

一种使用 SEM 广告展现数据优化投放策略的方法

常莹¹

1. 北京博雅立方科技有限公司，背景，cycufepku@126.com

摘要：本文主要介绍了一种使用公开的 SEM 广告展现数据，判断行业竞争态势及竞争对手动向、并进而优化 SEM 投放的方法。此方法由 SEM 第三方代理公司根据教育培训行业客户的通用需求研发，但具备通用性，可以便捷地复制到其他行业或供某一广告主使用。研究中所使用的主要数据是搜索引擎公开展示的 SEM 广告。整套方法由一个概念模型和一系列用于完成具体数据分析任务的描述性统计方法组成。分析结果包含一套刻画行业(也可以是子行业/业务线/竞争对手/行业重要关键词等维度)SEM 广告投放策略特征及变化趋势的图表，以及据此生成的对市场趋势的判断和优化投放建议。这种方法所需的数据容易获取、方法可扩展性好，产出可用来制定有针对性的投放策略和优化方案，是一种非常经济监控竞品以及提升广告投放效果的方法。

关键词：SEM 广告；品牌观察模型；市场监控；广告优化；教育培训

一、问题的引入

中国互联网信息中心(CNNIC)于 2016 年 1 月发布数据显示，我国网民规模在 2015 年底已达 6.9 亿，其中 82%的网民使用搜索引擎。互联网和搜索引擎成为终端消费者的重要聚集地，必然会带动企业将营销推广重心向线上及搜索引擎转移。有超过 1/3 的企业开拓有至少一个互联网营销渠道，其中近半数在进行 SEM 推广。SEM 是 Search Engine Marketing 的首字母简写词，是广告主在搜索引擎上投放的竞价关键字广告。这种广告形式带来的收入在互联网广告总体收入中的占比已接近 40%，是广告主进行线上营销最重要的推广方式。

教育培训行业是在营销上高度依赖搜索引擎的代表性行业之一。根据国内市场份额最大的搜索引擎百度发布的财报，该行业一直是 SEM 广告投入最高的五个行业之一。它表现出两个鲜明的特征：首先，行业快速发展，前景看好，据估计行业规模将在几年内超过万亿。其次，业内竞争者众多，领先品牌的优势仍然主要体现在市场占有率方面，各品牌的产品和服务差异化程度不高。这使得想要维持或扩大市场规模的竞争者高度依赖市场推广，SEM 广告作为其最为重要的推广

手段之一更是从业者彼此竞争的重点领域。各品牌的推广成本随着竞争激烈程度的提高不断飙升，单位客户成本在近两年内即接近翻番，给企业盈利造成了巨大压力。

激烈的竞争使得教育培训行业各品牌之间的攻防意愿非常强烈。广告主都希望能够全面了解市场大势及竞品动向，在知己知彼的基础上有针对性地制定自己的推广策略。但实际情况与广告主的期望有很大差距。绝大部分广告主(包括行业领先的厂商)，对市场趋势及竞争对手的监控都还停留在手工随机查看对手广告或零星搜集商业情报的阶段，缺乏系统的方法论和高效的实施方法。

笔者所服务公司是百度五星级代理商之一，在教育培训行业 SEM 代理服务方面占有绝对的市场优势。有感于客户对知己知彼及成本可控两个目标的迫切愿望，公司为教育培训行业定制了由复合转化模型与品牌观察模型组成的解决方案。前者通过打通在线推广与在线转化数据，帮助客户优化推广资源分配。后者帮助客户监控市场及竞品的重要动向，并提供推广优化建议。本文介绍品牌观察模型结构与实施。它所使用的数据较易获取，所需开发工作量相对可控，分析结果可供对 SEM 广告策略各主要方面进行优化，是一种性价很高的市场监控方式。

二、 问题与数据描述

为行文便利起见，以下将以笔者所在公司为客户 A 提供的一次品牌观察模型服务为例子进行说明。其他客户、多个搜索引擎以及类似广告形式的监控，皆可以类似的方法得到。

客户 A 是职业教育培训领域领先的上市公司，SEM 推广带来的客户在其客户总量中的占比约为 70%。因近年在搜索营销方面投入巨大，A 公司受到了来自资本市场的紧迫压力，急需控制推广成本、提升推广效果。同时，作为所在行业的标杆企业之一，业界众多竞品也一直在试图通过抢夺 A 的(潜在)客户资源来扩大自身的市场份额。为了守护并继续扩大自身的市场份额，A 需要及时根据市场情况做出应对。A 面临的挑战正是品牌观察模型所要解决的问题。

实施此方法需要准备两份数据：行业关键词样本、样本关键词激发的 SEM 广告。首先对 SEM 广告的基本业务逻辑做简要介绍。图 1 是百度提供的一个典型的、呈现了 SEM 广告的搜索结果页。SEM 广告以关键词为单位，广告主购买的每一个关键词都是一条广告。用户在页面最上端的搜索框(左侧上方方框)中输入搜

索词后会看到搜索结果页。如果这个搜索词恰好也是某些广告主购买的关键词广告的话，就会有广告被呈现出来，否则就只会呈现自然搜索结果。

广告以图 1 中右侧方框框出的“推广链接”为标识。广告语(左侧中间方框)被称为“创意”，由广告主根据关键词内容撰写，第一行为“标题”、余下为“描述”；创意中与搜索词匹配度最高的部分以红色显示，称为“飘红”；“显示 URL”(左侧下方方框)可用来识别购买广告的广告主。创意和显示 url 合并在一起，组成一条完整的关键词广告。广告主为每一个关键词设定一个“出价”，当多个广告主同时购买一个关键词，出价是决定各广告主广告排名的关键性因素。广告被点击时，广告主向搜索引擎支付一定费用。上述说明对 SEM 广告的业务逻辑进行了适度简化，有兴趣的读者请查阅相关文档。



图 1 百度搜索结果页 SEM 广告示意图

“行业关键词样本”，是自图 1 左侧上方方框中客户输入的搜索词中选取出来的一个样本。之后通过爬虫或搜索引擎提供的 API 接口，轮循抓取并记录搜索每个关键词时结果页上呈现的广告，即为“样本关键词激发的 SEM 广告”数据。具体的数据获取方式如下：

行业关键词样本对应的总体，是该行业所有广告主在百度购买过的关键词的总和；只有百度掌握这一数据，并且不会向任何代理商和广告主开放。品牌观察模型以笔者所在公司自行建立的行业词库中包含的全部关键词作为真实总体的替代。这一词库中的关键词来自于三个来源：一、公司服务过的该行业广告主购买过的关键词；二、通过搜索引擎提供的拓词接口增补的关键词；三、在相关网站抓取的文本中提炼出来的关键词。该词库是可获取到的、对真正总体最为近似的模拟。无论是真实总体还是公司行业词库，所包含的关键词数量都非常庞大；而抓取广告数据需要消耗一定资源，出于成本考虑，需要选取一个有代表性的关键词样本。

品牌观察模型采取分层抽样的方式选取样本。分层标准是行业词库中的分类及子类。表 1 提供了部分关键词示例。分类中的“通用词”指可以标识行业的关键词，“产品词”是能够标示某一种产品/服务的关键词，“品牌词”包含行业中某一公司的品牌名称。对每一分类都可进一步细分，通常至少会将品牌词按照关键词对应的品牌进行细分。一个成熟的 SEM 账户会涵盖各个类别的关键词，以便全面接触具有各种搜索习惯、处于购买决策各个阶段的潜在客户；这是品牌观察模型在对关键词进行抽样时选择分层抽样方法的主要原因。“搜索量”表示搜索此关键词的用户数量大小，此处有模糊化处理。搜索量由搜索引擎通过 API 接口提供，按照一定时间周期更新。

表 1 行业关键词样本示例

关键词	分类	子类	搜索量
职业培训	通用词		1000
职业技术培训	通用词		300
.....
java 培训	产品词	技术	3000
会计培训班	产品词	会计	4000
.....
达内培训	品牌词	达内	6000
中公	品牌词	中公	6000
华图	品牌词	华图	6000
华图教育	品牌词	华图	6000
.....
JAVA 培训哪家好	长尾词	疑问	10
北京会计实操培训班	长尾词	其他	10
.....

本案例共选择了 3,300 个关键词样本。在从百度抓取样本关键词激发的广告时，可供选择的方式有爬虫和百度提供的搜索实况 API 两种。前者适用于任何搜索引擎；但需自行开发部署抓取程序，其稳定性面临搜索引擎封禁等问题的挑战。

以外,过于频繁的抓取频率会降低关键词的点击率,进而影响关键词的排名位置。因此抓取页面的频率不可过于频繁。后者在开发成本和稳定性方面表现相对较好,不会影响关键词点击率;但不能移植到其他搜索引擎,且获取配额有一定门槛,需求量过大还需向百度支付费用。本案例使用公司现有的免费搜索实况 API,所得数据如被表 2 所示:

表 2 样本关键词激发的 SEM 广告示例

字段	取值示例
时间	2016/6/5 16:42
关键词	java 培训
显示 URL	www.sample.com
排名	2
创意标题	java 培训专注 java 培训 15 年
创意描述	北京 java 培训中关村 A, 专注 IT 教育, 50 万学员的共同选择. java 培训 A 金牌讲师授课 java 培训免费咨询电话:...
标题飘红	java 培训 java 培训
描述飘红	北京 java 培训 java 培训 java 培训

本案例在抓取广告时,将一天分为 4 个等长时段,每个时段对全部样本关键词抓取一遍其在北京(不同地区的广告展示结果可能不同)的广告展现结果,并令一个关键词在某一时段内的抓取时间服从均匀分布。以 2016 年 3 月 28 日至 4 月 3 日一周为例,共回收 67.9 万条广告,有 2,713 个广告主至少有一次广告展现,平均每次搜索展现 14.1 条广告。表 3 是抓取数据中回收到的每个广告主的广告数量、每个广告主购买的关键词个数这两个变量的四分位数。两个变量的分布都严重右偏,少数广告主占领了绝大多数广告展现资源,大量中小广告主仅进行非常小规模投放。中位数之下的广告主绝大部分并不从事职业教育培训,只是偶然购买了关键词样本中的一个关键词。例如代理会计服务行业广告主也会购买关键词样本中的关键词“会计”。删除这些非本行业的数据,留下 1,677 个广告主。

表 3 广告主投放广告基本情况

分位点	0%	25%	50%	75%	100%
分位数-广告数量	1	1	7	80	39,276
分位数-购买关键词数量	1	1	2	13	2,785

通常行业领先的广告主更关注与其规模相当或接近的竞争者，表 4 选择了包括客户 A 在内的、占有广告资源最多的 5 个广告主，展示它们占用广告资源的基本情况。A 是 5 个广告主中广告展现最多的一个，同时其关键词平均排名也远低于竞争对手。这通常意味着，广告主 A 为了维持更好的关键词展现位置，采用了普遍的高出价策略以使得大量关键词在更好的排名位置展现；它正是造成之前提到的、推广成本快速增加的主要原因。其他四个竞争对手在购买的关键词数量上与 A 相差不多，但出价策略相对保守，因此其展现位置明显偏低（平均排名数值更大）、同时会有更多广告不能展现的情况（每个关键词平均展现广告数较 A 明显偏低）。综上，A 以高额的广告费用维持了更充分地曝光，竞争者 B、C、D、E 的出价策略远不如 A 激进，有利于控制成本，但牺牲了部分曝光量和更好的展现位置。

表 4Top5 广告主基本情况

广告主	关键词个数	广告条数	每个关键词平均展现广告数	平均排名
A	2,707	39,276	14.5	2.4
B	2,785	25,888	9.3	4.5
C	2,731	25,031	9.2	6.2
D	2,683	24,269	9.0	7.2
E	2,581	23,906	9.3	6.3

三、 分析方法与实施

品牌观察模型主要用于帮助客户了解所在行业的整体竞争态势、剖析主要竞争者的策略特征与重要变动；并在此基础上，制定富有竞争力的投放策略与落实

计划。模型由行业层面和策略层面两个部分的分析组成。行业层面的分析主要完成如下几个任务：描绘投放市场竞争格局、圈定竞争对手、预判投放市场发展趋势。策略层面的分析主要围绕竞争者之间的攻防格局、各广告主购买关键词及资源配比情况、广告创意热点三个方面展开，为用户提供增删关键词、调整关键词价格、撰写创意素材建议三个方面的优化措施。品牌观察模型既可单独提取某段时间的数据进行一次性的分析，也可将各次分析的结果进行对比，前者是后者的基础。以下将主要陈述截面上的分析方法，加入时间轴后的分析主要是采取同比/环比/时间序列模型等手段定位变化异常的指标、广告主和关键词，不再一一赘述。

（一）行业层面分析

为方便说明计算逻辑，首先定义符号。假定分析中共涉及到 m 个广告主，对其按照抓取广告条数倒序排列，编号为 $1, 2, \dots, m$ 。关键词抽样共包含 n 个关键词，编号为 $1, 2, \dots, n$ ； n 个关键词的搜索量分别为 q_1, q_2, \dots, q_n 。 $I_{ij} (i=1, 2, \dots, m \quad j=1, 2, \dots, n)$ 是 0-1 变量，抓取到第 i 个广告主关于第 j 个关键词至少一条广告时取值为 1，否则为 0； $C_{ij} (i=1, 2, \dots, m \quad j=1, 2, \dots, n)$ 表示抓取到第 i 个广告主关于第 j 个关键词的广告条数。显然 $I_{ij}=0$ 时， $C_{ij}=0$ ； $I_{ij}=1$ 时， $C_{ij}>0$ 。对于 $C_{ij} \neq 0$ 的每一对 (i, j) ，以 $R_{ijk} (k=1, 2, \dots, C_{ij})$ 表示第 i 个广告主关于第 j 个关键词抓取到的各条广告的排名。于是对于每一个 $C_{ij} \neq 0$ 的

$$(i, j), \text{ 第 } i \text{ 个广告主关于第 } j \text{ 个关键词的平均排名为 } R_{ij} = \sum_{k=1}^{C_{ij}} R_{ijk} / C_{ij}。$$

行业层面的分析包含截面和时序两个部分，前者主要了解当前的推广格局，后者主要用于追踪和预测市场发展趋势。截面上的分析采用四象限分析法，可识别市场上众多竞争者在 SEM 推广上的竞争地位和基本推广策略，展现形式如图 2 所示。图中的纵轴“品牌推广度指数”是广告主购买的各关键词的搜索量的加权汇总，体现某一品牌希望在线上市场上取得的品牌地位。对于第 i 个广告主，该

$$\text{指数计算公式如下： } ip_i = \sum_{j \in (C_{ij} \neq 0)} \frac{C_{ij}}{R_{ij}} \times q_j \quad i=1, 2, \dots, m。$$

上式以广告条数和平均排名对广告主购买的关键词检索量进行了加权。这是因为，即使两个广告主都购买了某个关键词，但是在同样的抓取条件下，获得的广告数越多 (C_{ij} 越大)，说明客户在搜索关键词时看到该广告主广告的可能性越大，广告主在此关键词上的投入通常也越高。同理，平均排名越靠前 (R_{ij} 越小)，在此关键词上的投入通常也更高，因此平均排名以倒数的形式进行加权。

很多行业存在淡旺季，行业的推广力度会随时间整体性地发生明显变化。同时关键词表抽样也不是一成不变的，会随市场的发展有所增删。比如 2009 年之前，电商行业的关键词抽样中不会有“双十一”类的词，但是在双十一成为大促节点后，一个有代表性的关键词抽样必然要囊括此类关键词的代表。为便于比较各广告主在不同时段的推广力度的差异，通常要对原始的品牌推广度指数 ip_i ($i = 1, 2, \dots, m$) 进行标准化处理，得到 IP_i ($i = 1, 2, \dots, m$)。



图 2 行业层面-截面-示意图

图 2 的横轴“品牌知名度指数”代表某广告主当前被线上消费者认知的程度。获取这一数据方式之一是对样本关键词表中每个广告主对应的品牌词的检索量取平均值，即 $if_i = \frac{\sum_{j \in \text{广告主} i \text{ 的品牌词}}^n I_{ij} \times q_j}{\sum_{j \in \text{广告主} i \text{ 的品牌词}}^n I_{ij}}$ $i = 1, 2, \dots, m$ 。之后再对 if_i ($i = 1, 2, \dots, m$) 进行标准化，得到 IF_i ($i = 1, 2, \dots, m$)。对于那些行业公开数据较为丰富的行业，则可选取其他更能代表广告主整体实力的指标，例如市场份额、营收或利润等。

得到 IP_i ($i = 1, 2, \dots, m$) 和 IF_i ($i = 1, 2, \dots, m$) 之后，即可绘制如图 2 所示的

图形。广告主被分隔为四个区域：知名度高也大力推广的“领导者”，知名度高但是推广策略保守的“追随者”，以相对自己知名度要强的力度提升市场地位的“挑战者”，实力不高也无力支撑高额推广费用或专注于某一细分市场的“拾遗补缺者”。在具体使用过程中，这张图会为用户提供至少以下三种方面的重要信息：

第一、自己在哪个象限？与目标或预期是否有差距？差距在哪个方向？

第二、自己的对手和挑战目标是谁？要特别注意那些在图上离自己距离近的广告主，或者是那些表现与自己预期不符的广告主。

第三、位于挑战者区域的广告主，在以相对自身知名度更强的力度推广，它们通常在试图改变自己市场地位甚至是整个市场格局，值得所有竞争者关注。

最后提示一个特例，位于“追随者”象限的竞争者，通常是处于收缩或衰退阶段、因而推广策略相对保守的广告主，但也有例外。当一个品牌强大到成为一个行业里绝对的强势品牌时，即使它只投放很少的广告，消费者也会主动去搜索他们。这样的品牌通常出现在图形中非常偏右的位置，它们不是在追随他人，而是在领跑市场。

接下来通过各品牌的推广力度计算整个市场的推广力度指数。它通常受到两个因素的影响：整个行业的广告主数量、各广告主的推广力度。可采用如下公式计算： $im = m \times \sum_{i=1}^{100} ip_i$ 。使用未经标准化的品牌推广度指数，可以捕捉市场推广力度的变化。推广力度最大的 100 名品牌的推广力度指数之和表示市场总体推广力度，业内广告主数量 m 表示市场规模；两者中的任何一个扩大，都表示市场竞争激烈程度的增加。

当监控持续进行的时候， $IP_i (i=1,2,\dots,m)$ 、 $IF_i (i=1,2,\dots,m)$ 和 im 会扩展为包含时间维度的形式 $IP_{it} (i=1,2,\dots,m \ t=1,2,\dots)$ 、 $IF_{it} (i=1,2,\dots,m \ t=1,2,\dots)$ 和 $im_t \ t=1,2,\dots$ 。这些指数随时间推进发生的重大变化(如 im 或某个品牌的 IF 指数的剧烈变动)，也是需要重点关注的信息。

(二) 策略层面分析

SEM 投放涉及很多具体工作，例如应选取哪些关键词、如何设置账户结构、

如何撰写创意、如何为关键词出价和调价、如何设置匹配模式等。本节主要讨论如何通过公开的广告数据提炼竞争对手在以下三个方面的策略特征：

第一、品牌攻防策略：是否主动攻击竞品、攻击范围和力度、对来自对手的进攻是否进行了有效防守等

第二、关键词购买及出价：关键词配置是否能够合理有效地覆盖到潜在消费者、出价策略的激进程度、如何找到关键词购买和出价方面的提升空间等

第三、广告创意：客户的广告创意与其自我定位是否一致、主要竞品如何设定品牌形象、哪些创意点可以成为品牌特色等

品牌攻防策略部分以截面分析为主，具体实现方法如下：

1、选定分析的品牌范围(A、B、C、D、E)，在关键词抽样中的品牌词类别下，选择与 A 及 B、C、D、E 对应的全部品牌词

2、在广告数据中选择关键词为 1 中选定的品牌词、展示 url 为上述五个品牌之一的广告，得到如表 5 所示数据

3、在进行品牌攻防时，广告主既会购买包含自己品牌名称的关键词，也会购买包含竞品品牌的关键词以抢夺对方市场份额。广告主与品牌实际上是一一对应的同一批实体，在涉及购买关键词行为时，以“广告主”来称呼这一实体，而在同一实体作为被同行攻击的对象时，则称其为“品牌”

表 5 品牌攻防策略分析所使用的广告数据

关键词	品牌	广告主	排名	创意标题	创意描述
A 怎么样	A	A	1	A 怎么样?A 海 淀园中心—薪 资/口碑 NO.1	A 怎么样?A 海淀园中心，上市 IT 培训机构，TTS8.0 系统，24 小时 无缝学习!A 海淀园 PHP 中心…
A 怎么样	A	B	2	B 怎么样? BIT 培训中心，就 业后付款	B 专注泛 IT 职业教育培训，零基础 入学，4 个月入门到精通，独立开 发 App. B，无限用心，授课导师一 对一辅导…
.....
C	C	C	1	C—中国互联 网高端人才培	C 专注泛 IT 职业教育培训，零基础 入学，4 个月入门到精通，独立开

				养创新品牌	发 App. C，无限用心，授课导师一对一辅导...
C	D	D	2	D 全国移动互联网应用开发指定实训基地	html5/html5/HTML5 开发培训乱象终结，D 成为移动互联网官方指定培训基地!html5/html5/HTML5 开发培训再也不用怕
.....

4、假定分析涉及 $m_1 (m_1 \geq 2)$ 个广告主，关键词样本中关于每个品牌的关键词个数分别为 s_1, s_2, \dots, s_{m_1} ，第 $i (i = 1, 2, \dots, m_1)$ 个品牌对应的第 k 个品牌关键词的搜索量为 $q_{ik} (k = 1, 2, \dots, s_i)$ 。以 d_{ikj} 表示第 i 个品牌对应的第 k 个关键词是否被第 j 个广告主购买，是则取 1，否则取 0；则广告主 j 购买品牌 i 的品牌关键词在品牌 i 所有关键词中的占比即为 $D_{ij} = \sum_{k=1}^{s_i} d_{ikj} / s_i$ 。 D_{ij} 越大，表示广告主 j 对于品牌 i 的品牌词购买得越全面，从而对品牌 i 的竞争意图越强烈。

对于每一个 $d_{ikj} \neq 0$ ，以 a_{ikj} 表示第 i 个品牌对应的第 k 个关键词被第 j 个广告主所购买时，爬虫抓取到的广告条数；则广告主 j 在其所购买的所有品牌 i

关键词的词均广告展现次数为 $A_{ij} = \begin{cases} \sum_{k=1}^{s_i} a_{ikj} / \sum_{k=1}^{s_i} d_{ikj} & (\sum_{k=1}^{s_i} d_{ikj} \neq 0) \\ 0 & (\sum_{k=1}^{s_i} d_{ikj} = 0) \end{cases}$ 。这个值越大，表

示消费者在搜索品牌 i 的品牌词时，越有可能看到广告主 j 的广告，从而广告主 j 抢夺品牌 i 的市场份额的意图越强烈。

对于每一个 $d_{ikj} \neq 0$ ，以 b_{ikj} 表示第 i 个品牌对应的第 k 个关键词被第 j 个广告主所购买时，爬虫抓取到的广告的平均排名；则广告主 j 购买品牌 i 的品牌关

键词的广告展现平均排名为 $B_{ij} = \begin{cases} \sum_{k=1}^{s_i} b_{ikj} / \sum_{k=1}^{s_i} d_{ikj} & (\sum_{k=1}^{s_i} d_{ikj} \neq 0) \\ 0 & (\sum_{k=1}^{s_i} d_{ikj} = 0) \end{cases}$ 。这个值越小，表示

消费者在搜索品牌 i 的品牌词时，会同时看到广告主 j 的广告并且其排名位置越靠前，从而广告主 j 抢夺品牌 i 的市场份额的意图越强烈。

在分析时，可以直接以 D_{ij} 来表示广告主之间的竞争强度，也可以 A_{ij} 、 B_{ij} 对 D_{ij} 加权，构造综合指数 $U_{ij} = \frac{D_{ij} \times A_{ij}}{B_{ij}}$ 。 U_{ij} 综合了更多因素，可以更为全面地反映一个品牌的推广力度，但是这个指标的业务意义相对不如 A_{ij} 、 B_{ij} 、 D_{ij} 直接清晰；在实际应用中，通常会将它们结合在一起使用。

表 6 品牌攻防策略结果示意

③
↘

① ↓

品牌攻防策略		广告主(j)				
		A	B	C	D	E
② → 品牌(i)	A					
	B					
	C					
	D					
	E					

- 表 6 可从三个方向解读：
- 1、每一列表示一个广告主以哪些竞争者为攻击对象，以及攻击各竞争者的力度。如果 A 列在与 B、C、D、E 各行相交的位置数值都很高，则 A 在非常主动地抢夺四个竞争对手的用户；它攻击其他竞争者的意图非常强烈。
 - 2、每一行表示每个品牌遭到哪些竞争者的攻击，以及被攻击的力度大小。举例来说，如果 A 行在与 B、C、D、E 各列相交的位置数值都很高，那么 B、C、D、E 都在非常主动地想从 A 手中抢夺客户和市场份额；这样的品牌通常是市场上第一阵营中的品牌，也有可能是处于衰落过程中的品牌。
 - 3、对角线表示各个竞争者在攻击其他人之余，是否做好了防守。如果一个广告主在自己所在列的其他位置数值都很高，只是在列与对角线交叉的位置取值很小，那么这个竞争者在全力攻击竞争者，但是却遗漏了很多与自己品牌相关的

关键词。

上述分析的结论，可以被落实为具体的推广优化方案：广告主可以根据行业各主要品牌被攻击的集中程度，斟酌是否调整主攻品牌的范围。此外，广告主可以购买那些尚未购买的自身及主攻对手的品牌关键词（ D_{ij} 较小的品牌对应的 $d_{ikj} = 0$ 的词）；对于已购买的品牌词，可以考虑是否要调高那些展现机会较少（主攻的目标品牌对应的 a_{ikj} 较小的关键词）或排名较为靠后的关键词（主攻的目标品牌对应的 b_{ikj} 较大的关键词）的出价，让自己可以有更多的机会第一时间展现在潜在消费者面前。

品牌词抽样一旦确定一般会比较稳定，对于监控时间较长的品牌，此部分还会监控 D_{ij} 等指标在时间轴上的变化趋势，一旦发现市场上各主要品牌的攻防对象及强度的变化即随同发布相关广告主和品牌的详细变化情况。

第二个策略特征与关键词购买及出价有关，用于检查广告主对关键词的资源分配是否合理。由此衍生的优化建议主要用于帮助广告主调整关键词结构、调整关键词出价，进而控制投放成本；包含两个方向的截面分析。

第一个方向的分析用于判断广告主购买的关键词配比是否合理，分析结果以图 3 所示的方式展示。沿用行业层面分析中的符号定义。图 3 中横轴“流量覆盖率”衡量每个广告主覆盖的广告流量在全行业广告流量中的占比，对广告主 i ，

有 $iq_i = \sum_{j=1}^n l_{ij} \times q_j / \sum_{j=1}^n q_j$ ($i=1,2,\dots,m$)；纵轴“关键词覆盖率”衡量每个广告主

购买的关键词数量在全行业关键词中的占比，对广告主 i ，有

$ik_i = \sum_{j=1}^n l_{ij} / n$ ($i=1,2,\dots,m$)。

从表 1 提供的关键词表样本中，可以看到搜索不同关键词的消费者数量差异很大。一些行业强势品牌的品牌词、核心产品词搜索量很大，购买这样的关键词可以接触到大量的用户，而且它们通常也是消费者在决策关键阶段必然会搜索的词。行业内的每个广告主都会或多或少购买一些这样的词。而消费者每次搜索看到的广告条数是有上限的，广告主想要保证自己的广告可以被展现出来，就需要为它们出较高的价格。一个行业内，符合这一类特征的词，通常数量有限，广告

主对它们的竞争激烈，导致这些词的价格也是行业中最高的，它们被称为“热门词”。热门词之外，还有大量搜索量中等或较小、竞买广告主数量较少、价格也低的关键词，它们通常被用来捕获细分市场用户、聚集长尾消费者，因此被称为“长尾词”。

热门词保量，长尾词拉低平均成本，通常广告主购买关键词时会在两者之间保持一定平衡，分布在如图 3 中对角线所示位置附近。过多地购买长尾词、热门词不足的广告主会处于图中左上位置，他们可以将关键词点击成本控制在极低的水平上，但能够接触到的消费者数量受限、也可能错失一些消费者决策的关键环节。因此，这种策略通常不能够像广告主所期望地那样，带来较低的最终获客成本。与之相对，位于图中的右下角的广告主，长尾词不足，虽然能够有效接触到市场上绝大部分的需求和用户的主要购买决策环节，但成本通常较高、也易错失某些关键的细分市场。除了对每一个广告主购买的全体关键词进行此项分析外，还可以将分析的维度细化到某一个广告主所购买的每一类关键词上。

对于在整体或某一类关键词的投放上结构有所偏颇的广告主，图 3 所示的分析会为其指出关键词购买结构上的调整方向、并提供待添加的样例词作为参考。

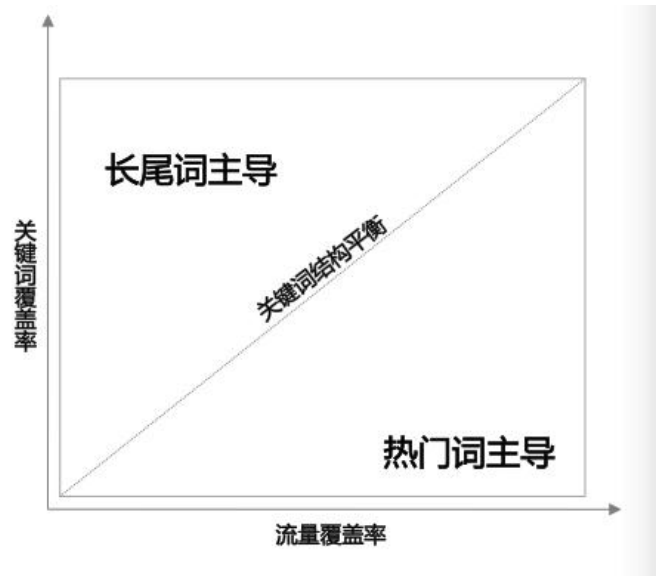


图 3 策略层面-截面-关键词购买结构-示意图

第二个方向主要考察广告主整体排名策略的激进程度，可用于推断竞争者的成本规模，并提供调整重点关键词排名的建议。首先，在整体投放的层面上，对各主要竞争对手的出价策略做出定位，具体参考图 4。图中每一条线代表一个广告主；横轴数字代表广告排名，多条广告从左侧最上方开始、从左至右从上至下

统一编号；纵轴表示抓取到的某个广告主的全部广告在各个排名上的分布比例。沿用行业层面分析部分的符号， $C_{ij} (i=1,2,\dots,m \quad j=1,2,\dots,n)$ 表示抓取到第 i 个广告主关于第 j 个关键词的广告条数， $R_{ikj} (k=1,2,\dots,C_{ij} \quad j=1,2,\dots,n)$ 表示抓取到的各条广告的排名。对广告主 i ，统计 R_{ikj} 的频数分布，生成序列 r_{i1}, r_{i2}, \dots 分别表示抓取到的广告主 i 所有广告中展现时排在第 1 位、第 2 位、……的广告条数，易知 $\sum_{j=1}^n C_{ij} = \sum_s r_{is}$ ；图 4 中各排名对应的纵轴指标计算方式为 $RR_{is} = r_{is} / \sum_{j=1}^n C_{ij} (s=1,2,\dots)$ 。

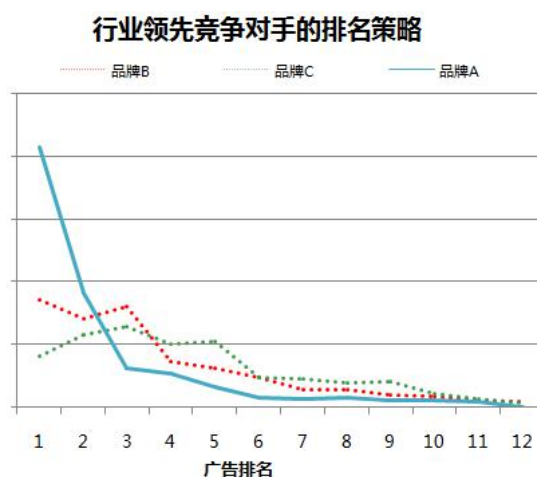


图 4 策略层面-截面-关键词整体出价策略-示意图

如前所述，广告排名越靠前，广告主需要支付的成本通常越高。图 4 中曲线左侧位置越高，表明该广告主为所购买的关键词设定的出价策略越激进，从而其推广成本越高。当然，推广成本还受到购买关键词规模等因素的影响，但是进入本部分分析的通常都是购买关键词规模相当的竞争者，因此在图 4 的分析中并未考虑这一因素。

图 4 帮助广告主对竞品的总体出价策略做出判断。当广告主需要针对竞品出价情况来调整自身的出价策略时，可参考图 5。图中最上方的条形代表使用品牌观察模型的客户 A，其下的每一条形代表它关注的一个主要竞品，条形的长度代表广告主购买的关键词个数。每一个竞品的条形由蓝色、红色和黄色三个色块组成：蓝色部分代表该竞争对手购买、而客户 A 没有购买的关键词个数；与之相对

应，竞品条形左侧的空白区域代表客户 A 购买、而该竞争对手没有购买的关键词个数。红色与黄色的部分合并在一起，表示客户 A 和该竞争对手同时购买的关键词个数，红色部分表示竞争对手的广告排位在客户 A 之上，黄色部分表示竞争对手的广告排位在客户 A 之下。每个色块均能对应到具体的关键词及其排名情况。

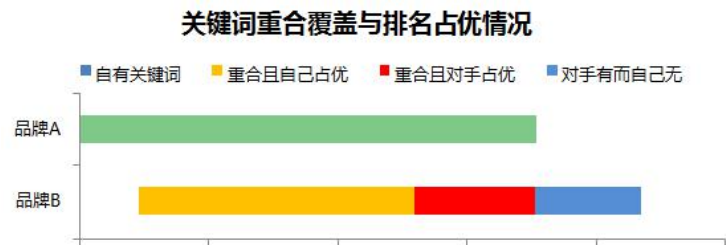


图 5 策略层面-截面-关键词排名占优情况-示意图

蓝色部分对应的关键词(尤其是那些多个竞争对手都购买的词)可能是客户 A 遗漏的行业重要关键词，可以考虑购买。黄色和红色部分的长度如果与客户 A 的条形长度非常接近的话，表示两者的业务和目标客户群非常相近。即使此竞争对手没有在主动竞争部分表现出很强的攻击意图，也应重点防范或主动加强攻击。黄色部分和红色部分的比例，以及落在两个部分的关键词分布是否合理，也是值得关注的问题。搜索黄色部分关键词，客户 A 的广告展现位置更靠上，付出的成本一般也更高。在这个区域内应仅保留那些对于促成购买或加深消费者的品牌印象有重要作用的关键词，为它们争取最好排名。

第三个策略层面的应用，与广告创意有关。产出结果可以用来帮助客户了解自身的广告创意与品牌自我设定是否一致、主要竞品如何设定自身品牌形象以及是否有变化，总结市场上可供参考的创意点等。

创意是整个 SEM 策略中最难以量化的部分，一个好的创意要言简意赅，同时富有感染力。投放人员在编写创意时，经常会通过搜索关键词，参考其他广告主创意的方式来寻找灵感。在实际工作当中，这些工作大多由投放人员手工进行，不但耗时多、效率低，而且限于搜索规模其采样的典型性和代表性都没有保证。品牌观察模型对创意文本的处理，模拟了分析师的工作，但大幅提升了其自动化程度、采样代表性更好、并且分析结果更容易解读，具体处理方式如下：

1、将抓取的广告数据，抽取掉时间及排名，将展现 url 替换为对应的广告主名称，按照“广告主+关键词+创意标题+创意描述+标题飘红+描述飘红”去重，并生成计数；再按照关键词，将关键词抽样中的关键词检索量补充到结果中。得

到如表 7 所示的数据。

表 7 创意处理示意

字段	取值示例
关键词	java 培训
广告主	A
创意标题	java 培训专注 java 培训 15 年
创意描述	北京 java 培训中关村 A，专注 IT 教育， 50 万学员的共同选择. java 培训 A 金牌 讲师授课 java 培训免费咨询电话:...
标题飘红	java 培训 java 培训
描述飘红	北京 java 培训 java 培训 java 培训
广告记录数	32
关键词检索量	3000

表 7 的数据可供使用者按照在市场上的展现机会多少来筛选创意，但是对于需要掌握整个市场和主要竞争对手创意热点的使用者来说，它的数据量仍然过于庞大。以本例所使用的抽样词表和抓取规则为例，即使只考虑少量主要竞争对手，完整的表 7 也有数万行记录。同时，这种以整条创意为单位的统计结果，也不利于阅读者总结提炼。以表 7 中的创意为例，其中“专注 IT 教育”主打消费者对“专业”培训机构的诉求；“50 万学员的共同选择”、“金牌讲师”则从用户规模和师资力量角度打消消费者对于“教学质量”的疑虑。也就是说，每一条创意都有可能糅合了多个诉求点，从创意中拆解出的诉求点及其对应的具体表达方式比完整的创意文本更适合阅读。接下来讨论如何对创意文本进行拆分、归类 and 汇总。

2、拆分文本的关键问题之一是决定拆分的粒度。以字或单词为单位的拆分通常不能完整表达一个意思。以上例子中的“专注 IT 教育”如果被拆分成了“专注”、“IT”和“教育”，阅读者将难以解读出“专业”的诉求点。因此字或单词不是合适的拆分维度。但如前讨论，拆分粒度也不可过粗，否则不利于分类汇总、总结规律。品牌观察模型采取的方式是：将创意文本(标题+描述)中的飘红部分

替换为分隔符，将创意文本中与关键词一致的部分替换为分隔符，将创意文本中所有标点符号替换为分隔符。将创意文本以分隔符与飘红及关键词合并。合并后的每一条文本，按照分隔符拆开，计算素材的频数，再以广告记录数和搜索量加权得到素材的曝光指数。最后，去掉广告主，按照素材对曝光指数进行加总。数据加工过程如表 8 所示：

表 8 创意拆分和素材汇总示意

1、表 7 中的数据	字段	取值示例
	关键词	java 培训
	广告主	A
	创意标题	java 培训专注 java 培训 15 年
	创意描述	北京 java 培训中关村 A，专注 IT 教育，50 万学员的共同选择. java 培训 A 金牌讲师授课 java 培训免费咨询电话:...
	标题飘红	java 培训 java 培训
	描述飘红	北京 java 培训 java 培训 java 培训
	广告记录数	32
	关键词检索量	3000
2、将标题和创意中飘红和标点替换为分隔符	创意标题	专注 15 年
	创意描述	中关村 A 专注 IT 教育 50 万学员的共同选择 A 金牌讲师授课 免费咨询电话
3、将关键词、标题、创意、飘红以分隔符连接在一起	文本	java 培训 专注 15 年 中关村 A 专注 IT 教育 50 万学员的共同选择 A 金牌讲师授课 免费咨询电话 java 培训 java 培训 北京 java 培训 java 培训 java 培训
	广告记录数	32
	关键词检索量	3000
	广告主	A

4、按照分隔符 拆分文本，统 计曝光指数	素材	曝光指数	广告主
	java 培训	=5X32X3000=480000	A
	专注	=1X32X3000=96000	A

5、抽取掉广告 主后，按“素材 ”对“曝光指 数”分类加总	素材	曝光指数	
	java 培训	65300000	
	专注	8210000	
	

3、对表 8 中第 5 步数据按照曝光指数降序排列，即可截取到那些最多地被展现在消费者面前的素材；将第 4 步数据按广告主分开，对每个广告主素材按照曝光指数倒序排列，截取每个广告主最集中使用的素材。将两部分素材合并后去重。在初次使用品牌观察模型时，需要对这些素材按照诉求进行归类，每一类诉求称为一个“创意点”；此后每次使用时，只需将新增的高频素材加入到分类中。新增高频素材本身代表了市场的异常变化，如有此类情况发生，应随同产生明细报表。

将素材分类补充到表 8 第 4 步的数据中，按照“广告主+创意点”对曝光指数进行加总，得到如图 6 所示的图形。图 6 中每一个柱体的高度代表某一个广告主在创意中使用某一类创意点的力度大小。每一列表示一个广告主如何塑造自己的品牌形象，每一行表示某一创意点是否是市场上的宣传热点或盲区。

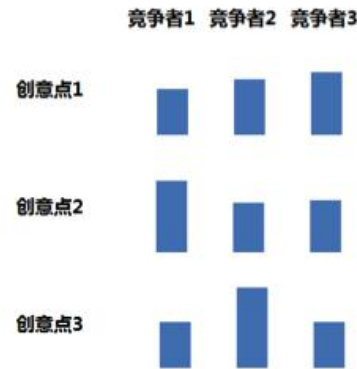


图 6：策略层面-截面-创意-示意图

四、 总结与讨论

本文主要介绍了一种使用公开的 SEM 广告展现数据,推断广告主所在行业的 SEM 推广市场格局、评估各主要竞争对手竞争策略、并根据市场情况和竞争对手策略优化自身投放的方法。此类需求是众多重度依赖 SEM 广告以维持或扩大业务规模的广告主的通用需求,但业界此前并未有解决此类问题的系统性方法。本文所介绍的品牌观察模型根据 SEM 广告的基本业务逻辑,系统地回答企业在市场和竞品监控上的主要问题。在开始的研发和准备阶段过后,绝大部分后续工作都可以自动执行,为广告主提供全面的市场监控、基本的策略判断和有的放矢的优化建议,从而有效地提升了 SEM 投放优化工作的效率。文中所述事例虽然来自职业教育培训行业,但方法本身具有通用性。

本方法后续的改进方向有两个:首先,进一步提升方法的自动化程度。重度依赖 SEM 的行业一般都发展迅速、竞争激烈,行业的快速变化导致新的关键词、新的竞争对手、新的创意点都层出不穷;而与之相对应的关键词抽样、重点竞争对手遴选、创意点归类正是本方法中需要人工介入最多的部分,也是产出结果客观、及时的必要保证。这几个环节的自动化程度,是决定此项服务是否可规模化的关键因素。其次,加强市场环境及竞品相关数据与企业内部数据(推广的投入产出、用户的属性与行为特征等)的整合。营销是一个涉及到企业多个职能的漫长链条,SEM 投放是其中非常靠近起始位置的一个环节,它和其他的环节都应对营收、投入产出比、市场份额等企业的核心业绩指标负责。在实际工作中,这些本应相互呼应配合的环节却经常是各自为政甚至互相对立。究其原因,没有层级清晰、关联合理的指标体系是一个重要的原因。因此需要从起始的环节开始逐步打通各业务环节的数据,并由最终结果反推投放优化方向。

笔者所在的公司已使用品牌观察模型为多个行业、超过一百家大量投放 SEM 广告的客户提供过市场竞争环境评估和策略优化服务,得到了他们的一致认可。但是从数据分析的专业角度看,这一方法所使用的数据结构非常简单,分析手段也无非是简单的描述性统计方法。“雕虫小技”被推崇备至的同时,笔者也多次遇到过从数据分析专业角度看非常出色的工作却得不到客户的欣赏和认可的情况。这些都并非是偶然现象。从国内企业对于数据的认知、积累和使用能力来看:一方面,近几年“大数据”这一概念受到了广泛的关注,业界也确实有一些标杆

企业完成了非常出色的案例与实践。但同时，即使是在互联网等所谓的前沿领域当中，大部分企业的问题仍然是数据收集不够完备准确，或者有了数据却不知该如何使用。这些企业在数据应用的问题上，正处在一个“看上去很美”，却不知道该如何系统开展的阶段。类似本案例所使用的方法，虽然仅仅只涉及到了最简单的数据分析方法，但是对于客户来说它带来的几乎是从零到一的改变。它不但足够贴合客户的痛点，解决了某个一直困扰他们的具体问题，而且在某种程度上也帮助他们迈出了系统性地开展数据分析工作的第一步。

类似的经历促使笔者思考一个问题：作为一个数据分析从业者，我们到底应该做什么以及怎么做，才能更快更好地帮助更多的行业和企业具备将数据转化为价值的能力，同时也让数据分析成为企业中必不可少的部门，增强数据分析人员在企业中的话语权？笔者相信，只有当数据的价值在各行各业各种规模的企业中都被充分地发挥出来的时候，才真正意味着商务领域数据分析行业全盛时代的到来。今天我们有幸站在这个时代的开端，它是否到来、何时到来就由每一位同行的努力来决定。因此上，谨以此问与诸位同行共勉。

参考文献

- [1] CNNIC(中国互联网络信息中心)，《中国互联网络发展状况统计报告(2016 年 1 月)》，2016
- [2] 易观智库.《中国新媒体市场数据盘点专题研究报告 2015 年第 1 季度》，2015
- [3] 百度历年财报，
<http://ir.baidu.com/phoenix.zhtml?c=188488&p=irol-reportsAnnual>
- [4] 张跃志.《2014 中国教育市场发展报告》，2014
- [5] 胡锦涛.《中国教育培训市场供需细分与市场前景分析》，2014