

基于大数据平台的精准广告系统研究与设计

张 建¹, 孙 铭¹, 段 娟²

(1. 北京工业大学软件学院, 北京 100124; 2. 北京邮电大学计算机学院, 北京 100876)

摘 要 : 互联网广告是网络营销的最主要手段之一。在当今大数据时代的背景下, 精准广告系统作为一种新的广告投放模式, 与传统广告投放有很大区别。文章通过搭建大数据平台, 在基础数据仓库中存储网络用户访问行为数据, 并使用 cookie 技术标识和追踪访客, 或者与其他数据平台方通过 cookie mapping 来实现用户 cookie 数据的交换, 从而获得更多访客身份特征描述信息。依据网站内容的关键词库和垂直领域的专业词库, 对人群进行标签化, 对拥有不同标签人群根据其特点, 依据用户和广告的匹配算法进行不同类别广告的精准投放, 从而提高广告投放的效果和转化率。

关键词 : 大数据平台; 精准广告; 广告系统; cookie 映射

中图分类号 : TP393 **文献标识码** : A

Targeted Advertising System Research and Design Based on Big Data Platform

ZHANG Jian¹, SUN Ming¹, DUAN Juan²

(1. School of Software Engineering, Beijing University of Technology, Beijing 100124, China;

2. School of Computer Science, Beijing University of Posts and Telecommunications, Beijing 100124, China)

Abstract : Network advertising is one of the most important means of network marketing. Under big data time background, targeted advertising, as a new advertising model, is extremely different from traditional advertising. Through building the big data platform, network user access behavior data is stored by cookie or session to distinguish and track someone. In addition, we can exchange more user data with other data platform based on cookie mapping technology. According to topic word, website keyword or other domain noun, people are divided different class with tags. Based on this conditions, system uses matching algorithm to connect user and advertising, to achieve targeted advertising of different categories, improve the effect of advertising and CTR (click-through-rate). Advertisers cost at the same time, improve the utilization rate of media advertising resources.

Key words : big data platform; targeted advertising; advertising system; cookie mapping

传统的互联网广告投放系统分为硬性广告和软性植入广告。硬性广告从售卖方式来划分, 常见的有 CPC、CPM、CPS、CPA 等^[1]。对于同一网站的访问用户来说, 如果是传统的广告投放系统, 那么所有的用户看到的都是同一个广告主投放的广告。然而并不是所有的用户都喜欢看这个广告, 或者该广告展示的内容并不适合当前用户, 反而给用户带来了较差的用户体验, 浪费了一次广告的展示, 这并不是媒体和广告主期望的结果。

精准广告系统改变了传统广告系统的投放方式。精准的含义就是通过对广告的受众甄别筛选, 区分出

不同的人群来展示不同的广告。因此精准广告投放系统所要展示的广告完全是由访客的属性来决定的。对于访客的刻画也由多个因素组成, 可以包括性别、年龄、行业、地理区域、关注话题、浏览器版本、操作系统、访问设备等。精准系统就是依据这些因素来决定对访客展示不同内容的广告。

设计和采用精准广告投放系统的目的在于使得投放的广告更加精准, 对广告主而言节约投放成本, 获得更好的展示效果, 还可以进一步提高效果广告的及时转化率, 如注册、购买、评论等行为。

收稿日期 2015-04-23

基金项目 北京市自然科学基金(项目编号: 7142014)

作者简介 张建(1965-) 男, 四川人, 副教授, 主要研究方向: 企业 IT 管理与创新、人工智能与知识工程、大型信息系统建设; 孙铭(1989-) 男, 山西人, 硕士研究生, 主要研究方向: 信息服务、计算广告; 段娟(1986-) 女, 陕西人, 硕士研究生, 主要研究方向: Web 安全、Webshell。

1 基于大数据的精准广告研究现状

从 2008 年开始 ,一批新兴的技术相继出现 ,移动计算、物联网、云计算等技术的发展和以及在媒体社交网络、虚拟化服务等新型模式中得到的广泛的应用使得全球的数据信息量呈现了爆发式的增长趋势 ,数据的复杂性也随之上涨^[2]。从客观上来看 ,这也催生了新的数据分析方法和数据挖掘技术来提取数据的内在价值。大数据技术应运而生 ,并得到了飞速的发展。如今已经步入了大数据时代 ,大数据的发展引起了许多领域的变革。

广告通过对受众群体传播,创造、交付优质的顾客价值来获得顾客、挽留顾客和提升顾客数量,而整个过程是依赖于对数据的搜集和分析,数据的使用贯穿在广告的传播过程中。在互联网广告领域内,大数据的兴起和全方位的发展,使得广告的投放能更加洞察消费者的需求,更加精准化,也为精准广告营销带来了前所未有的机遇。

目前国内外对大数据和精准广告的研究理论甚多,但是利用大数据平台和技术,通过研究用户行为特征,简历大数据用户特征描绘和精准广告营销策略的研究则相对较少。随着互联网企业自身实力的不断壮大,硬件部署、服务和终端的布局、自身生态环境的构建都在迅速发真,而互联网广告技术也在不断革新。RTB 产业链逐渐完善、ADX 与 DSP 平台的涌现、注重用户数据挖掘与内容创意的原生广告等新兴网络营销形式将进一步挖掘网络媒体的营销价值^[3],这些都将成为未来国内互联网广告市场的发展带来动力。

2 广告系统框架设计

整个系统业务框架围绕精准广告投放和传统广告投放相结合的业务逻辑为基础进行设计，兼备两种业务模式的广告投放方式。

2.1 业务流程处理

2.1.1 精准广告业务流程

用户通过浏览器对某一页面发起 http 请求, 精准广告投放系统所在服务器收到请求以后, 先在缓存服务器中查询用户属性信息, 如果没有查询到, 则在缓存服务器中基于该次请求创建该访客的属性特征; 如果查询到, 则更新该访客的属性信息。精准广告系统根据已经获得的用户属性和匹配算法, 向精准广告排期系统数据库、素材管理系统及数据库分别发起查询请求, 经过匹配算法, 最终向用户所访问的页面输出广告素材, 完成本次广告投放^[4]。系统同时会记录该次投放, 用

于结算广告费用和跟踪广告投放效果。

2.1.2 传统广告业投放

如果本次请求访问的用户与精准系统中的广告没有匹配,或者在无法对用户进行标签化时,或者客户没有精准广告投放业务时,系统将把请求转发到传统的广告投放系统中,系统对本次访问展示一般客户的广告,或者展示网站内部广告和默认广告等。

整个广告业务系统基于业务流程的基本架构如图 1 所示。

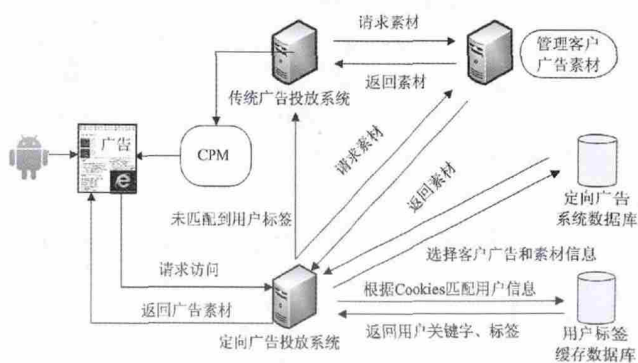
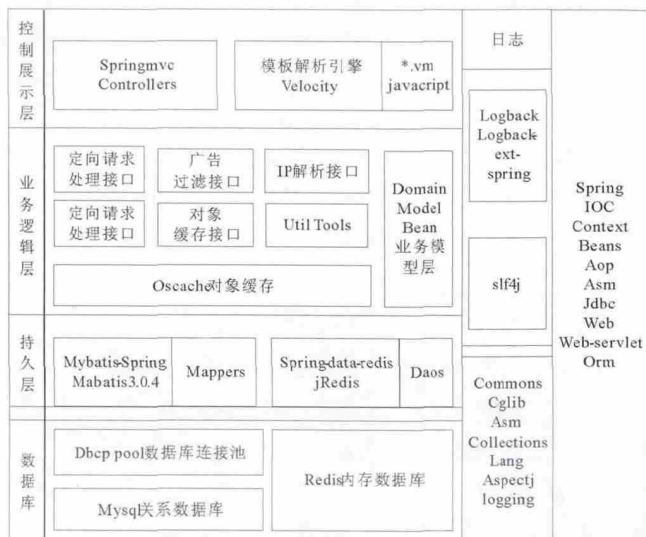


图 1 业务流程基本架构

2.2 技术框架设计

系统的技术架构如图 2 所示。



3 核心技术关键点设计

3.1 精准广告系统核心子系统

整个系统的架构依据功能划分为 7 个核心子系统:业务子系统、存储子系统、传输子系统、检索子系统、反作弊子系统、计费子系统、统计子系统。

3.2 广告匹配算法设计

广告的在线投放过程如图 3 所示。

3.2.1 广告在线匹配问题研究

当系统获取到用户的身份描述信息,通过标签化,

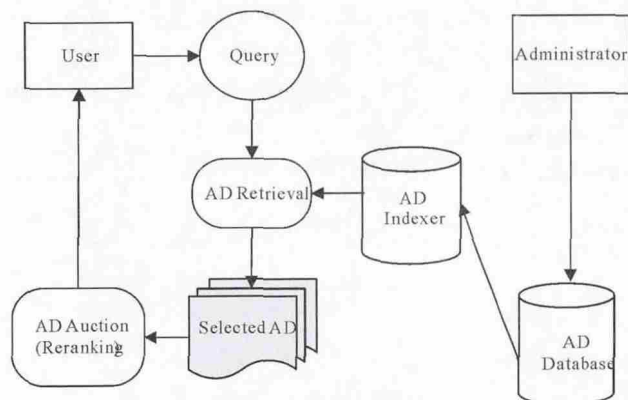


图 3 广告的在线投放过程

将其基本属性和兴趣关注点统一作为广告匹配的查询条件信息(Query),随后需要在广告库(AD database)中匹配到广告主投放的相关广告。系统的任务就是在用户信息集合和广告集合中进行最优的匹配,最终筛选出广告进行投放^[5]。该匹配的任务描述如下:

Vocabulary $V = \{w_1, w_2, \dots, w_N\}$

Query $q = q_1, \dots, q_m$, where $q_i \in V$

AD $adi = di_1, \dots, di_{m_i}$, where $dij \in V$

Ad Collection $C = \{ad_1, \dots, ad_k\}$

Set of relevant documents $R(q) \subseteq C$

Task = compute $R'(q)$ an "approximate $R(q)$ "

对于上述的广告投放任务是有约束条件的,需要考虑网站访客的用户体验和广告商的用户体验,即: user utility per search $> \alpha$; advertiser ROI per search $> \beta$;

3.2.2 广告匹配模式

广告搜索匹配模式通常有 3 类,最简单的是字符串匹配(string matching),在引入了 topic model/semantic web/ontology 等概念之后,又产生了语义匹配模式(semantic matching),而随着大数据平台技术的发展,通过使用大规模数据挖掘、推荐技术、personalized 等实现了意图匹配模式(intent matching)。为了实现用户和广告内容的匹配,需要提取用户特征、网页特征和广告特征^[6]。

(1)用户特征的提取方式主要有历史 Query 特征、历史浏览特征、用户分类特征。

(2)网页特征的提取主要包括基础特征提取^[7]和 Term 赋权。

(3)广告特征的提取方式主要有拍卖词特征^[8]、广告创意特征、广告落地页特征、广告素材特征。

4 广告投放效果跟踪及分析

精准广告是对传统广告投放的优化投放,互联网广告的用户角色有 3 种,分别为广告主、媒体和广告受众,所以精准广告的最终目标就是实现各方的最优目

标^[9]。

对广告主来说,希望投放效果最优,即:

$$\max \sum_{i=1}^M CRT(a, u, c) \cdot (Value(a, u, c) - Price(a, u, c)) \quad (1)$$

$$s.t. \sum_{i=1}^M Price(a, u, c) \leq B_i, UE_j \geq UE_{j0} \quad (2)$$

对于媒体来说,希望收益最大,即:

$$\max \sum_{j=1}^M \sum_{i=1}^T CTR(a, u, c) \cdot Price(a, u, c) \quad (3)$$

$$s.t. \sum_{i=1}^M \sum_{j=1}^T UE_{ij}(a, u, c) \geq UE_0 \quad (4)$$

对社会来说,希望社会效率最优(即广告受众获得最大的广告价值信息),即:

$$\max \sum_{j=1}^N \sum_{i=1}^T (CTR(a, u, c) \cdot Value(a, u, c) + UE(a, u, c)) \quad (5)$$

媒体可以直接收集到原始数据,所以从媒体的角度来分析,为了对比精准广告投放效果和传统广告投放效果,可以做几组对照实验。分别找两个日访问量相近、访问用户组成相似、入口位置的流量相近的两组页面,或者是同一页面上左右两个广告位等,然后分别投放精准广告和普通广告。通过一定周期内的数据采集,可以对广告投放用户 IP/PV 比、用户点击率、落地页二跳率、落地页跳出率、停留时间、新老用户 cookie 覆盖率等指标进行对比分析。对于 cps 类型的广告,还可以对其转化率、投入产出比和用户忠实度进行多维度的分析^[10]。

5 结束语

精准广告的投放,是今后广告主和媒体都愿意采用的,并将逐渐代替现有的 CPM 和 CPD 广告投放方式。随着大数据的普及,更多互联网企业开始积累原始用户信息,并对用户信息进行数据挖掘,甚至和一些第三方数据平台通过 Cookie Mapping^[11]的方式交换共享用户数据,互联网广告的投放也将逐渐基于大数据挖掘技术以一种更加高效的方式投放。无论从广告主的角度还是媒介,到最终的广告受众,采用精准投放会更好迎合多个角色的需求。

参考文献:

- [1] 林宏伟. 网络广告运作的若干关键问题研究[D]. 成都:电子科技大学, 2013.
- [2] 魏丽宏. 关于我国手机传媒产业发展的研究[D]. 北京:中国社会科学院, 2012.
- [3] 张之怡. 大数据时代背景下的广告营销模式研究[D]. 苏州:苏州大学, 2014.
- [4] 李引, 袁峰. 基于领域驱动设计的应用系统模型[J]. 计算机工程与应用, 2013, 49(16): 1-8.
- [5] 陈震, 刘洪健. 基于倒排列表的网流索引检索与压缩方法

- [J]. 信息网络安全 2014(4) 25-30.
- [6] 许卓然. 广告搜索中的查询处理算法[D]. 上海:上海交通大学 2011.
- [7] 陈涛,薛丽敏,宋庆帅. 基于 DOM 的信息检索研究[J].信息网络安全 2014(5) 82-86.
- [8] 解忠乾,常笑,姬东鸿. 融合广告主行为的拍卖词实时触发[J]. 计算机应用 2014 (9) 2566-2570.
- [9] 熊亚洲.网络广告投放效果影响因素的实证分析[J].黄石理工学院学报(人文社会科学版) 2012 29(2) :18-21.
- [10] 张泽云. 基于多层 B/S 结构的网友访问行为的分析系统[D]. 华南理工大学,2011.
- [11] 蒋海婷.网络广告实时竞价(RTB)模式研究[D].东北师范大学,2014.

(上接第 31 页)

表现视觉传达内涵,可通过编程来满足某种特定的需要,在个人 PC 上制作完成一个作品,为其加上背景音乐,输出视频作品,以光盘的形式将色彩、音乐及图像的互动效果展示给人们,这种多维度、多元化的信息给人一种身临其境的特殊感观场景,将产生良好的信息传达效果。

4 设计理念和思维的创新

将视觉传达设计与传播学、社会学、心理学等多学科知识融合,通过艺术设计、多媒体及网络等技术,引导设计者的思维和情感表达,创造理想的视觉语言,展现多元化、个性化、的内涵。作品将表现为形式多样、题材新颖、动静结合的照片和视频,能够通过计算机辅助视觉传达设计对视频图像的处理产生所需的情景特效,结合音频使作品同构通过与终端交互更深入了解作品的内涵^[4]。这种视觉传达设计能够使人们体验到视听作品的魅力,能够达到引人入胜、与之交互的目的。人们在生活中凭借媒体与工具的辅助,根据实际需要和表达内心的想法,动手创作一些简单有趣的作品(微博、微视频、微电影等)视频拍摄与剪辑等,不乏很多好的创意。

视频拍摄与剪辑逐渐作为视觉传达设计中的重要方式与手段。随着数字技术的成熟,将来的动态影像将变成视觉传达设计必不可少的重要素材。视觉传达设计的动态影像具有动态性,最容易引起观众的注意和共鸣,同时,为了防止由于过多使用没有深度的动态影像而引起的视觉疲劳,需要考虑观众视觉感受过后,对作品内涵的认同程度。因此,在视频传达设计中需要营造出内涵体会的组成要素及形成手段,需要观众的共鸣,促使其达到视觉传达设计中的主题传播效果,这种共鸣感将作为视觉传达设计中中视频影像设计的主要

目标。新科技和新媒介推动了视觉传达不断向前发展,我们需要通过思维、设计理念和传达方式进行创新,关注视觉传达设计中的多元化、多学科融合的设计需要,进行个性化设计。移动智能终端的普及,使人们可以不受时空限制进行获取和发布信息,新媒体的视频图像的把多种元素融合起来,动态过程能够自然传达信息,通过网络使观众轻松的获取主题内容并融入设计成果。新的元素扩展了视觉传达的表现手法,使表达内容更加丰富多彩,人们对信息的认知更直接,使新媒体艺术在人们的生活中发挥更大的作用。

5 结束语

数字技术与新媒体的发展、多元化需求,计算机辅助设计者通过使用新技术丰富动态影像、创新思维,为观众带来全新的视觉感受。数字化使视觉传达设计的内容包含了更多的独特内涵,动态影像的视觉传达设计必须调动多元化、交互性因素,充分呈现文化内涵,调动情感,成功传递设计者想表达的信息,使观众产生共鸣。视觉传达设计人才的培养需要跟上数字时代,进行创新和尝试,视觉传达设计思维方式的更新将对创造性设计人才的培养产生重要的意义。

参考文献:

- [1] [日]中川作一.视觉艺术的社会心理[M].上海人民美术出版社,1991.
- [2] [美]阿历克斯·伍·怀特著,黄文丽,文学武译.平面设计原理[M].上海人民出版社,2005.
- [3] 邹攀宇. 在视觉传达设计中要注重以人为本 [J]. 美术大观, 2011(01):152-153.
- [4] 艾敬园. 视觉传达设计中的个性研究[J]. 电影评介, 2013(19): 105-106.