

用户画像很重要，那你知道是怎么画出来的吗？

我们看过应该不下10篇关于用户画像的干货。但是依旧不知道应该怎么做一份用户画像出来。干货里告诉我们用户画像的价值、用户画像应该有的数据，用户画像应该包含的内容。好的，我都按大神们的教导收集好数据了。麻烦谁能告诉我到底这些数据怎么处理啊！喂！



今天我就抛砖引玉的说一下我曾经经手的一份用户画像及其每一步的建立步骤吧！当然我觉得那次的结果很粗糙，原因在于还是没人告诉我每一步要怎么做，但是起码应该是一个粗线条的步骤原型了。希望今天之后大家能在收集齐全大神的数据建议之后，终于知道每步要怎么做了！不要像我一样，太惨了。

step1 我们还是要说一下数据源的事情

虽然我从来都秉承着能百度到的东西基本不废话，但是这张图我觉得还是有必要贴出来的。这个图不是我的，原文作者：郭志金。知识产权还是要保护的，大家可以百度一下郭老师关于如何构建用户画像的文章找到更加详细的论述。不多说了。上图。





这里能就是郭老师说的需要准备好的所有的相关用户数据。好的，我觉得这些数据对于大家来说并不那么困难吧。所以收集到了之后呢，就开始真正的处理数据了。

step2 数据收集后的处理步骤

第一步：清洗数据

本来我不打算说这个，但是觉得还是有必要提一句，收集到的数据很多都不能直接使用。需要对数据进行编码和分类，还要去掉异常值和补全缺失值这些。我这里将会用到的分析软件是跟随我五六年的SPSS（用了这么多年依旧不是高手）。我就假设这一步大家都做完了。

第二步：因素分析

本来我也不知道为什么要做因素分析。直到我最近看书才了解，给大家解释一下做因素分析的原因。用户画像就是把一堆数据分类，分到同一组的数据所代表的人就是一类。那么聚类就需要考虑把哪些数据拿出来聚类呢？比如说个人收入和家庭总收入。这是两列数据，最后可能都会决定我是屌丝还是白富美。要把这两个数据放进去一起聚类吗？恐怕不是！聚类是不建议把高度相关的数据都扔进去的。为什么？因为高度相关了就基本等于同一个数据了，你把同一个数据放进去两遍，不就是加重了这个数据在所有数据里的权重了么！这里一些专业统计词我就不过分解释了，这次真的可以百度。

知道了为什么要因素分析，就谈谈因素分析的作用，因素分析就是把N个数据先归归类，特别相同的就是一类的了，我们选其中一些数据就可以了。这样因素分析就可以帮我们把N列数据减少到几列数据。至于因素分析怎么做，我也不在这里废话了。有一个神奇的老师叫吴明隆，他有一本SPSS实操，那里手把手教学。各位可以去自行查阅。

第三步：相关分析

我们已经把数据N列减少到重要的几个了。在开始聚类之前，还要做一件事就是看一下这些数据的

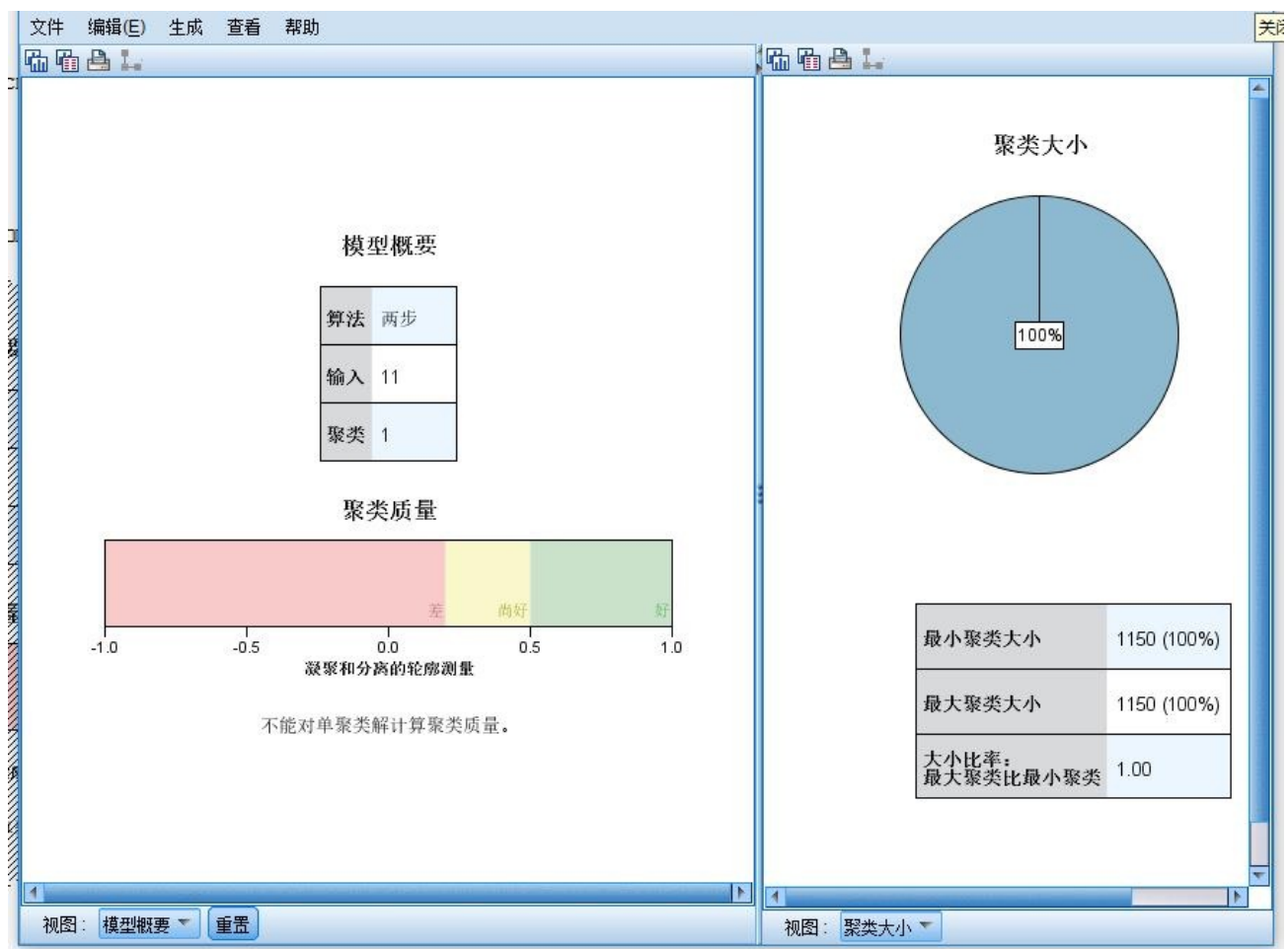
相关性。一般我们选择中等相关的数据。太相关不能用原因见上一条。太不相关了也别放进去了，毕竟个人收入的数据和隔壁老王的体重数据放在一起也没什么大用。相关分析怎么做也请见吴老师的宝书。

第四步：聚类分析

好的，我们终于说到我们的主角了，聚类分析。我曾记得当初我做的时候一个恩师姐姐建议我，聚类都做一遍。虽然我还不不懂她在说什么。但是就我现在的理解，给大家谈一谈。首先SPSS里有三种聚类方式。见图。



上面看到啦，这是SPSS里的3中聚类方式。我会首先做一下两步聚类。为啥呢，因为两步聚类不需要动脑子。为什么这么说。就是你丢一堆数据进去，你不需要预设你要聚类成几类。就是你可以不知道到底有几类啊，看SPSS大大的心情啊。两步聚类把数据类型分分开就可以了，别把连续的放在类别的里面。就可以了。聚类之后你会得到这样一个结果。上图。



酱紫的。具体的解释和怎么看。建议去看沈浩老师博客的聚类分析，百度一下你值得拥有。然后我们把一些重要程度很低的变量可以试着剔除一些了。这里就是第一次聚类。

然后到了第二次聚类，K均值聚类。第一步已经知道大概的聚类类别了。我这个是聚出了1类。K均值的时候有一个聚类数需要你填写。你就把两步聚类里的聚类数填进去就好啦。然后就得到了K均值聚类的结果。我们这样不断调试删除增加变量的原因是为了让聚类结果更加稳定。

最后的最后，系统聚类。你可以看到一个长得很酷的图，然后可以看看结构什么的。也是为了确定聚类结果的稳定性。具体操作也自行解决吧。

第五步：判别分析

聚类好了之后，我们是有态度有素质的团队嘛。就要看看自己聚类出来的结果准不准。就需要判别分析了，把已经聚类好的数据和待验证的数据放进去看下结果，来评判一下聚类的效果。本来我想上个图，觉得要保护我司的数据安全。就放弃了。大家百度一下贝叶斯判别，很多效果图那边。这里只讲步骤。

好啦。最后的结果如果稳定且验证后效果良好。那你的用户就真的画完了。很多人在说每个标签权重的事情。我个人是这么理解的，看贵司用户占这些分类的多少比较重要。说的更直接点，加权重给每个画像这个事情，我暂时还没有知道要怎么做。如果有牛人用R或者什么统计软件完爆我，我

真的认真的说，请收我为徒。我真的很想要一个老师。为了实现我做最好的用研的梦想。

作者：王蕊（微信号Xiaolan_psy），斗鱼用户研究员。1年互联网用户研究经验，曾主导同程旅游多款项目的产品调研与数据分析工作。

人人都是产品经理（woshipm.com）中国最大最活跃的产品经理学习、交流、分享平台