

# 基于数据仓库的篮球教练辅助决策支持系统<sup>①</sup>

方心飞, 叶 枫, 周 钧

(浙江工业大学 经贸管理学院, 杭州 310023)

**摘 要:** 现在的篮球教练做决策时候还仅仅依靠原先的经验和专业技能, 这样的决策有时候往往会产生不好的效果导致球队输球, 教练自己下岗。就这个问题笔者认为可以采用设计一个篮球教练辅助决策支持系统的方法给予一定程度的解决, 具体在实现该系统时运用数据仓库和数据挖掘技术的方法, 然后还以 NBA 球队火箭队球员在助攻维度的数据进行实验分析。从得出的实验结果即频繁项集发现, 结果可以为教练如何排兵布阵提供有效的帮助, 所以笔者相信该系统肯定可以在一定程度上为篮球教练决策时候提供一些帮助和参考。

**关键词:** 篮球; 决策支持系统; 数据仓库; 数据挖掘

## Basketball Coach's Assistant DSS Based on Data Warehouse

FANG Xin-Fei, YE Feng, ZHOU Jun

(Zhejiang University of Technology, Hangzhou 310023, China)

**Abstract:** Now many basketball coaches only rely on their experience and specialty Skill when they do the decision. So in this case the coach's decision may be wrong and the team may lose the game. Even the coach may be fired. Concerning this problem, we can design this DSS (Decision Support System) to solve it. Here we use Data Warehouse and Data Mining skill to realize the DSS. And we use the assist data of Rockets players to experimental analysis. From the experimental results that frequent itemsets, we found the results were helpful for coaches on how to provide effective help. So we believe to some extent that the system can certainly provide some help and reference for the basketball coach.

**Keywords:** basketball; decision support system; data warehouse; data mining

众所周知, 一个篮球队要取得好的成绩甚至拿到总冠军, 除了要有好的明星球员, 还要有一个好的教练。我们可以从教练的工作任务看出一个教练对一个球队的作用有多重要。教练的工作主要是两个方面: 第一个方面是计划安排球队的训练计划, 还有组织球队球员训练。通过这个训练的过程使球员的身体素质和篮球技能达到一定的程度。第二个方面是: 每场比赛的时候针对不同的对手如何去排兵布阵。用最优的球员轮换上场方式来帮助球队获胜。所以总的来说, 教练就是一个决策者。以往的教练在决策是就是根据自己的经验和自己在篮球方面的专业知识来决定一个周期里训练的强度, 训练的项目以及技战术如何安排。

但是随着管理科学、运筹学、决策分析技术、数理统计的进步, 一个光靠以往经验和专项知识的教练比不上一个会利用决策支持系统来辅助自己决策的教练。因为决策支持系统是在搜集了各方面可用的数据, 然后通过一定的数据处理模型和技术对数据进行专业的分析, 然后得出比较可靠的经验。所以每个俱乐部最好有一个自己的决策支持系统, 那样可以帮助球队取得更好的战绩

## 1 篮球教练辅助决策支持系统设计

### 1.1 系统设计目标

系统的设计目标总的来说就是辅助篮球教练员在

<sup>①</sup> 收稿时间:2010-05-10;收到修改稿时间:2010-06-06

平时训练时候和比赛时候做决策。具体的讲即通过建立数据仓库，然后通过数据挖掘技术等等技术来处理收集来的数据，为决策者解决下面一些方面的问题：

(1) 球队训练的安排：这里包括训练周期的划分和训练强度的确定。一个球队要提高战绩并不是说训练的越勤快，训练强度越大越好。如果训练过于频繁或者训练强度太大的话，那么球员们会产生抵触心理。同样的如果训练太少或者周期间隔太长的话，对于球员保持竞技状态也是不好的。所以球队训练的时间和强度也是一个篮球教练员的战略性的决策。

(2) 球员的安排的安排：每个球员的身体素质和篮球技能是不同的。所以教练在安排每一个球员的训练项目时要根据个体的差异性有针对性。例如通过技术分析出某一个球员投篮技术不怎么好，那么教练就要为这个球员多安排一些投篮练习。我们还可以分析 3 分和 2 分的命中率的差别，决定这个球员是该多练习中投还是多练习 3 分。

(3) 比赛时候得分问题：一个球队要想赢得比赛，肯定要得比对方更多的分数，所以利用该决策系统可以帮助教练分析出自己球队球员与对手球队某一个球员对位时候可以得更多的分数或者得分比较困难。那么就可以安排相应的球员来对抗对手。

(4) 比赛时候防守问题：在 NBA 几年来来总冠军球队的防守往往不会差，例如太阳队这样的炮轰球队一般不易得到总冠军。因为炮轰球队不会防守。所以防守对于一个球队是很重要的。系统可以分析出针对对方球队的不同选择不同的防守方式。到时候是选择选择人盯人，还是联防等等系统可以给出一个优先的方式。然后具体到每一个球员的防守时候，可以分析出派谁防对方某某球员更好。

(5) 团队化学反应：比赛不是一个人的。不是光靠球星的。团队之间的配合很重要，不论是防守时候的配合还是进攻时候的配合都很重要。系统可以分析出哪些球员在一起打球容易配合，容易产生化学反应。

总的来说，系统的目标就是在球队训练时和比赛时给予篮球教练提供帮助。

### 1.2 系统功能设计

结合上面讲的系统需要解决的问题，该系统主要实现的功能有两点，一是给予教练训练场的决策辅助，另一个功能是具体到每一场比赛，给予教练决策支持。

所以该系统主要包括两个子系统：一个是训练决

策支持子系统，包括根据这段时间球队训练的频率，强度和球队战绩之间的关系来确定球队最佳的训练频率，最佳的训练强度。还有某某球员在比赛时候在摸个项目上暴露出来的问题确定他这段时间的主要训练项目，以及记录下球员训练时候的状态以备比赛时候参考。

另一个是赛场决策支持子系统，包括球队在比赛时候如何进行球员轮换，针对对方球员如何排兵布阵。首先，结合对方球队和球员的比赛特点，确定出比赛时候要想赢球的关键点。例如，对方球队 3 分命中率很高，那就不能采取联防的防守策略，而要采用人盯人的防守策略。然后具体到比赛时候可以参考系统在防守维度(主要是篮板，抢断，盖帽)给出的统计分析如何派人去防守对方的某某球员。然后参考系统在助攻维度通过分析助攻发生在哪些球员之间来安排哪些球员一起上场比赛可以实现更好的配合，参考得分维度把球权主要交给谁还有分析出每个球员的投篮热区(就是命中率比较高的范围)等等。

### 1.3 系统结构设计

该系统的设计是基于数据仓库的，结合了数据仓库、联机分析处理和数据挖掘设计的。其层次系统结构框图 1 所示<sup>[1]</sup>。它主要由 3 部分(层次)组成：

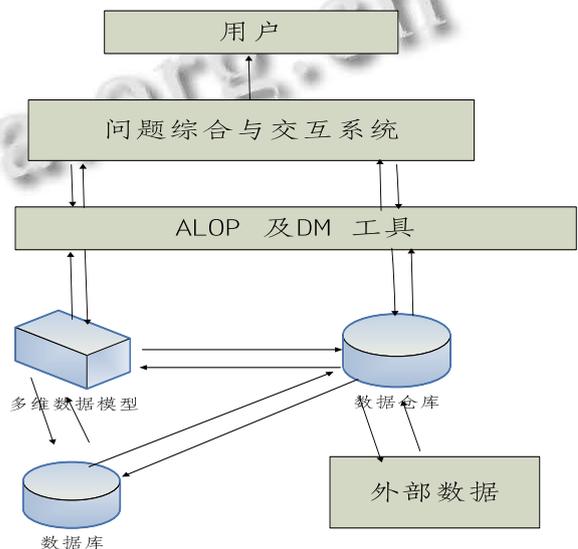


图 1 基于数据仓库的 DSS 的体系结构

(1) 数据仓库。它是 DSS 的决策基础，是这个系统的数据的一个集合<sup>[2]</sup>。

(2) 数据分析处理工具和数据挖掘工具。前者是以数据仓库的数据为基础，按一定的模型进行分析，主要以人机交互为操作手段。后者是对数据仓库的数据进行归纳总结和分析，最好形成规则性的知识<sup>[3]</sup>。

(3) 人机交互的模块。这里主要是指将分析结果以表格、直方图、曲线图、饼图展示给用户。

## 2 系统的数据仓库设计

### 2.1 确定主题

数据仓库是一种管理技术，所以数据仓库是面向主题的，我们根据主题来对数据进行归类<sup>[4]</sup>。在该系统中我们对主要用户教练进行需求分析后，我们觉得可以分析出教练员最关心下列一些指标。如图 2 所示。

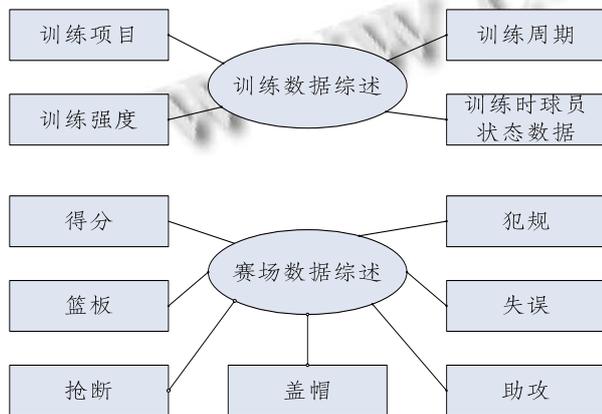


图 2 系统数据综合描述

该系统数据主要包括两个方面：一方面是训练场的数据主要包括训练项目、训练强度、训练周期以及球员训练时候表现出来的状态。另一方面是比赛场上球员的数据情况，主要包括一些基本的技术统计：得分、篮板、抢断、犯规、失误、盖帽、助攻。

### 2.2 概念模型设计

概念模型是通过计算机里的模型和语言对可观世界里的问题进行一些具体的描述，起到沟通主观与客观之间的桥梁作用<sup>[5]</sup>。概念模型的设计方法如下：

(1) 确定事实：决策者决策时候的目标数据就是事实。在该系统中主要是以各个球员的一些详细数据作为度量值。

(2) 确定维度：维度是考察事实的角度，包括了事实的属性信息。在该系统中主要包括下列几个维度：

①训练周期维度。主要记录球队训练周期的安排情况。

②训练项目维度。主要记录球队球员训练的项目情况。

③训练强度维度。主要记录球队训练强度的情况。

④训练时候球员状态维度。主要记录球员在训练时候表现出的状态。

⑤得分维度：记录球员在赛场上比赛时候的得分情况。

⑥篮板维度：记录球员在赛场上比赛时候的篮板情况。

⑦助攻维度：记录球员在赛场上比赛时候的助攻情况。

⑧抢断维度：记录球员在赛场上比赛时候的抢断情况。

⑨盖帽维度：记录球员在赛场上比赛时候的盖帽情况。

⑩犯规维度：记录球员在赛场上比赛时候的犯规情况。

⑪失误维度：记录球员在赛场上比赛时候的失误情况。

### 2.3 逻辑模型设计

#### 2.3.1 相关表的设计

球员表 (球员 ID, 身高, 体重, 位置), 训练项目表(项目 ID, 一次训练时间, 间隔训练时间), 训练周期表(训练 ID, 间隔训练时间, 训练强度等级), 球员训练状态表(球员 ID, 训练 ID, 状态评价), 得分表(得分 ID, 球员 ID, 分数, 得分方式, 得分位置, 得分时刻), 篮板表(篮板 ID, 球员 ID, 篮板数, 前场篮板数, 后场篮板数, 抢篮板时刻), 抢断表(抢断 ID, 球员 ID, 抢断数, 抢断对手球员位置, 抢断时刻), 盖帽表(盖帽 ID, 球员 ID, 盖帽数, 盖帽时刻), 助攻表(助攻 ID, 球员 ID, 助攻数, 被助攻球员, 助攻时刻), 失误表(失误 ID, 球员 ID, 失误数, 失误类型, 失误时刻), 犯规表(犯规 ID, 球员 ID, 犯规数, 被犯规球员位置, 犯规时刻), 对手表(球员 ID, 得分, 篮板, 盖帽, 抢断, 助攻, 失误, 犯规)



图 3 训练子系统关系图



图 4 赛场子系统关系图

### 2.3.3 维度设计

由于这里有两个子系统，所以两个子系统分开考虑，

#### 2.3.3.1 训练子系统

##### (1) 事实表的设计

事实表(球员 ID, 训练 ID, 项目 ID)

##### (2) 各个维度的设计

在该子系统中主要考虑的维度有训练强度、训练项目、训练周期以及训练时候球员状态。在这里每个维度涉及的表是和维度同名的表。

#### 2.3.3.2 赛场子系统

##### (1) 事实表的设计

事实表(得分 ID, 篮板 ID, 抢断 ID, 助攻 ID, 失误 ID, 犯规 ID, 盖帽 ID)

##### (2) 各个维度的设计

在该子系统中主要考虑的维度有得分维度，篮板维度，抢断维度，助攻维度，失误维度，犯规维度，盖帽维度。在这里每个维度涉及的表是和维度同名的表。

## 2.4 物理模型设计

这是该系统设计的最后一个阶段，主要有下列的任务：

(1) 选择开发工具：选用 SQL2005 作为开发平台。因为它自带了数据挖掘的功能。

(2) 创建数据库表和多维数据集。根据上面的分析建立相应的表和数据集。

(3) 创建适当的索引。由于数据比较庞大。建立适当的索引可以优化检索速度。

## 3 例证分析

### 3.1 分析原理

这里以助攻维度为例，来说明系统是如何通过数据挖掘相关技术来发现球员之间的化学反应的。如果两个球员之间助攻记录比较多，那么他们的化学反应就比较好，那么教练就应该多让他们一起出现在场上。我们把发生一次助攻记为一条记录，主要设计到的就是助攻球员 ID 和被助攻球员 ID，运用关联规则对其进行数据挖掘。在关联分析时候我们把助攻球员和被助攻球员作为一个项集。则每条记录这样记为：(idk, 助攻 ID, 被助攻 ID)其中 k=1, 2, 3...，助攻 ID 和被助攻 ID 是球员的号码。

在数据仓库理论中如果项集满足最小支持度，那么这个项集就是频繁项集<sup>[6]</sup>。但是针对该系统，我们在这可以不去管最小支持度是多少，而只须考虑哪些项集比其它项集频繁，那么这些项集就是我们所谓的“频繁项集”，针对与助攻来说就是这两个球员在一起容易产生化学反映，所以他们的号码组成的项集比较频繁的出现。通过这样找到球队中最好的球员组合。通俗的说就是我们要找到某两个球员他们一起出现在记录中次数最多的那对。

### 3.2 分析实践

#### 3.2.1 数据准备

我们针对 NBA 球队火箭队的 09-10 常规赛前 10 场比赛记录，例如具体数据各个体育网站都有。第一场比赛，2009 年 10 月 28 日火箭队对阵开拓者队的比赛发生的助攻记录如下表，其余 9 场比赛的记录在各种体育网站都有，这里由于篇幅问题就不一一列举。

表 1 2009 年 10 月 28 日火箭队对阵开拓者队的比赛发生的助攻记录

助攻 ID	助攻者 ID	被助攻者 ID	助攻 ID	助攻者 ID	被助攻者 ID	助攻 ID	助攻者 ID	被助攻者 ID
1	0	1	2	0	1	3	0	4
4	0	44	5	1	7	6	0	7
7	1	44	8	44	31	9	4	44
10	7	1	11	13	14	12	7	0
13	7	13	14	7	13	15	7	10
16	7	10	17	7	0	18	7	1

### 3.2.2 数据分析

我们通过数据挖掘得到以下的信息, 球员 ID 为 0 和球员 ID 为 4 的球员之间发生了 14 次助攻, 而且经计算机统计是最大的, 那么我们就说(0, 4)的频数最大, 为 14, 之后类似的分别是(0, 1)频数为 13, (1, 4)(7, 14)频数都为 12, (1, 44)(7, 10)频数为 10. 之后小于 10 的我们就不考虑。上面这些项集我们可以把他叫做频繁项集。

### 3.2.3 分析结论

根据(0, 4)是频繁项集的话, 教练员应该多让 0 号球员和 4 号球员同时在上, 又根据(0, 1)是频繁项集的话, 教练员应该多让 0 号球员和 1 号球员同时在上, 后面那些频繁项集结论类似可得。总之就是让记录中是频繁项集对应的那些球员同时出现在场上打球。

## 4 结束语

该系统的设计是为了帮助篮球教练更好的在球场和比赛时候做出决策。在系统设计时候主要以 SQL2005 为平台, 以数据仓库和数据挖掘为主要技术。希望教练们以后做决策时候不光依靠自己的经验和专业知识, 使用一些现代化的技术来辅助自己。

### 参考文献

- 1 李志刚. 决策支持系统原理与应用. 北京: 高等教育出版社, 2005. 323—326.
- 2 陈文伟, 廖建文. 决策支持系统及其开发(第三版). 北京: 清华大学出版社, 2008. 194—196.
- 3 周根贵. 数据仓库与数据挖掘. 杭州: 浙江大学出版社, 2004. 21—22.
- 4 于宗民, 刘义宁, 祁国辉. 数据仓库项目管理实践. 北京: 人民邮电出版社, 2006. 124—125.
- 5 池太崑. 数据仓库结构设计与实施. 北京: 电子工业出版社, 2005. 161—165.
- 6 邵峰晶, 于忠清, 王金龙, 孙仁诚. 数据挖掘原理与算法(第二版). 北京: 科学出版社, 2009. 90—92.