



链滴

# 为什么这部电影的豆瓣评分我不认可? —— 一点数据分析的视角

作者: [cqcn1991](#)

原文链接: <https://ld246.com/article/1500697467612>

来源网站: 链滴

许可协议: [署名-相同方式共享 4.0 国际 \(CC BY-SA 4.0\)](#)

源代码: <https://github.com/cqcn1991/movie-compare>

文中涉及的交互式散点图: <https://cdn.rawgit.com/cqcn1991/movie-compare/master/clusters.ht>

之前其实写过这个问题 [1], 不过没有在这里分享过, 这里简单写一下, 希望能对大家有所帮助

## 1. 我们所讨论“评分”到底是什么?

拿刺客聂隐娘来说



评分虽然不错, 但能看到评价差异很大的评价, 而且都有不少人赞

-  小猪大侠 看过 ★★★★★ 2015-08-18 4844 有用  
其实挺不希望聂隐娘在内地上映的, 想到侯导一把年纪了, 功成名就拍了一辈子电影还要被一群不学无术又自以为是的傻逼指着鼻子骂装逼的情景就心酸。
-  taoxinxu释凡 看过 ★★★★★ 2015-08-24 3123 有用  
以为穿套古人服, 大镜头不动, 闷骚, 老念叨古文就叫艺术, 只能证明编导看《资治通鉴》《新旧唐书》走火入魔, 形式大于内容, 以为法国拿杯子, 八个评委搞定四个, 就华语艺术片骄傲? 那不过脱离市场自嗨; 以为水军刷8.3分, 装清高都吹嘘, 就比《绣春刀》牛逼? 那不过点评人都眼睛装清高, 奸商操盘阴谋
-  蒂耶里·福茂 看过 ★★★★★ 2015-05-21 2139 有用  
青鸾舞镜。《聂隐娘》讲得其实是孤独, 一个人身处琼楼玉宇, 无人知, 无人懂, 只能一条道走到底。孤独不仅是隐娘, 也是田季安, 也是侯导。如一幅“坐看云起时”的文人画, 第一次, 华人从光影的缝隙里寻回那遗失已久的古典美学与仪式。
-  扭腰客 看过 ★★★★★ 2015-08-24 1740 有用  
自己没拍明白, 让一群傻逼来猜, 这就是大师。看电影的时候, 刚开始全场都在屏摄, 中间全场都在酣睡, 最后主题曲一出来, 全场都在抖腿。唱念做打均无动机, 几句文言文不是什么禅意, 也不是什么“这才是中国人的武侠片”。侯导辛苦, 但实在难负盛名。

于是我们会问: 这部电影的豆瓣评分靠谱吗? 上面这些人他们谁是对的, 谁是错的? 是不是有水军?

然而, 同样的评分, 却发现其他的电影并不如此, 比如我11



虽然也有好评、中评, 但差评明显少多了



西市大官人 看过 ★★★★★ 2012-05-18

860 有用

难得的国产好片，但这也未能阻止坐我旁边的那个臭傻逼不停的玩手机，我操你妈的高清屏真他妈的亮把你那臭脸照的更傻逼了，影院里这样的傻逼真多，可怜的傻逼们，被手机劫持的傻逼们，都没有两个小时属于自己的时间。臭傻逼！



罗弘霉素 看过 ★★★★★ 2012-04-30

141 有用

别名：那年贾樟柯11。片段事件有些散乱，但是儿时对于记忆本就如此。父亲对儿子关于绘画的教育大概影射了当时社会现状，遗憾的是为了上映有些放不开手脚。闫妮适合演这类角色，所以还是远离骗钱的黄百鸣导演吧...



内陆飞鱼 看过 ★★★★★ 2012-05-19

83 有用

感情上值得尊敬，环境还原得不错，看来下了大功夫，人物选得挺灵，演得很扎实。但是，叙事方式老套，拍摄手法流于庸常，从头到尾都是电视剧一样的视点，扁平慢不凝练，缺乏活络的神采，没有印象深刻的段落和触点，像一顿夹生饭。从电影语言上看，导演真没什么大进步，甚至比起先前略有倒退。



悍客罗 看过 ★★★★★ 2012-04-20

68 有用

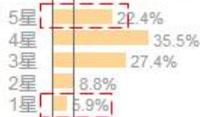
王小帅的个人回忆，同时带出文革后期的时代感。画面很美，风景绝佳，小时候的细节动人。有《阳光灿烂》+《西西里传说》的影子，堪称西双版纳的美丽传说。不过，那小孩太像贾樟柯了实在。。

### 仔细观察，会发现两者的评分分布并不相同



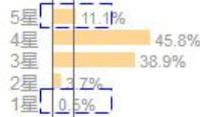
刺客聂隐娘 (2015)

7.2 ★★★★★  
122990人评价



我11 (2012)

7.3 ★★★★★  
42954人评价



- 刺客聂隐娘：同时有不少人觉得不错/较差，在1星和5星有不少的分布。
- 我11：大家都觉得比较不错，所以评分集中在了3星-4星

也就是说，尽管两者（平均）分数相同，但是背后的看法非常不同，评分差异很大，这也正好对应了面截然相反的热评的情况。

我们可以用方差来衡量评分分布的差异，计算方法如下

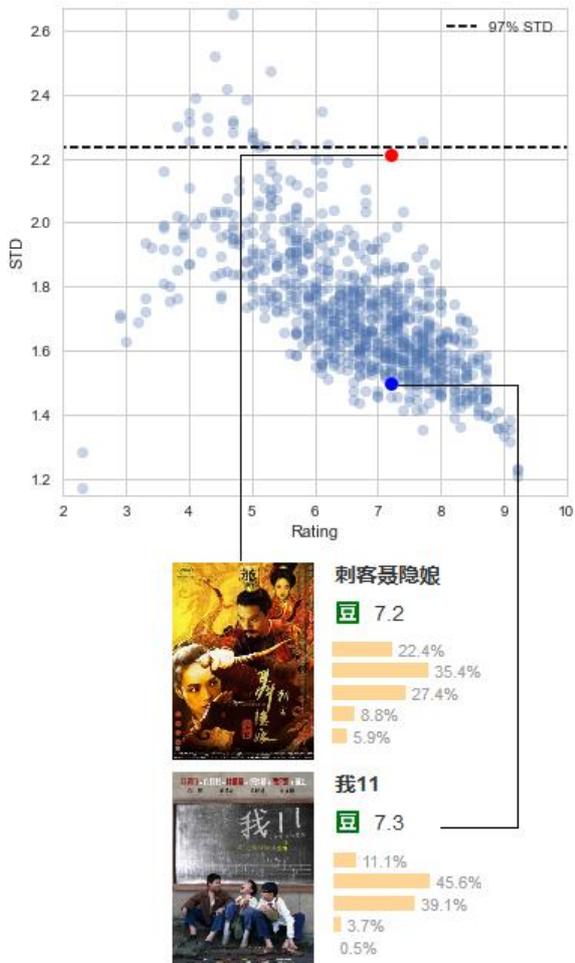
$$\text{方差} = p_2(2 - \text{Avg})^2 + p_4(4 - \text{Avg})^2 + \dots + p_{10}(10 - \text{Avg})^2$$

Avg 为豆瓣（平均）分数

$p_2, p_4, \dots, p_{10}$  为评分中 1-5 颗星所占的比例

也就是计算 **评分偏离平均分的程度**。下文使用标准差（STD），由方差开方即可。

用近几年的国内上映的电影，做出标准差（STD） - 豆瓣评分（Rating）散点分布图。为了便于比较做标准差97%范围线。



散点图可见 <https://cdn.rawgit.com/cqcn1991/movie-compare/master/clusters.html>

可以看到两者STD差别确实很大。仔细看上面的评论，也能发现分歧的原因——刺客聂隐娘因为画水平和讲故事的水平相差很大，大家侧重不同，所以也就产生了评价差异。

也就是说，我们所说的豆瓣评分，**只是一个平均分，并不能反映内在评价差异很大的情况**

另外，在这幅图上也能发现其他有争议、STD很高的电影，比如大鱼海棠、后会无期等等，这里不再述，具体分析可见[1]。

## 2. 这意味着什么？

这意味着什么？不妨从另一个例子来理解

日赚1.6亿！腾讯员工人均月薪6.3万 ...以此计算，截至2017年3月31日止3个月，腾讯员工平均月酬金 742300万/(39258\*3)，约为6.3万元。作为一家人员接近4万的公司，腾讯这样大手笔给员工发薪酬发业界震惊。一位腾讯内部人士则指出，酬金不能只看平均的。腾讯的雇员也包括腾讯CEO马化腾、裁刘炽平等高管，你要是把万达董事长王健林的收入和你平均一下，你的收入也高很多。 <http://digi.63.com/17/0520/06/CKS04PC2001680NE.html>

这样其实大家也就明白了

豆瓣评分是平均分。当评分分布的差别很大的时候，这个平均分的意义就有限了，就像上面这个人均入的例子一样。

不过，为什么很多时候，我们觉得豆瓣电影的评分有用呢？因为大部分时候，大家的看法是比较一致（都觉得好看/不好看），就像上面的我11一样。

明白这点之后，我感觉轻松了不少

因为以前我总纠结电影的分数。因为自己的看法和别人不同，所以担心自己是错的、不敢表达自己的法。现在，我明白原来还有很多人的看法和我一样（评分分布上那些打其他分的比例），只是人比较而已。

单单谈论平均分高低的问题也在于此——预先假设了对于一部电影，其分数应该是唯一的。但实际上，大家对电影可以有不同的看法，不是说你是对的，那我就是错的。

### 3. 总结

最后，回答开头的问题——你是否认可豆瓣的评分？

大多数时候，我是认可的，因为大多数电影，大家的看法很接近，平均分是很有效的。但是当内在评差别很大的时候，平均分的意义有限，也就无从讨论对其认可不认可了。因为对电影的看法本身可以多元的。

那么，有什么办法能解决这个问题，帮大家更好的理解豆瓣评分呢？

一种可能的做法，是给豆瓣评分旁边加上一个小标签。比如，对STD特别大的电影，在旁边加个“分提醒”标签，注明“这部电影的评价差异水平达到了前3%，平均分的参考意义较为有限”，进一步可以分开展示好评/差评，向用户解释评价差异具体如何。这样或许能减少一些人们对（平均）评分疑虑。

### 4. 其他问题补充

(1) 没考虑水军

对于评分，当然应该考虑水军的影响。

但是我们所谓的水军，\*\*究竟是虚假帐号评分，还仅仅只是别人的评价和我不同而已？\*\*就如有的，确实是喜欢刺客聂隐娘的画面，而对故事没那么看重。更重要的是，只考虑水军，没有办法解释剔水军之后，分歧依然很大的现象。

(2) 豆瓣不是简单的平均分，是加权吧？

这并没有本质影响。因为问题核心在于，单个指标（无论是平均数/中位数/加权）难以表现内部差异。

可以假想一部电影，在剔除水军之后，50% 1星，50% 5星，怎么理解？用哪个分数，都说明不了问。又比如上面人均收入的例子，用算法再怎么算，也没办法用一个工资同时表达高管和新入职员工的平。

最后，关于这个问题更详细的分析，见 [1]。

[1] [为什么这部电影的豆瓣评分我不认可？ -- 一点数据分析的视角 - 知乎专栏](#)