

# The Analysis of Heroes in DOTA2 Games Based on Multivariate Statistics

Dongyang Peng, Xiao Han\*, Xinghu Teng

Army University of Engineering, Nanjing Jiangsu  
Email: [h.x.good@163.com](mailto:h.x.good@163.com)

Received: Feb. 24<sup>th</sup>, 2020; accepted: Mar. 18<sup>th</sup>, 2020; published: Mar. 25<sup>th</sup>, 2020

---

## Abstract

Dota2 is a popular moba game, through the control of the game characters and cooperation with each other to win the game. It is popular with gamers. In this paper, we select 116 heroes in the game and 9 relevant indicators as samples and use SPSS to analyze the hero data. Firstly, hierarchical cluster method is used to analyze the degree of correlation among indicators, then the indicators are divided into two categories. After that, heroes are clustered by k-means under different indexes meanwhile the characteristics of some of the heroes are analyzed. Finally, the method of factor analysis is used to find two main factors influencing 9 indicators.

## Keywords

DOTA2 Hero, SPSS, Hierarchical Clustering, K-Means Clustering, Factor Analysis

---

# 基于多元统计的DOTA2游戏中英雄的分析

彭冬阳, 韩笑\*, 滕兴虎

陆军工程大学, 江苏 南京  
Email: [h.x.good@163.com](mailto:h.x.good@163.com)

收稿日期: 2020年2月24日; 录用日期: 2020年3月18日; 发布日期: 2020年3月25日

---

## 摘要

Dota2是一个热门的moba类游戏,通过对游戏角色的控制,互相间配合取得游戏的胜利,深受游戏玩家喜欢。本文选取了游戏中的116个英雄,9项相关指标作为样本,利用SPSS统计软件对英雄数据进行统计分析。首先,使用系统聚类法分析各个指标间的关联程度,将指标分为两大类。然后,在不同类指标下分别对英雄进行k均值聚类,并分析了其中一些英雄的特点。最后,使用因子分析的方法,找到两个影响9个指标的主因子。

\* 通讯作者。

## 关键词

DOTA2英雄, SPSS, 系统聚类, K-均值聚类, 因子分析

Copyright © 2020 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

Dota2 (Defense of the Ancients 2)是一个热门的 moba 类游戏, 双方各 5 人, 通过对游戏角色的控制, 互相配合取得游戏的胜利。Dota2 中可使用的英雄众多, 共 100 多位, 他们的技能、属性、定位等各不相同。本文试通过对这些英雄的数据进行分析, 来比较分析结果是否和我们对这些英雄的特点的认知相吻合。

本文数据来源于 max+网站, 获取了 Dota2 游戏中 116 个英雄的 9 个指标的数据, 部分样本数据如下

图 1:

	name	kda	胜率	出场次数	经济	经验	英雄伤害	建筑伤害	正补	反补
1	Anti-Mage	2.52	52	4055442	575.35	635.24	449.39	109.75	78.66	13.11
2	Axe	2.40	54	5630097	429.24	548.60	589.14	24.66	38.04	3.80
3	Bane	2.05	46	767818	278.77	436.41	266.82	11.74	9.23	5.48
4	Bloodseeker	2.55	52	3867957	500.39	593.57	612.81	74.94	50.69	12.66
5	Crystal Maiden	2.04	53	3671090	342.34	430.52	345.37	11.42	15.24	2.43
6	Drow Ranger	2.20	51	3781721	462.16	549.92	420.73	118.76	49.27	10.82
7	Earthshaker	2.46	51	6270128	356.64	486.69	433.32	16.82	23.64	4.01
8	Juggernaut	2.66	51	4439774	524.41	603.65	542.08	94.73	62.70	11.78
9	Mirana	3.30	51	4352379	422.50	513.69	451.08	50.60	37.90	7.12
10	Morphling	2.82	47	1917394	513.17	602.57	563.31	117.87	56.44	19.79
11	Shadow Fiend	2.49	45	3188205	520.59	615.70	599.01	81.03	62.14	19.99
12	Phantom Lancer	2.52	48	2622579	481.74	567.19	484.13	75.24	55.05	8.92
13	Puck	2.36	39	527579	437.99	538.37	489.70	25.09	38.05	7.89
14	Pudge	2.45	53	12117994	365.91	512.94	457.67	12.40	15.73	2.69
15	Razor	2.10	43	532306	422.56	514.37	525.98	49.78	40.41	10.46
16	Sand King	2.17	44	1003347	352.61	471.99	367.93	8.84	27.24	2.73
17	Storm Spirit	3.05	50	4214835	500.60	627.00	651.72	31.80	51.82	8.96
18	Sven	2.41	51	2052130	491.47	582.04	582.09	71.56	55.47	7.64
19	Tiny	2.56	43	1589341	414.45	520.63	531.79	53.46	36.50	9.80
20	Vengeful Spirit	2.28	54	2139738	330.71	448.94	388.53	28.67	18.06	5.90
21	Windranger	2.69	47	4267990	416.62	528.62	507.50	51.81	34.94	8.27
22	Zeus	3.70	55	4929606	418.48	576.19	919.58	10.57	33.89	4.61
23	Kunkka	2.71	48	2744614	447.64	554.91	523.11	36.14	44.92	7.63
24	Lina	2.12	42	2469755	455.77	556.96	528.90	61.82	44.69	6.11
25	Lion	1.85	48	3462002	306.03	419.07	258.10	9.58	10.57	3.46
26	Shadow Shaman	1.74	53	3735255	334.32	431.84	307.06	75.80	19.91	4.70
27	Slardar	2.22	47	1437103	346.26	477.47	298.09	20.09	24.01	6.54

Figure 1. Sample data

图 1. 样本数据

## 2. 描述性分析

对游戏中 116 个英雄的 9 个指标进行描述性统计分析。这 9 种指标分别为 KDA (Kill Death Assist), 胜率, 出场次数, 经济(每分钟), 经验(每分钟), 英雄伤害(每分钟), 建筑伤害(每分钟), 正补(每 10 分钟), 反补(整场), 分析的内容包括范围、最小值、最大值、均值和方差。如下表所示:

通过图 2 的结果可以得出以下结论:

1) KDA ( $KDA = (\text{杀敌数} + \text{助攻数}) / \text{死亡次数}$ )和胜率这两个指标的方差比较小, 最大值和最小值的差异也比较小。这是符合游戏规则的设定的, 因为要保证游戏的公平性, 所以游戏官方要保证每个英雄

的胜率都差不多，维持在 50%左右；也要保证每个英雄不会过于强势或劣势，所以各个英雄的 KDA 的方差也会很小。

2) 其余 7 个指标的方差还是较大的，这说明该游戏中不同英雄的差异性还是比较大的。这体现出了该游戏中英雄的多样性和趣味性。

Descriptive Statistics					
	N	Minimum	Maximum	Mean	Std. Deviation
KDA	116	1.74	3.70	2.4859	0.36186
胜率	116	0.38	0.56	0.4852	0.03904
出场次数	116	143931	12117994	2157623.59	1841914.110
经济	116	273.03	748.88	415.6407	78.72848
经验	116	401.88	815.30	518.2596	69.93268
英雄伤害	116	181.82	924.99	473.8532	136.20754
建筑伤害	116	3.71	163.16	48.9025	38.34499
正补	116	9.23	78.66	35.5781	16.69837
反补	116	0.79	21.49	7.3792	3.96296
Valid N (listwise)	116				

Figure 2. Descriptive statistics  
图 2. 描述统计

### 3. 系统聚类分析

聚类分析是一种建立分类的多元统计分析方法，能将一批样本或指标数据根据其特征，按照在性质上的亲疏程度在没有先验知识的情况下进行自动分类，产生多个分类结果。同类内部个体特征具有相似性，不同类间个体特征的差异性较大[1] [2]。

#### 3.1. 对指标的系统聚类

利用 SPSS(Statistical Product and Service Solutions)对英雄的各项指标进行聚类分析，以得出各项指标之间的相关关系，聚类结果如下图所示：

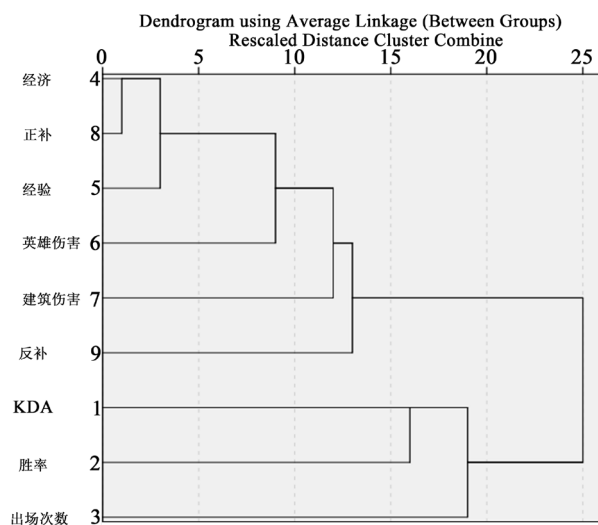


Figure 3. System clustering results  
图 3. 系统聚类结果

分析谱系图 3 中可以得到以下结论:

根据指标的相关性对指标进行分类可以分为两类, 第一类为: 经济、正补、经验、英雄伤害、建筑伤害、反补, 第二类为: KDA、胜率、出场次数。

第一大类可以看成是与玩家获取的金钱和经验直接相关的指标, 可以看成英雄发育程度, 但是不对游戏胜利有直接影响。

而第二大类中与胜率绑定在一起的有 KDA 和出场次数, 这个指标的高低可以反应在游戏中玩家所处的状态, KDA 越高对游戏局势支援程度越高, 获胜概率也就越大; 而出场次数和胜率间的关系说明, 英雄胜率越高, 玩家越愿意选择玩这个英雄, 也就是我们常说的版本热门英雄。

### 3.2. 对样本的 K-均值聚类

在系统聚类的基础上, 我们根据这两大类指标分别对英雄进行聚类。为了后面对一些英雄进行代表性的分析, 不妨选取 50 位英雄作为样本进行 K-均值聚类[3]。

第一大类指标下的英雄的聚类情况如下:

Number of Cases in each Cluster		
Cluster	1	2.000
	2	16.000
	3	10.000
	4	13.000
	5	9.000
Valid		50.000
Missing		0.000

Figure 4. Number of cases in each cluster  
图 4. 每个聚类中样本的个数

Cluster Membership				Cluster Membership			
Case Number	name	Cluster	Distance	Case Number	name	Cluster	Distance
1	Anti-Mage	5	126.204	26	Shadow Shaman	3	58.294
2	Axe	4	84.998	27	Slardar	3	46.442
3	Bane	3	75.446	28	Tidehunter	3	83.883
4	Bloodseeker	4	20.363	29	Witch Doctor	3	47.595
5	Crystal Maiden	3	32.429	30	Lich	2	97.005
6	Drow Ranger	5	80.773	31	Riki	2	75.205
7	Earthshaker	2	71.804	32	Enigma	2	59.598
8	Juggernaut	4	75.976	33	Tinker	1	48.743
9	Mirana	2	52.999	34	Sniper	4	74.450
10	Morphling	4	68.318	35	Necrophos	4	69.775
11	Shadow Fiend	4	48.171	36	Warlock	2	95.887
12	Phantom Lancer	5	9.361	37	Beastmaster	2	99.092
13	Puck	2	62.512	38	Queen of Pain	4	42.649
14	Pudge	2	53.107	39	Venomancer	2	116.726
15	Razor	2	51.998	40	Faceless Void	5	41.423
16	Sand King	3	60.919	41	Wraith King	5	31.209
17	Storm Spirit	4	78.249	42	Death Prophet	2	28.653
18	Sven	4	19.333	43	Phantom Assassin	4	50.639
19	Tiny	2	54.343	44	Pugna	2	108.895
20	Vengeful Spirit	3	66.103	45	Templar Assassin	5	57.217
21	Windranger	2	43.878	46	Viper	4	71.941
22	Zeus	1	48.743	47	Luna	4	100.600
23	Kunkka	5	69.064	48	Dragon Knight	5	38.435
24	Lina	5	55.992	49	Dazzle	3	73.000
25	Lion	3	74.071	50	Clockwerk	2	74.452

Figure 5. Cluster member  
图 5. 聚类成员

分析上图 4 和图 5，可以得到以下结论：

1) 第二至第五这四类中的英雄的共性还是比较明显的，比如第三类中各个指标值都是相对较低的，通过分析其中的英雄可以发现，第三类中的英雄大都是打 5 号位的较多，可以理解为辅助。这些英雄在对局中很少拿经济，对敌方的伤害也较低，主要提供对团队的保护作用。

2) 第四类中各个指标值都是相对较高的，通过分析其中的英雄可以发现，第四类中的英雄大都是打 1 号位的较多，可以理解为主力输出。这些英雄在对局中拿经济最多的经济，通常是核心输出位置。

3) 第一类中只有两个英雄，它们的特点是对英雄的伤害极高，但对建筑的伤害很低。这两个英雄为 Zous 和 Tinker，他们的主要特点都是在队伍中扮演着强力输出以及 Ganker (指负责偷袭抓人的队友) 的角色，属于那种到处游走拿人头的人。所以，他们经常会对方英雄造成巨额伤害，却又很少进行推塔。

4) 通过分析可以看出，根据第一类的指标对英雄进行 K-均值聚类的结果和实际游戏中英雄的特点是吻合的。

第二大类指标下的英雄的聚类情况如下：

Number of Cases in each Cluster		
Cluster	1	5.000
	2	17.000
	3	12.000
	4	1.000
	5	15.000
Valid		50.000
Missing		0.000

Figure 6. Number of cases in each cluster

图 6. 每个聚类中样本的个数

Cluster Membership				Cluster Membership			
Case Number	name	Cluster	Distance	Case Number	name	Cluster	Distance
1	Anti-Mage	5	82014.267	26	Shadow Shaman	5	238172.733
2	Axe	1	482591.200	27	Slardar	2	530205.118
3	Bane	3	8786.417	28	Tidehunter	3	115262.583
4	Bloodseeker	5	105470.733	29	Witch Doctor	2	81016.118
5	Crystal Maiden	5	302337.733	30	Lich	2	15345.118
6	Drow Ranger	5	191706.733	31	Riki	1	291807.200
7	Earthshaker	1	157439.800	32	Enigma	3	255481.583
8	Juggernaut	5	466346.267	33	Tinker	5	154691.733
9	Mirana	5	378951.267	34	Sniper	1	345437.800
10	Morphling	2	49914.118	35	Necrophos	5	67841.733
11	Shadow Fiend	5	785222.733	36	Warlock	3	231820.583
12	Phantom Lancer	2	655270.882	37	Beastmaster	3	516190.417
13	Puck	3	249025.417	38	Queen of Pain	2	420525.118
14	Pudge	4	0.000	39	Venomancer	2	536649.118
15	Razor	3	244298.417	40	Faceless Void	5	62589.733
16	Sand King	3	226742.583	41	Wraith King	2	976829.882
17	Storm Spirit	5	241407.267	42	Death Prophet	3	451133.417
18	Sven	2	84821.882	43	Phantom Assassin	1	271520.800
19	Tiny	2	377967.118	44	Pugna	2	535474.118
20	Vengeful Spirit	2	172429.882	45	Templar Assassin	3	196248.583
21	Windranger	5	294562.267	46	Viper	2	130294.118
22	Zeus	5	956178.267	47	Luna	2	47070.882
23	Kunkka	2	777305.882	48	Dragon Knight	2	538786.118
24	Lina	2	502446.882	49	Dazzle	3	287275.583
25	Lion	5	511425.733	50	Clockwerk	3	156602.583

Figure 7. Cluster member

图 7. 聚类成员

分析上图 6 和图 7, 可以得到以下结论:

1) 第一类中各个指标值都是相对较高的, 通过分析其中的英雄可以发现, 第一类中的英雄大都是那种版本强势的英雄。在经过游戏官方的版本更新后, 这些英雄的某些属性得到提高, 因此有着较高的胜率和出场率。

2) 第三类中各个指标值都是最低的, 通过分析其中的英雄可以发现, 第四类中的英雄都是版本弱势英雄。因此, 拥有最低的 KDA, 胜率和出场率就是很合理的了。

3) 第四类中只有一个英雄 Gudge, 它的特点很明显, 有着极高的出场率, 而且胜率也不低。这与该英雄的特性有关: Gudge 的特有神技钩子决定了他的高出场率, 它最大的作用在于战场的分割与救人, 对团战胜败有极高的决定作用。其次, 这个技能有很强的娱乐性, 适合各个级别的玩家使用。

4) 通过分析可以看出, 根据第二类的指标对英雄进行 K 均值聚类的结果是也是符合游戏实际的。

#### 4. 因子分析

从上文的聚类分析结果来看, 所选取的 9 个指标存在一定的相关性, 下面进行因子分析, 提取影响 9 个指标值的公共因子, 我们称为主因子。SPSS 分析数据结果如下:

		Correlation Matrix								
		KDA	胜率	出场次数	经济	经验	英雄伤害	建筑伤害	正补	反补
Correlation	KDA	1.000	0.473	0.336	0.344	0.461	0.498	0.054	0.288	0.328
	胜率	0.473	1.000	0.448	0.072	0.154	0.184	0.085	-0.022	-0.004
	出场次数	0.336	0.448	1.000	0.196	0.281	0.347	-0.001	0.115	0.139
	经济	0.344	0.072	0.196	1.000	0.860	0.695	0.682	0.932	0.584
	经验	0.461	0.154	0.281	0.860	1.000	0.722	0.616	0.837	0.591
	英雄伤害	0.498	0.184	0.347	0.695	0.722	1.000	0.319	0.630	0.515
	建筑伤害	0.054	0.085	-0.001	0.682	0.616	0.319	1.000	0.735	0.501
	正补	0.288	-0.022	0.115	0.932	0.837	0.630	0.735	1.000	0.619
	反补	0.328	-0.004	0.139	0.584	0.591	0.515	0.501	0.619	1.000

Figure 8. Correlation matrix

图 8. 相关系数矩阵

由上图 8 可以得出结论: 各指标之间还是有比较大的相关性的, 比如经济 and 正补, 经济和经验。

KMO and Bartlett's Test		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		0.811
Bartlett's Test of Sphericity	Approx. Chi-Square	770.378
	df	36
	Sig.	0.000

Figure 9. KMO and Bartlett's test

图 9. KMO 和 Bartlett 半球检验

由上图 9 可以得出结论: KMO 的值为 0.811, 大于阈值 0.5, 所以说明了变量之间是存在相关性的, 符合要求; Bartlett 球形检验的结果, 在这里只需要看显著性这一项, 其值为 0.000, 小于 0.05。两个结果均说明了该数据是可以进行因子分析的。本文主要采取主成分法提取公共因子。

由下图 10 可以得出结论: 前 2 个主成分对变量的表达已达到了 70% 左右, 可作为影响各指标的公共因子, 完全可以用于后面相关的分析。

Total Variance Explained									
Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	4.625	51.388	51.388	4.625	51.388	51.388	4.278	47.537	47.537
2	1.749	19.438	70.826	1.749	19.438	70.826	2.096	23.289	70.826
3	0.755	8.392	79.218						
4	0.654	7.263	86.482						
5	0.519	5.762	92.244						
6	0.314	3.492	95.736						
7	0.178	1.982	97.718						
8	0.149	1.653	99.371						
9	0.057	0.629	100.000						

Figure 10. Each principal component explains the total variance of the original variable  
图 10. 各主成分解释原始变量总方差的情况

Rotated Component Matrix <sup>a</sup>		
	Component	
	1	2
KDA	0.282	0.744
胜率	-0.064	0.800
出场次数	0.073	0.753
经济	0.932	0.150
经验	0.879	0.296
英雄伤害	0.686	0.448
建筑伤害	0.789	-0.117
正补	0.956	0.033
反补	0.735	0.103

Figure 11. Rotated component matrix  
图 11. 旋转后的载荷矩阵

由上图 11 可以得到结论:

1) 在第一主因子  $F_1$  主要由经济、经验、英雄伤害、建筑伤害、反补、正补 6 个指标决定。这 6 个指标代表了英雄在游戏中的发育情况，在游戏中英雄绝大多数的行为都与这两个系数直接相关，对敌方单位和中立单位的击杀都能够同时带来经济和经验的提高，这两个指标越高也能代表英雄在游戏中获得的资源越多；总体来说，第一主成分代表了英雄在游戏中的发育情况。

2) 在第二主因子  $F_2$  主要由 KDA、胜率和出场次数决定，代表了英雄的强势和弱势情况，以及玩家们对强势英雄更加热衷的状态。总体来说，第二主成分代表了英雄在游戏版本的强弱势情况。

## 5. 总结

1) 从描述性分析中可以得出，为保证游戏的公平性，游戏官方要保证每个英雄的胜率都差不多，维持在 50% 左右；也要保证每个英雄不会过于强势或劣势，所以各个英雄的 KDA 的方差也会很小。

2) 通过聚类分析，将英雄的 9 个指标划分为两个类别。第一类是游戏中英雄的发育程度相关指标。第二类是与游戏胜率有关的指标，反映了英雄的强势弱势情况。

- 
- 3) 通过  $K$  均值聚类分析的结果与实际中的英雄特性的比较, 分析出  $K$  均值聚类的结果是比较合理的。
  - 4) 通过因子分析, 得出两个比较具有代表性的主因子: 经济和胜率。这也与之前系统聚类分析得到的结果相是一致的。

### 基金项目

陆军工程大学基础学科培育基金项目(KYJBJQZL1922); 陆军工程大学基础学科培育基金项目(KYJBJQZL1921)。

### 参考文献

- [1] 何晓群. 多元统计分析[M]. 北京: 中国人民大学出版社, 2012.
- [2] 薛毅, 陈立萍. 统计建模与 R 软件[M]. 北京: 清华大学出版社, 2007.
- [3] 杨虎, 杨玥含. 金融大数据实证分析[M]. 北京: 科学出版社, 2018.