

数据挖掘在精确营销中的应用

作者：段勇

华院分析技术(上海)有限公司 技术总监

引言

美国前邮政部长、美国百货商店之父约翰·华纳梅克曾这样感叹到：“我在广告上的投资有一半是无用的，但问题是我不知道是哪一半。”

“现代营销之父”菲利普·科特勒先生也曾指出：“促销费用的大部分都打了水漂，仅有 1/10 的促销活动能得到高于 5% 的响应率，而这个可怜的数字还在逐年递减。”

这是专家们对传统市场营销体系及理论缺陷的高度概括。

那么是什么原因使曾经风靡一时的“4P”，“4C”等传统营销理论在实战中大打折扣了呢？我们认为日趋激烈的市场竞争是其中一个因素，但更重要的因素是客户日趋个性化的偏好与需求。面对客户的多样化、层次化和个性化的偏好与需求，传统大众化的营销就失去了优势。本文将就基于客户需求的精确营销的概念和方法以及其中数据挖掘的应用展开详细的阐述。

精确营销与大众营销的对比

那么基于客户需求的精确营销与大众营销相比有什么特点和优势呢？

首先，我们说精确营销是真正以客户为导向，侧重于满足客户个性化需求，通过分析每个客户的消费行为和偏好，进而解决两个问题：哪些用户是某个产品或者营销活动的目标用户？每个用户最适合给他推荐的产品是什么？而传统营销则要笼统的多，虽然也是基于用户细分来设计营销组合，但是并不知道每个客户个体的偏好以及适合推荐的产品，最终的营销是针对某几个用户群体，这样也就忽视了细分用户群体里的差异化、个性化的需求，效果自然也就没有精确营销好。

其次，精确营销是一个基于数据分析的量化过程，是以事实为依据，是对用户使用行为和偏好的精准衡量和分析，从而实现对客户精准定位和对不同客户不同业务内容的精确推荐。而传统营销更多采用市场调研方式了解客户消费行为及偏好，定性分析和主观因素更多，而且客户某些潜在的需求和间接的偏好是无法通过调研得出所有答案的。

第三，精确营销的推广销售群体是有针对性的一部分的目标用户，而传统营销则面对的是所有大众。我们来看一下这个例子：某企业有客户群 25 万人，希望对他们做一次邮寄的促销活动，每一个用户邮寄成本为 1.5 元，如果客户对促销活动响应，平均能带来 200 元的

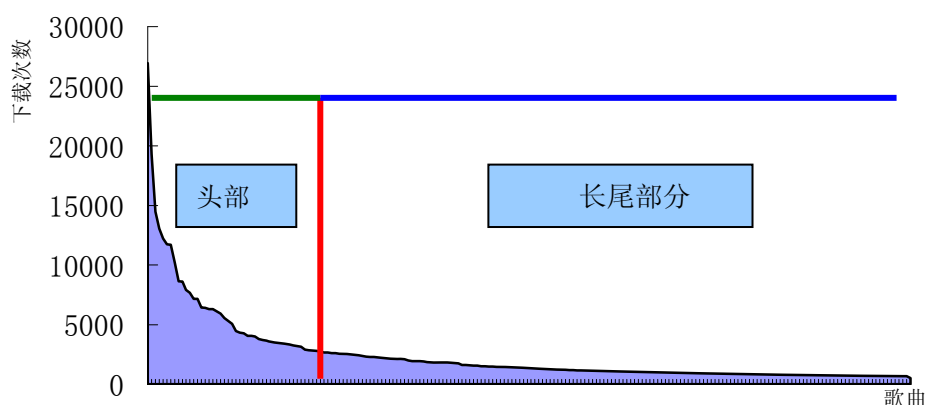
利润。对 25 万用户全部邮寄，如果响应率在 1% 左右，那么收益 = $250000 * 1\% * 200 - 250000 * 1.5 = 125000$ ；通过精确的目标用户筛选，选择 2.5 万用户（是总用户的 10%），如果响应率达到 5%（是原来的 5 倍），那么收益 = $25000 * 5\% * 200 - 25000 * 1.5 = 212500$ ，比对全体用户邮寄的收益提高了 87500 元。通过这个例子我们可以看到精确营销在节约营销成本，提高利润水平上无疑比传统营销更具优势。

精确营销与长尾理论

在传统营销中，由于资源的限制和对高营销效率的追求，企业更关注重要的人和重要的事，即重点针对创造 80% 利润的 20% 的客户做营销，这就是我们通常所说的“二八”定律。但“二八”定律对于以用户需求为中心的精确营销来讲是不适用的。因为我们对用户分析的目的就是要找出每一个个体在需求上的差异，进而针对这种差异所产生的个性化需求进行定向的精确营销。

此外，用户需求的差异化和产品种类的丰富性也使用户的选择趋于多样化。以移动电话运营上的彩铃业务为例，可以供客户下载的歌曲有上万首，这样用户便面临着无限的选择，而其中的每一首歌曲都有可能被用户下载，尽管绝大部分歌曲下载的需求和实际下载量并不高，但这些处于长尾部分的下载量占总下载量的比例加在一起却可能超过正态曲线分布中处于头部位置主流歌曲的比例，也就是说那些不流行的占绝大多数的彩铃相对于流行的少数主流彩铃所创造的收入和利润要更多，这就是目前颇为流行的“长尾理论”。

图 1：“长尾理论”示意



长尾理论告诉我们，不仅要关注处于传统需求曲线上那个代表“畅销品”的头部；更要关注所谓“冷销品”的长尾部，这就需要我们更深入地研究目标客户群体和个体之间的需求差异。

数据挖掘在精确营销中的应用

让我们回到前面某企业做邮寄促销的例子，如何才能获得筛选 10%的用户，响应率提高 4 倍（是原来的 5 倍）的效果呢？这就是数据挖掘的威力。

精确营销解决的问题是：哪些用户是某个产品或者营销活动的目标用户？或者每个用户最适合给他推荐的产品是什么？前者是为产品或者营销活动筛选目标客户；后者是对用户进行产品推荐，从本质上来说，两者是类似的。数据挖掘正是通过对客户消费行为数据和历史规律的挖掘与分析，进而可以找到目标用户的特征，实现以客户为中心的精确营销。

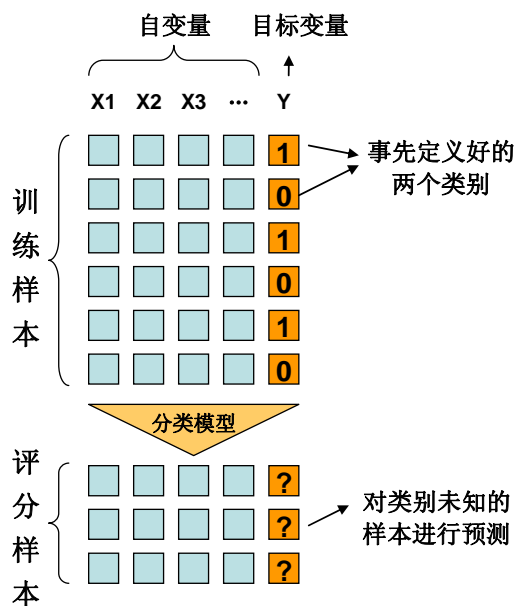
数据挖掘是对一系列对数据进行分析 and 挖掘的方法的统称，在精确营销领域，最常用的数据挖掘方法主要包括以下三类（完整来说数据挖掘范围会更广，但本文仅侧重描述如下三类）：分类、聚类、关联。

→分类(Classify)

分类是数据挖掘应用最广泛的应用之一，分类属于预测性模型，分类模型解决的问题是对类别未知的用户进行预测，以判断其属于哪个类别的概率比较高。例如事先定义用户的信用状况分为两类：信用好、信用坏。对于一个信用状态未知的用户，如果需要预测其最有可能属于哪个类别（信用好/信用坏），这就需要构建一个分类模型。

分类模型的构建需要一个“类别已知”的历史样本，我们称之为“训练样本”。由于训练样本中每一个个体的类别都是明确的，因此可以通过分类的算法找出能显著区隔不同类别的典型特征，这些特征就是分类模型的结果。特征变量一般称为“自变量”又叫“预测变量”，类别变量称为“目标变量”。通过训练样本找出来的特征，对新样本（又称“评分样本”）进行预测，以判断满足不同特征的用户属于不同的类别，如图 3 所示。

图 2：分类模型示意图

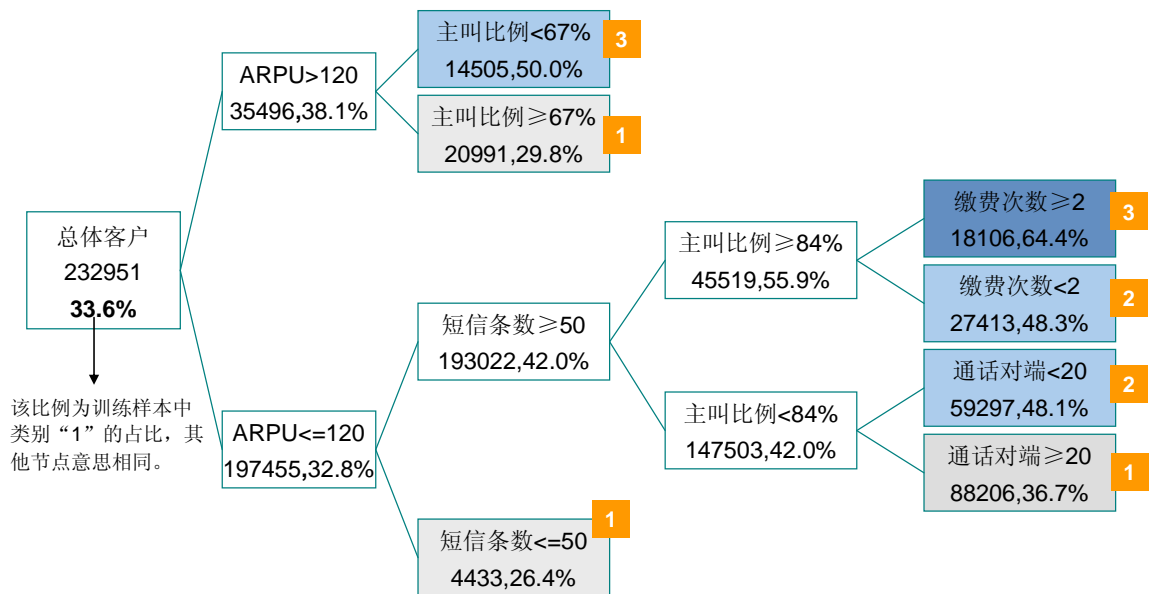


分类在精确营销中的应用

某电信运营商正在推销某种增值业务，需要寻找有购买潜力的目标用户特征，通过数据挖掘分类模型可以非常直观准确的找到目标用户：如图 3 所示：

训练样本中有 33.6%的用户已经订购了该增值业务（定义为目标变量取值“1”），其余 66.4%的用户均未订购（定义为目标变量取值“0”）。采用“决策树”方法构建分类模型，可以直观看到满足“ $ARPU > 120$ 并且 主叫比例 $< 67%$ ”特征的用户中有 50%订购了该增值业务，显著高于总体中的 33.6%，因此可以认为满足该特征的用户购买该增值业务的可能性比较高。同样，我们还可以看到“ $ARPU \leq 120$ 并且 短信条数 ≥ 50 并且 主叫比例 $\geq 84%$ 并且 缴费次数 ≥ 2 ”的用户购买可能性会更高，达到了 64.4%。

图 3：通过“决策树”方法构建的分类模型



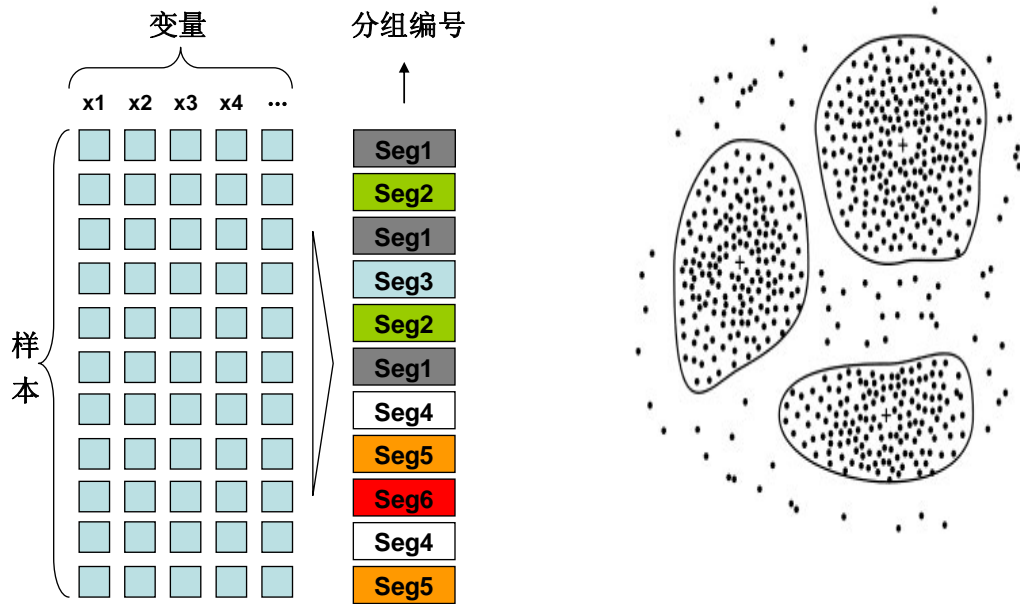
决策树是分类模型中最常用的方法之一，具有预测精度高，预测结果稳定性高，结果易理解等优点。除了决策树之外，Logistic 回归、神经网络、判别分析等方法也可以构建分类模型。

→ 聚类(Clustering)

与分类模型有着本质的区别，聚类模型属于非预测模型（描述型模型）。聚类模型解决的问题是对用户进行分组（或者叫分群），特征相似用户在一个组内，特征不同的用户分在不同的组。

聚类模型不需要“目标变量”，只需要给定自变量，聚类模型就可以自动的对用户进行分组，输出每个样本对应的组编号，如图 4 所示。

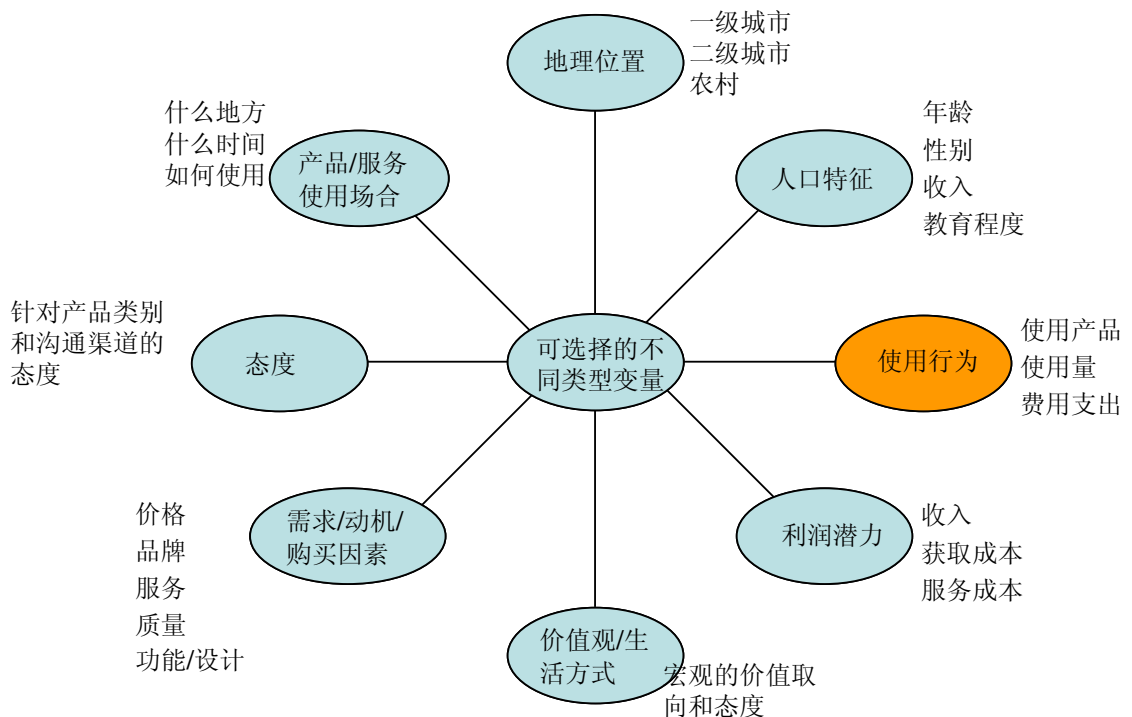
图 4：聚类模型示意图



聚类在精确营销中的应用

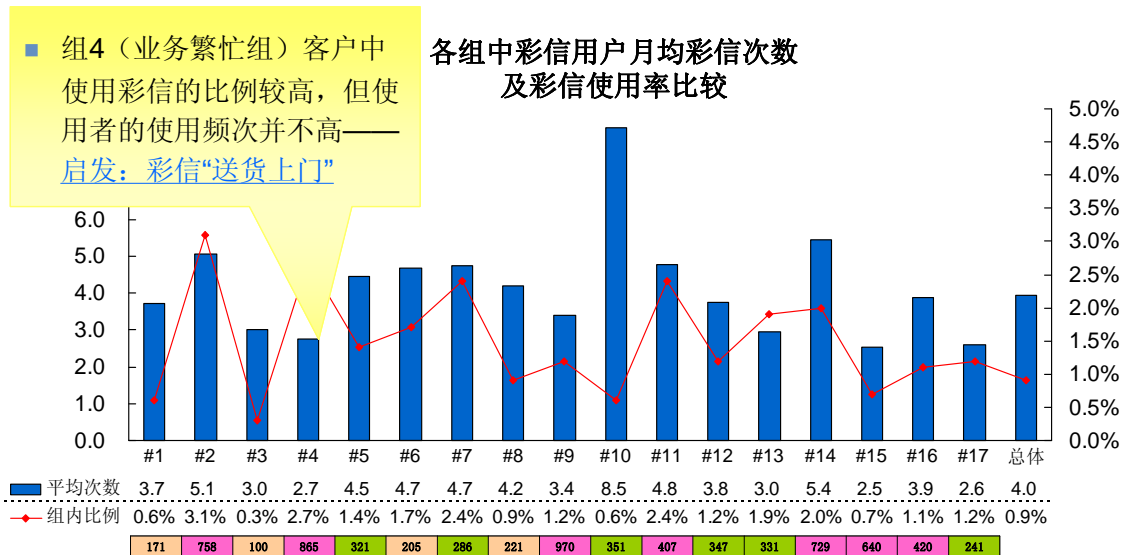
选择聚类所需的变量是构建聚类模型最关键的工作,变量的选择往往取决于应用的目标要求,一般来说,有8种不同类型的变量,如图6所示:

图6: 不同类型的变量。



选择符合应用需求的变量来构建聚类模型,对用户进行分组,如下图所示,针对聚类生成的17个细分群体,可以进一步统计其产品使用情况,进而可以找到针对不同群体实施精确营销的机会。

图7: 针对聚类生成的不同细分群体产品使用情况统计



注释：左轴表示本组内使用彩信者的月均彩信次数，右轴指这些客户占本组客户总数的百分比。

→关联(Association)

关联模型与聚类模型都属于非预测模型，其主要解决的问题是研究产品购买的关联性，即买 A 产品的同时是否会对 B 产品也很感兴趣。经典的数据挖掘案例“啤酒和尿布”就属于关联模型。

关联模型又叫“购物篮分析”，这个名字很形象，在超市购物时一个购物车中往往会放多种不同的商品，通过对大量的购物车进行分析，这些商品之间可能会存在众多意料之中或意料之外的关联性。关联模型中度量两个产品关联性强弱主要用三个指标：

针对两个商品 A, B，研究 $A \rightarrow B$ 的相关性，三个指标的定义如下：

支持度(Support): 表示 A、B 同时购买的人数占总购买人数的比例。支持度越高，表示 A,B 商品同时购买的人数越多，这两个商品越主流。

可信度(Confidence): 表示在购买 A 商品的人中同时购买了 B 商品的比例。可信度越高，表示购买了 A 商品后再购买 B 商品的可能性就越大。

提升度(lift): 可信度除以总用户中购买过 B 商品的用户占比。提升度越高，表示购买了 A 商品对购买 B 商品的影响度就越大，也即他们之间的相关性就越强。

关联模型的示意图如下所示：

图 8：关联规则模型示意图



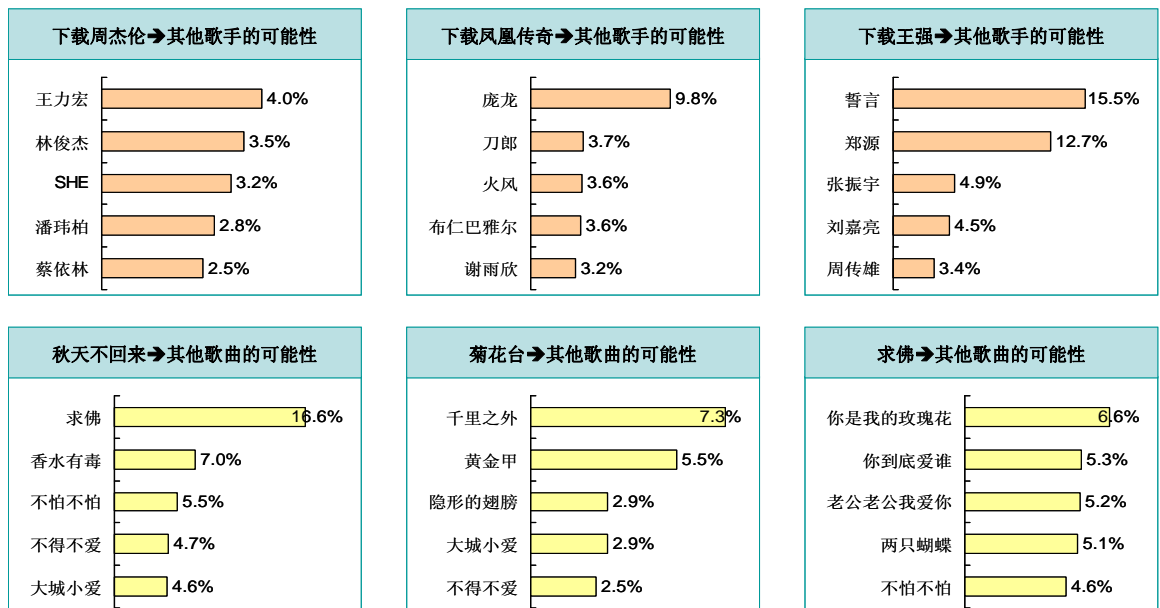
<u>Rule</u>	<u>Support</u>	<u>Confidence</u>
$A \Rightarrow D$	2/5	2/3
$C \Rightarrow A$	2/5	2/4
$A \Rightarrow C$	2/5	2/3
$B \& C \Rightarrow D$	1/5	1/3

关联模型在精确营销中的应用

关联模型主要可以解决两大类问题：1、如何对用户进行商品推荐？2、把哪些商品捆绑在一起销售更好？前者我们叫“交叉销售”的问题，后者叫“捆绑销售”。

以电信运营商的彩铃业务为例，我们把歌曲或者歌手当做商品来研究，用户在订购歌曲或者某个歌手的歌曲时的关联性如下图所示：

图 9：彩铃关联规则分析



从上图可以看出：下载过周杰伦歌曲的用户中，还下载过王力宏的比例最高，林俊杰次之。因此可以针对下载过周杰伦歌曲的用户推荐王力宏或者林俊杰的歌曲（交叉销售）或者把周杰伦、王力宏、林俊杰的歌曲捆绑在一起打折销售给客户（捆绑销售）。如下图所示：

图 10：基于关联规则的交叉销售和捆绑销售示例

1

基于铃声关联性的 交叉推荐

下载了周杰伦→王力宏、林俊杰、SHE、潘玮柏
下载了王强→誓言、郑源、张振宇、刘嘉亮
下载了求佛→你是我的玫瑰花、你到底爱谁
下载了秋天不回来→求佛、香水有毒、不怕不怕

2

基于铃声关联性的 捆绑销售

求佛、秋天不回来、香水有毒捆绑销售3元
菊花台、千里之外、黄金甲捆绑销售3元

写在最后

商业竞争的日益激烈和信息技术的突破性进展，营销界正在爆发一场意义深远的革命，无论是营销理论还是实践都面临着一种结构化的转型：从传统的、大众的和粗糙的方法跃变到深度化、细分化和精确化的模式。任何公司要在这场革命中占领先机，都必须坚定不移地聚焦客户，并开始向精确营销转型。

数据挖掘技术作为支撑精确营销的重要手段，随着企业对精确营销认识的提升和需求的加强，以及其本身算法的不断完善，必将在未来的营销领域中发挥强大的作用。