

机器学习算法系列（37）：外卖O2O的用户画像实践

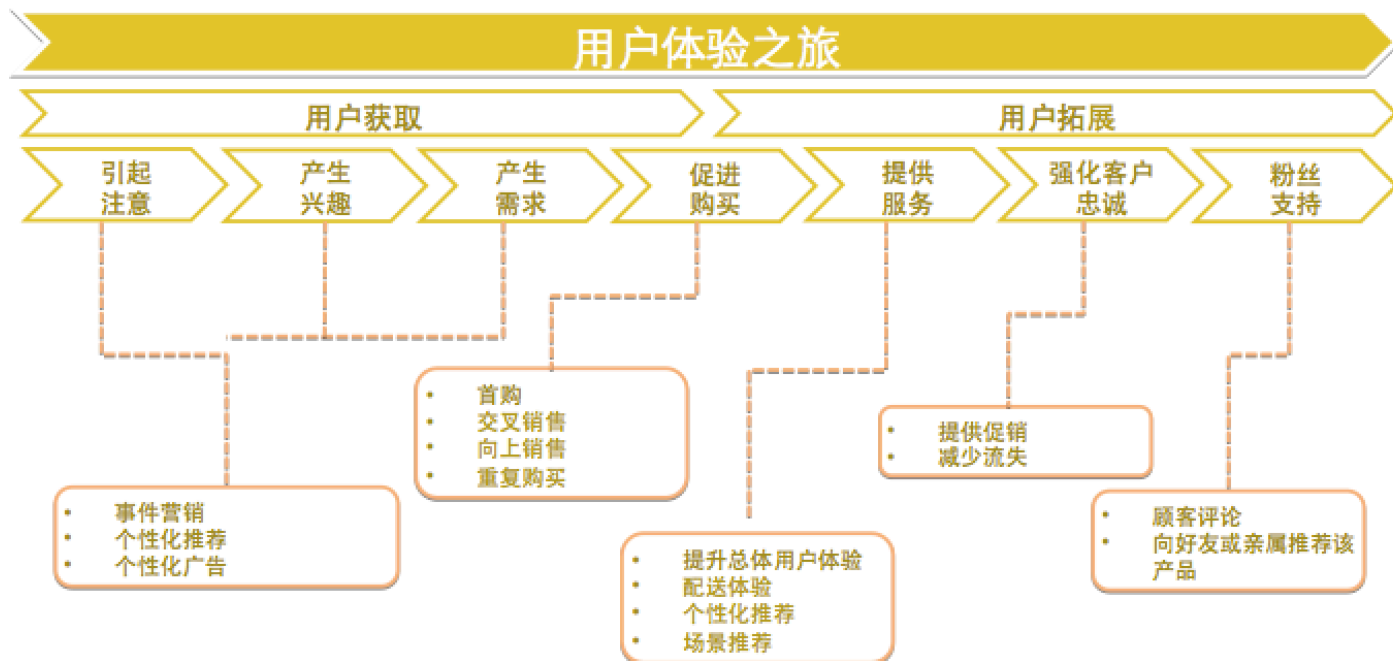
美团外卖经过3年的飞速发展，品类已经从单一的外卖扩展到了美食、夜宵、鲜花、商超等多个品类。用户群体也从早期的学生为主扩展到学生、白领、社区以及商旅，甚至包括在KTV等娱乐场所消费的人群。随着供给和消费人群的多样化，如何在供给和用户之间做一个对接，就是用户画像的一个基础工作。所谓千人千面，画像需要刻画不同人群的消费习惯和消费偏好。

外卖O2O和传统的电商存在一些差异。可以简单总结为如下几点：

- 1) 新事物，快速发展：这意味很多用户对外卖的认知较少，对平台上的新品类缺乏了解，对自身的需求也没有充分意识。平台需要去发现用户的消费意愿，以便对用户的消费进行引导。
- 2) 高频：外卖是个典型的高频O2O应用。一方面消费频次高，用户生命周期相对好判定；另一方面消费单价较低，用户决策时间短、随意性大。
- 3) 场景驱动：场景是特定的时间、地点和人物的组合下的特定的消费意图。不同的时间、地点，不同类型的用户的消费意图会有差异。例如白领在写字楼中午的订单一般是工作餐，通常在营养、品质上有一定的要求，且单价不能太高；而到了周末晚上的订单大多是夜宵，追求口味且价格弹性较大。场景辨识越细致，越能了解用户的消费意图，运营效果就越好。
- 4) 用户消费的地理位置相对固定，结合地理位置判断用户的消费意图是外卖的一个特点。

一、外卖产品运营对画像技术的要求

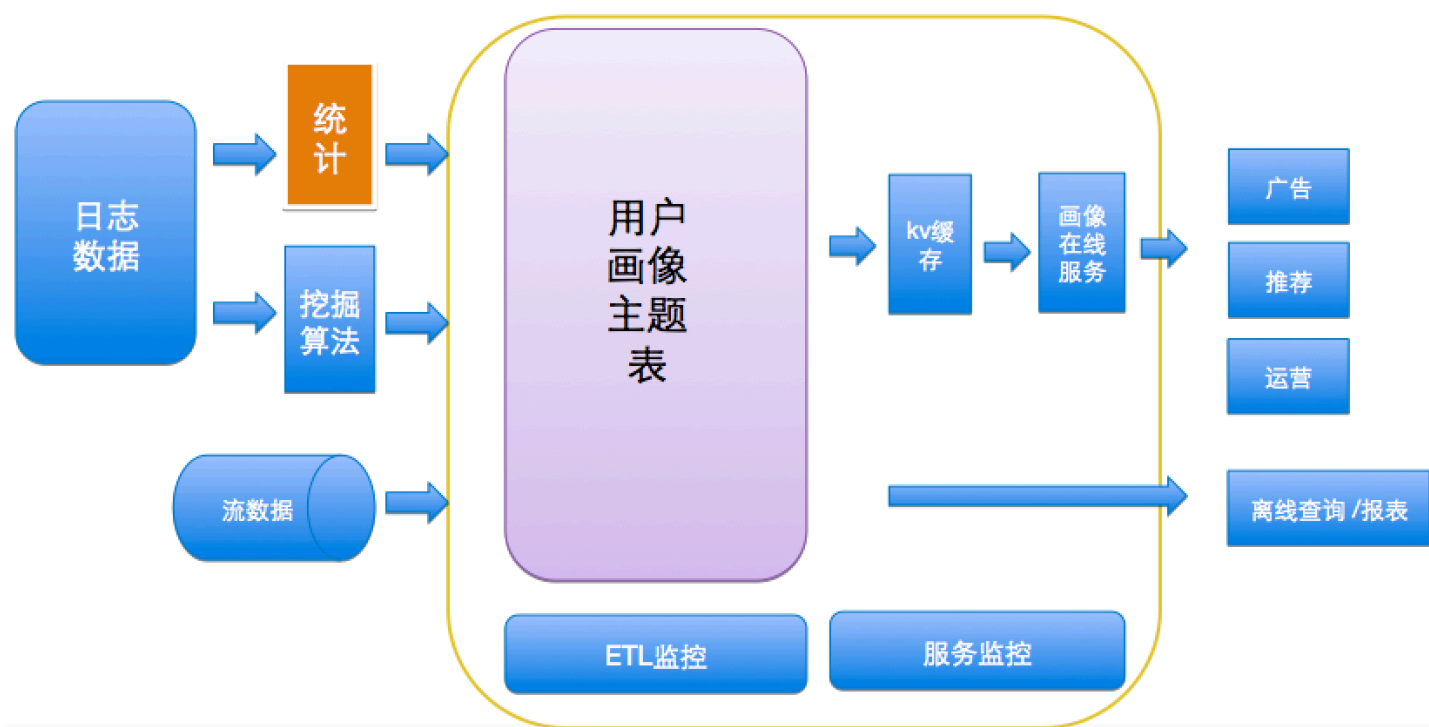
如下图所示，我们大致可以把一个产品的运营分为用户获取和用户拓展两个阶段。在用户获取阶段，用户因为自然原因或一些营销事件（例如广告、社交媒体传播）产生对外卖的注意，进而产生了兴趣，并在合适的时机下完成首购，从而成为外卖新客。在这一阶段，运营的重点是提高效率，通过一些个性化的营销和广告手段，吸引到真正有潜在需求的用户，并刺激其转化。在用户完成转化后，接下来的运营重点是拓展用户价值。这里有两个问题：第一是提升用户价值，具体而言就是提升用户的单均价和消费频次，从而提升用户的LTV（life-time value）。基本手段包括交叉销售（新品类的推荐）、向上销售（优质高价供给的推荐）以及重复购买（优惠、红包刺激重复下单以及优质供给的推荐带来下单频次的提升）；第二个问题是用户的留存，通过提升用户总体体验以及在用户有流失倾向时通过促销和优惠将用户留在外卖平台。



所以用户所处的体验阶段不同，运营的侧重点也需要有所不同。而用户画像作为运营的支撑技术，需要提供相应的用户刻画以满足运营需求。根据上图的营销链条，从支撑运营的角度，除去提供常规的用户基础属性（例如年龄、性别、职业、婚育状况等）以及用户偏好之外，还需要考虑这么几个问题：1）什么样的用户会成为外卖平台的顾客（新客识别）；2）用户所处生命周期的判断，用户是否可能从平台流失（流失预警）；3）用户处于什么样的消费场景（场景识别）。后面“外卖O2O的用户画像实践”一节中，我们会介绍针对这三个问题的一些实践。

二、外卖画像系统架构

下图是我们画像服务的架构：数据源包括基础日志、商家数据和订单数据。数据完成处理后存放在一系列主题表中，再导入kv存储，给下游业务端提供在线服务。同时我们会对整个业务流程实施监控。主要分为两部分，第一部分是对数据处理流程的监控，利用内部自研的数据治理平台，监控每天各主题表产生的时间、数据量以及数据分布是否有异常。第二部分是对服务的监控。目前画像系统支持的下游服务包括：广告、排序、运营等系统。



三、外卖O2O的用户画像实践

3.1 新客运营

新客运营主要需要回答下列三个问题：

- 1) 新客在哪里？
- 2) 新客的偏好如何？
- 3) 新客的消费力如何？

回答这三个问题是比较困难的，因为相对于老客而言，新客的行为记录非常少或者几乎没有。这就需要通过一些技术手段作出推断。例如：新客的潜在转化概率，受到新客的人口属性（职业、年龄等）、所处地域（需求的因素）、周围人群（同样反映需求）以及是否有充足供给等因素的影响；而对于新客的偏好和消费力，从新客在到店场景下的消费行为可以做出推测。另外用户的工作和居住地点也能反映他的消费能力。

对新客的预测大量依赖他在到店场景下的行为，而用户的到店行为对于外卖是比较稀疏的，大多数的用户是在少数几个类别上有过一些消费行为。这就意味着我们需要考虑选择什么样的统计量描述：是消费单价，总消费价格，消费品类等等。然后通过大量的试验来验证特征的显著性。另外由于数据比较稀疏，需要考虑合适的平滑处理。

我们在做高潜新客挖掘时，融入了多方特征，通过特征的组合最终作出一个效果比较好的预测模型。我们能够找到一些高转化率的用户，其转化率比普通用户高若干倍。通过对高潜用户有针对

性的营销，可以极大提高营销效率。

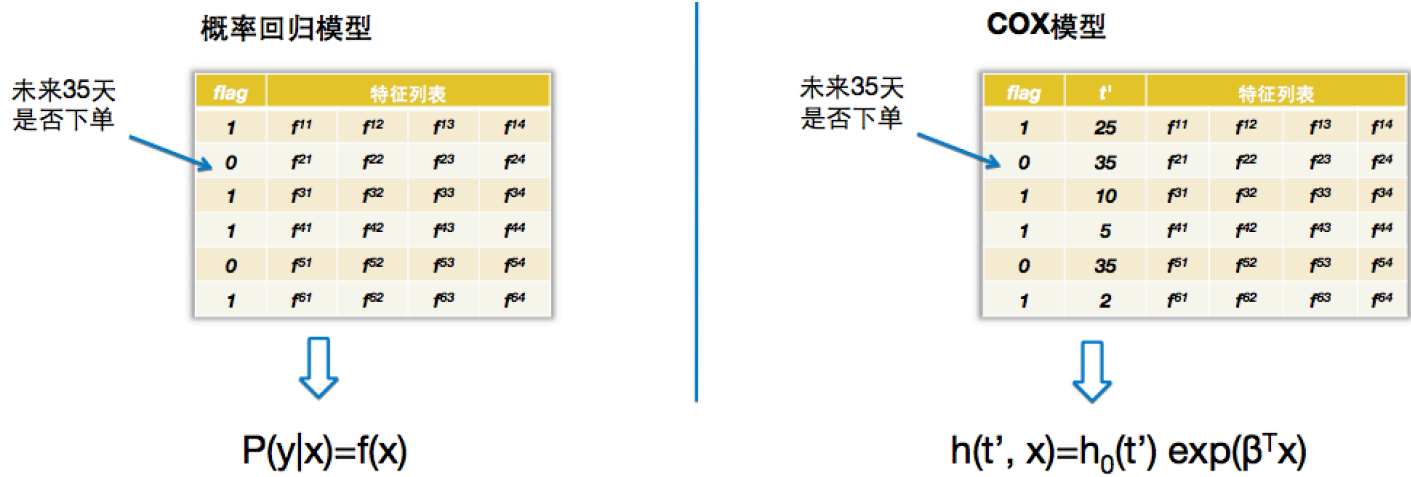
3.2 流失预测

新客来了之后，接下来需要把他留在这个平台上，尽量延长生命周期。营销领域关于用户留存的两个基本观点是（引自菲利普.科特勒 《营销管理》）：

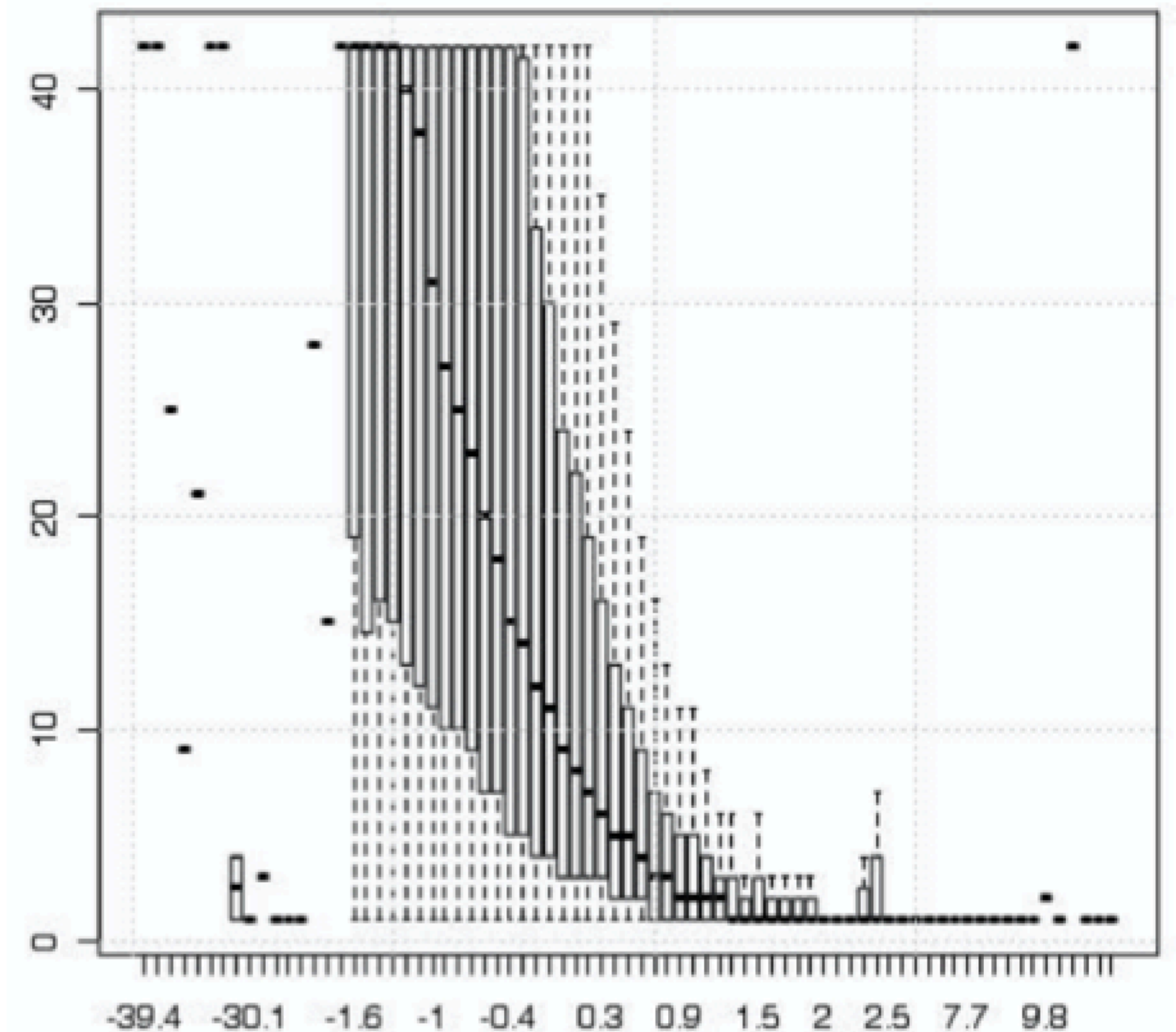
获取一个新顾客的成本是维系现有顾客成本的5倍！

如果将顾客流失率降低5%，公司利润将增加25%~85%

用户流失的原因通常包括：竞对的吸引；体验问题；需求变化。我们借助机器学习的方法，构建用户的描述特征，并借助这些特征来预测用户未来流失的概率。这里有两种做法：第一种是预测用户未来若干天是否会下单这一事件发生的概率。这是典型的概率回归问题，可以选择逻辑回归、决策树等算法拟合给定观测下事件发生的概率；第二种是借助于生存模型，例如COX-PH模型，做流失的风险预测。下图左边是概率回归的模型，用户未来T天内是否有下单做为类别标记y，然后估计在观察到特征X的情况下y的后验概率 $P(y|X)$ 。右边是用COX模型的例子，我们会根据用户在未来T天是否下单给样本一个类别，即观测时长记为T。假设用户的下单的距今时长 $t < T$ ，将t作为生存时长 t' ；否则将生存时长 t' 记为T。这样一个样本由三部分构成：样本的类别(flag)，生存时长(t')以及特征列表。通过生存模型虽然无法显式得到 $P(t'|X)$ 的概率，但其协变量部分实际反映了用户流失的风险大小。



生存模型中， $\beta^T x$ 反映了用户流失的风险，同时也和用户下次订单的时间间隔成正相关。下面的箱线图中，横轴为 $\beta^T x$ ，纵轴为用户下单时间的间隔。

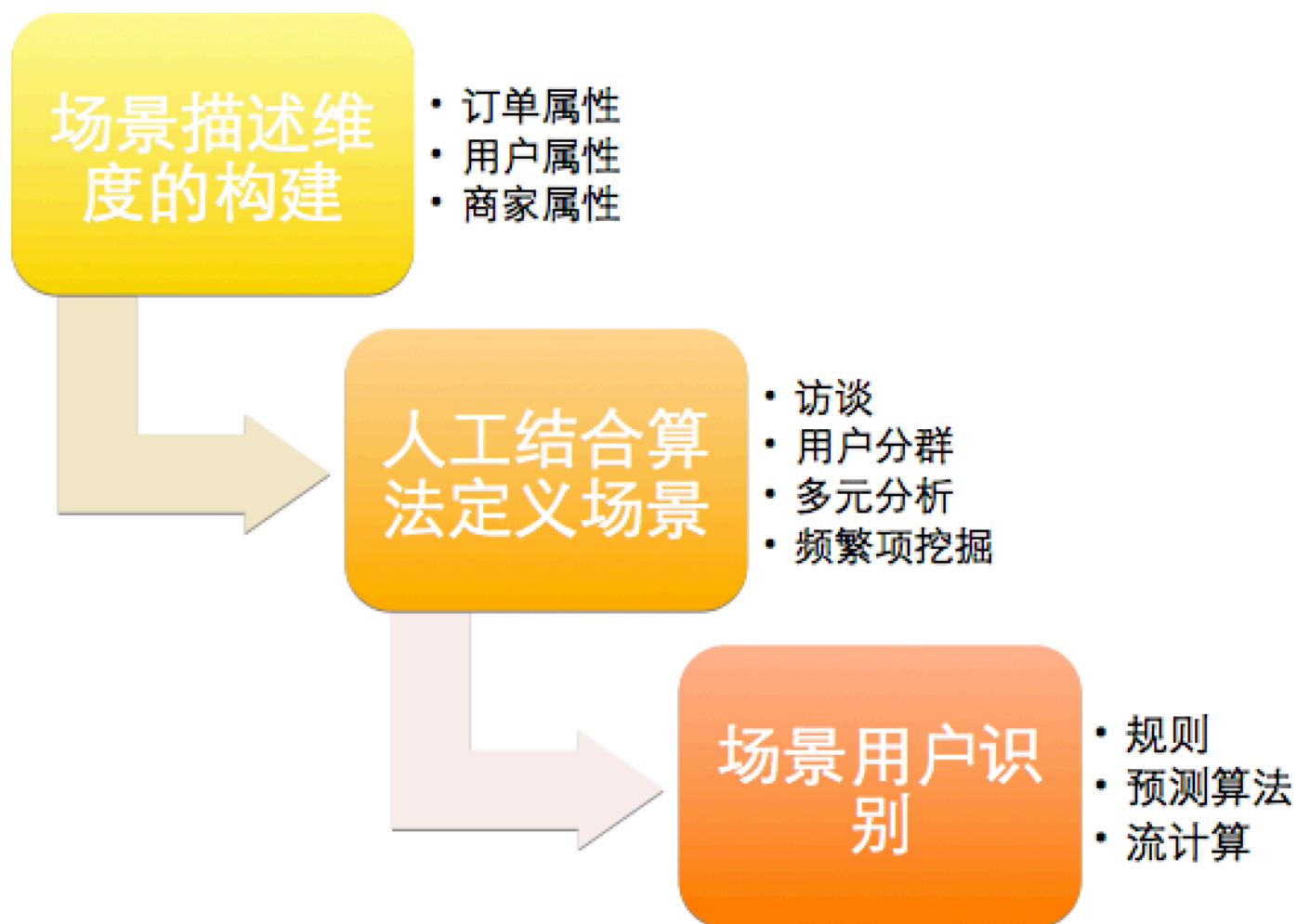


我们做了COX模型和概率回归模型的对比。在预测用户XX天内是否会下单上面，两者有相近的性能。

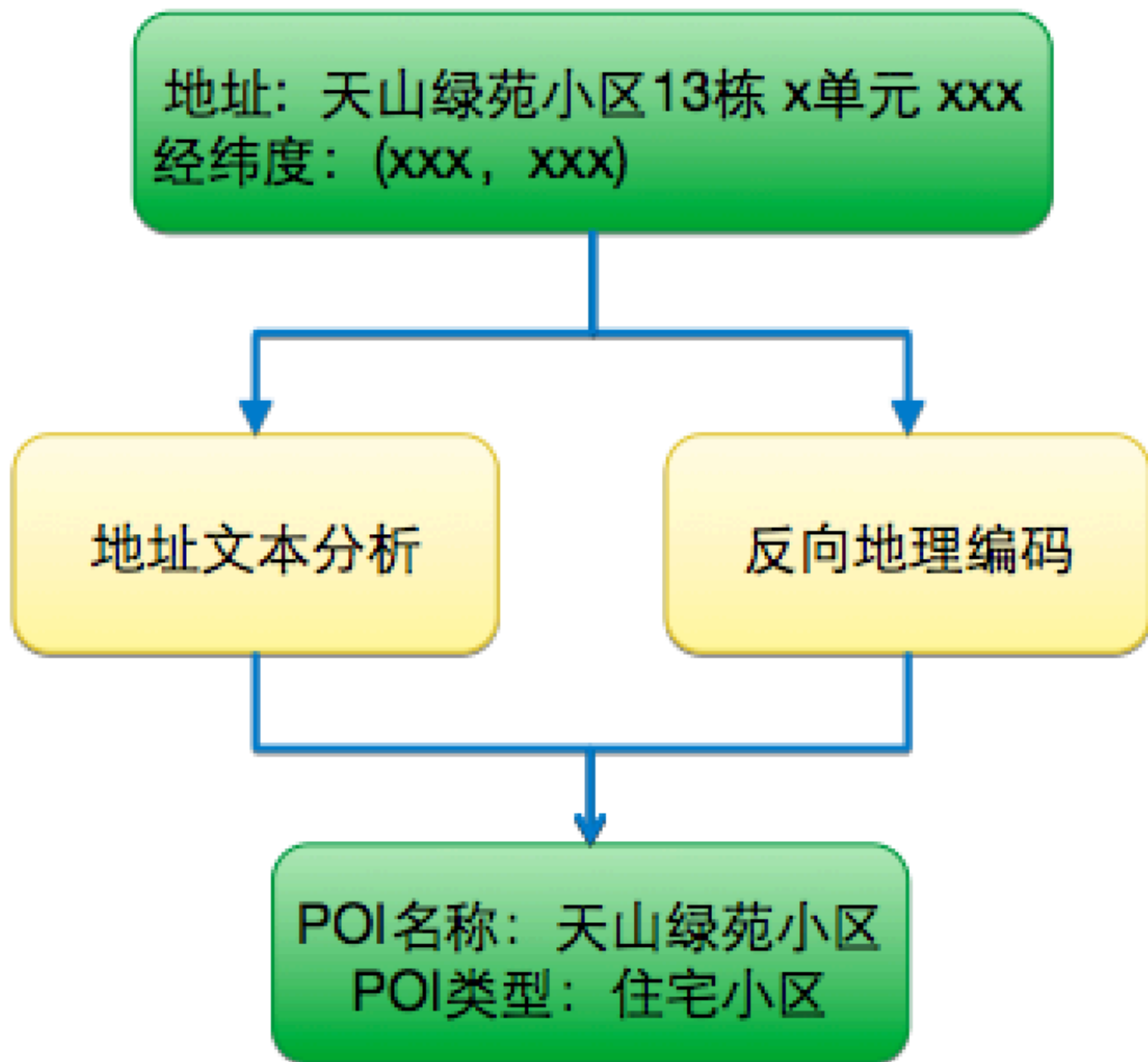
美团外卖通过使用了用户流失预警模型，显著降低了用户留存的运营成本。

3.3 场景运营

拓展用户的体验，最重要的一点是要理解用户下单的场景。了解用户的订餐场景有助于基于场景的用户运营。对于场景运营而言，通常需要经过如下三个步骤：



场景可以从时间、地点、订单三个维度描述。比如说工作日的下午茶，周末的家庭聚餐，夜里在家点夜宵等等。其中重要的一点是用户订单地址的分析。通过区分用户的订单地址是写字楼、学校或是社区，再结合订单时间、订单内容，可以对用户的下单场景做到大致的了解。



上图是我们订单地址分析的流程。根据订单系统中的用户订单地址文本，基于自然语言处理技术对地址文本分析，可以得到地址的主干名称（指去掉了楼宇、门牌号的地址主干部分）和地址的类型（写字楼、住宅小区等）。在此基础上通过一些地图数据辅助从而判断出最终的地址类型。另外我们还做了合并订单的识别，即识别一个订单是一个人下单还是拼单。把拼单信息、地址分析以及时间结合在一起，我们可以预测用户的消费场景，进而基于场景做交叉销售和向上销售。

四、总结

外卖的营销特征，跟其他行业的主要区别在于：

外卖是一个高频的业务。由于用户的消费频次高，用户生命周期的特征体现较显著。运营可以基于用户所处生命周期的阶段制定营销目标，例如用户完成首购后的频次提升、成熟用户的价值提升、衰退用户的挽留以及流失用户的召回等。因此用户的生命周期是一个基础画像，配合用户基

本属性、偏好、消费能力、流失预测等其他画像，通过精准的产品推荐或者价格策略实现运营目标。

用户的消费受到时间、地点等场景因素驱动。因此需要对用户在不同的时间、地点下消费行为的差异做深入了解，归纳不同场景下用户需求的差异，针对场景制定相应的营销策略，提升用户活跃度。

另外由于外卖是一个新鲜的事物，在用户对一些新品类和新产品缺乏认知的情况下，需要通过技术手段识别用户的潜在需求，进行精准营销。例如哪些用户可能会对小龙虾、鲜花、蛋糕这样的相对低频、高价值的产品产生购买。可以采用的技术手段包括用户分群、对已产生消费的用户做 look-alike 扩展、迁移学习等。

同时我们在制作外卖的用户画像时还面临如下挑战：

- 1) 数据多样性，存在大量非结构化数据例如用户地址、菜品名称等。需要用到自然语言处理技术，同时结合其他数据进行分析。
- 2) 相对于综合电商而言，外卖是个相对单一的品类，用户在外卖上的行为不足以全方位地描述用户的基本属性。因此需要和用户在其他场合的消费行为做融合。
- 3) 外卖单价相对较低，用户消费的决策时间短、随意性强。不像传统电商用户在决策前有大量的浏览行为可以用于捕捉用户单次的需求。因此更需要结合用户画像分析用户的历史兴趣、以及用户的消费场景，在消费前对用户做适当的引导、推荐。

面临这些挑战，需要用户画像团队更细致的数据处理、融合多方数据源，同时发展出新的方法论，才能更好地支持外卖业务发展的需要。而外卖的上述挑战，又分别和一些垂直领域电商类似，经验上存在可以相互借鉴之处。因此，外卖的用户画像的实践和经验累积，必将对整个电商领域的大数据应用作出新的贡献！