

# SurveyRecommendersystem

Wednesday, June 15, 2016 8:37 PM

推荐系统是一个当前非常火热的领域，通过推荐系统，商家可以发掘客户的潜在需求，提高商铺的效率。尽管推荐系统和系统软件的关系不大，但是已经在一个非常火热的领域，文章《When do Recommender Systems Work the Best? The Moderating Effects of Product Attributes and Consumer Reviews on Recommender Performance》第一次探讨了之前的文章中提出的各种推荐算法的效率，揭露了不同的算法不同的参数所造成的影响如何。思考问题的方式和实验验证的方法值得学习。

推荐系统的种类有很多，算法也有很多，参数也有很多。同时推荐系统已经无处不在，通过推荐系统的部署，普遍认为可以提高客户的转化率，即会有更多的客户购买东西。但是没有一个详细而系统的解释各个参数对算法效果的影响，那么到底如何调整推荐系统才能让系统达到最好的状态，即能够提高用户的转化率？

作者通过选取一个经典的推荐系统：“购买了这个产品的人同时也购买了...”来进行试验。首先作者通过对产品的属性进行分析，提出了六个假设：

## 1. 享乐型和实用型

享乐型的产品指的是他的主要作用是给客户快乐，如一个玩具娃娃。而实用型的产品则通常指一种产品在现实的生活中有明显的用处，例如扳手。基于作者的推测给出两个假设：

- 在线系统中，实用型的产品转化率会变高
- 推荐系统下，享乐型的物品的转化率要高于实用型的产品

## 2. 搜索型的和经验型的

搜索型即指一种产品可以通过一些非常明确的关键字来进行搜索，比如说显示器可以通过一个尺寸进行搜索，内存条可以通过内存的大小和带宽频率进行搜索。而经验型则相反，难以通过一两个确定的词汇找到自己想要的产品。基于这个属性，作者给出两个假设：

- 在在线系统中，搜索型的物品的转化率会变高
- 当推荐系统部署后，经验型的物品的转化率所受到的好处大于搜索型物品。

## 3. 客户的评论好坏，评论数量的多少

基于评论，作者得出两个假设：

- 有着高评论分的产品在在线系统中会得到更高的转换率
- 在推荐系统下，高评分带来的好处会被限制。

为了能够对上述假设进行假设检验，作者对产品属性进行建模。为了消除在线系统对推荐效果的影响，得出一个基础转化率，模型基于一个双重差分模型。

$$\begin{aligned} P(\text{conversion})_{iu} = & \beta_0 + \beta_1 \text{PRICE}_i + \beta_2 \text{REC}_u \\ & + \beta_3 \text{UTILHEDO}_i + \beta_4 \text{SEARCHEXP}_i \\ & + \beta_5 \text{DURABILITY}_i + \beta_6 \text{BRAND}_i + \beta_7 \text{DESLEN}_i \\ & + \beta_8 \text{AVGRATING}_i + \beta_9 \text{RATINGNUMB}_i \\ & + \beta_{10} \text{PRICE}_i \times \text{REC}_u + \beta_{10} \text{UTILHEDO}_i \times \text{REC}_u \\ & + \beta_{11} \text{SEARCHEXP}_i \times \text{REC}_u + \beta_{10} \text{DESLEN}_i \times \text{REC}_u \\ & + \beta_{10} \text{AVGRATING}_i \times \text{REC}_u \\ & + \beta_{11} \text{RATINGNUMB}_i \times \text{REC}_u + \epsilon_u \end{aligned}$$

虽然作者采用了一个线性回归模型，但是由此带来的误差作者认为可以通过大数据基数消除，与此同时，线性模型的可解释性更好，效率更高。

以下是结果

**Table 5: Main Results Table:**

$t^*$ , =  $p$ -value < 0.05,  $t^{**}$ , =  $p$ -value < 0.01,  $t^{***}$ , =  $p$ -value < 0.001

Variables	Estimate	Std Error
Constant	0.034771 <sup>***</sup>	0.001175
PRICE	-0.000019 <sup>***</sup>	0.000003
REC	0.002797 <sup>***</sup>	0.001042
DESLEN	-0.000001	0.000001
AVGRATING	0.002013 <sup>***</sup>	0.000153
RATINGNUMB	-0.000002	0.000003
UTILHEDO (UTIL=1)	0.005120 <sup>***</sup>	0.000677
SEARCHEXP (SEA=1)	0.003207 <sup>***</sup>	0.000677
BRAND	0.001941 <sup>**</sup>	0.000681
DURABILITY	-0.004763 <sup>***</sup>	0.00018
REC X PRICE	-0.000010 <sup>*</sup>	0.000004
REC X DESLEN	0.000005 <sup>*</sup>	0.000002
REC X AVGRATING	-0.000772 <sup>***</sup>	0.000215
REC X RATINGNUMB	0.000011 <sup>*</sup>	0.000004
REC X UTILHEDO	-0.003064 <sup>**</sup>	0.000944
REC X SEARCHEXP	0.000148	0.000945

通过对图中的P-value对原假设进行检验，结果如下

**Table 7: Hypotheses and Results**

Attribute Construct	Hypotheses	Supported
Hedonic-Utilitarian	The base conversion rate for utilitarian goods will be higher in online settings	YES
<b>Hedo-Util × Rec</b>	<b>The increase in conversion rate under the use of a recommender will be higher for hedonic goods, compared to utilitarian goods</b>	<b>YES</b>
Search-Experience	The base conversion rate for search goods will be higher in online settings	YES
<b>Sea-Exp × Rec</b>	<b>The increase in conversion rate under the use of a recommender will be higher for experience goods, compared to search goods</b>	<b>NO</b>
Avg Review Rating	The base conversion rate will be increased for products with higher average review ratings	YES
<b>Avg Review Rating × Rec</b>	<b>The positive impact on conversion from high average review ratings will be lessened under the presence of a recommender system</b>	<b>YES</b>

其中可以看到，大部分作者的假设都被验证为真，只有假设4失败。作者认为可能是因为推荐系统提供的信息不足以说服客户进行购买，或者说在这种口口相传的基于经验的商品上，客户并不一定相信别人的喜好。

个人总结，我觉得这篇survey很有意思，通过一系列的实验，对推荐系统的不同的参数进行剖析。然而作者只用了一种推荐系统和两个推荐算法，可能还是不能完全消除其结果的片面性。尽管如此，这篇文章的目的应该已经达到了。