

优化离网预警模型的几个方面

华院分析技术（上海）有限公司

数据挖掘经理 段勇

高级咨询顾问 宋向平

1. 合理的目标定义

离网预警模型的目标是预测当前在网用户在未来离网的概率，离网往往分为主动离网和被动离网。在模型构建的时候往往需要根据历史数据对用户是否离网做一个标志，因此合理的离网定义对模型起着至关重要的作用。例如要用 1、2、3 月均在网的用户行为来预测这些用户 4 月份是否离网，那么 4 月份的离网需要有明确的定义，可以使用下面这种定义

- 4 月份主动消号的用户
- 4 月份 5 月份连续 2 个月无行为

2. 合理的建模目标选择

并非所有的用户都需要进行建模分析的。

当前已经表现出手机业务状态不正常的用户，未来离网的可能性很大，就没有必要作为建模的目标用户，如果选择了反而会影响到对其他正常用户的预测。因为相对那些不正常用户来说，正常的用户未来离网的概率小得多，预测的时候也许就不认为这些用户离网概率高了。

除了不正常用户之外，还有一些内部员工、数据卡、公话、特殊用户等都可以考虑不纳入模型中

3. 对建模目标用户的细分

并不是对所有的用户放在一起构建一个模型，这样可能会忽略一部分重要用户的预测，就是说有很多高端用户预测离网概率很低，因为他和低端用户放在一起，模型认为它离网概率的确是比低端用户低很多。

往往来说，确定建模目标用户之后，需要根据一些业务上的情况来对用户做一个进一步的细分，在对每个细分群体进行建模。可以考虑从 ARPU 的高低来细分，也可以考虑从使用某些套餐或者其他方面来细分。

4. 合理定义模型构建的时间窗口

常见的模型定义时间窗口是用 1、2、3 月的行为预测 5 月是否离网，中间空一个月作为活动月。这样的模型的结果跟经验模型预测的结果自然就很难重合，因为经验模型的时间性很强，也许有这个特征在未来 1 周内就有可能离网。而数据挖掘模型更强调一种客户的提前预警和维系。不能单纯的理解预警模型出来的

用户就马上要离网。

此外，如果考虑到市场环境变化很快，用户 3 个月的行为不足以预测他是否在下一个离网，可以考虑用 1、2、3、4 周的行为来预测下 2 周是否离网。这样做对数据处理以及离网挽留执行力的要求就比较高，挽留更加具有及时性。

5. 提高模型的准确度

在前面 4 个方面都做得很合理之后，模型的准确度的提高效果往往是有限的。原因是客户的离网规律千差万别，我们所能获得仅仅是用户的消费数据，而基本属性和心理特征方面的数据很难获得或者误差很大。

6. 对于经验离网模型与数据挖掘模型对比

经验离网模型通常是通过用户已经表现出的较为明显的离网行为特征进行判断，而数据挖掘离网预警模型可以对于当前仍然比较正常的用户的离网可能性进行判断。也正因为如此，经验离网模型的准确率可以高于数据挖掘离网预警模型，但通过经验离网模型抓取的用户比例（查全率）较少。

从对于实际维系挽留工作的指导性来看，由于经验模型的滞后性，在实际维系挽留工作中，用户的联系成功率低，联系成本较高，及时联系成功的，挽留成功率也比较低。

通过数据挖掘离网预警模型获取的用户名单，联系成功率则要高得多，但由于提升度有限，为了挽留目标用户，需要联系更多数量的用户。

从对于一线工作人员的要求来看，经验模型抓取的目标用户状态明确（一般是通过单变量确定），一线工作人员更容易直接切入。通过数据挖掘离网预警模型抓取的目标用户，状态不够明确（多变量协同判定），而且目前一般正常在用，对于一线工作人员的要求较高，需要提供一定的指导。

基于以上特点，可以将基于数据挖掘模型的离网预警工作与主动营销工作相结合。通过数据挖掘模型指导，将维系挽留工作，融入日常客户接触工作中，大大提高工作的效率，降低成本。

因此，我们认为经验模型与数据挖掘离网预警模型各有利弊，互为补充。在利用数据挖掘模型评分之前，可以首先通过经验模型对于评分目标用户进行筛选，对于已经具有较为明确地离网倾向的目标用户加以标记。将两者综合应用，各取所长，才能全方位地做好客户维系挽留工作。