

基于SEM和AHP的推广数据采集与决策方案优化

林之博, 刘媛华

(上海理工大学管理学院, 上海 200093)

摘要: 企业信息化的普及使广告投放和宣传手段从以往的人工宣讲、传单发放等方式, 到目前的SEM(搜索引擎推广)技术应用, 突破了时间空间限制, 将产品服务展现给更广阔的市场。部分企业在网络营销过程中常因数据采集方案落后、决策因素过于狭窄等导致收益降低。本文分析案例公司A存在的不合理业务流程, 并基于SEM接口技术、AHP(层次分析法)对相关流程进行建模优化, 设计了一套可行的系统和决策优化方案。

关键词: 广告投放决策; 评价模型; 信息系统; 业务流程分析; 工作简化

中图分类号: TP399 **文献标识码:** A

Promotion Data Acquisition and Decision-making Scheme Optimization Based on SEM and AHP

LIN Zhibo, LIU Yuanhua

(Business School, University of Shanghai for Science and Technology, Shanghai 200093, China)

Abstract: With the popularization of enterprise informatization, the advertising and promoting means have broken the restrictions of time and space, expanding from the previous manual preaching, leaflet distribution and so on to the current application of SEM (search engine promotion) technology. The new SEM technology shows the products and services to a broader market. In the process of online-marketing, some enterprises often suffer from returns reduction due to outdated data collection schemes and limited decision-making factors. This paper analyzes the unreasonable business process of case company A. Then based on SEM interface technology, AHP (analytic hierarchy process), this paper models and optimizes the relevant processes, and designs a set of feasible system and decision optimization scheme.

Keywords: advertising decision-making; evaluation model; information system; business process analysis; work simplification

1 引言(Introduction)

近年来我国各省市互联网发展指数持续上升, 利用SEM营销技术, 相比传统的人工宣传、现场散发传单等广告方式, 能在使用更低成本、更少人力物力的条件下高效广泛宣传^[1]。但部分企业由于未能优化原有的信息手段、决策方案, 导致收益下降、效率过低, 无法完全发挥出SEM的效果。A公司是广州市一家网络教育培训公司, 为推广其服务, 设立BI部门进行销售数据分析和信息管理工作。以该公司为例, 针对广告投放、信息采集与分析决策流程链进行分析, 找出不合理环节并进行改进优化, 提供可选择的高效率、低成本运作方案^[2]。

2 技术与理论(Introduction to technology and theory)

SEM接口技术即搜索引擎营销接口技术^[2], 利用java API编程进行网络POST请求获取Response, 提取json进行数据清

洗、整理、规范、存储, 最终提供给BI部门规范化的数据用于决策分析^[3]。AHP即层次分析法, 分析复杂问题中的各种因素, 得出评分构成的矩阵, 利用特征值矩阵运算, 得到相应每个因素的最佳权重, 设计决策方案能达到较满意决策^[4]。

3 流程与决策分析(Process and decision mode analysis)

3.1 流程分析

广告投放组员工进行初始化检查, 逐个登录推广数据中心复制过去一天的推广状态数据到Excel, 通过QQ或微信发送给流量部门, 由之逐个下载打开提取数据到总表后, 进行决策分析。BI部门逐个下载、打开文件, 复制数据到总表汇总。该过程存在循环性重复人工工作。上述操作方式会导致员工整体工作积极性下降、效率和准确度低等问题^[5]。

3.2 人机分析

数学方法分析人机配置状态, 模型如下:

$$\min T = \sum_{i=1}^k \sum_{j=1}^p r \cdot (t_{ij}^C + t_{ij}^H) \quad (1)$$

$$\min Z = \frac{\sum_{i=1}^k \sum_{j=1}^p r \cdot t_{ij}^H}{\sum_{i=1}^k \sum_{j=1}^p r \cdot t_{ij}^C} \quad (2)$$

$$\sum_{j=1}^p (t_{ij}^C + t_{ij}^H) < S$$

$$k \geq 1, p \geq 1, r \geq 1$$

$$i=1, 2, 3, \dots, k; j=1, 2, 3, \dots, p$$

其中, t^C 和 t^H 分别为子流程中计算机运行时间和人工操作时间, k 为流程步数, p 、 r 分别为子流程步数和循环次数。模型要求所有人工重复子流程时间总和与人机时间利用率应向最小化方向进行优化, 且总时间不超过可以接受的值 S 。经过抽样统计分析, 该公司员工数据采集汇总时间利用情况如表1所示。

表 1 时间分析总表

Tab.1 Time analysis summary

数据采集阶段	人工操作时间	计算机利用时间	员工等待时间	计算机等待时间	总时长
	49.316	16.606	1.29	34	50.606
分析决策阶段	人工操作时间	计算机利用时间	员工等待时间	计算机等待时间	总时长
	68.21	54.88	0	13.33	68.21

用于分析决策下一阶段广告投放方案的时间 $\geq 2h$ 。数据采集阶段计算机利用率约为32%, 数据分析决策阶段计算机利用率约为80%。利用人机配置分析法进行方案设计、比较和模型验证可知应改善平台登录、数据复制、文件传输等重复工作项, 在保持计算机利用率的条件下减少总时长^[6]。

3.3 原有决策模式分析

BI部门将来自点击广告的客户登记信息称为“名片”, 并将具有较大转化为订单可能性的“名片”称为“机会名片”^[7]。BI部门根据 N_{O_i}/C_{C_i} 计算转化率降序排序, 更倾向在转化率更高的地区投放更多广告。但当作为样本的名片数量、机会名片数量若相比其余地区过小, 数据将不具有统计意义上的说服力。

此外, 不全面考虑社会因素、经济因素、地理因素等多方面带来的影响, 可能导致偏差。例如原有的参考指标并不能更好地反映对应地区对该公司业务的需求程度。

4 优化方案设计与实现(Design and implementation of optimized scheme)

4.1 自动化SEM接口数据采集系统设计实现

百度、360等第三方推广平台实际上已经提供了非常方便的数据接口文档, 例如《百度搜索推广API V4开发文档》, 能够很方便地查找获取指定数据集应构造的HEADER、POST结构、API-URL和方法^[7], 结合阿里公司提供的JSON数据处理模块, 通过快速原型法开发可扩展Web系统, 则可以使用java语言, 依托Tomcat虚拟服务器, 使用Myeclipse引

入HttpsUtil外部类编写POST功能块, 引入fastjson-1.2.7工具包处理获取的JSON数据, 最后存储在MySQL数据库中。如图1所示为系统数据库E-R图。

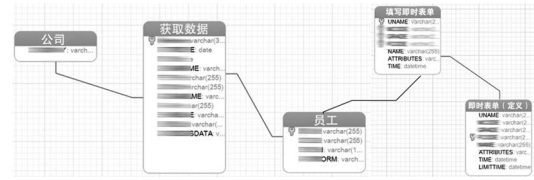


图1 数据库E-R图

Fig.1 Database E-R diagram

使用java结合POST工具包HttpsUtil实现调用百度接口推广数据方法部分实例代码如下:

```
String url="https://api.baidu.com/json/sms/service/ReportService/getRealTimeData";
getdatafrombaidu DA=new getdatafrombaidu();
byte[] result=HttpsUtil.post(url,DA.appendReportServiceJson(//写入需要查询的字段
User,PWD,token,...).toString(),"UTF-8");//将返回得到的数据存储为字节码
jsonResult = new String(result);//将字节码字符串化
Data=DA.getData(jsonResult,performanceData.length);//json解包处理
```

设计一套类似Json的数据压缩方案, 利用特殊的符号连接表单列名储存在静态数据库^[8], 员工访问时进行分割产生属性数组作为在线表单生成依据, 由员工在规定时间内填写。BI部门管理端进行一键采集, 将数据导出到Excel中即可。如图2为功能模块试运行效果。



图2 推广数据获取系统

Fig.2 Promotion data acquisition system

4.2 层次分析法模型

4.2.1 多因素影响因子选取

通过组织BI部门工作人员使用头脑风暴方法列举了各类应考虑在内的潜在影响因素, 得出了相应的影响因素提及频度。专家、管理层人员采取名义小组方法对员工调研结果图表、数据进行分析评价, 确定 P_{A_i} (地区发达指标)、 P_{TC_i} (地区转化与成本指标)、 P_{S_i} (地区战略倾向指标)、 P_{D_i} (地区需求指标)、 P_{G_i} (地区政策支持指标)等五项具有重大影响力的影响因素。

4.2.2 转化与成本指标

以货币为统一度量单位, 将订单数量换算为实际订单成本, 用“转化花费”表现转化为订单的点击数, 其中 C_{O_i} 、 C_{C_i} 分别为 i 地区订单成本和单次点击成本, C_{A_i} 、 N_{O_i} 、 C_{T_i} 分别为 i 地区广告投放总成本、订单数量和转化花费。

$$P_{TCi} = (P_{OCi} + P_{Ti}) / 2, i = 1, 2, 3, \dots, n \quad (4)$$

$$C_{Ti} = N_{Oi} \times C_C, i = 1, 2, 3, \dots, n \quad (5)$$

以订单成本作为基准, 计算订单成本指标, 要求订单成本应与指标呈反比例关系, 使用“订单数量/机会名片数量”作为“伪转化率” R_{FTi} 进而计算该地区转化率。采用区间映射法将数据规范化到[0, 10]。 P_{OCi} 和 R_{Ti} 分别为*i*地区订单成本指标和转化率, P_{Tj} 为*i*地区转化指标。

$$P_{OCi} = 10 \times (1 - C_{Oi} / \text{Max}\{C_{Oj}\}) \quad (6)$$

$$i = 1, 2, 3, \dots, n; j = 1, 2, 3, \dots, n$$

$$R_{Ti} = R_{FTi} \times (C_{Ai} / \sum_{i=1}^n C_{Ai}) \quad (7)$$

$$P_{Tj} = 10 \times (R_{Tj} / \text{Max}\{R_{Ti}\}) \quad (8)$$

$$i = 1, 2, 3, \dots, n; j = 1, 2, 3, \dots, n$$

最终可以得到核心指标之一, 转化与成本指标:

$$P_{TCi} = (P_{OCi} + P_{Ti}) / 2, i = 1, 2, 3, \dots, n \quad (9)$$

4.2.3 需求指标

式(10)至式(12): 需求比例 S_{Di} 能够反映该地区有相关需求受众的占比, 需求强度指标 I_{Di} 用于反映有需求部分人口中有机会转化为订单部分人口的占比, 并进行数据规范化处理, 可计算出需求指标 P_{Di} 。用 N_{Ci} 作为*i*地区机会名片数量, H_i 为*i*地区总人口数量。

$$S_{Di} = (C_{Ai} / C_C) / H_i, i = 1, 2, 3, \dots, n \quad (10)$$

$$I_{Di} = N_{Ci} \times C_C / \sum_{i=1}^n C_{Ai} \quad (11)$$

$$P_{Di} = ((S_{Di} \times I_{Di}) / \text{Max}\{S_{Dj} \times I_{Dj}\}) \times 10 \quad (12)$$

$$i = 1, 2, 3, \dots, n; j = 1, 2, 3, \dots, n$$

4.2.4 其他指标

地区发达指标反映该地区经济发展状况与互联网发展程度。以年度全国人均GDP发展排名 L_i 数据为基准^[9], 将所有省市的人均GDP数据规范到10分评价区间中作为地区发达指标(13)。将该公司在各地区战略发展倾向和地区政策对公司相关业务的支持度归类得到战略倾向指标和政策支持指标(14)。

$$P_{Ai} = (L_i / n) \times 10, i = 1, 2, 3, \dots, n \quad (13)$$

$$P_{Si} = \begin{cases} 10, & \text{优先} \\ 9, & \text{持续} \\ 7, & \text{维持} \\ 5, & \text{衰退} \\ 0, & \text{剥离} \end{cases} \quad P_{Ci} = \begin{cases} 10, & \text{鼓励} \\ 5, & \text{支持} \\ 0, & \text{不支持} \end{cases} \quad (14)$$

$$i = 1, 2, 3, \dots, n$$

4.3 基于层次分析模型的案例计算

组织A公司人员进行近期工作情况汇报和数据反馈, 由该公司专家审阅评价, 得到判断矩阵如图3所示(正互反矩阵)。

	地区发达指标	地区政策支持指标	需求指标	转化与成本指标	公司战略倾向指标
地区发达指标	1.00	0.17	0.33	0.25	0.20
地区政策支持指标	6.00	1.00	4.00	3.00	2.00
需求指标	3.00	0.25	1.00	0.50	0.25
转化与成本指标	4.00	0.33	2.00	1.00	0.25
公司战略倾向指标	5.00	0.50	4.00	4.00	1.00

图3 A公司评价矩阵

Fig.3 AHP evaluation matrix

导入Matlab计算最大特征值 $\lambda_{\max}$ 和一致性指标 CI , 其中 $CI = (\lambda_{\max} - n) / (n - 1)$ 。查询平均随机一致性指标表可得

平均随机一致性指标 RI ^[4]。计算 CI/RI 得一致性比例 CR , $CR < 0.1$, 则矩阵通过一致性检验。分别用算术平均法(15)、几何平均法(16)和特征值法求解权重向量^[4]如表2所示。

$$\omega_i = \frac{1}{n} \sum_{j=1}^n \frac{a_{ij}}{\sum_{k=1}^n a_{kj}} \quad (15)$$

$$\omega_i = \frac{\left(\prod_{j=1}^n a_{ij} \right)^{\frac{1}{n}}}{\sum_{k=1}^n \left(\prod_{j=1}^n a_{kj} \right)^{\frac{1}{n}}}, i = 1, 2, 3, \dots, n \quad (16)$$

表2 三种方法计算影响因子权重结果

Tab.2 The weight of influence factors calculated with three methods

指标	算术法	几何法	特征法
地区发达指标	4.33%	4.65%	4.67%
地区政策支持指标	35.53%	40.66%	40.27%
需求指标	11.10%	9.37%	9.21%
转化与成本指标	16.83%	13.85%	13.80%
公司战略倾向指标	32.20%	31.47%	32.04%

取均值或取其中任意一组作为五项影响因子的权重。假设: 某平台一次点击收费0.3元, 经系统获取得到五个案例地区数据如表3计算可得综合评价指标如表4所示。

表3 五个案例地区数据

Tab.3 Datasets from five case area

省份	名片	机会名片	订单数	人口总数	人均GDP
新疆	90	90	90	21813300	7567
广东	15325	7289	617	104300300	13058
上海	4325	2312	666	23013910	20421
四川	9000	6982	456	80418200	7387
西藏	500	123	21	3002100	6500

表4 多因素计算结果

Tab.4 Multi factor calculation results

省份	公司战略倾向	地区发达指标	转化与成本指标	需求指标	地区政策支持指标	综合评价	备注
新疆	7	2.65	5.15	0.41	10	4.5	酌情投放
广东	10	9.71	5.00	6.96	5	7.6	优先投放
上海	5	7.65	8.49	10.00	5	7.5	保持投放
四川	9	7.35	3.29	8.64	5	7.2	保持投放
西藏	0	0.88	0.54	4.08	10	2.4	放弃投放

5 结论(Conclusion)

案例A公司部署优化后的信息管理系统并采纳多因素计算方案后系统具有信息结构形态扁平、信息存储管理更便捷等
(下转第15页)