

# 基于关联模式数据挖掘的篮球比赛技战术分析

屈静力, 李大永, 宋国强

(河北工程大学 公共体育部, 河北 邯郸 056038)

[摘要] 主要探讨关联规则挖掘技术在篮球比赛技战术分析中的应用。具体分为两个环节, 第一在数据采集方面, 通过寻找频繁脚本提高了采集速度, 满足了临场采集的实时性要求; 第二, 该技术能够解析利用关联规则挖掘中的 Apriori 算法, 解决了篮球比赛中寻找关键动作转换过程的问题, 寻找比赛中技战术之间的关联关系, 从而为教练员的指导和决策提供科学依据。

[关键词] 数据挖掘; 关联规则; 技战术分析

doi:10.3969/j.issn.1673-9477.2013.01.037

[中图分类号] G841 [文献标识码] A [文章编号] 1673-9477(2013)01-0118-04

## 一、引言

数据挖掘也被称为知识发现, 到目前为止仍然是数据库领域和人工智能研究的热点, 数据挖掘就是从数据库大量的数据中发掘出隐含的, 以前未知的和潜在的有用信息的过程, 数据挖掘也是以一种帮助决策者决策的决策支持过程, 这主要是基于模式识别, 人工智能, 机器学习, 数据库, 统计学, 等, 精确分析赛场的数据, 归纳推理, 挖掘潜在的模式, 帮助决策者调整赛场战略, 做出正确的决策。

数据挖掘的方法并不限于只在科研商业应用, 特别是竞技体育, 通常不仅需要运动员的水平高, 同时教练员的战术也很重要, 有时战术甚至起决定性作用。数据挖掘在国外竞技体育运动中已得到广泛应用。例如, 美国 NBA 篮球队教练, 使用公司提供的数据挖掘工具进行辅助决策和替换, 并取得良好效果。安佛尼·哈德卫(Anfernee Hardaway)和伯兰·绍(Brian Shaw)是魔术的先发阵容中的两个后卫, 数据挖掘系统分析表明, 他们在前两场比赛中被评为-17分, 这意味着它们的存在使得球队输了17分。然而, 当替补后卫达尔利·阿姆斯创(Darrell Armstrong)和哈德卫合作时, 魔术队的得分是14分。在接下来的比赛中, 魔术队增加了阿姆斯创比赛时间。此着果然见效。魔术队借助数据挖掘技术赢得了打满5场比赛的机会, 直到最后才决出胜负。同样, 数据挖掘技术可以分析排球, 足球和其他类似的对抗性运动, 在比赛中找出对手的弱点, 从而为教练员提供更有效的策略。

基于关联模式的数据挖掘过程主要由两个阶段构成: 第一阶段: 必须先从历史数据构成的集合中找出所有与比赛相关的高频项目组(Frequent Itemsets), 第二阶段: 利用这些高频项

目组的数据, 产生关联规则(Association Rules)。在数据挖掘关联规则中有一种最优影响的算法就是 Apriori 算法。该算法是一种递推算法, 其核心是依赖于两阶段频集思想。本文研究了一种基于关联规则的数据挖掘算法, 并将其应用在篮球比赛中, 结合篮球比赛介绍其应用。

## 二、基本概念

关联规则定义为: 设定  $I = \{i_1, i_2, \dots, i_m\}$  是项的集合, 我们选定一个用于交易的数据库  $D = \{t_1, t_2, \dots, t_m\}$ , 该数据库中的每个事务(Transaction)  $t$  是  $I$  的一个非空子集, 即  $t \in I$ , 我们赋予每一个交易一个唯一的标识符 TID(Transaction ID), 该标识符合该事物一一对应。关联规则是形如  $X \Rightarrow Y$  的蕴涵式, 其中  $X, Y \in I$  且  $X \cap Y = \emptyset$ ,  $X$  和  $Y$  分别称为关联规则的先导(antecedent 或 left-hand-side, LHS) 和 后继(consequent 或 right-hand-side, RHS)。我们把  $D$  中事务包含  $X \cup Y$  的百分比, 即概率  $P(X \cup Y)$  称为关联规则  $X \Rightarrow Y$  在  $D$  中的支持度(support); 把包含  $X$  的事务中同时包含  $Y$  的百分比, 即条件概率  $P(Y | X)$  称为置信度(confidence)。关联规则是有意义的是指满足最小支持度阈值和最小置信度阈值。这些阈值由用户或者专家设定。

数据挖掘中关联规则挖掘过程由两部分构成:

第一部分: 该规则必须从原始数据库集合中, 找出所有的频繁项集(Large Itemsets)。高频是指某些集合出现的次数在相关记录中, 必须达到一定的水平。所谓的支持度(Support)是指一项目发生的频率, 如果这种支持度大于或等于最小支持度阈值, 则  $\{A, B\}$  称为频繁项集。一个满足最小支持度的  $k$ -itemset, 则称为高频  $k$ -项目组(Frequent  $k$ -itemset),

[投稿日期] 2012-10-10

[基金项目] 邯郸市科学技术研究与发展计划项目(编号: 1234201101-2); 国家自然科学基金(编号: 41104005)

[作者简介] 屈静力(1977-), 男, 河北遵化人, 讲师, 硕士, 主要研究方向: 篮球训练与教学。

一般表示为 Large k 或 Frequent k。该算法将从 Large k 的频繁项集中再产生 Large k+1，直到它不再能发现更长的频繁项集为止。

第二部分：利用所得到的频繁项集产生关联规则(Association Rules)。该步骤是利用前一步骤的频繁 k-项目组来生成规则，在最小信赖度(Minimum Confidence)的条件下，如果一个规则得到的信赖度满足最小信赖度，那么该规则被称为关联规则。篮球技术动作的关联规则挖掘根据 Apriori 算法篮球技术动作关联规则挖掘方法如下：

Apriori 算法，使用逐层迭代找出频繁项集。

输入：事务数据库 D；最小支持度阈值 min\_sup。

输出：D 中的频繁项集 L。

- 1)  $L_1 = \text{find\_frequent\_1\_itemsets}(D)$ ;
- 2) for ( $k=2$ ;  $L_{k-1} \neq \emptyset$ ;  $k++$ ) {
- 3)  $C_k = \text{apriori\_gen}(L_{k-1}, \text{min\_sup})$ ;
- 4) for each transaction t D { //扫描 D 用于计数
- 5)  $C_t = \text{subset}(C_k, t)$ ; //得到 t 的子集，它们是候选

- 6) for each candidate  $C_t$
- 7)  $c.\text{count}++$ ;
- 8) }
- 9)  $L_k = \{C_k \mid c.\text{count} \geq \text{min\_sup}\}$
- 10) }
- 11) return  $L = \cup_k L_k$ ;

Procedure apriori\_gen ( $L_{k-1}$ :frequent(k-1)-itemsets)

- 1) for each itemsets  $l_1, l_{k-1}$
- 2) for each itemsets  $l_2, l_{k-1}$
- 3) if ( $(l_1[1]=l_2[1]) \wedge (l_1[2]=l_2[2]) \wedge \dots \wedge (l_1[k-2]=l_2[k-2]) \wedge (l_1[k-1] < l_2[k-1])$ )

- then{
- 4)  $c=l_1l_2$ ; // 连接步：产生候选
  - 5) if has\_infrequent\_subset( $c, L_{k-1}$ ) then
  - 6) delete c; // 剪枝步：删除非频繁的候选
  - 7) else add c to  $C_k$ ;
  - 8) }
  - 9) return  $C_k$ ;

Procedure has\_infrequent\_subset (c:candidate k-itemset;  $L_{k-1}$ :frequent(k-1)-itemsets) //使用先验知识

- 1) for each(k-1)-subset s of c
- 2) If  $s \in L_{k-1}$  then
- 3) return TRUE;
- 4) return FALSE;

据以上算法，把动作 A → 动作 B 这个关联规则作为挖掘最常用技术动作关联的依据，因此，可以

为球员和教练设计最佳的训练。

### 三、关联规则的预处理

本文以体育比赛中的篮球比赛为例，应用关联规则中的 Apriori 算法，对相关篮球技术动作之间的关系进行挖掘，并得到了有益的输出。因为篮球比赛数据库数量比较大，本文只选取其中的一部分进行分析。假设对篮球比赛事物数据表如表 1 所示， $D = \{A, B, C, D, E, F, G, H, I\}$ ，其中 A 代表技术动作助攻，B 代表 2 分球得分，C 代表篮板，D 代表抢断，E 代表 3 分球得分，F 代表罚球，G 是突破，H 是挡拆，I 是盖帽。

$S = \{S1, S2, S3, S4, S5, S6, S7, S8, S9, S10\}$

假设最小支持度为 30% 及其以上，最小置信度为 40%。

### 四、关联规则挖掘

对数据表进行了上述分类和符号化后，得到篮球比赛事务数据表。该表的相关信息字段记录了比赛出现的技术动作及其相关信息。篮球比赛事务数据表如表 1 所示：

表 1 篮球比赛事务数据表

TID	出现的技术动作
S1	A, B, G, H
S2	A, B, C, E
S3	A, B, C, D, F
S4	C, G, H
S5	A, B, C, G
S6	C, D, H
S7	B, E, F
S8	B, C, F, I
S9	C, H
S10	B

根据 Apriori 算法，我们设定最小支持度为 45%，对表 1 中的相关数据进行挖掘，得到包含元素和病变的各个大项目集。第 1 大项目集如表 2 所示：

表 2 C1

项目集	支持度	项目集	支持度
A	65%	E	32%
B	85%	F	31%
C	66%	G	52%
D	15%	H	18%

其中第一条记录 A 的支持度为 67%，这是因为在表 1 中共有 4 条记录包含 A，表 1 中共有 6 条事件记

录,所以A的支持度为 $4/6=67%$ ,大于预设的最小支持度30%,所以保留A作为第1大项目集中的一条记录。表2中的其它记录也可以类似得到。对C1的筛选,可得L1。

表3 L1

项目集	支持度	项目集	支持度
A	65%	C	66%
B	85%	G	50%

L1 经过“连接”和“剪枝”,得到C2。如表4所示

表4 C2

项目集	支持度	项目集	支持度
AB	65%	BC	51%
AC	53%	BG	51%
AG	54%	CG	33%

其中第一条记录{AB}的支持度为50%,这是因为在表1中共有4条记录包含{AB},且表1中共有6条事件记录,所以{AB}的支持度为 $4/6=67%$ ,且大于预设的最小支持度。经过筛选大于最小支持度的如表5所示:

表5 L2

项目集	支持度	项目集	支持度
AB	67%	BC	50%
AC	50%	BG	50%
AG	50%	CG	33%

同理可得C3, L3 如表6, 7 所示:

表6 C3

项目集	支持度	项目集	支持度
AB	67%	BC	50%
AC	50%	BG	50%
AG	50%		

表7 L3

项目集	支持度	项目集	支持度
ABC	50%	ACG	33%
ABG	50%	BCG	33%

如上可知,在数据库中用的最多的是{ABC}和{ABG},即{助攻,2分球得分,篮板}和{助攻,2分球得分,突破}。

## 五、结论

通过对篮球比赛现场丰富的数据进行挖掘,实现比赛现场数据的实时展现,辅助教练员进行现场分析,对于教练员的战略部署具有重要的参考价值。如果能对对方的动作和战术做出合理的预测,就能达到一种“知己知彼”的状态,占据比赛中的有利形势。

## 参考文献:

- [1]茅洁,单曙光:篮球技术统计软件的研[J].武汉体育学院学报,2012(2):70-73.
- [2]Usama Fayyad, Gregory Piatetsky-Shapiro, and Padhraic Smyth, Knowledge discovery and data mining: Towards a unifying framework. Proceedings of the International Conference on Knowledge Discovery and Data Mining, pages, 1996:82-88.
- [3]Agrawal R, Srikant R. Fast algorithms for mining association rules [A]. Proceedings of the 20th International Conference on Very Large Databases [C]. Santiago: Morgan Kaufmann, 1994:487-499.
- [4]Agrawal R, Srikant R. Mining sequential patterns. In Proc, 1995 Int, Conf. Data Engineering (ICDE'95), pages 3-14, Taipei, Taiwan, Mar, 1995.

[责任编辑 王云江]

## Data mining algorithm of association rule was used to analysis basketball matches technical-tactics

QU Jing-li, LI Da-yong, SONG Guo-qiang

(Public Physical Education Department, Hebei University of Engineering, Handan 056038, China)

**Abstract:** This paper discusses the application of data mining algorithm of association rule in basketball match. The process includes two steps. The first is data collection. Finding out the focal process of action transition improves the speed of data collection, meeting the demand of the instantaneity of of data collection. The second step is that the apriori algorithm is applied to find the association rules in the technical-tactics of the match to provide a scientific basis for coaches' guidance and decision.

**Key Words:** data mining; association rule; technical-tactics analyzing