

6. 国足在亚洲属于几流？ - 1stOpt 聚类工具箱案例应用介绍

近日中国足协主席陈戌源在上海召开的中国足球职业联赛专项治理工作会议上怒斥“我们俱乐部的投入，是 J 联赛俱乐部的三倍多，是 K 联赛的 10 倍多，球员薪酬是 J 联赛的 5.8 倍，是 K 联赛的 11.7 倍，这些数字触目惊心。金元足球吞噬健康的足球肌体，现在有些人还不觉醒，难道良心已死吗？”

国足高薪低能早已耳闻，但球员薪酬是日本的 5.8 倍，韩国的 11.7 倍，还是足够震撼。

国足水平低是不争的事实，那在亚洲属于几流水平？如何科学的评判？下面以亚洲 20 国足球三次排名成绩，基于 1stOpt 的聚类工具箱进行聚类分析演示，科学地评判国足在亚洲的地位。

6.1 亚洲 20 国足球成绩数据

表 6.1 国足数据

样本编号	国家	2019 国际排名	2018 世界杯排名	2015 亚洲杯排名
1	中国	73	40	7
2	日本	60	15	5
3	韩国	61	19	2
4	伊朗	34	18	6
5	沙特	67	26	10
6	伊拉克	91	40	4
7	卡塔尔	101	40	13
8	阿联酋	81	40	6
9	乌兹别克斯坦	88	40	8
10	泰国	122	40	17
11	越南	102	50	17
12	阿曼	87	50	12
13	巴林	116	50	11
14	朝鲜	110	50	14
15	印尼	164	50	17
16	澳洲	40	30	1
17	叙利亚	76	40	17
18	约旦	118	50	9
19	科威特	160	50	15
20	巴勒斯坦	96	50	16

6.2 1stOpt 聚类工具箱

1stOpt 聚类工具箱的聚类算法包括：K 均值算法（K-Mean），模糊 C 均值算法（Fuzzy C-Means），基于优化的 K 均值算法，基于优化的模糊 C 均值算法，以及建立在投影寻踪技术上的前述四种算法，使用非常简单。



图 6-1 聚类工具箱设定界面

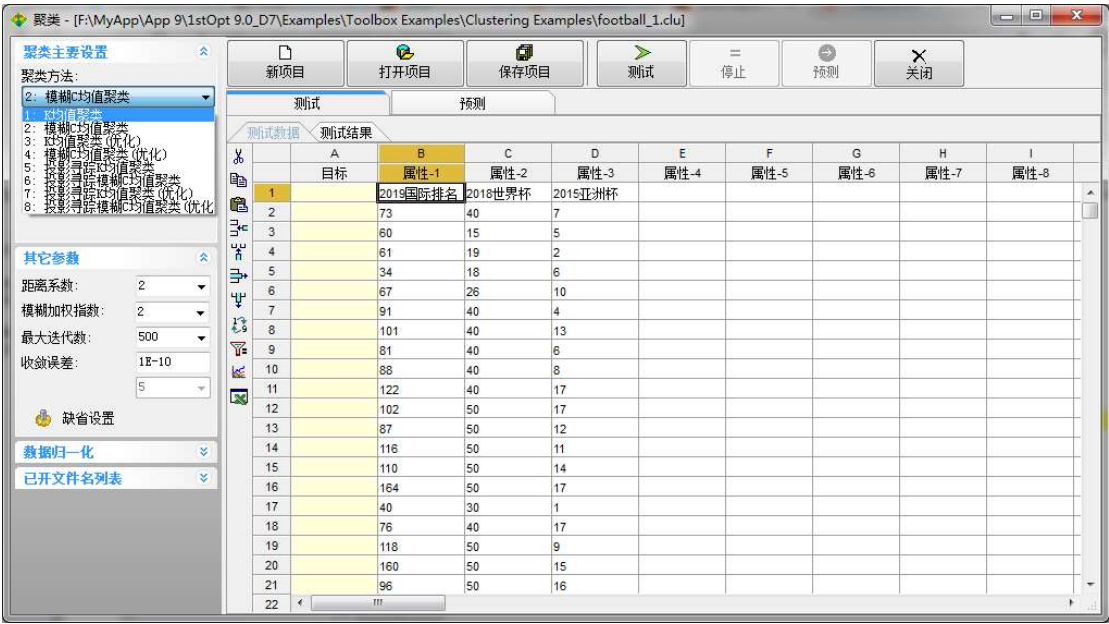


图 6-2 聚类工具箱数据录入界面

如上图，将相关数据输入，分类数设定为 3（即将 20 过聚类成 3 类），算法可先选择 K-Mean 或 Fuzzy C-Mean，很容易得到下列结果

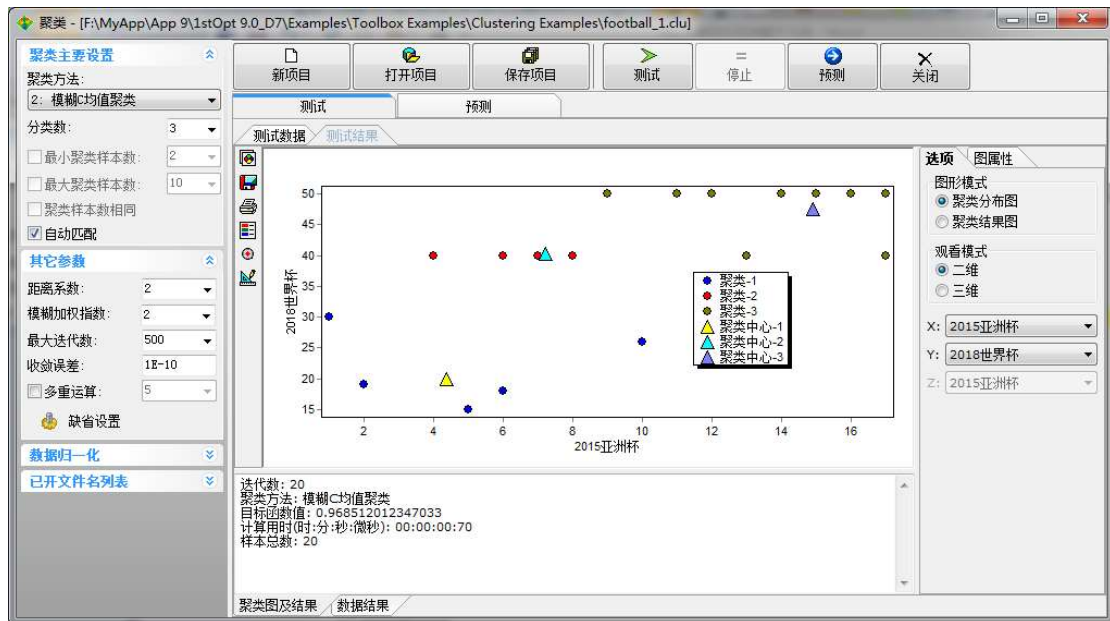


图 6-3 聚类结果二维图形

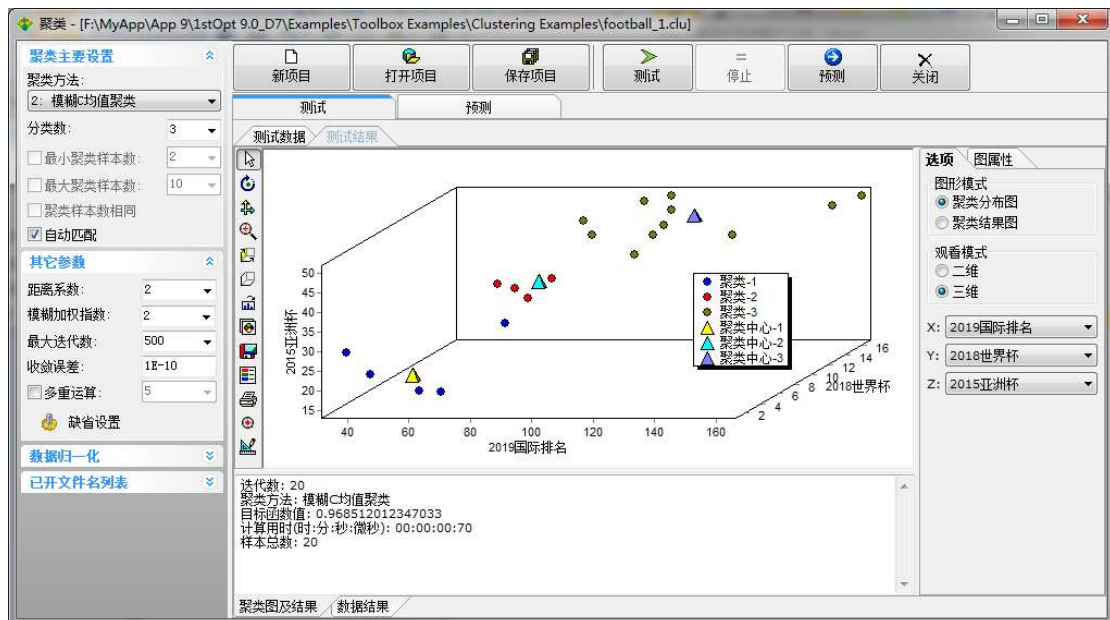


图 6-4 聚类结果三维图形

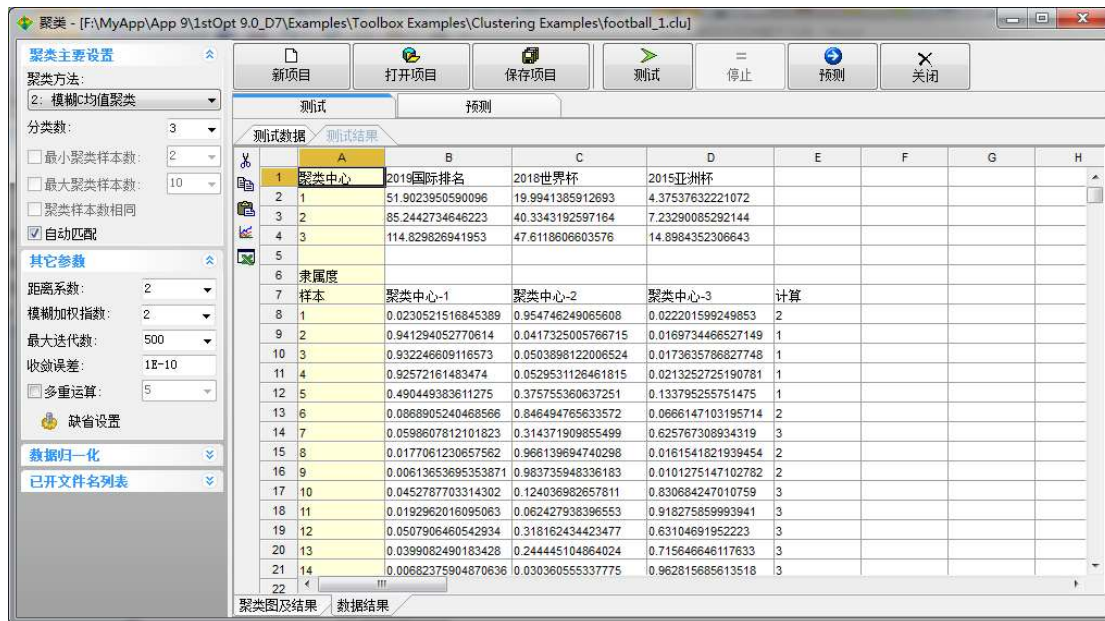


图 6-5 聚类数值结果

表 6.2 聚类计算结果

样本编号	计算类别标签号	对应国家
1	2	中国
2	1	日本
3	1	韩国
4	1	伊朗
5	1	沙特
6	2	伊拉克
7	3	卡塔尔
8	2	阿联酋
9	2	乌兹别克斯坦
10	3	泰国
11	3	越南
12	3	阿曼
13	3	巴林
14	3	朝鲜
15	3	印尼
16	1	澳洲
17	3	叙利亚
18	3	约旦
19	3	科威特
20	3	巴勒斯坦

根据上表计算结果，20 国聚类如下：

- 第一类：沙特，日本，韩国，伊朗，澳洲
- 第二类： 中国，伊拉克，阿联酋，乌兹别克斯坦
- 第三类：泰国，越南，阿曼，巴林，朝鲜，印尼，卡塔尔，叙利亚，约旦，科威特，巴勒斯坦

根据“物以类聚”，上述计算结果表明国足在亚洲也只能是二流水平。

6.3 聚类工具箱使用注意事项

标准 K-Mean 算法容易陷入局部解，结果不稳定，使用时可多计算几次，选择目标函数小的作为最终结果；Fuzzy C-mean 算法结果相对要稳定，比 K-Mean 算法更值得推荐使用；而基于优化算法的 K-Mean 和 Fuzzy C-Mean，效果都好且稳定，此外还可以设定最小聚类数等，应用场景更多。

另外注意聚类与分类的区别。聚类是事先不知道类别标签，只把属性相近的归为一类并赋予一标签值，该标签值只是一个类别符号，不存在好坏或高低之分，比如上述聚类计算，每一次的计算类别标签号有可能是不一样的：中国，伊拉克，阿联酋，乌兹别克斯坦四国的聚类标签号有可能是 1，也有可能是 2 或 3，其它国家的聚类标签号也相应不同，比如某一组结果如下：

- 第一类：中国，伊拉克，阿联酋，乌兹别克斯坦
- 第二类：泰国，越南，阿曼，巴林，朝鲜，印尼，卡塔尔，叙利亚，约旦，科威特，巴勒斯坦
- 第三类：沙特，日本，韩国，伊朗，澳洲

上述结果就能说明国足为一流吗？非也，只是说明国足的聚类标签是一类。那如何判断国足属于几流？同理，根据物以类聚，这时就要看大家的普遍共识了：沙特，日本，韩国，伊朗和澳洲都是大家公认的亚洲足球强国，此时的第三类自然是属于一流，国足所处的第一类最多也就是二流。

6.4 结论

数据不会说谎，国足在亚洲最好也就二流，这也符合“大众常识”。