decision of each example in the training set is classified correctly as shown in Figure 5. The decision tree leafs are category name. Other nodes are composed by the feature attribute of examples and their different value of each attribute corresponds to a branch. If we want to classify the entities, we will begin form the root node. The node is test along the branch to enter lower nodes. This process continues until arrives at the leaf node. Then the entity is determined to belong to the class that this leaf belonging to that one.

TABLE 2 Comparison of improved algorithm and previous algorithm in gain

Attribute	Gain of original algorithm	Gain of improved algorithm
Height	0.244	1
Strength	0.152	1.540
Flexibility	0.031	1.659
Speed	0.048	1.933

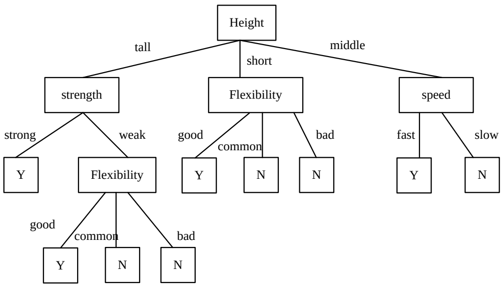


FIGURE 5 ID3 decision tree and improved decision tree

It can be seen from the results that the improved decision tree structure is reasonable, that is, the depth makes little difference and the final rule has no sex attribute node. It is consistent with its low importance and it shows the advantage of improved model.

5 Conclusion

This paper uses data warehouse technology and data mining technology to establish the assistant decision support system of sports training for athletes. It provides organic integration of athletes' different aspects and applies scientific training theory and advanced training method into long-jump athletes' sports training management. This system is based on athletes' sports scores and physical checklist in one university. It produces new knowledge rule to enrich the knowledge base by data mining. Corresponding models are selected according to users' input. Meanwhile, a set of reasonable sports training projects are gradually generated with rules in knowledge base. We study and improve the data mining association rule algorithm, and mainly discuss the implementation of association rule model in sports decision support system. It can instruct the sports training project generation of athletes by generated training rules which are mined by multi-dimension dataset in data warehouse. It also provides reference for establishing physical education course and PE elective course, which also enriches the knowledge base.

Acknowledgement

This work is supported by Science and Technology Department of Henan Province, No: 132300410446, "Under the environment of the particle swarm optimization algorithm application research of cloud computing".

References

[1] Zhang W, Yang B, Song W 2006 Review of Multi-relational Data Mining *Computer Engineering and Applications* **2** 1-6

[2] Ali Khan Sharjeel, Manarvi Irfan 2009 Selecting a sports car through data mining of critical features *In proceedings of International Conference on Computers and Industrial Engineering* 1480-4

[3] Zhang Z 2005 Data Mining Technology and its Application to University Decision Support System *Distance Education Journal* **6** 32-6

[4] Bhatt Chidansh A, Kankanhalli M S 2011 Multimedia data mining: State of the art and challenges *Multimedia Tools and Applications* **51**(1) 35-76

[5] Wang J, Zhao J, Zhang C 2010 Application of data mining in the customization design of the sport bicycle *Journal of Beijing University of Technology* **36**(6) 742-7

[6] Jiang M, Tang Y 2008 Research on Sorting of Student's Achiements in College Computer Teaching by Using ID3 algorithm *Computer & Digital Engineering* **36**(5) 51-4

[7] Dalal M K, Zaveri M A 2012 Automatic text classification of sports blog data *In proceedings of Computing, Communications and Applications Conference* 219-22

[8] Tang Y, Qin Y 2011 A review on classification algorithm based on data mining *Journal of Bohai University* **32**(4) 372-5

[9] Lustigova Z, Dufresne A 2010 Courtemanche François, Mining physiological data for automated feedback in virtual learning environments *In Proceedings of the International Conference on Information Technology Interfaces* 295-300

[10] Wang X, Jiang Y 2011 Analysis and improvement of ID3 decision tree algorithm *Computer Engineering and Design* **32**(9) 3069-76

[11] Rajesh KA B, Phani R C, Madhusudhan E 2012 Threshold extended ID3 algorithm *In Proceedings of The International Society for Optical Engineering* 83-94

[12] Yu D, Zhong Y,Yu Y 2010 Application of Data Mining Technology in Human Muscle Power Data Analysis-Taking the Testing Data of Muscle Power of Grip Strength as Example *China Sport Science* **30**(2) 70-4

Authors

Tao Lu, 1980.5, Qingdao, Shandong Province, P.R. China.

Current position, grades: undergraduate the lecturer of Qingdao Vocational and Technical College of Hotel Managment, China.
University studies: B.P.E. in Physical Education from Xi'an Physical Education University in China, M.PE. from Tianjin University of Sport in China.
Scientific interest: physical education and teaching.
Publications: more than 3 papers.
Experience: experience of 12 years, completed 3 scientific research projects.

Xiaojun Zhang, 1980.01, Zhengzhou County, Henan Province, P.R. China.

Current position, grades: the lecturer of computer Department, Henan Institute of Education of China.
University studies: Master of Science in Multimedia Systems and Computer Graphics (2005) from National University of Technology in Moscow.
Scientific interest: different aspects of network engineering and data mining.
Publications: more than 12 papers.
Experience: teaching experience of 9 years, completed 4 scientific research projects.