

# 基于大数据的成都大学生网络游戏画像研究

覃亿淦 杨 杉

四川大学锦城学院计算机与软件学院 四川 成都 611731

**【摘要】**以四川省的10所高校的在校大学生为研究对象,对从这10所高校进行问卷调查的到的数据做画像研究。采用spss statistics,spss modeler,对不同玩家的学校层次,专业类型,玩家接触网络游戏的年限,喜欢的游戏类型,每天游玩游戏的时间,每月在游戏消费情况,在电脑外设方面的消费情况进行了画像。画像结果能直观看成都在校大学生喜欢的游戏类型,所认为的网络游戏各方面的优点和缺点,游玩时间和消费程度等各项基本情况,从而让游戏公司更有针对性的进行改善或新游戏的开发。

**【关键词】**大数据;网络学习;画像研究;行为研究

## 1 引言

2020年以来,在线娱乐需求受疫情影响迎来增长,游戏业逆势“红火”。在“宅经济”的推动下,各类手游迎来了新一轮集中爆发;移动游戏市场收入规模、用户规模都有了较大的增长幅度。《2020年中国游戏产业报告》显示,2020年,国内游戏市场实际营销总额为2786.87亿元,同比增长20.71%;国内游戏用户规模达6.65亿,同比增长3.7%。整体而言,2020年国内游戏走向规范化、精品化,更加注重研发、创新;电竞游戏、云游戏等迅速升温,逐渐在游戏市场占据重要地位。此外,游戏厂商则纷纷争相发展海外市场,正在加速抢占全球市场份额。

国产游戏加速出海,新秀崭露头角。随着市面上的游戏越来越多,玩家们对于游戏也是越来越挑剔。从问卷调查中得来的关于大学生玩家的游戏行为数据进行画像,可以让各大游戏厂商了解大学生们对游戏的偏好,从而更有针对性的进行新游戏的开发。也能够观察现有的游戏的优点和缺点,对游戏的机制或者画面,进行更加偏向玩家爱好的改良和提高。

## 2 研究思路

以四川省的10所高校的在校大学生为研究对象,对从这10所高校进行问卷调查的到的数据做画像研究。采用spss statistics,spss modeler,对不同玩家的学校层次,专业类型,玩家接触网络游戏的年限,喜欢

的游戏类型,认为游戏的优点或缺点,每天游玩游戏的时长,每月在游戏消费情况,在电脑外设方面的消费情况进行了画像。

## 3 数据说明

### 3.1 数据来源

数据来源于四川省10所高校进行的问卷调查,覆盖了各种学校层次,专业类型游戏行为数据,数据表中包含了性别,学校层次,专业类型,接触网络游戏的年限,认为游戏的缺点和优点,每天游玩游戏的时长,每月在游戏消费,在游戏外设方面的消费等51个字段共1000条数据。

### 3.2 数据清洗

数据表中,很少的数据对于我们研究大学生网络游戏行为画像时是没有用的,因此对数据进行了数据预处理[10],筛选出了我们需要的数据,把不需要的数据进行了剔除,其中主要对填写网络问卷的时间过长或者过短进行筛选。删除无效字段:该字段对分析帮助和影响不大。删除异常记录:问卷多选题选择的选项与前面已填写的选项有冲突。

## 4 学生画像

### 4.1 不同玩网络游戏年限情况下每月游戏消费和游戏外设消费描述统计

#### 4.1.1 描述

2-4年	两年内	8年以上资深玩家						
平均值 <sup>1</sup>	3.26 <sup>1</sup>	.062 <sup>1</sup>	平均值 <sup>1</sup>	3.30 <sup>1</sup>	.068 <sup>1</sup>	平均值 <sup>1</sup>	2.54 <sup>1</sup>	.214 <sup>1</sup>
平均值的95%置信区间 <sup>2</sup>	3.13 <sup>1</sup>		平均值的95%置信区间 <sup>2</sup>	3.16 <sup>1</sup>		平均值的95%置信区间 <sup>2</sup>	2.11 <sup>1</sup>	
下限 <sup>3</sup>	3.38 <sup>1</sup>		下限 <sup>3</sup>	3.43 <sup>1</sup>		下限 <sup>3</sup>	2.98 <sup>1</sup>	
上限 <sup>4</sup>	3.25 <sup>1</sup>		上限 <sup>4</sup>	3.31 <sup>1</sup>		上限 <sup>4</sup>	2.52 <sup>1</sup>	
5%剔除后平均值 <sup>5</sup>	3.00 <sup>1</sup>		5%剔除后平均值 <sup>5</sup>	3.00 <sup>1</sup>		5%剔除后平均值 <sup>5</sup>	2.00 <sup>1</sup>	
中位数 <sup>6</sup>	.774 <sup>1</sup>		中位数 <sup>6</sup>	.781 <sup>1</sup>		中位数 <sup>6</sup>	1.608 <sup>1</sup>	
方差 <sup>7</sup>	.880 <sup>1</sup>		方差 <sup>7</sup>	.884 <sup>1</sup>		方差 <sup>7</sup>	1.269 <sup>1</sup>	
标准差 <sup>8</sup>	1 <sup>1</sup>		标准差 <sup>8</sup>	1 <sup>1</sup>		标准差 <sup>8</sup>	1 <sup>1</sup>	
最小值 <sup>9</sup>	5 <sup>1</sup>		最小值 <sup>9</sup>	5 <sup>1</sup>		最小值 <sup>9</sup>	5 <sup>1</sup>	
最大值 <sup>10</sup>	4 <sup>1</sup>		最大值 <sup>10</sup>	4 <sup>1</sup>		最大值 <sup>10</sup>	4 <sup>1</sup>	
全距 <sup>11</sup>	1 <sup>1</sup>		全距 <sup>11</sup>	1 <sup>1</sup>		全距 <sup>11</sup>	3 <sup>1</sup>	
四分位数 <sup>12</sup>	.238 <sup>1</sup>	.170 <sup>1</sup>	四分位数 <sup>12</sup>	.010 <sup>1</sup>	.187 <sup>1</sup>	四分位数 <sup>12</sup>	.128 <sup>1</sup>	.396 <sup>1</sup>
偏度 <sup>13</sup>	.185 <sup>1</sup>	.342 <sup>1</sup>	偏度 <sup>13</sup>	.368 <sup>1</sup>	.371 <sup>1</sup>	偏度 <sup>13</sup>	-1.396 <sup>1</sup>	.778 <sup>1</sup>
峰度 <sup>14</sup>			峰度 <sup>14</sup>			峰度 <sup>14</sup>		

4-6年	2-8年	两年内			
平均值 <sup>1</sup>	3.23 <sup>1</sup>	.074 <sup>1</sup>	平均值 <sup>1</sup>	2.76	.076
平均值的95%置信区间 <sup>2</sup>	3.14 <sup>1</sup>		平均值的95%置信区间 <sup>2</sup>	2.61	
下限 <sup>3</sup>	3.33 <sup>1</sup>		下限 <sup>3</sup>	2.91	
上限 <sup>4</sup>	3.22 <sup>1</sup>		上限 <sup>4</sup>	2.79	
5%剔除后平均值 <sup>5</sup>	3.00 <sup>1</sup>		5%剔除后平均值 <sup>5</sup>	3.00	
中位数 <sup>6</sup>	.676 <sup>1</sup>		中位数 <sup>6</sup>	.982	
方差 <sup>7</sup>	.822 <sup>1</sup>		方差 <sup>7</sup>	.991	
标准差 <sup>8</sup>	1 <sup>1</sup>		标准差 <sup>8</sup>	1	
最小值 <sup>9</sup>	5 <sup>1</sup>		最小值 <sup>9</sup>	4	
最大值 <sup>10</sup>	4 <sup>1</sup>		最大值 <sup>10</sup>	3	
全距 <sup>11</sup>	1 <sup>1</sup>		全距 <sup>11</sup>	2	
四分位数 <sup>12</sup>	.235 <sup>1</sup>	.144 <sup>1</sup>	四分位数 <sup>12</sup>	-.052	.187
偏度 <sup>13</sup>	.473 <sup>1</sup>	.287 <sup>1</sup>	偏度 <sup>13</sup>	-1.214	.371
峰度 <sup>14</sup>			峰度 <sup>14</sup>		

2-4年				4-6年				6-8年			
平均值		2.73	.069	平均值		2.70	.057	平均值		2.48	.078
平均值的 95% 置信区间	下限	2.59		平均值的 95% 置信区间	下限	2.58		平均值的 95% 置信区间	下限	2.33	
	上限	2.87			上限	2.81			上限	2.64	
5% 剪除后平均值		2.76		5% 剪除后平均值		2.72		5% 剪除后平均值		2.48	
中位数		3.00		中位数		3.00		中位数		2.00	
方差		.952		方差		.914		方差		.871	
标准差		.976		标准差		.956		标准差		.933	
最小值		1		最小值		1		最小值		1	
最大值		4		最大值		4		最大值		4	
全距		3		全距		3		全距		3	
四分位距		2		四分位距		2		四分位距		1	
偏度		-.189	.172	偏度		-.037	.144	偏度		.104	.203
峰度		-.996	.342	峰度		-1.039	.287	峰度		-.843	.403

图 1 不同玩网络游戏年限情况下每月游戏消费描述统计

由图 1 可知：0-2 年的每月消费的偏度为 0.01，峰度为 0.368。2-4 年的每月消费的偏度为 0.238，峰度为 0.185。4-6 年的每月消费的偏度为 0.235，峰度为 0.473。6-8 年的每月消费的偏度为 0.145，峰度为 0.35。在游戏外设消费中，0-2 年的偏度为 -0.52，峰

度为 -1.214，2-4 年偏度为 -0.189，峰度为 -0.996，4-6 年的偏度为 -0.037，峰度为 -1.039，6-8 年的偏度为 0.104，峰度为 -0.843，8 年以上骨灰级玩家的偏度为 0.795，峰度为 -0.692

4.1.2 数据可视化展示

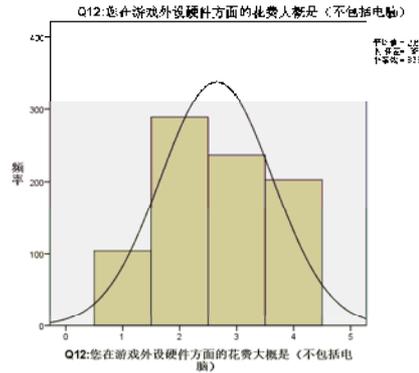
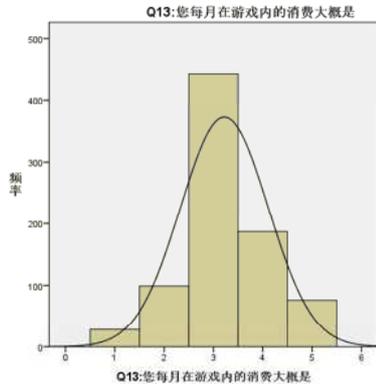


图 2 不同接触网络游戏的年限，每月在游戏中的消费分布可视化展示

根据图 2 可知：由两个个图表可以明显看出，所有的频率分布都是中间部分较高，两边部分较低，说明

在每月游戏内消费，游戏外设消费等方面，大多数人在都比较的适中

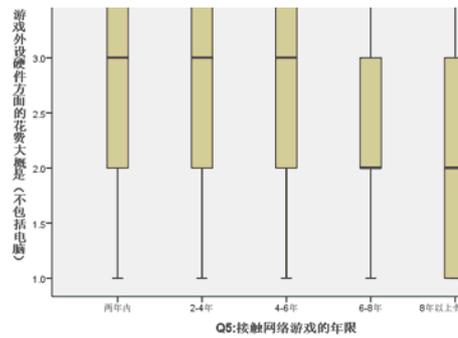
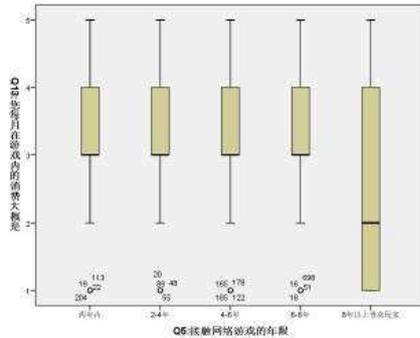


图 3 不同接触网络游戏的年限下每月游戏内消费和游戏外设消费箱型图展示

根据图 3 可知：触网络游戏的年限为 0-8 年，每月在游戏中的消费大部分都集中在 100-500 元，8 年以上玩家的消费为 0-500 元。同时在对应的 5 个年限区间内，偏度和峰度都为正，说明数据为右偏，在游戏中的消费，在中位数右边的极端数值较多，大部分人的平均每天游戏时长其实都是要小于于平均值的。接触网络游戏的年限为 0-8 年，在电脑外设方面的消费大部分都集中在 100 元以上，中位数在 500-1000 元，6-8 年以上玩家的消费 100-1000 元，8 年以上的玩家消费为 0-1000 元，中位数为 100-500 元。同时在 0-6 年区间内，偏度为负，说明数据为左偏，在中位数左边的极端数值较多，大部分人的游戏外设消费其实都是要大于于平均值的。6 年以上的区间，偏度为正，说明数据为右偏，游戏外

设方面的消费，在中位数右边的极端数值较多，大部分人的平均游戏外设消费其实都是要小于于平均值的。同时在对应的 5 个年限区间内峰度都为负，都不为尖峰分布由此可以得出如下结论：在接触网络游戏的年限为 0-8 年的人群数量最多，他们在游戏中的消费能力也比 8 年以上的骨灰玩家强，游戏厂商应开发出玩法更加新颖，机制更加先进，符合接触网络游戏年限不长的玩家的口味，吸引更多的玩游戏时间不长的玩家游玩，游戏厂商们能获得更大的利益。

4.2 以接触网络游戏年限为分组变量，以每月在游戏内的消费和在游戏外设方面的消费为检验变量进行独立样本 t 检验

数据分析目的：针对线上平台的学习数据对学生

网络学习的状况进行分析，目的在于调查网络学习的教学方法是否能够帮助到学生的学习，同时为教学方法的不断完善和改进提供一定的思路。

数据分析方法：主要独立样本 t 检验的方法对不同的接触网络游戏年限下玩家游戏每月消费和游戏外设方面的消费进行分析，从而帮助厂商们对于不同的玩家们做出有针对性的游戏开发策划，和提供不同的游戏外设套餐的推广。

使用的特征变量：分组变量：接触网络游戏年限。检验变量：每月在游戏内的消费和在游戏外设方面的消费。

使用分析工具：Spss statistics

### 4.2.1 分组依据

Q5:接触网络游戏的年限		
个案数	有效	833
	缺失	0
平均值		2.61
平均值标准误差		.039
中位数		3.00
众数		3

图 4 接触网络游戏年限统计指标

根据图 4 可知：平均值为 2.61，所以我们以选项 3（4-6 年）为分组的依据

### 4.2.2 分析结论

组统计					
	Q5:接触网络游戏的年限	个案数	平均值	标准差	标准误差平均值
Q13:您每月在游戏内的消费大概是	>= 3	464	3.18	.899	.042
	< 3	369	3.27	.881	.046
Q12:您在游戏外设硬件方面的花费大概是（不包括电脑）	>= 3	464	2.57	.979	.045
	< 3	369	2.74	.982	.051

图 5 组统计

独立样本检验										
		莱文方差等同性检验			平均值等同性 t 检验					
		F	显著性	t	自由度	显著性 (双尾)	平均值差值	标准误差差值	差值 95% 置信区间	下限
Q13:您每月在游戏内的消费大概是	假定等方差	.331	.565	-1.526	831	.127	-.095	.062	-.217	.027
	不假定等方差			-1.530	796.119	.126	-.095	.062	-.217	.027
Q12:您在游戏外设硬件方面的花费大概是（不包括电脑）	假定等方差	.003	.960	-2.508	831	.012	-.171	.068	-.306	-.037
	不假定等方差			-2.507	788.359	.012	-.171	.068	-.306	-.037

图 6 独立样本检验

由图 6，图 7 可知：

H0：游戏游玩年限为 0-4 年和 4 年以上的玩家，每月在游戏内的平均消费是相同的

H1：游戏游玩年限为 0-4 年和 4 年以上的玩家，每月在游戏内的平均消费是不相同的

结论 1：对于 13 题，由于方差同等性检验的显著性为 0.565 大于 0.05，接受原假设（方差具有齐次性）。在假定等方差的情况下，显著性为 0.127 大于 0.05，接受原假设，所以游戏游玩年限为 0-4 年和 4 年以上的玩家，每月在游戏内的平均消费是相同的。

H0：游戏游玩年限为 0-4 年和 4 年以上的玩家，在游戏外设硬件方面的平均花费是相同的

H1：游戏游玩年限为 0-4 年和 4 年以上的玩家，在游戏外设硬件方面的平均花费是不相同的

结论 2：对于 12 题，由于方差同等性检验的显著性为 0.96 大于 0.05，接受原假设（方差具有齐次性）。在假定等方差的情况下，显著性为 0.012 小于 0.05，拒绝原假设，所以两种不同游玩年限的玩家在游戏外设硬件方面的平均花费是不相同的。

### 4.3 对于游戏偏好为 DOTA 类网游和游戏偏好为 ACT 动作类网游情况决策树分析

数据分析目的：针对戏偏好为 DOTA 类网游和游戏

偏好为 ACT 动作类网游情况进行分析，目的在于调查偏好这两类游戏的玩家的大致特征。

数据分析方法：主要采用决策树分析的方法对这两类游戏的偏好进行分析，从而帮助厂商了偏好这两类游戏的玩家的基本特征。

使用分析工具：Spss modeler

### 4.3.1 分析过程

如图 7 所示，建立起决策树的分析模型，再从分析出的规则集中挑选出游戏偏好为 DOTA 类网游和游戏偏好为 ACT 动作类网游，并且正确率 >60% 的规则集

### 4.3.2 分析结果

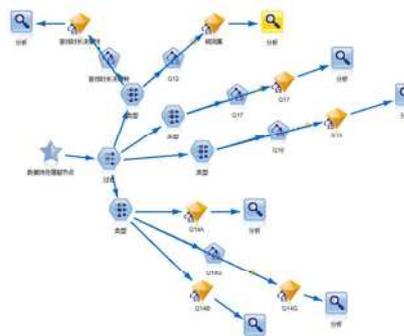


图 7 决策树分析流程图

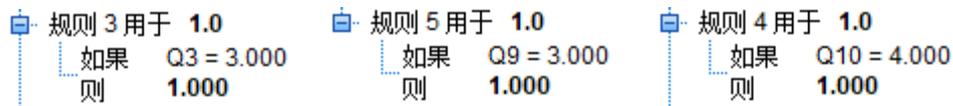


图 8 游戏偏好为 DOTA 类网游规则

由图 8 可知: 特征为游戏偏好是 DOTA 类网游的人: 第 3 题选择了 C 选项, 也就是说专业类型为医学类的人游戏偏好是 DOTA 类网游。

特征为游戏偏好是 DOTA 类网游的人: 第 9 题选择了 C 选项, 也就是说平均每天玩游戏的时间为 4-6 小时



图 9 游戏偏好为 ACT 动作类网游规则

由图 9 可知: 特征为游戏偏好是 ACT 动作类网游的人: 第 3 题选择了 B 选项, 第 9 题选择了 A 选项, 第 10 题选择了 C 选项, 也就是说专业类型为理工类, 平均每天玩游戏的时间为 0-2 小时, 在游戏中处于高水平的人游戏偏好是 ACT 动作类网游。

特征为游戏偏好是 ACT 动作类网游的人: 第 2 题选择了 C 选项, 第 10 题选择了 D 选项, 也就是说所在学校的层次为专科, 在游戏中处于万中无一天之骄子水平的人游戏偏好是 ACT 动作类网游。

特征为游戏偏好是 ACT 动作类网游的人: 第 10 题选择了 D 选项, 也就是说在游戏中处于万中无一天之骄子水平的人游戏偏好是 ACT 动作类网游。

特征为游戏偏好是 ACT 动作类网游的人: 第 10 题选择了 C 选项, 也就是说在游戏中处于高水平的人游戏偏好是 ACT 动作类网游。

总体上来看: 游戏水平处于高水平以上的玩家都爱玩 dota 类网游和 act 动作类网游, 医学类的学生, 和每天玩游戏时长为 4-6 小时的学生更加偏好与 dota 类网游, 理科类的, 每天玩游戏 0-2 小时的学生更加偏好于 act 动作类网游。

## 5 结论及建议

不同的接触网络游戏年限的玩家在游戏每月消费和游戏外设消费方面都有所不用。不同游戏偏好的人的

的人游戏偏好是 DOTA 类网游。

特征为游戏偏好是 DOTA 类网游的人: 第 10 题选择了 D 选项, 也就是说在游戏中处于万中无一天之骄子水平的人游戏偏好是 DOTA 类网游。

特征也具有什么明显的差异。游戏厂商开发的游戏想要在市场上有立足之地, 就必须考虑开发出的游戏的类型所针对玩家的特征, 再根据特征来优化游戏的机制和游戏整体的氪金程度, 可以起到让游戏的生态良好发展的效果。

## 【参考文献】

- [1] 张莹, 朱丹红. 网络游戏行为的影响因素: 一个文献分析 [J], 兰州大学学报 (社会科学版), 2019, 47, (01): 134-142
- [2] 刘晓. 网络游戏行为及其影响因素研究 [D], 武汉大学, 2018: 1-65
- [3] 崔宇娜, 周蓓蓓. 新电子产品时代大学生网络游戏行为分析 [J], 中国管理学科, 2014, 22, (S1): 829-833
- [4] 王荣华. 基于 SPSS 的大学生参与网络游戏影响因素分析 [J], 中国多媒体与网络教学学报 (上旬刊), 2020, (05): 24-26
- [5] 张定坤. 网络游戏的消费者行为分析 [J], 武汉大学, 2018, (32): 200-201