

Study on the Rule of European Soccer Players Transfer Based on Hidden Markov Model

Xiaomeng Chen, Beier Jin, Cheng Zhu, Yuzhou Gao

Department of Mathematics, Shanghai Jiao Tong University, Shanghai
Email: chenxm94@outlook.com, bejintinkie@gmail.com, arrowrowe@gmail.com, gaoyuzhou1995@163.com

Received: Mar. 11th, 2016; accepted: Mar. 26th, 2016; published: Mar. 31st, 2016

Copyright © 2016 by authors and Hans Publishers Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY).
<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

A more accurate illustration of the mechanism of European soccer players transfer market can be achieved by taking "hidden" factors into consideration and simulating the transfer model in each period through HMM (Hidden Markov Model). With the underlying mechanism of European transfer market better illustrated, it is possible to provide hints for the future path of Chinese soccer market.

Keywords

HMM, Player Transfer, Transfer Mechanism

基于隐Markov链模型的欧洲足球运动员转会规律的研究

陈晓蒙, 金贝尔, 朱成, 高宇洲

上海交通大学数学科学学院, 上海
Email: chenxm94@outlook.com, bejintinkie@gmail.com, arrowrowe@gmail.com, gaoyuzhou1995@163.com

收稿日期: 2016年3月11日; 录用日期: 2016年3月26日; 发布日期: 2016年3月31日

摘要

通过隐Markov链模型描绘欧洲足球运动员在每一个转会期的球员转会过程, 模型中所考虑的各种隐藏

文章引用: 陈晓蒙, 金贝尔, 朱成, 高宇洲. 基于隐 Markov 链模型的欧洲足球运动员转会规律的研究[J]. 统计学与应用, 2016, 5(1): 48-57. <http://dx.doi.org/10.12677/sa.2016.51006>

因素的影响使得模型对转会规律能够做出更为准确的刻画。实证结果所揭示的欧洲足球运动员转会市场所隐藏的规律，对中国足球今后的发展具有一定的借鉴意义。

关键词

隐Markov链，球员转会，转会机制

1. 课题背景及意义

伴随着体育运动职业化的产生，运动员转会成为一个越来越被体育爱好者所熟知的现象。1876年，苏格兰足球运动员詹姆斯·兰转会到英格兰俱乐部，是历史上记载最早的转会事件。此后，运动员转会特别是欧洲职业足球运动员转会逐渐成为普遍现象，大大推动了职业足球联赛中的人才流动和技术流动。

经济上我们经常研究市场的竞争性和外部效应，在欧洲足坛，相对成熟的俱乐部体制下也有一个类似的市场环境。在博斯曼法案¹实行近20年后的今天，转会也变得更加有规律可循。我们注意到在当前的市场上，2010年王红娟[1]、2007年吴育华等[2]和刘月花[3]等文献从制度法律、社会学等角度切入探究这一现象，却极少有类似的数学模型来研究足球运动员转会，而我們希望能更理性地来分析这一现象。因此本小组旨在运用数学模型寻找球员转会的潜在规律，并分析近年转会形势，希望能在将来对中国足球市场的运作有所借鉴。

2. 模型构造

2.1. 数据来源

为探究欧洲足球运动员转会的定价机制，需要大量的数据，经过调查研究，最终将数据来源确定在SQUAWKA (<http://www.squawka.com/>)。数据收集范围为从2012年夏季转会期(7、8月)开始到2014年夏季转会期(7、8月)共计六个转会期。其中球员收集着重于与欧洲五大联赛球队有关联的球员，总计179名。球队主要也来源于欧洲五大联赛，共计99队。经过筛选，去除有所不足的部分数据，最终有效数据为：42支球队在3个赛季中的122条记录。每条记录含12项数据，总计1464项；85名球员，每名球员13项数据，总计1105项。

其中，每条球队记录包括如下属性：各赛季排名、总进球数、定位球得分、角球、点球、SA(射门精度)、TCC(创造机会总数)、助攻、失球数、平均对抗成功率、拦截成功率、FoulsS/CRate(被犯规与犯规比例)、TakeOn成功率、HeadedDuel(争顶)成功率、场均防守动作、控球率、平均传球精度。

接下来，在研究交易情况时，我们的数据包括了以下19个属性：被交易球员的司职位置、类型(攻、防、综)、出场次数、场均时间(分钟)、场均进球、场均助攻、场均射门、射门准确率、威胁球、过人、对抗成功率、解围、总共防守失误、黄牌、红牌、传球成功率、场均防守动作、买家、卖家。

2.2. 模型结构概述

隐马尔科夫模型是一个双内嵌式随机过程。它包含两个过程，一个是状态转移序列，另一个是转移时输出的符号组成的符号序列。其中状态转移随机过程不可观测，只能通过另一个随机过程的输出观测序列观测。由于模型的状态不为外界所见，只能根据观测序列推导出来，因而称为隐马尔科夫模型(隐马

¹1990年夏，比甲RFC列日队中场球员让·马克·博斯曼(Jean-Marc Bosman)在赛季末合同到期，而俱乐部准备削减60%年薪，故希望转会到法国敦刻尔克俱乐部，但敦刻尔克无力支付列日俱乐部开出的高额转会费，导致转会失败。旧有转会体制下，球员合同到期要转会必须向原俱乐部支付转会费。1995年12月15日，欧盟法院作出了有利于博斯曼的裁决，此后，欧盟范围内球员流动加速，小俱乐部球员能等到合同期满加盟豪门。

尔科夫模型相关理论推导参考[4]-[7])。

在我们的模型中，事先给定了若干赛季中的球员交易记录，根据每次转会方向作允许重边的有向图来模拟马尔科夫链。效果如图 1，例如如果一名球员由曼联转向皇家马德里，则作一条由曼联指向皇家马德里的有向线段。

其中，顶点与球队一一对应：

- 1) 有向边与该赛季中的转会球员一一对应，自转出队指向转入队；
- 2) 从图中取若干链，将每条链视为上述一般的隐马尔可夫链模型中所说的过程；
- 3) 对于某条链上的某个点，称它为一个状态，并具有如下定义的里状态和表状态。

这样，对于长度为 n 的链，里状态 X_1, \dots, X_n 组成隐马尔可夫链模型的状态转移序列，表状态 Y_1, \dots, Y_n 组成转移时输出符号组成的符号序列，并且，里状态描述了所对应的球队按本身属性被划分为的类别(共划分了 9 类)，即，里状态取值范围为 $1, 2, \dots, r$ ，其中 $r = 9$ 。

表状态描述了该球队按买入情况被划分为的类别(共划分了 10 类)，即，表状态取值范围为 $1, 2, \dots, s$ ，其中 $s = 10$ 。

2.2.1. 模型要素

- 初始概率分布 π ，即 X_1 服从的概率分布

$$\pi = (\pi_i = P(X_1 = i), 1 \leq i \leq r)$$

为该赛季中球队属性的分布，这里隐含了“该赛季内每个球队具有自己确定的属性”的假定。

- 转移矩阵 P ，即时齐马尔可夫链 $\{X_i, 1 \leq i \leq n\}$ 的转移矩阵

$$P = (p_{i,j} = P(X_2 = j | X_1 = i), 1 \leq i, j \leq r)$$

描述了球员在不同属性球队间的流入流出规律。

- 发射矩阵 Q ，即，给定里状态时，表状态所服从的确定的概率分布

$$Q = (q_{i,j} = P(Y_1 = j | X_1 = i), 1 \leq i \leq r, 1 \leq j \leq s)$$

描述了各属性球队的买入规律。

2.2.2. 状态算法的解释

以下为表里状态判定算法的细节。

- 里状态算法

里状态是球队评分坐标点的分类。对每个球队，在进攻、防守、综合三个维度上分别评分，从而使其对应到三维空间上的一个点。对所有球队的坐标点进行分类，每个点所属的类别，就是其所对应的球队的里状态。

这里考虑了整合不同赛季时的排名加权，高排名赛季数据减弱，低排名赛季数据加强。

- ◇ 进攻维评分与总进球数、定位球得分、角球数、点球数、射门准确率、总共创造机会数、助攻数正相关。
 - ◇ 防守维评分与失球数负相关，与平均对抗成功率、场均防守动作正相关。
 - ◇ 综合维评分与总共创造机会数、助攻数、控球率、平均传球精度、平均对抗成功率正相关。
- 这三个评分的具体参数将在第 2.5 节模型实现中得出。

- 球员定位算法

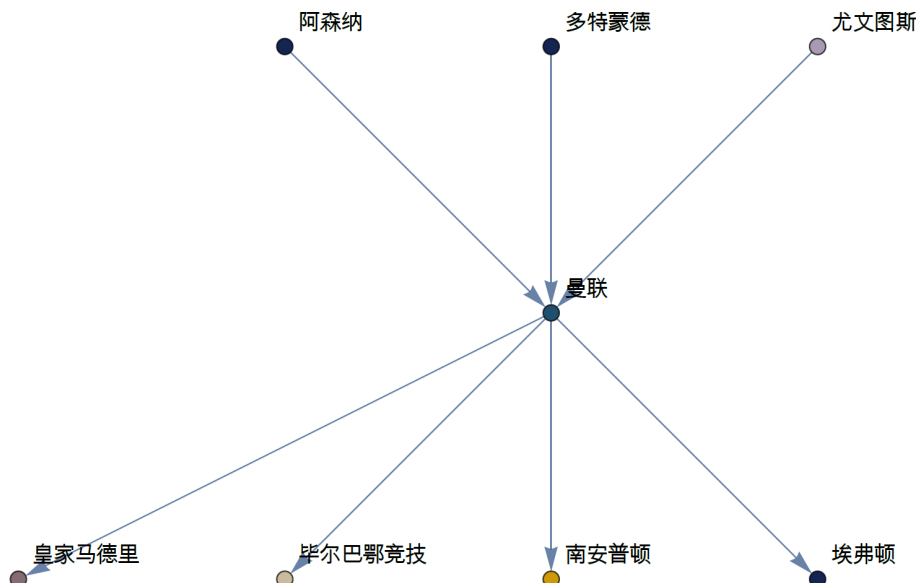


Figure 1. The given directed graph transfer records in several seasons

图 1. 给定若干赛季中的球员交易记录允许重边的有向图

球队的买入情况是由其买进的球员体现的。这里，球员定位是将球员分配到一个维度并在这个维度上评分。对每个球员，认为其属于进攻、防守、综合三个维度之一，并具有一个评分。

首先，根据司职位置确定球员所属维度(不考虑 GK 司职)：

- **进攻型球员**包括：CF, F, LWF, RWF, SS, WF
- **防守型球员**包括：CB, LB, RB, WB
- **综合型球员**包括：AMC, CMF, DMF, RMF

对各类数据都按照出场次数和出场时间做了平均再参与计算。

进攻型球员的评分与场均进球、场均助攻、场均射门、射门准确率、威胁球、过人次数正相关。

防守型球员的评分与对抗成功率、解围正相关，与防守失误、黄牌、红牌负相关。

综合型球员的评分与场均助攻、威胁球、传球成功率、对抗成功率、场均防守动作正相关。

这三个评分的具体参数同样也在第 2.3 节模型实现中得出。

• 表状态算法

表状态是对球队买进坐标的分类，这里买进坐标就是它所买进球员的评分的平均。

具体而言，对每个球队，在进攻、防守、综合三个维度上分别评分，评分标准为该维度上买进球员的平均。

2.3. 模型实现

2.3.1. 里状态算法

根据我们确定的算法所需各参数，绘制出的三维点阵能明显看出同类点的靠拢，绘制的函数就是将三个评分函数组成一个由球队到坐标点的映射，用 Manipulate 函数包裹以限定各参数取值，效果如图 2。

基于这个交互界面，逐步调整参数直到三维空间中聚类效果满意，然后记录下分类，即得里状态，详细描述可参考[8]。

2.3.2. 表状态算法

对原始数据每行所代表的的球员逐个评分为确定一个球队的坐标，对每个维度，取它所购买的该维度

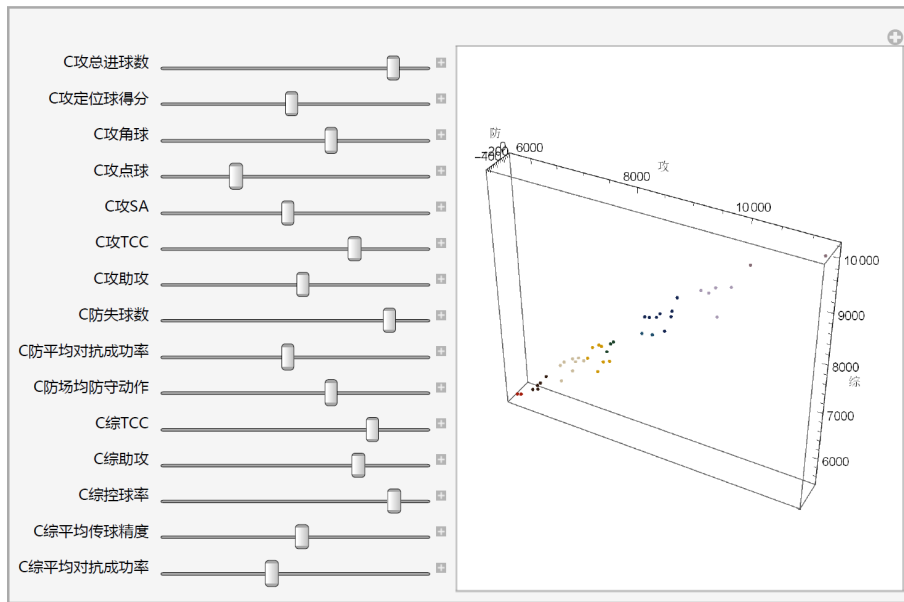


Figure 2. Three dimensional inner status lattice simulation effect diagram
图 2. 里状态三维点阵模拟效果图

的所有球员的均值。

以 Manipulate 函数包裹参数后效果如图 3，基于此的表状态获得过程与第 2.3.2 节一致，不再赘述。

2.3.3. 里状态分布、转移矩阵、发射矩阵的计算

已经获得里状态和表状态后，其分布和相互转换的概率的计算都没有实质困难了。可以观看一下里状态分布，效果如图 4。

里状态分布、转移矩阵、发射矩阵的计算公式如下：

$$\text{转移矩阵}_{i,j} = \frac{\text{从第}i\text{类球队到第}j\text{类球队的球员数}}{\text{第}i\text{类球队卖出的球员总数}}, i = 1, \dots, 9, j = 1, \dots, 9$$

$$\text{发射矩阵}_{i,j} = \frac{\text{第}i\text{类球队中买入情况为第}j\text{类的个数}}{\text{第}i\text{类球队总数}}, i = 1, \dots, 9, j = 1, \dots, 10$$

$$\text{里状态分布}_i = \frac{\text{第}i\text{类球队数}}{\text{球队总数}}, i = 1, \dots, 9$$

具体所得数据在以下第 3 节模型分析结果中分析解释。

3. 模型分析结果

在上述模型中需要解释的主要有里状态分布 π 、转移矩阵 P 、发射矩阵 Q 和表状态分布 πQ 。

我们通过收集球队的各项数据，对球队进行一个评分，以方便对球队进行评级。如前所述，球队本身的数据可以对应到一个三维向量，各维度分别表示球队在前场、中场以及后场方面的表现。最终将所有球队分为 9 类，如图 5 中颜色区分所示。

如图，从左下角到右上角依次为第一至第九类，其中同种颜色代表同一类球队。我们发现，各支球队的防守对于分类的影响并没有进攻和综合能力影响大。在进攻和综合的二维平面上，球队几乎呈现出了一个线性分布。其中，拜仁慕尼黑和皇家马德里凭借出色的进攻能力和综合实力处在直线的最上方，英超球队普遍处于整条直线的中间部分。这是由于英超整体实力相较于其他联赛更加强大，球队在联赛中

较难打出高数据表现，使得在由数据占主导的分析模型中处于中游位置。意甲球队除尤文图斯外均集中于英超球队的下游，这是由于意大利足球风格较为沉闷、强调防守，使得在进攻层面上表现不佳。西甲球队巴塞罗那和皇家马德里处于整体最靠上的一个分类中，其余西甲球队零散分布于整条直线上。这是由于西甲两极分化严重，除开巴塞罗那和皇家马德里其余球队水平参差不齐。

最终我们得到里状态分布向量：

$$\left[\frac{1}{21}, \frac{5}{42}, \frac{4}{21}, \frac{1}{6}, \frac{1}{14}, \frac{1}{21}, \frac{4}{21}, \frac{5}{42}, \frac{1}{21} \right]$$

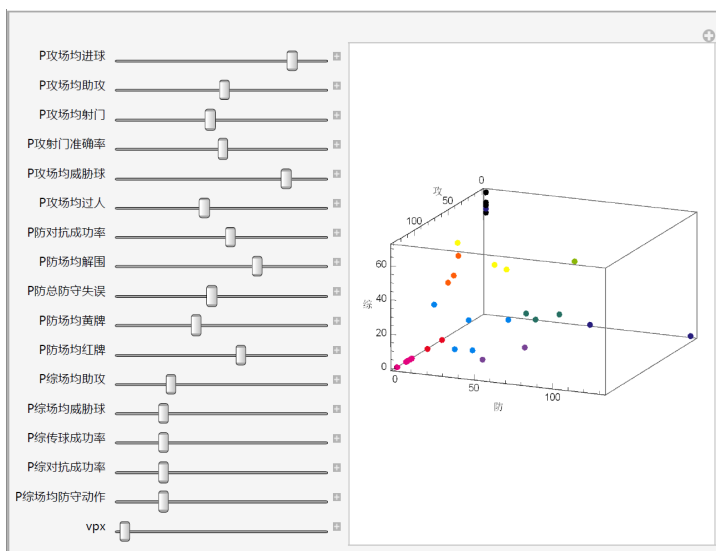


Figure 3. Three dimensional outer status lattice simulation effect diagram
图 3. 表状态三维点阵模拟效果图

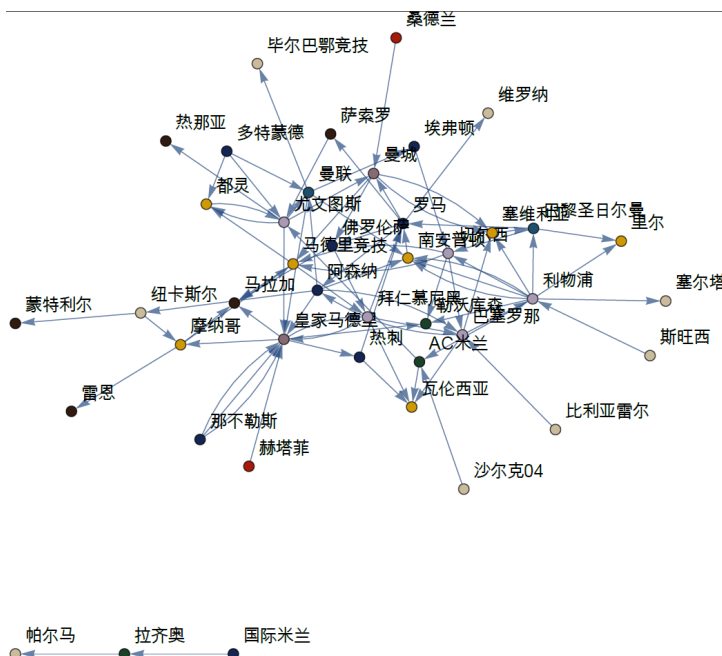


Figure 4. Transfer effect graph
图 4. 转会效果图

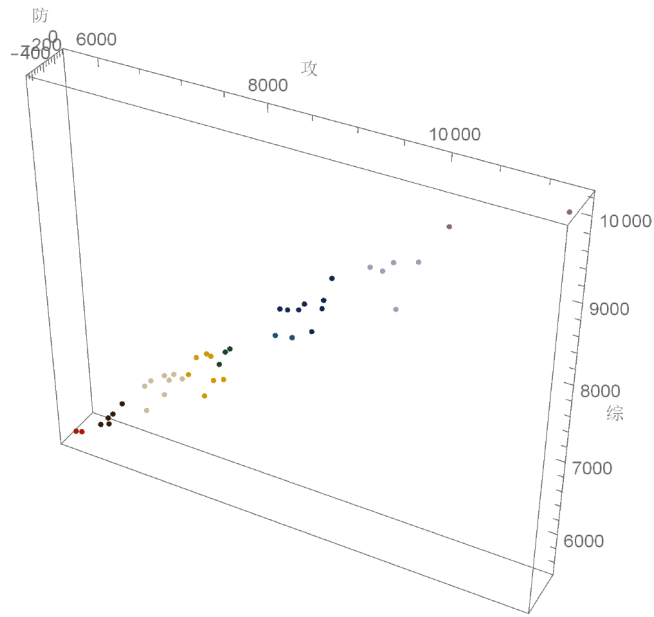


Figure 5. Simulation effect graph of team rating
图 5. 球队评级模拟效果图

在得到里状态的分布后，我们希望了解里状态间互相转移的机制，也就是里状态的转移矩阵。矩阵如下：

$$\begin{bmatrix}
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
 0 & \frac{1}{5} & 0 & \frac{1}{5} & \frac{1}{5} & 0 & 0 & \frac{2}{5} & 0 \\
 0 & \frac{1}{3} & 0 & \frac{2}{9} & 0 & \frac{1}{9} & \frac{1}{9} & \frac{2}{9} & 0 \\
 0 & 0 & \frac{1}{3} & \frac{1}{3} & 0 & 0 & 0 & \frac{1}{3} & 0 \\
 0 & 0 & \frac{1}{7} & \frac{2}{7} & 0 & 0 & \frac{2}{7} & \frac{1}{7} & \frac{1}{7} \\
 0 & \frac{1}{11} & \frac{1}{11} & \frac{3}{22} & \frac{1}{22} & \frac{1}{11} & \frac{3}{22} & \frac{2}{11} & \frac{5}{22} \\
 0 & \frac{1}{27} & \frac{1}{27} & \frac{11}{27} & \frac{1}{9} & \frac{2}{27} & \frac{2}{27} & \frac{4}{27} & \frac{1}{9} \\
 0 & \frac{1}{9} & 0 & \frac{4}{9} & \frac{1}{9} & 0 & \frac{2}{9} & \frac{1}{9} & 0
 \end{bmatrix}$$

这是一个 9×9 的矩阵，其中第 i 行第 j 列表示，对于一支里状态为 i 的球队，卖出球员时买家里状态为 j 的概率。如第三行第二列为 $1/5$ ，里状态为 3 的球队共有五支，分别为英超球队纽卡斯尔和斯旺西，西甲球队比利亚雷尔和毕尔巴鄂竞技，德甲球队沙尔克 04，其中只有纽卡斯尔将球队球员卖给了里状态为 2 的球队。

通过观察矩阵我们发现，分类为 1、2 的球队极少卖出球员。这是由于在模型中，分类为 1 和 2 的球队在综合表现和进攻上均处于一种较弱的位置，而我们的样本取自欧洲五大联赛，因此，分类为 1、2 的球队在联赛中处于一种排名倒数的地位，其所拥有的球员很难达到其他球队的能力要求，故很难将球员卖给模型中的其他球队。分类为 7、8 的球队拥有最多的买卖记录，这是因为处于这两个分类的球队大多

在各自联赛中处于上半分段的球队，需要频繁地买卖球员更新阵容以保持竞争力，并且其球员也能符合大多数同类别球队的能力要求，故转会相对频繁。

在这里，里状态转移矩阵描述的就是某类球队卖出球员时，买家球队的类别的分布。

我们以几乎同样的方法对球队的表状态进行了处理，所不同的是我们选取决定里状态的量是球队在过去三个赛季中的各项数据所取的加权平均，而我们认为球队购买的球员能够准确地决定球队目前的表状态，故决定表状态的是球队所购买的球员在被买入前一个赛季中的数据。同样用一个三维向量来表示球队的表状态，并且在一个空间直角坐标系中表示出来，如图 6 所示。

其中同种颜色的点表示点所对应的球队是同一种类。图中共分了十类点，也就是表状态共有十个。结合之前已经计算出的球队里状态，我们能得出球队的发射矩阵：

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ \frac{1}{2} & \frac{1}{4} & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{4} & 0 \\ 0 & 0 & \frac{1}{5} & 0 & 0 & \frac{1}{5} & 0 & 0 & \frac{1}{5} & \frac{2}{5} \\ \frac{1}{3} & \frac{1}{3} & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{3} & 0 \\ 0 & 0 & \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{8} & \frac{1}{8} & 0 & 0 & \frac{1}{8} & \frac{1}{4} & 0 & \frac{1}{8} & \frac{1}{4} & 0 \\ 0 & 0 & 0 & \frac{1}{5} & \frac{2}{5} & \frac{2}{5} & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 & \frac{1}{2} & 0 & 0 \end{bmatrix}$$

这是一个 9×10 的矩阵，第 i 行 j 列表示已知里状态为 i 时，表状态为 j 的概率。如第一行第七列是

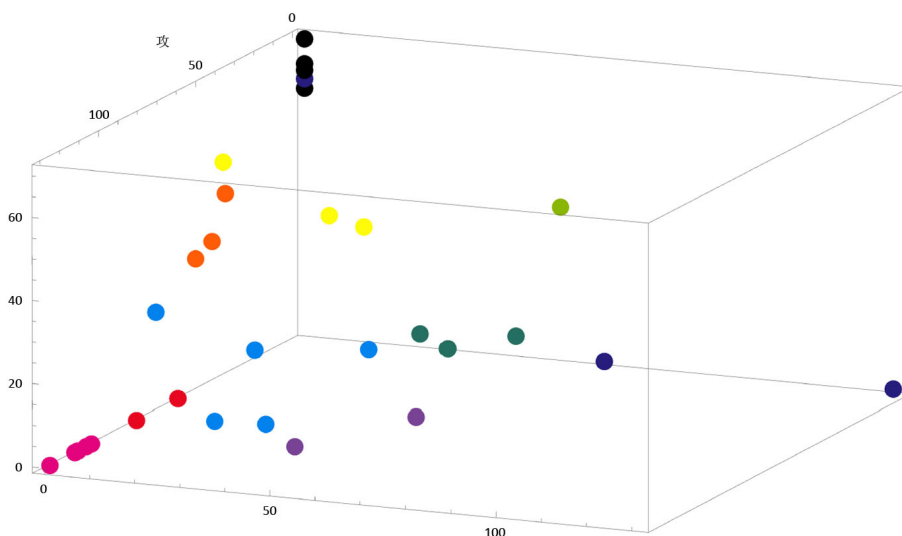


Figure 6. Outer status distribution with weighted value
图 6. 加权平均后的球队表状态分布

1/2, 意思是由里状态 1 表现为表状态 7 的概率是 1/2, 结合实际情况, 里状态为 1 的球队有两只, 分别是表状态为 7 的英超球队桑德兰和表状态为 9 的西甲球队赫塔菲, 与我们的理论相吻合。

观察矩阵易知, 表状态为 9 的球队最多, 而 9 是代表买入了综合型球员如中场, 这代表了**当今足坛最稀缺的球员类型是中场组织者**, 相较于十年前靠中锋打开局面的情况, 当今足坛的风格随着第五次足球革命发生了翻天覆地的变化, 球队更多选择靠中场球员控球来寻找机会打开局面。表状态数量排在 9 后则是标准的进攻型球员即中锋, 即使经历了如此大的变革, 中锋的终结能力仍然是每个球队所需要的。同时观察图表可以发现, 大多数球队的点处于坐标平面上, 也就是说, **大多数球队只购买了单一类型或某两种类型的球员**, 这是由于对于大多数球队财政预算不足, 只能将转会经费花在球队最需要的地方。反之像一些大球队比如拜仁慕尼黑, 其所对应的点并不在任何坐标平面上, 是因为其球队财政预算充裕, 除开急需补强的位置, 仍有余力对其他位置进行一些小的修缮, 使球队保持强大的实力。

4. 主要结论

本文主要运用隐 Markov 模型的概念刻画了 2012 夏季转会期开始到 2014 夏季转会期间欧洲足球运动员转会的各类因素, 并基于这些因素研究其中的潜在规律。根据模型, 我们实现了对球员和球队属性的划分, 所设计的程序可以在一定程度上实现依据前一时间段的数据对球员下一阶段转会情况的预测。总体来说, 对任意符合相关性检验的数据, 我们都可以实现“计算出其转移和发射矩阵, 并直接输出相关球队属性和球员转会倾向, 实现未来转会情况的预估”这样的过程。

通过对现在欧洲足球市场数据的实证分析, 我们也得出了以下结论:

- 1) 相对于防守类型来说, 进攻类型是一个更好地区分球队的方式。
 - 2) 当今足坛最稀缺的球员类型是中场组织者, 随着足球运动员技能的不断成熟, 较大的控球率显然更为主流球队所热衷。
 - 3) 排名越靠前的球队之间球员的转会越是频繁, 在联赛中一直处于倒数排名的球队, 很难将球员卖给其他球队。
 - 4) 限于经费的原因, 球队往往只购买单一类型或两种类型的球员来使球队实力实现最大效益的提升。
- 总体而言, 用隐 Markov 模型来研究足球市场的转会机制是合理的, 我们的模型为这种受多种因素影响而呈现复杂表征的对象的研究提供了一种可能的方法, 该程序的实现也使得球队的分类更可视化, 能满足更多人群的使用需求。

致 谢

本文作者感谢本课题组项目组指导教师林建忠副教授的指导和大力支持!

基金项目

该项目为第十期上海交通大学大学生创新实践项目、上海市大学生创新实践项目。

参考文献 (References)

- [1] 王红娟. 中国职业足球运动员转会制度研究[D]: [硕士学位论文]. 北京: 中国政法大学, 2010: 2-4.
- [2] 吴育华, 杨顺元, 叶加宝. 中国、欧洲足球运动员转会制度分析[J]. 武汉体育学院学报, 2007, 41(9): 19-22.
- [3] 刘月花, 赵鹏. 社会学视角下欧洲足球运动员转会行为分析[J]. 吉林体育学院学报, 2012, 28(2): 34-36.
- [4] Heller, A. (1965) On Stochastic Processes Derived from Markov Chain. *The Annals of Mathematical Statistics*, **36**, 1286-1291. <http://dx.doi.org/10.1214/aoms/1177700000>
- [5] Leroux, B.G. (1992) Maximum-Likelihood Estimation for Hidden Markov Models. *Stochastic Process and Their Ap-*

plications, **40**, 127-135. [http://dx.doi.org/10.1016/0304-4149\(92\)90141-C](http://dx.doi.org/10.1016/0304-4149(92)90141-C)

- [6] Leroux, B.G. and Puterman, M.L. (1992) Maximum-Penalized-Likelihood Estimation for Independent and Markov-Dependent Mixture Model. *Biometrics*, **48**, 545-558. <http://dx.doi.org/10.2307/2532308>
- [7] 胡淑兰. 隐藏马尔科夫模型及其在金融中的应用[M]. 湖北: 华中师范大学出版社, 2014: 37-75, 159-170.
- [8] 龚光路, 钱敏平. 应用随机过程教程及其在算法与智能计算中的应用[M]. 北京: 清华大学出版社, 2003: 96-283.