



(12) 发明专利申请

(10) 申请公布号 CN 115860833 A

(43) 申请公布日 2023. 03. 28

(21) 申请号 202211150268.4

(22) 申请日 2016.07.25

(30) 优先权数据

62/196,637 2015.07.24 US

62/317,440 2016.04.01 US

(62) 分案原申请数据

201680055868.9 2016.07.25

(71) 申请人 安普视频有限公司

地址 美国加利福尼亚州

(72) 发明人 D·雷 R·麦克雷 D·古洛

J·普拉萨德

(74) 专利代理机构 广州嘉权专利商标事务所有

限公司 44205

专利代理师 黄晓升

(51) Int. Cl.

G06Q 30/0251 (2023.01)

G06Q 30/0273 (2023.01)

H04L 65/611 (2022.01)

H04L 65/612 (2022.01)

H04N 21/258 (2011.01)

H04N 21/2668 (2011.01)

H04N 21/45 (2011.01)

H04N 21/472 (2011.01)

H04N 21/643 (2011.01)

H04N 21/81 (2011.01)

H04N 21/845 (2011.01)

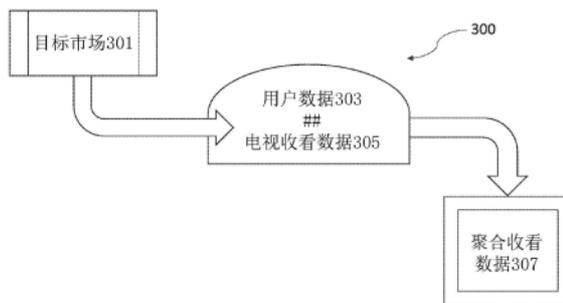
权利要求书1页 说明书29页 附图11页

(54) 发明名称

基于消费者在线行为的电视广告时段定位

(57) 摘要

在广告购买平台中,广告购买者可将广告内容导入与目标细分市场相匹配的电视/OTT/视频点播(VOD)广告库存。所述广告购买者从可用的细分市场列表中进行选择,并且系统将所述细分市场与媒体消费者的匿名数据集配对(利用散列在线ID的数据集);然后通过散列函数标识映射方法将所述散列在线ID的数据集映射至外部原始设备制造商(OEM)ID(如,来自智能电视),使得广告主可将所述广告内容定向投放给电视受众群。本文还描述了程序化电视内容的竞价方法。



1. 一种用于向电视受众群定向投放广告内容的方法,所述方法包括:

接收第一消费者群体的程序化电视数据,其中,所述程序化电视数据包括:所述第一消费者群体中的消费者可收看到的每台电视的至少一个IP地址和电视收视数据列表;

去除每台电视的IP地址标识,为每台电视创建与所述第一消费者群体中的每个消费者相关联的散列ID;

将每台电视的散列ID与消费者数据图中的消费者标识符相关联,其中,所述消费者数据图由第二消费者群体的在线行为和人口统计学数据构建而成;

基于所述第二消费者群体中的消费者的一个或多个人口统计学因素以及与所述消费者相关联的电视收视数据列表,识别电视广告库存的一个或多个时段;

购买与所述第二消费者群体相关的电视广告库存时段;以及

指示媒体渠道将广告内容投放给所述库存的一个或多个时段中的所述第二消费者群体。

2. 根据权利要求1所述的方法,其中,所述电视数据包括程序化电视数据。

3. 根据权利要求1所述的方法,其中,所述购买包括通过如实时竞价和程序化电视内容竞价等对库存出价。

4. 根据权利要求1所述的方法,进一步包括:通过浏览器或网页界面提供细分市场选项,使广告主可从多个细分市场类别中进行选择。

5. 根据权利要求1所述的方法,所述消费者数据包括可变化和不可变化数据。

6. 根据权利要求1所述的方法,进一步包括:通过机器学习过程预测细分市场的库存时段。

7. 根据权利要求1所述的方法,其中,去标识采用了单向散列函数。

8. 根据权利要求1所述的方法,其中,购买的电视广告库存时段在程序化电视或模拟电视上投放。

基于消费者在线行为的电视广告时段定位

[0001] 分案申请

[0002] 本申请为申请号2016800558689、申请日2016年7月25日、题为“基于消费者在线行为的电视广告时段定位”的分案申请。

[0003] 优先权申明

[0004] 本申请基于申请号为62/196,637、申请日期为2015年7月24日的美国临时专利申请和申请号为62/317,440、申请日期为2016年4月1日的美国临时专利申请提出,并根据《美国法典》第三十五编第119条(e)款主张其优先权,上述两个申请的全部内容在此引入本申请作为参考。

[0005] 相关申请的交叉引用

[0006] 本申请涉及于2016年7月25日提交的美国专利申请(名称:“跨屏广告投放的优化(CROSS-SCREEN OPTIMIZATION OF ADVERTISING PLACEMENT)”);代理人案号:2792-00-004U01)、2016年7月25日提交的美国专利申请(名称:“广告绩效的跨屏测量精度(CROSS-SCREEN MEASUREMENT ACCURACY IN ADVERTISING PERFORMANCE)”);代理人案号:2792-00-005U01)、2016年7月25日提交的美国专利申请(名称:“广告内容跨媒体设备的顺序投放(SEQUENTIAL DELIVERY OF ADVERTISING CONTENT ACROSS MEDIA DEVICES)”);代理人案号:2792-00-006U01)、2015年7月24日提交的申请号为62/196,618的临时申请、2015年7月24日提交的申请号为62/196,898的临时申请、2015年7月24日提交的申请号为62/196,592的临时申请、2015年7月24日提交的申请号为62/196,560的临时申请、2015年12月8日提交的申请号为62/264,764的临时申请、2016年1月14日提交的申请号为62/278,888的临时申请以及2016年2月2日提交的申请号为62/290,387的临时申请,以上专利申请的全部内容在此引入本申请作为参考。

技术领域

[0007] 本说明书中描述的技术通常涉及基于受众的在线收看习惯数据有效匹配广告内容与电视及其他设备上的电视节目受众,更具体地涉及程序化电视设备的广告库存的竞价方法。

背景技术

[0008] 视频广告是最高端、最复杂和费用最高昂的广告内容形式之一。除视频内容本身的制作成本外,通过广播和有线网络传输视频内容所需的费用也相当可观,其部分原因是由于电视(TV)时段在当今的经济背景下仍然具有较优质的广告宣传空间。此外,电视也不再是媒体市场的一个独立分支,电视内容的收看数据可通过不同方式进行跟踪。现在,消费者不仅能够通过传统的电视和DVR拓宽视频内容,尤其是优质内容的观看渠道,而且还可通过智能电视、游戏机、移动设备和传统电视获得OTT视频资源和视频点播服务。

[0009] 总之,随着收看现场直播的观众人数比例减少,而收看点播视频(VOD)和流式视频文件的观众人数比例增加,电视收看的模式正转变成为数字化分布式收看。

[0010] 将在线消费添加到任何指定消费者可选择的选项列表中只会使得在协调投放视频广告给相关行业的公众时的流程变得更为复杂。这就意味着当前用于优化广告内容投放的任务量已远远超出了传统优化的任务量,并且也超出了之前在投放优化方面有经验人士的能力范畴。由于每个个体和家庭均可通过网络设备以迥然不同方式来收看越来越多的媒体,因此,完全了解某个指定消费者所需的数据通常是非常零散的。总之,当前的复杂环境需要采用专门量身定制的技术解决方案,并且需要由技术人员通过可以以人类无法达到的速度处理大量因素的计算方法来做出决策。例如,人工分析已被用于基于诸如台式设备用电子数据表及其他静态数据管理工具等来引导(并在某些情况下继续引导)广告库存的选择。但是,这将会导致选择效率低下,并且对市场趋势的响应迟缓。这些方法也无法快速、准确地整合消费者在其所有设备上的行为信息。

[0011] 因此,影响广告主选择广告库存和内容投放类型的重要因素有很多,比如,广告的播放时段、期望的展示次数、期望广告接触的受众类型以及广告时段的价格等。

[0012] 然而,广告主在决定内容投放位置和时间时高度依赖于媒体渠道和面板数据供应商提供的信息,并在针对后续广告策略做决定时也是高度依赖于媒体渠道和面板数据供应商提供的信息来评估广告投放的有效性。内容的投放方式及形式尤其受到内容供应商可获得的收看数据信息的影响。例如,内容供应商可根据时段和节目内容通知广告主哪些人群可能是某个指定节目的受众。然而,在当今的富媒体环境下,在决定广告内容的投放时间和投放设备时需要考虑的因素更多。

[0013] 此外,在当今的广告环境下,想要同时确保内容投放的灵敏性和灵活性既重要也复杂:广告主希望能够对市场状况或特定事件(如,新闻发布或大型体育赛事)快速做出反应;广告主还希望在意识到最初的策略不佳时采取行动,并且希望消费者访问多个不同的收看平台。

[0014] 另外,媒体渠道相互隔绝,使得它们无法同时通过电视媒体和数字视频平台协调广告活动。广告主在每个渠道中的团队和工具通常是不同的。例如,互联网企业——谷歌和脸书被视为是媒体渠道,因为它们拥有自己的平台,可向专门的消费者群体发送广告内容。所有这些企业均限制相互交换数据。同样地,这也不易于广告主在脸谱等社交媒体网络与DirectTV等电视内容供应商之间协调广告内容的投放。因此,许多广告商将广告预算分为电视投放预算和在线投放预算。

[0015] 然而,目前整合内容所有者、设备及媒体格式的数据花费昂贵而不切实际,并且在某些情况下,根据政府隐私法规的规定,还是非法的。由于存在这些限制条件,目前的广告库存采购在很大程度上依赖于有限的数据库对广告和目标细分市场进行不完整匹配。

[0016] 因此,目前的工具无法对看过特定广告的收看者做精确跟踪,这是因为无法通过该人士收看过内容的所有可用媒体渠道来聚集其信息。

[0017] 大多数电视广告的购买决定都是基于目标受众的面板数据(如,由尼尔森公司提供的数据)做出;根据定义,面板数据涉及由评级公司选择用于代表大多数人群的固定消费者群体。群体中的成员报告其看过的电视内容。然后,生成样本数据和报告度量的统计模型,通过这些统计模型推断国家或区域的收视趋势及行为。例如,广告主从电视面板公司(评级机构)接收数据,并将这些数据信息用于决定其广告活动的设计及实施方式。另外,在线收视数据还可由Comcast、Netflix和YouTube等内容供应商跟踪。因此,广告主还可接收

由康姆斯科、尼尔森和凯度等在线跟踪受众的线上面板公司提供的数据。此外,有线电视运营商也会向订购者出售其收视数据。但是,这些不同来源的数据需要由广告主和品牌经理独立分析。基于面板的模型的另一缺点是其依赖于相对较少的固定参数来对观众进行表征。

[0018] 虽然目前的信息工具已经可以通过计算受众在每台媒体设备的浏览数或供应商提供的浏览数来量化受众的参与度,并可基于可用的人口普查数据来推断个体可能收看的广告类型,但是,通过多个媒体渠道收集到受众确切行为在准确度或速度上无法满足要求。因此,为了使广告接触到所期望的受众,广告主需要通过多个媒体渠道不断重复播放同一个短片,或者通过精选的流行媒体渠道多次不断重复播放同一个内容或其非冗余版本。然而,在预测哪些观众实际收看内容的次数(频次)时仍然存在难题。

[0019] 评估用户是否观看了电视推送内容是历来就存在的一项难题,这是因为很难确定一个人是否实际上观看了电视上播放的节目或片段。

[0020] 然而,三星、LG和Vizio等原始设备制造商生产的“智能电视”(也被称为联网电视)的出现为这些数据的测量提供了更为可靠的手段,这是因为智能电视可允许消费者选择在线连接电视。电视连接至互联网和馈送资源(如,有线电视公司),使其可通过互联网实时发送用户在特定电视上观看过的节目的馈送数据。

[0021] 智能电视提供的数据可用于生成测量结果,其中所含的信息量至少可与尼尔森公司相媲美,并且由于种种原因,前景更为光明,例如,智能电视接收到的数据中的信息量相比于仅仅只有指定观众是否收看特定节目的简单响应(是/否)更为丰富;智能电视的流通量甚至超过了评级公司部署的最大面板数,并且这一数字还将继续增涨;以及智能电视的数据还可与指定消费者的其他相关数据相关联。这就意味着依赖消费者组验证模型的老式技术将不再有意义。

[0022] 然而,谷歌等在线媒体经销商并不具备来自智能电视的数据。因此,现有的广告策略技术在不同媒体之间有所不同。例如,数字广告能够基于已知的在线行为定向投放,而电视广告策略却基于人口普查数据专注于使特定人群接触到广告。此外,虽然目前数据驱动和自动化电视广告(被称为程序化“PTV”电视广告)的花费占有所有电视广告费用的7%,并且每年增长一倍,但是,电视领域的购买决策反馈时间与在线内容的决策反馈时间仍存在差距。因此,另外专门针对程序化电视内容提供购买广告库存的竞价方法是有用的,这使得广告主能够可靠地接触到增长部分的人群,其中的增长数据通过程序化电视的收视数据获得。

[0023] 由于没有更加细化的消费者分类,因此,利用现有的统计方法对消费者行为进行预测可能是无效的。消费者分类数据是广告主和品牌经理购买广告库存时的决策核心。

[0024] 目前,尚无法对观众的电视收看行为和在线收看数据进行匹配。单独的面板数据并不能为分析师提供所有已经收集到的消费者行为,并且分析师也不具备一对一分析的能力。目前,如果可跨设备和媒体供应商采集数据,则可对用户的特定行为做出更好解释。然而,隐私法的规定制约了从媒体、设备和服务运营商收集和共享用户个人身份数据以及商业化用户个人身份数据的能力。在某些情况下,收集受隐私法保护的数据还是违法的。因此,这些数据资料中不得含有可导致消费者身份面临非法暴露风险的信息。利用散列标识符可在保护用户隐私的同时采集数据,但是,目前的方法无法为广告购买平台提供快速有

效且安全的散列识别。

[0025] 因此,需要一个可提供统一的视频广告库存购买经验,并且能够可靠地定位相关消费者群体的系统。

[0026] 本说明书中对背景技术的探讨是为了对技术的背景进行解释。不能将其视为是承认本说明书中所述及的任何资料信息是公开的、众所周知的,或者是在所附任何权利要求的优先权日已是公知常识的一部分。

[0027] 在本申请的整个说明书和权利要求书中,词语“包括(comprise)”及其变体(如,“comprising”和“comprises”)并非旨在排除其他附加物、组件、整数或步骤。

发明内容

[0028] 本公开旨在解决与跨显示设备投放广告内容的优化有关的消费者及广告库存的处理问题。尤其是,本公开包括由计算机或计算机网络执行的用于解决与跨显示设备投放广告内容的优化有关的消费者及广告库存处理的方法。本公开进一步包括用于执行该方法的计算装置和具有执行该方法的指令的计算机可读介质。本公开的装置及流程尤其适用于在线及电视媒体的视频内容。

[0029] 本发明中的方法使用消费者的在线行为和类型数据(如,直接测量获得的数据)作为购买目标电视库存的基础。通过数据映射可在无需任何优化(如,约束优化或拟合)的情况下对潜在受众人群的成员进行合理推导。例如,可对600万用户做在线使用识别,其中,200万用户拥有智能电视。可检索这200万用户的电视收视数据,从而看出哪些电视节目最适合投放广告内容。该方法解决了目前无法针对消费者定向投放电视广告内容的局限性。但是,如果能够识别出高关注度的潜在电视观众,则对广告主而言更为有利。

[0030] 如本文中所述,使用深度学习方法就能够了解各条内容之间的关系。

[0031] 该技术包括通过绘制个体与视频内容元素之间的关系图来弥合在线与离线视频广告视频购买之间的差距的方法。

[0032] 该技术包括媒体购买平台,该平台通过结合数字定位及报告功能与离线策略来弥补在线视频策略与离线视频策略之间存在的差距。

[0033] 本文的技术将媒体消费与个人属性相关联,从而为广告主在电视消费者数据与在线消费者数据之间提供了路径导航。

[0034] 本技术领域的一个问题是从电视和有线电视公司接收到的数据只包含正在观看的节目信息,而不包含人口统计学信息。通过将在线观看及观看其他数字内容获得的信息与电视收看数据耦合可获得一个结果,即基于群体中的指定部分人群的在线观看行为确定该部分人群选择观看或可能选择观看的电视节目。

[0035] 本文的技术进一步包括一种用于在通过细分市场检索数字广告时散列身份的方法。在电视/OTT/VOD广告购买平台中,购买者可选择与目标细分市场相匹配的电视/OTT/VOD库存(如,节目和播放内容)。购买者从提供的列表中选择细分市场,而本发明的系统及方法则在处理环境(平台供应商的服务器)中匹配用户选择与媒体消费者(带在线散列ID)的匿名数据集。

[0036] 然后,通过平台的身份散列方法将在线散列ID映射到外部OEM的ID(如,三星的在线用户ID等)。目标细分市场可通过在线数据资源推导出,如,第三方社交媒体数据、第三方

在线购买历史记录、第三方返回链接历史记录、第三方链接历史记录、第一方获得的cookie数据及其他数据(如,包括浏览、社交媒体内容和购买等在内的互联网行为数据)。

[0037] 该方法可进行匿名匹配或细分市场描述,而不会向任何终端用户或第一方平台供应商提供实际在线受众的身份或IP地址信息。媒体策划人、广告商和电视广告主等终端用户被屏蔽,以避免暴露目标细分市场的身份及其元数据。同样地,系统可存储和处理散列身份信息,使得可在不暴露受众身份的情况通过竞价等购买广告库存。

[0038] 本公开进一步提供了一种用于向电视受众群定向投放广告内容的方法,该方法包括:接收第一消费者群体的程序化电视数据,其中,该程序化电视数据包括:第一消费者群体中的消费者可收看到的每台电视的至少一个IP地址和电视收视数据列表;去除每台电视的IP地址标识,为每台电视创建与第一消费者群体中的每个消费者相关联的散列ID;将每台电视的散列ID与消费者数据图中的消费者标识符相关联,其中,消费者数据图由第二消费者群体的在线行为和人口统计学数据构建而成;基于第二消费者群体中的消费者的一个或多个人口统计学因素以及与该消费者相关联的电视收视数据列表,识别电视广告库存的一个或多个时段;购买与第二消费者群体相关的电视广告库存时段;以及指示媒体渠道将广告内容投放给所述库存的一个或多个时段中的第二消费者群体。第一群体虽然并非一开始就是第二群体的子集,但最终通常都是第二群体的子集。

[0039] 本公开进一步包括用指令编码的计算机可读介质,该指令用于执行本文中所述的方法以及通过一个或多个适当配置的计算机进行处理。

[0040] 本公开还包括计算装置,配置用于执行诸如存储在计算机可读介质上用于执行本文中所述方法的指令。

附图说明

[0041] 图1是投放广告内容的广告主、广告交易平台、媒体渠道及消费者等各方之间的关系示意图;

[0042] 图2是帮助广告购买者识别电视市场中的广告点位的概述信息的示图;

[0043] 图3是消费者图;

[0044] 图4是节点图;

[0045] 图5A和图5B是创建消费者图的步骤的示图;

[0046] 图6是在识别电视市场中广告点位的过程中使用的去除消费者数据标识的方法示图;

[0047] 图7是多维数据集示图;

[0048] 图8是消费者设备与DSP之间的信标事件数据的通信示图;

[0049] 图9是本说明书中描述的过程的执行装置的示图;以及

[0050] 图10是广告购买领域中各实体之间的关系示图。

[0051] 在各附图中,相同的参考符号是指相同的元部件。

具体实施方式

[0052] 本技术涉及允许广告主基于节目观看行为将广告内容定向投放给电视受众群的计算机系统和方法,其中,节目观看行为通过结合电视台和智能电视的电视收看数据以及

从第一方和第三方中的一方或两方获得的在线和人口统计学数据推导得出。数据的组合方式使得个人身份信息得到保护,不会在各数据资源之间共享。

[0053] 广告功能

[0054] 在广告内容的购买、投放及消费业务中,实体之间的关系如图1中所示。如图所示,广告生态系统是一个复杂的、涉及多个不同实体、存在多种不同关系的体系。

[0055] 广告主101是广告库存109的购买者。广告主可能是直接对广告功能实施控制的企业,也可能是管理一个或多个客户(通常是企业实体)的广告要求的广告商。广告主旨在通过每个消费者的一台或多台设备107使广告内容103(也可称为“广告”)可让一个或多个消费者105(通常是消费者群体)获得。

[0056] 对于指定消费者,设备107包括下列装置中的一个或多个:电视机(包括智能电视机)、移动设备(手机、智能电话、媒体播放器、平板电脑、笔记本电脑和穿戴设备)、台式计算机、影音图像类软件、机顶盒、游戏机、流媒体设备、可在“物联网”内运行的家用电器等设备(如,冰箱等)以及恒温器和报警器系统等其他家用联网监控设备。

[0057] 广告内容103通常由广告主101或其签约的第三方创制,并且通常包括旨在促进特定产品或服务的销量或消费者意识的视频、音频和/或静态影像。广告内容103通常通过一个或多个中介方投放至消费者(本文中将进一步描述)。

[0058] 广告内容通常分为两种不同的类型:品牌营销和直接响应营销。这两种类型的时间表也不甚相同。品牌营销旨在提升品牌意识,而直接响应营销旨在形成即时响应。例如,汽车制造商可能会将直接响应营销材料投放到市场中,并通过查看谁在看到广告之后去找过经销商或访问过网站的方式来衡量市场反应。本文中所述的方法均可适用于这两种类型的广告内容,但对这两种类型广告内容的有效性的衡量却不同。例如,品牌营销的有效性通过总收视率GRP(在本文其他地方将做进一步描述)衡量,而直接响应营销的效果则通过如网站访问量等来测量。

[0059] 另外或替代地,在将广告内容103投放至电话或平板电脑等移动设备时,可采用文本/SMS信息、电子邮件,或者警报、横幅或徽章等通知信息的形式。在将广告内容103投放至台式计算机、笔记本电脑或平板电脑时,可在应用程序或浏览窗口中以弹窗形式展示,或者可设计成在下载或缓存其他所需视频内容时以视频形式播放。

[0060] 消费者105是广告内容103的浏览者和潜在浏览者,也可能是之前已购买过广告中产品或服务的消费者,也可能是正通过观看广告内容103首次了解广告中产品或服务的潜在消费者(这对广告主而言是有利的)。

[0061] 广告库存109(也可称为广告位或可用库存)包括跨多个媒体界面或媒体渠道111的广告可用时隙或时段117,其中,消费者可通过媒体界面或媒体渠道访问信息和广告内容。这类媒体界面包括电视、收音机、社交媒体(如,LinkedIn、Twitter和Facebook等在线网络)、数字广告牌和移动应用程序等。媒体渠道111可生成自己的内容113,或者播放由一个或多个其他内容供应商或发行商115提供的内容。例如,有线电视公司就是一个可投放多个电视频道制片商和内容发行商的内容的媒体渠道。媒体界面也通常被称为内容供应商,这是因为他们可将媒体内容113(电视节目和电影等)投放给消费者105。本文技术的一个方面包括:从一种以上类型的媒体界面或内容供应商聚集广告库存109。媒体渠道111还可将已购买在时段117投放的广告内容103投放给消费者105,供其在多台设备107上观看。发行商

115通常是内容所有者(如,BBC和ESPN)。

[0062] 时段117即时间,通常用一天中的特定时间(中午、下午16:30等、下午14:00~16:00或下午21:00~次日12:00)或者特定广播频道(如,电视台或社交媒体馈送)上的特定播放(如,电视节目)期间的窗口(1分钟和2分钟等)表示。可用时段是指广告库存中可供广告主购买用于投放广告内容的时间段。该时段还可用通常是因为另一位广告主尚未购买该时段。在本说明书中的进一步描述中,时段还可受某些特定限制条件的约束,如,特定类型的广告内容103是否可在特定时段进行投放。例如,运动设备制造商可能已经购买了由特定频道上的特定时间限定的特定时段,并且还可能已经购买了排除其他运动设备制造在某个范围内(在第一个制造商的时段内)购买相同频道上的时段的权利。在这种情况下,“硬约束”是指法律限定或强制性限制在特定时段或特定媒体内投放广告。“软约束”是指希望限制(非强制性限制)在特定时段或特定媒体内投放广告。“约束满足”是指找到一组约束的解决方案的过程,这些约束可以是施加变量必须满足的条件。因此,该解决方案是一组满足所有约束的变量值。

[0063] 从广义上而言,信息是指消费者可通过电视机或计算机等的屏幕,平板电脑、智能手机和笔记本电脑等移动设备的显示器,智能手表和健身监测器等穿戴设备,或者车载或机载显示屏收看、阅读、收听或以其组合的方式获得的任何内容。信息可由电视台或无线电台等媒体界面111、有线电视运营商(如,Comcast)等多频道视频节目经销商(MVPD),或者雅虎(Yahoo)或脸书(facebook)等在线网络提供。

[0064] VOD是指视频点播系统,用户可通过该系统选择和收看/收听其选择的视频或音频内容,而无需在预定的播放时间收看。互联网技术通常用于将点播视频传输给电视和个人电脑。电视VOD系统可通过机顶盒、计算机或其他设备流传输内容,使得内容可被实时观看或下载到计算机、数字录像机(也被称为个人用视频录像机)或便携式媒体播放器等设备上以供随时观看。

[0065] 广告主与媒体渠道之间的通信可由包括要求方平台(DSP)123、广告交易平台119和供应方平台121在内的多个实体进行管理。广告交易平台119(本文中也称为平台)是广告主可对可用媒体库存进行竞价的环境。该媒体库存可以通过互联网在线传送或者通过SiriusXM等数字无线电传送的数字库存,或者是通过ESPN、CNN、Fox或BBC等电视频道或FM/AM无线电广播传送的模拟库存。广告交易平台119通常专门运营特定类型的内容。例如,SpotX专注于数字内容,而WideOrbit专注于程序化电视。

[0066] 供应方平台(SSP)121是从媒体渠道111获取库存109并(可选地,通过平台119)提供给要求方平台(DSP)123的中介机构,使得广告主在定位广告内容103时可购买或竞价购买库存。根据媒体渠道是否能够限制访问库存的广告主的身份和数量,SSP有时又可被分为公开SSP或私有SSP。在有些情况下,如果广告内容的广告主所依赖的广告交易平台功能由DSP和SSP中的一个或二者执行,则SSP可直接与DSP交互,而无需广告交易平台的参与。本文技术尤其适于由适当配置的DSP来执行和实现。

[0067] 在一个配置中,广告交易平台119是供应方平台(SSP)121与要求方平台(DSP)123之间的连接接口。该接口的作用包括接收来自一个或多个SSP 121的库存109,并将其提供给DSP,然后接收来自DSP的库存的竞价125,并将这些竞价125提供给SSP。因此,DSP使得广告主可对SPotX或WideOrbit等特定SSP提供的库存进行竞价。在一些配置中,DSP承担了广

告交易平台的大部分或全部功用。

[0068] 在本文技术的一个实施例中,DSP向广告主提供广告活动时间表,如果获得广告主许可,则DSP必须代表广告主购买广告活动,并安排广告活动的实施。SSP控制投放广告内容至媒体渠道。

[0069] 广告活动(或活动)是广告主将广告内容投放给特定消费群体的计划。活动通常包括广告内容(如,特定形式或各种形式的广告或以特定顺序投放的一系列相关广告)和活动运行时间段(如,1周、1个月和3个月等)的选择。广告主通常将活动描述127发送给广告交易平台119或DSP 121,并反过来接收可用的库存列表109。活动描述127可包括单条广告内容103和用于定向投放的一种或多种设备107,并且包括用于跨一台或多台设备107顺序投放两条或多条广告内容103的时间表。活动描述127还可包括目标受众描述,其中,目标受众由一种或多种人口统计学因素定义,而人口统计学因素包括但不限于年龄范围、性别、收入和住址。

[0070] 然后,DSP 123提供一个界面,广告主101可通过该界面使其活动描述127与库存109对齐,并购买或竞价购买库存中的各种时段117。DSP 123或平台119可提供一组以上与指定活动描述127匹配的库存。在本文中,每组与指定活动描述匹配的库存被称为广告对象129。广告主101可从广告对象列表或多个对象中选择想要购买的对象。一旦特定对象购买成功,将通知SSP 121,并发送投放指令137给各个媒体渠道111,指示其在适当时段117或在内容113选择期间将广告内容103向相关消费者投放。

[0071] 指定时段的购买并不仅仅是一个直接的定价售卖过程,而是通过竞价实现。DSP将针对多个时段竞价,并为每个时段确定一个投标价格,提交给SSP。SSP将竞价成功的广告内容交付给媒体渠道,并最终向消费者投放。具体定向的出价通常比总体定向的价格高。

[0072] 竞价过程部分取决于广告内容的类型。电视内容可提前预定,而在线内容通常采用“准时制”竞价结构,即在特定消费者在线时才会投放广告。本文的方法通常独立于竞价过程,并且适用于通常使用的任何竞价方法,包括实时竞价和基于详细的程序化电视数据的竞价方法。

[0073] 通过视频播放器广告接口定义(VPAID) (<https://en.wikipedia.org/wiki/Mixpo>) 或视频广告投放模板(VAST:<http://www.iab.com/guidelines/digital-video-adserving-template-vast-compliance/>) 等协议为指定的在线广告添加标签,可采集到包括消费者是否点击或浏览广告内容在内的数据。其中,标签通常包含多个与消费者与广告内容互动的方式有关的数据项。可将该数据项返回SSP和/或DSP,用于为广告的投放提供反馈信息。例如,数据项可包括与用户是否在线点击视频有关的数据。某些数据项由于其相对于广告主而言所具备的显著性而与行业中被称为“信标”的事件相对应。例如,信标事件可包括用户在完成视频片段浏览之前已停止视频片段播放的事实。

[0074] 生成广告对象的过程还取决于一个或多个活动要求。本文中使用的活动要求是指由广告库存的广告主或其他购买方制定的预算等财务限制以及目标消费者数量等性能指标。在购买或竞价购买库存时,活动要求信息与活动描述一起使用。

[0075] DSP 123还可向广告主101提供自各种资源汇聚形成的消费者及设备相关数据。该数据可帮助广告主选择最适合其目标的库存、时段及媒体渠道。

[0076] DSP使用的数据可包括人口普查数据131或特定消费者及设备的数据133。人口普

查数据131包括可用于优化库存的购买的人口数据。因此,人口普查数据131可包括特定收看区域内的群体的年龄分布、收入变化和婚姻状况等人口统计学数据,其中,收看区域与群体成员实际浏览的媒体界面无关。人口普查数据131可源自多个资源渠道,如,州郡记录及美国人口普查局。

[0077] 数据管理平台(DMP)135可提供其他类型的关于消费者及其在DSP上使用的设备的第三方数据。DMP通常提供具有嵌入式功能的数据仓库服务。DMP可下载数据,并可执行各种功能,包括分类、存储、处理和进行匹配算法,以及向购买方和订购方提供数据输出。DMP的示例包括KruX、Exelate、Nielsen和Lotame。第三方运营商发送给DSP的消费者及设备数据133可用于补充媒体渠道提供的其他消费者及设备数据143。与广告主有关的消费者及其使用设备的数据包括浏览习惯以及可通过媒体渠道直接检索到的特定行为数据。例如,如在本文中其他地方所做的进一步描述,媒体渠道在向消费者投放广告时可采集到用户访问广告的方式的信息。由于所涉数据量的原因,媒体渠道在相对较短的时间(如,14天)内无法提供特定消费者的任何信息。在这种情况下,DSP可从第三方(如,DMP)获取用户数据。第三方也可以离线获取数据。本文中使用的离线事件是独立于互联网浏览或电视收看而发生的事件。例如,离线事件可包括从商店购买物品以及广告主认为重要的其他类型的位置事件。数据可通过普遍接受的任何数据共享和传输文件格式在本文中所述的实体之间(如,DMP与DSP之间、DSP与SSP之间或者媒体渠道与SSP/广告交易平台之间)共享。其中,这些格式包括但不限于:JSON、CSV、Thrift以及任何适合格式的文本文件。

[0078] 一次展示(印象)是指广告接触消费者的任何机会。对于电视广告而言,假设电视中正在播放广告,且已知的电视拥有者或定期收看电视的观众已经接触到广告,则本次显示被计为一次展示。如果同一家庭中有多名成员,则展示次数等于可观看电视的人数。在在线环境下,如果消费者正在浏览(如,网页),并且广告以弹窗形式展示在网页上,或者用户点击了可触发广告运行的链接,则计为一次展示。

[0079] 细分受众是利用Cookie同步过程或其他方法消除个人识别信息后的消费者列表,其中的消费者属于同一类型(收入、性别、地域及位置等)或者与一种行为相关联(如,购买、电视收看和网站浏览行为等)。

[0080] Cookie同步是指一种可使数据在DMP的SSP与DSP之间,甚至是通常在内容发行商与广告购买方之间交换的过程。Cookie是移动设备或台式计算机用于存储和恢复与特定用户或设备有关的信息的文件。Cookie中存储的信息受保护,因此,通常只有创建Cookie的实体才可检索其中的信息。Cookie同步是一种允许一个实体从另一个实体创建的Cookie中获取消费者信息(而无需获取消费者确切身份信息)的方法。因此,在通过Cookie同步从媒体渠道中检索特定消费者的特定信息时,可从DMP中获得该消费者的其他信息。

[0081] 对于移动设备,特定的设备具有唯一识别的设备ID,而智能电视具有散列IP地址。设备ID信息可用于将一组设备链接至特定消费者,以及将多个消费者(如,指定家庭)链接至特定设备。通过结合可增量更新“cookie”数据的移动设备ID和电视地址,DSP可收集一段时间内累积的存储数据。

[0082] 跨屏是指经由指定消费者的多个设备分发包括广告内容在内的媒体数据。其中的设备如电视屏幕、计算机屏幕、移动设备显示屏(如,笔记本电脑、智能手机和笔记本电脑等)、智能手表和健身监测器等穿戴设备、车载或机载显示屏,以及冰箱等联网家用电器的

显示屏。

[0083] 到达率(reach)是指在特定时期内不同人群至少一次接触广告的总次数。

[0084] 在跨屏广告或媒体活动中,同一消费者可通过其使用的不同设备(如,电视、台式计算机或移动设备)多次接触到一个广告。非重复到达率是指不同人群接触广告的次数,而不管其使用的是何种设备。例如,如果特定消费者通过其电视、台式计算机及一台或多台移动设备接触过广告,则该消费者的到达率仅为1次。

[0085] 增量到达率是指一项活动的额外非重复到达率,即超过活动开始之前(如,以前的活动)的到达率。在本文的一个实施例中,一种活动可包括电视分机。在这种情况下,即使广告主已经在电视上推出了活动,但是活动的收益也到达其递减点。广告主希望找到可修改活动计划的方式,使其适合数字市场,以提高到达率。通过这种方式,DSP可继续运行已经在一个或多个媒体渠道上运营的活动。

[0086] 除了电视节目内容和投放至台式计算机及移动设备上的在线内容外,还可在OTT内容中投放广告。OTT(源自术语“过顶传球”)是指通过互联网传送音频和视频,而无需MVPD参与内容的控制或分发。因此,OTT内容与任何特定机顶盒或设备不存在关联。例如,Netflix或HBO-Go提供的OTT内容使消费者不需要通过特定的设备进行观看。与之相反,如传送给电缆或机顶盒的MVPD内容,其电缆或机顶盒由Comcast、AT&T和DirecTV等电缆或卫星运营商控制,因此,其不能被称为OTT。OTT尤其是指来自诸如Sling TV、YuppTV、亚马逊实时视频(Amazon Instant Video)、Mobibase、Dramatize、Presto、DramaFever、Crackle、HBO、Hulu、myTV、Netflix、Now TV、Qello、RPI TV、Viewster、WhereverTV、Crunchyroll和WWE Network等第三方的内容,其直接被传送给终端用户设备,互联网服务运营商(ISO)仅负责传输IP数据包。

[0087] 此外,OTT设备是指连接至互联网且可访问多个内容的任何设备。例如,Xbox、Roku、Tivo、Hulu(以及可在电缆上搭载运行的其他任何设备)、台式计算机和智能电视机即是OTT设备的实例。

[0088] 总收视点(GRP)是根据所涉及的时间表和媒体渠道衡量广告活动投放量的计量单位,用每个目标受众的接触百分比表示(因此,GRP可大于100)。例如,如果有30%的洛杉矶人群接触过广告4次,则 $GRP = 120$ 。(例如,可由尼尔森公司团队组织1000名洛杉矶市民参与测量数据)。

[0089] 目标受众收视率(TRP)是指基于样本人口的每个目标受众成员的展示次数。该展示次数涉及到个人,如,广告主希望在洛杉矶有25岁及以上年龄的男性受众。如果在洛杉矶小组中有100名符合条件的男性受众,且其中70%的受众看过广告,则 $TRP = 70\% \times$ 观看人数。

[0090] 实时是指实时计算,其定义为计算机系统可接收和处理数据,并可非常迅速地(如,几秒内)返回分析结果,而不会给依靠该分析结果做决定的一方造成延迟。如本文中所述,假设允许广告主选择、竞价和购买广告库存的过程可被实时执行。

[0091] “设备图”是指用户与用户用于消费媒体的设备之间的关系图。

[0092] 跨屏广告内容的投放与优化

[0093] 本说明书中所述的技术使广告主可跨多个媒体渠道(包括电视和在线媒体)将广告内容定向投放给消费者。广告主可在两种情况下向消费者定向投放广告。在1:1情况下,

如果消费者与目标参数相匹配,则DSP可仅用实际细分人群和/或实际细分人群的模拟数据实时决定是否投放广告。在索引法中,当无法1:1定向投放且无法动态插播广告或实时决策时,系统将会考虑预计访问时段(如,电视节目或VOD节目)的观众的集中度,然后向目标消费者集中度最高的时段定向投放广告。

[0094] 在优选的实施例中,广告主可对广告内容的分配进行控制,这是因为广告主可通过一个统一界面访问系统,其中的界面可显示广告库存相关信息、管理广告库存竞价,以及提供符合活动描述和广告主预算的潜在广告对象列表。然后,系统与供应侧供应商等通信,以确保能够购买到所需的时段(通常通过竞价过程)以及投放或促使广告内容的投放。

[0095] 消费者数据

[0096] 消费者数据可分为两类:种族和性别等不可变化的特征,以及年龄、职业、住址、婚姻状况、收入、趣味和偏好等可变化的特征。职业等各种可变特征可随时间发生变化,而年龄等其他特征的变化速率却是持续一致的。目前,用于跟踪广告内容目标消费者信息的数据系统无法用于广泛地跟踪两种类型的消费者数据。大多数数据系统包含消费者分类信息均是静态的、同质化的。例如,两年前购买汽车的29岁人士的数据将会成为一个消费者数据点,不会随时间更新或增添。虽然存储在系统中的个人的年龄分类信息可随时间进行调整,但是婚姻状态或生活方式的变化等其他可变化的特征却不在调整的范围之内。

[0097] 确定电视收看数据

[0098] 本说明书中所述的技术使广告主可跨多个媒体渠道(包括电视和在线媒体)将广告内容定向投放给消费者。广告主可在两种情况下向消费者定向投放广告。在1:1情况下,如果消费者与目标参数相匹配,则DSP可仅用实际细分人群和/或实际细分人群的模拟数据实时决定是否投放广告。在索引法中,当无法1:1定向投放且无法动态插播广告或实时决策时,系统将会考虑预计访问时段(如,电视节目或VOD节目)的观众的集中度,然后向目标消费者集中度最高的时段定向投放广告。

[0099] 图2从终端用户(如,广告主)的角度示出了本文的方法300的操作概图。优选地,该方法可在DSP或广告交易平台上或由DSP或广告交易平台执行。在一个实施例中,提供了一个界面,广告主可通过该界面从可用的目标市场数据列表中选择目标市场301的属性,如,位置、年龄、收入水平和性别。一旦选定,这些属性就可与配对的用户数据的数据库相匹配,然后就可推导出电视收看数据305,其中的数据库包括从在线受众数据中获得的消费者数据303。消费者数据和电视收看数据可通过本文的方法做散列处理,以保护个人用户的身份信息。由于通过在线数据识别的指定消费者可被映射到特定的电视收看数据组中,因此,可在两组数据之间建立关联。

[0100] 广告主接收所选目标市场的聚合收看数据307,其中,该数据可包括安排有广告内容的电视节目的列表。这些就是目标市场成员可能集中收看的电视节目。

[0101] 本文中的方法描述了其在各个阶段以特定方式对消费者群体中的每个消费者的处理:例如,为计算机编程,用于分析数据库中的每个消费者的数据,以确定哪些消费者(若有的话)观看了特定电视节目或访问了特定网站;或者,可执行比较分析,将一类人群中的每个用户的属性与另一类人群中的每个消费者的属性进行对比。每个群体中可包含数千、数十万,甚至数百万个个体。在本文中,假设配置在适合的计算资源上的方法能够对群体中的每个成员执行所述的计算和操作。然而,在本文的方法中,群体中的“每个消费者”也可能

是指群体中的大多数消费者,或者群体中可进行所述计算操作的所有消费者。例如,如果群体中有一个或多个指定消费者由于个人数据不足而未进行特定计算,为了获得有意义的计算结果,则此次分析的人数将被视为不足。因此,数百万潜在消费者群体中的“每一个”并不一定是指群体中的每一个确切成员,而是指可进行指定计算且足以生成计算结果的大量的、实际合理的消费群体人数。

[0102] 消费者图

[0103] 消费者图是每个节点代表一个消费者(或个人用户)的图。该技术利用加权图来呈现各个实现方式,其中,消费者(节点)之间的关系用相似度(边)界定。本文中的消费者图用于分类、存储和聚合大量不可被识别的消费者个人数据,且允许DSP等实体通过指定的消费者设备上的数据在用于构造消费者图的数据与电视观看数据等其他数据之间建立关联。消费者图的优点在于可通过潜在不限数量的输入参数来构建。将需要考虑的数据量扩展到除固定数量的消费群属性(年龄、收入和性别等)外的持续更新的使用行为数据(如,个人消费者的具体观看和购买历史记录、人口统计学数据和位置相关数据源)使得消费者之间的比较变得更为复杂。

[0104] 其中一种方式是使用确定性关系数据构图,另一种是使用每个节点的属性进行概率性的推论。在一些情况下,可组合使用确定性法和概率法。在确定性法中,其方法相对简单,主要基于消费者的确切数据,如,来自运营商的登录信息。因此,如果一个人用相同的ID多次登录不同的设备,则可确定其身份匹配。但是,如此准确的信息并不是总是存在的。相反地,在概率法中,需要进行推论:例如,如果同一设备在相同的地方出现,或者类似的行为在不同的时间出现在指定设备上,则可推断该设备属于同一用户。

[0105] 在本文的一些实施例中,采用了机器学习算法、贝叶斯算法和回归算法来探测消费者之间的共性。这些方法在参数数量有限的情况非常有用。在其他一些实施例中,深度学习技术在发现消费者相似性及构造消费图方面更为有用。机器学习是精确匹配信息片段的首选技术,例如,两个消费者是否访问过相同的网站。但是,深度学习技术可探知特定视频或电视节目的详细信息(例如,通过分析自然场景统计),从而可确定被一个指定消费者浏览过的两则广告中除其主题内容外是否还存在其他相同点。例如,两则广告中可能包含消费者喜欢的同一个演员,尽管其展示的产品却无共同之处。

[0106] 在优选实施例中,本文中的设备图基于概率数据。在用概率法构图时,使用了行为数据来匹配用户。

[0107] 在一些实施例中,DSP等实体可构建设备图,而在其他实施例中,可从DMP等其他实体获取(如,购买)设备图,但在购买时,优选地,设备图根据本文的方法来扩增。

[0108] 在本文的多个实施例中,设备图和消费者图两者均可以以一种方式一起运行,即允许将移动数据绑定到电视数据的方式。

[0109] 本说明书中使用的术语“图”是数学意义上的图形,是一个连接节点对的节点(N)和边(E)的集合 $G(N,E)$ 。图G是节点之间的关系图。根据一些准则,由边连接的两个节点彼此相似,并且边的权重定义了相似程度。不符合相似性准则的节点对不能通过边连接。图3是图形的概念图,示出了6个节点 $N_1 \sim N_6$,其中,3对节点通过边连接。

[0110] 在本文的图形的实施方式中,节点N是属性集合为A的实体或对象。如图3所示,每个节点都有一个与之关联的属性数组,以节点 N_i 的 A_i 表示。

[0111] 在本文图形的实施方式中,两个节点之间存在的边E表示两个节点之间的关系,或者两个节点之间的相似度超过预定阈值。边的权重 w_E 是两个节点之间的相似度。在图3中,边的权重以厚度示意图示出(在这种情况下, $w_{E_{12}} > w_{E_{34}} > w_{E_{15}}$)。

[0112] 在消费者图中,一个节点表示具有一组属性(如,个体的性别和年龄、电视节目观看记录和网站访问记录等)的个体或者含有两个或多个个体的家庭。

[0113] 图4是消费者图的节点结构示例图。每个节点具有包括类型和行为在内的属性集合,其中,数据从第一方和第三方资源连续采集获得。如果新的消费者信息变为可用,并且如果属性的数量(如,针对指定消费者存储的不同属性的数量)可在采集到新的消费者数据时随时间增长,则许多属性是可变化的。这使得广告主可在活动运营期间对其进行定制和调整。在本文的技术的一个方面中,可通过输入指定消费者的潜在的无限数量的数据(如,在线、离线、行为及人口统计学数据)进行构图。这些输入数据可随时间更新,并且可完善指定消费者的数据,以及扩充数据可用的消费者群体的数量。事实上,本文对可用数据的类型和特征并无限制,这就意味着本文的方法优于那些依赖静态数据库和固定人群的企业所采用的方法。

[0114] 其中一些数据采集资源如下。

[0115] 类型数据是通常不可改变的消费者分类数据,如,不可变数据。行为数据是基于消费者近期活动持续更新的数据。

[0116] 每个节点包括一个或多个设备的分组(台式计算机、移动设备、平板电脑和智能电视)。对于每个设备,基于设备的用户类型数据都是从第三方和第一方资源采集。

[0117] 表1是按类别和资源划分的数据示例。

[0118]

	第一方	第三方
不可变化		人口普查数据(政府) 家庭收入教育水平(如,从 Exceleate 采集) 性别(如,从 Nielsen 和 DAR 采集)
可变化	行为(在线) 电视收看 可见度(如,可在线持续观看的广告数量) 访问的在线网站位置事件	线下行为 零售购买 到店访问(药店、影院和汽车经销商等)

[0119] 表1

[0120] 第一方数据包括用户行为数据(如,购买、收看及网站访问等),以及收入和性别等类型数据,其中,类型数据可由运营商直接提供,以提高其活动定位和报告质量。例如,可口可乐公司可向DSP提供一份在社交媒体上“喜欢”可口可乐产品的用户列表,以提升其视频广告活动质量。第一方类型数据通过直接投放给设备的广告以及设备采集信息(如,一个或多个IP地址)进行收集。第一方类型数据包括IP地址的位置、移动设备的地理位置以及设备

是否处于商业和住宅区的信息。

[0121] 第三方类型数据由外部供应商提供。外部供应商可通过一对一Cookie同步或设备同步提供有关Cookie或设备的信息,例如,外部供应商可包括KruX (<http://www.krux.com/>)、Experian(提供购买行为数据)和Adobe等DMP。示例数据包括消费者占领的细分市场,如,年龄范围、性别、收入水平、教育税、政治面貌和喜好(如,消费者在社交媒体上喜欢或关注的品牌)。此外,外部供应商可根据设备的近期购买记录提供类型数据。第三方数据包括性别和收入等信息,因为外部供应商无法直接采集到这类信息。第三方数据可在不投放广告的情况下进行采集。观看的电视节目和购买记录属于第三方数据。

[0122] 第一方数据通常由DSP生成,例如,DSP从投放广告或品牌商/广告商提供的数据中采集到的数据。第一方数据包括依赖于已投放广告是否已被观看的数据。

[0123] 行为数据可通过第一方和第三方资源从设备上采集。行为数据通常是第一方数据,并且是可变的。

[0124] 第一方行为数据可从直接投放给设备的广告中采集。这包括网站访问行为以及通过设备观看电视节目、OTT内容或点播视频(VOD)的行为。

[0125] 第三方行为数据从外部供应商,通常为DMP处获得,如,Experian、KruX、Adobe、Nielsen、Comscore或Brightroll、potX、FreeWheel及Hulu等广告交易平台和网站。示例数据包括上个月内设备上的电视节目收看记录、网站访问记录(个人计算机、笔记本电脑或移动设备),以及移动设备的位置事件记录(如,设备是否位于星巴克)。在某些情况下,同类型的数据可从第一方和第三方实体两者中获得。

[0126] 消费者图中的节点之间的边表示消费者具有阈值相似性或者彼此交互。例如,如果节点的物理位置相近或者基于属性相似性的概率,可确定性计算边。使用的概率法包括但不限于:K均值聚类算法和基于图遍历算法的连通区域分析(基于图遍历算法包括在两个顶点之间构建遍历图的路径)。由于属性是可变化的,因此,也可通过加权,或者通过创建或删除(如果节点对的相似度分数改变)来改变边。因此,图并非是静态的,是可以随时间变化的。在一些实施例中,变化是动态的。相似度分数可随节点属性的更新而被不断地重新计算。

[0127] 属性和数据(在获得时)通常动态添加。可每周根据新的属性和数据重新构图,从而为边建立新的权重,并对新连接或重新连接的设备进行识别。(例如,可通过将计算分发给计算机网络上的许多处理器或者分发给存储在受DSP控制的数据中心上的处理器,在云端构建或重新构建图。)

[0128] 节点 N_1 和 N_2 之间的相似度 S 根据相似性度量来计算,相似度 S 是距离函数的倒数, $f(N_1, N_2) : N_1, N_2 \rightarrow S$,基于两个节点的属性定义其相似度。

[0129] 在消费者图中,相似度表示两个个体在人口统计学特征及其观看喜好方面的相似性。可按属性计算相似度,然后加权并组合个体的相似度属性,以生成节点对的总相似度分数。

[0130] 当用二进制向量表示两个节点之间的属性时,可采用多种度量方法基于节点属性来定义节点对之间的相似度。这些度量方法中的任何一种均适于和本文技术一起使用。在一些实施例中,为了确保存储效率,二进制向量可用比特串或比特串的阵列表示。

[0131] 当使用相似性度量法(距离函数的倒数: $f(N_i, N_j)$)进行计算时,距离函数为零,

则表示两个节点的类型和行为相同。相反地,距离函数的值越大,表示两个节点越不同。距离函数的一个示例即为欧几里得距离:

$$[0132] \quad f(N_i, N_j) = \|A_i - A_j\|^2$$

[0133] 式中, A_i 和 A_j 是表示节点 N_i 和 N_j 的属性的稀疏向量,而距离等于每个向量的对应分量差值的平方和。

[0134] 二进制向量或比特串的比较可根据多个相似性度量中的一个或多个进行,其中,最常用的是余弦系数。其他常用的度量方法包括但不限于:余弦距离、Dice系数、欧几里得距离、曼哈顿距离、城市街区距离、海明距离和Tversky理论。另一种可用的距离度量是隐含狄利克雷分布(LDA)。另一种定义距离比较算法的方法是深度学习嵌入,其可以学习到最佳的距离形式,而不是将距离固定,如,固定为余弦距离。一种方法示例为流形学习。

[0135] 余弦点积是一种可用于定义消费者图中两节点之间的相似度的优选度量方法。余弦相似度(即 A_i 和 A_j 的点积)的计算公式如下:

$$[0136] \quad f(N_i, N_j) = A_i \cdot A_j$$

[0137] 在这种情况下,归一化每个向量,使其长度为1.0。余弦相似性度量的值为1.0时表明两个节点相同。相反地,余弦相似性度量的值越接近0.0,则两个节点越不同。相似度量可通过从1.0中减去相似度量值转换为类似距离的度量:

$$[0138] \quad f(N_i, N_j) = 1 - A_i \cdot A_j$$

[0139] 一个更加复杂的距离函数示例是参数化核函数,如,径向基函数。

$$[0140] \quad f(N_i, N_j) = \exp(-\|A_i - A_j\|^2 / s^2),$$

[0141] 式中, s 表示参数。

[0142] 在更为一般的情况下,如果比特串是包含除1和0以外数字的向量(如,包含百分比或非标准化数据),则可基于向量值之间的距离度量来计算相似度。其他度量也适用,如,马氏距离。

[0143] 相似度分数 s 通常在0~100之间,也可采用其他归化方案,如,0~1.0、0~10或0~1000。计分系统也可能是非标准化的,仅仅以一个与两个消费者之间的相似度计算值成正比的数字表示。

[0144] 在一些实施例中,在计算相似度分数时,可用表示贡献因子重要性的系数对每个贡献因子加权。例如,一个人的性别权重高于其观看的特定电视节目。这些权重可通过启发式法进行初始设置,并最终通过对广告活动的效果的统计分析得出,其中,广告活动效果是随时间持续更新的。其他用于确定特定属性对相似度分数的贡献度的加权系数的方法包括:回归或特征选择(如,最小绝对值收缩和选择算子“LASSO”)。或者,也适合采用“标注数据”,如,登录数据。在一些实施例中,当系统尝试不同组合或特征时,可利用留存测试数据集推导出可生成较高准确率/查全率的组合或特征(其中的特征未用于构图)。

[0145] 另一种用于推导特征相似度分数的方法是利用下列方法对通过持续比较广告活动与消费者反馈获得的数据进行分析:机器学习、神经网络及其他多层感知机、支持向量机、主成分分析、贝叶斯分类器、Fisher判别法、线性判别法、最大似然估计、最小二乘估计、逻辑回归分析、高斯混合模型、遗传算法、模拟退火算法、决策树、似然的投影算法、k近邻分类算法(KNN)、判别函数分析、基于规则集成的预测性学习、自然语言处理、状态机、规则系统、概率模型、期望最大化算法,以及隐性马尔可夫模型和最大熵马尔可夫模型。这些方法

中的每一种均可用于评估消费者的特定属性的相关性,从而测量广告活动的效果,并为每个属性加权定量。

[0146] 表示

[0147] 为了对整个消费者群体做出评估,需要存储大量节点。或者,需要大量可表示节点类型及行为的属性集合。存储大量节点属性集合存在挑战,这是因为节点的数量可能会高达数亿个。高效存储数据也很重要,这是因为如果将节点存储数据在内存中就可以快速、高效地完成图的计算。

[0148] 在一个优选实施例中,属性可由稀疏向量表示。为了完成属性的向量表示,将用于指定类型的所有可能的节点的联合属性存储在属性库中。然后,用二进制稀疏向量表示每个节点的类型或行为,其中,1和0分别表示属性是否存在。由于指定类型的可能属性的数量庞大,因此,指定消费者的大多数条目均为0。因此,只需存储属性非零的地址,并且每个稀疏向量都可被有效存储,其存储空间通常低于全向量存储空间的百分之一。

[0149] 例如,用属性对指定消费者上个月观看的电视节目编码。系统列举出库中存储的全部可能的不同电视节目,其中,节目数可高达100000个。无论消费者上个月是否观看了节目,每个节点均表示为1,否则为0。

[0150] 如果属性表示不同收入水平,则列举出多个收入水平,而1表示属于特定收入水平的消费者(其他所有条目为0)。

[0151] 因此,年收入在30000~60000美元之间,且上个月观看过“疯狂汽车秀(Top Gear)”的消费者*i*,其计算公式如下:

[0152] (电视_库 = {“行尸走肉”,“权利游戏”,……,“疯狂汽车秀”})

[0153] $TV_i = [0, 0, \dots, 1]$

[0154] TV_i 可仅存储为[4],只有向量的第四个元素为非零。同样地,收入公式为:

[0155] 收入_库 =

[0156] $\{<\$30000, \$30000-\$60000, \$60000-\$100000, >\$100000\}$

[0157] $收入_i = [0, 1, 0, 0]$

[0158] $收入_i$ 可仅存储为[2],因为只有向量的第二个元素为非零。

[0159] 因此,节点*i*的所有属性均可由稀疏向量有效表示。稀疏向量所需的内存比稠密向量低2~3个数量级。

[0160] 构图

[0161] 图5A和图5B是构建消费者图的步骤流程图。

[0162] 首先,图是映射到消费者的设备的集合。多个设备(平板电脑、移动设备和电视等)可通过多个数据资源分组给单个消费者。这通常需要采用聚集技术。为了使单个设备(如,智能电视)归属于多个消费者,可采用细分技术。

[0163] 通过聚集法,可将多个设备分组给单个消费者(或图节点)。其中,使用的数据资源包括但不限于:

[0164] IP地址:多台设备归属同一个IP地址则表示单个消费者或家庭。

[0165] 地理位置:经、纬度接近的多台设备可归属于单个消费者。

[0166] 发行商登录:如果同一消费者通过多台设备登录,则这些设备与该消费者存在关联。

[0167] 在这个过程中,消费者的身份信息通过掩码掩盖,以避免侵犯隐私。最后,获得与特定设备链接的单个消费者的ID。

[0168] 假设 $P(d_i, d_j)$ 是两台设备 d_i 和 d_j 归属同一节点(消费者或家庭)的概率,则根据从不同类别的设备获得的多个数据集,概率可表示为:

[0169] $P(d_i, d_j) =$

[0170] $w_{IP} \times P(d_i, d_j | IP) \times w_{Geo} \times P(d_i, d_j | Geo) \times w_{Login} \times P(d_i, d_j | Login) / Z$

[0171] 式中,“ \times ”表示“乘”; w_{\cdot} 表示加权因子; $P(d_i, d_j | Y)$ 表示条件概率(如果设备 i 和 j 的 Y 值相同,则观看设备 i 和 j 的概率归属于同一用户);而 Z 表示归一化因子。因此, Y 可以是IP地址。

[0172] (条件概率值可能是0.80)。每个数据资源获得一个不同的加权因子,例如,登录数据的权重可高于IP地址。权重可以是固定的,也可以从独立的验证数据集学习。

[0173] 一旦多个设备被分组给单个节点,来自各个设备的类型和行为将被聚集到单个节点的属性。例如,来自移动设备(如,位置事件)和台式设备(近期购买)的属性(以及对应的稀疏向量)将被聚集。这样就可提供更加全面的信息,以便对节点做更准确和更有意义的推论。

[0174] 由于与这些设备相关联的数据是各媒体渠道已知的,因此,可将设备与指定消费者相关联。例如,智能电视可存储位置信息及其播放内容的订阅信息。这种信息既可用于共享,也可从有线电视公司等其他实体处获得。同样地,平板电脑和智能电话等移动设备与智能电视一样可以与相同的WiFi网络(家庭网络)之间建立关联。因此,可将位置相关信息与手机运营商及移动设备订阅内容的播放方等共享。本文图形法的一个重要方面是允许跨越不同设备和媒体平台(通常彼此分离)链接消费者信息。尤其是,本文的图形能够将来自在线和离线购买及浏览资源的消费者数据与电视收看数据相链接。

[0175] 通过细分法,单个设备(如,智能电视)可与多个消费者(或图节点)相关联,例如,这些消费者拥有的设备均与智能电视连接的WiFi网络相同。

[0176] 假设节点 n 被分配了多台设备,则各属性将被聚类到通过公共IP地址连接至多台设备的较小设备组,例如,电视ID。将电视收视数据与来自所有设备的属性一起聚集。可用聚类算法(如, k 均值聚类)将设备分类到较小簇中。聚类簇数 k 一般设为设备数(默认 $k = \text{设备数} \# / 4$)。有时也只以家庭为单位采集聚合数据。例如,一个家庭中可能拥有20台设备。但是,通过行为数据可以确定20台设备只有4个主簇,假设每簇5台设备,则簇数对应于同一家庭中的不同个体。因此,虽然有两类设备(共享和个人),但将行为数据归属于用户仍然很重要。

[0177] 一旦将共享设备归属于多个节点,则从该设备上采集的数据也可被归属于这些节点。例如,智能电视的收看数据可从OEM处采集。通过属性归属可将电视收看数据添加到节点的属性集合中。最后,可将智能电视归属于同一家庭中的不同个人。

[0178] 通过学习距离函数对相似人群建模

[0179] 假设图 $G(N, E)$ 、定义相似性度量的函数形式与种子节点集,则可生成与种子节点相似的“相似”节点集,其中,相似度由固定函数或学习函数定义。这在确定新消费者(对相同或相似内容感兴趣的消费者)是否属于广告主已知的消费者群体方面很有用。在根据消

费者群体的历史数据预测消费者可能的观看行为时,也可使用类似的理论方法。

[0180] 种子节点可以是通过使用固定或学习的相似性度量生成相似节点集,如,家庭或个人。例如,种子节点可被定义为观众群(如,观看某些特定节目的用户列表)。这对于为观众群中每个成员确定可能具有相似观看习惯的其他观众成员列表是有用的,即使他们并没有观看与种子节点完全相同的节目。

[0181] 假定图(及其属性)中的种子节点集,则相似人群建模的输出结果是基于固定或学习相似性度量的、与种子节点相似的节点集(包括种子节点)。

[0182] 可用多个不同向量确定相似人群模型。其中一个向量是全部电视节目的向量。该向量的长度可达40k。另一个向量是观看了特定节目(如,辛普森一家“The Simpsons”)的消费者列表。收看指定电视节目的观众的向量长度可高达10M,这是因为每个消费者包含一个元素。还有一个向量是网站访问向量(如,向量长度达100k)。另外还有一个向量是基于浏览的在线视频的向量(长度也可达到100k)。

[0183] 电视节目对比数据通常需要存取10M用户群。在线数据可用于识别更多的潜在受众,如,150M。应当理解的是,电视数据可通过各种电视消费设备累积,包括但不限于:线性设备、时移设备、传统设备以及程序化设备。

[0184] 2个不同节点之间的相似度根据其属性计算得出,用稀疏向量表示。假设距离函数 $f(N_i, N_j)$ 和种子节点集 N_S ,则可计算出种子节点的每个元素(N_S 中的 n)以及除种子节点外的其他所有节点 n' 之间的成对距离。也就是说,可计算得出所有的量值 $f(n, n')$ 。

[0185] 在计算所有成对相似度之后,只选择 $f(n, n') < T$ 的节点。 T 是最大距离阈值,低于最大距离阈值的节点将被视为相似。或者,降序排列 $f(n, n')$ (其中, n 不是 n'),并选择顶层 t 的节点对。无论在何种情况下, T 和 t 是预设参数(适用于本方法),或者从标注数据或验证数据中学习。符合上述标准的全部节点集 n' 形成了“相似节点”集。

[0186] 图论推理

[0187] 假设图 $G(N, E)$,可基于图中相邻节点的属性推断出节点的相似属性。这在指定的消费者的信息不完善,但却足以进行推断的情况下有用。例如,节点 n 的电视收看属性缺失(通常,无论用户观看了节目,还是不知用户是否观看了节目,都应该有正信息),则可使用图中相邻节点 n' 和 n'' 的属性。节点 n, n' 和 n'' 包含了其他所有属性,如,收入水平和访问过的网站。

[0188] 在另一个示例中,假设 n' 和 n'' 两个节点都观看了“行尸走肉”节目,则可计算与节点 n 相关联的消费者观看“行尸走肉”的概率。如果相似度 w' 和 w'' 分别为0.8和0.9(相似度由 n, n' 和 n'' 之间的边的权重给出),并且节点 n 基于其属性观看节目的似然相似度为0.9,那么概率可由下式给出:

[0189] $P(n \text{ 观看“行尸走肉”})$

[0190] $= [0.8 \times 0.9 + 0.9 \times 0.9] / [0.8 \times 0.9 + 0.9 \times 0.9 + (1 - 0.8 \times 0.9) + (1 - 0.9 \times 0.9)]$

[0191] $= 0.765$

[0192] 在根据消费者群体的历史数据预测消费者可能的观看行为时,也可使用类似的理论方法。

[0193] 准确性

[0194] 随着接收新数据,图像将得到不断完善。在一个实施例中,使用了诸如机器学习等

技术来随时间提升图形的质量。可定期完善图形,如,每周。图论法与本文方法一致的地方在于,图形会经常随新的消费者数据不断更新。这使广告主可在接收到新数据之后随时对活动进行调整。

[0195] 为了确保图的准确性,可将准确率和查全率与验证数据集进行比较。验证数据集通常是一个(子)图,其中,设备和节点之间的关系是已知和确定的。例如,来自在线网络(如,eHarmony)的登录信息表示同一用户从不同台式设备(办公室电脑和笔记本电脑)和移动设备(智能手机和平板电脑)登录到该网站的时间。因此,经常用于登录该网站的所有设备都将被绑定至同一个消费者,从而绑定至个人的图形节点。该信息可用于验证构图是否将这些设备绑定至同一个节点。

[0196] 如果D是验证集中的设备集,则假设Z(D)是根据设备集D构建的由节点集组成的图。数据集和图形构建方法不同获得的Z(D)结果也可能不同。

[0197] 可计算Z(D)集合的真正类(TP)率、假正类(FP)率和假负类(FN)率。TP是Z(D)中的所有节点,也是验证集中的节点。FP是N(D)中的所有节点,但不属于验证集中的节点集。FN是属于验证集,但不属于Z(D)的所有节点。

[0198] 查准率是被正确分组为消费者节点的检索设备的分数值,定义为 $TP/(TP+FP)$ 。

[0199] 查全率是正确分组的消费者节点的分数值,定义为 $TP/(TP+FN)$ 。

[0200] 根据当前的应用程序,查准率与查全率之间会存在不同的交替损益。在构建消费者图时,优选获得高查准率及查全率,以用于比较不同的消费者图。

[0201] 验证数据集不得用于构图,因为这会造成查准率与查全率之间出现偏差。

[0202] 学习相似性度量:

[0203] 图的另一个特征是可随更多数据的引入而调整基础相似性度量。度量通常长时间(如,图的5~10次迭代)固定,并且也不得以与准确性评估相同的频率进行重新评估。

[0204] 在距离函数为非固定值的情况下,可向该函数族学习特定距离函数的参数,或者从该函数族中选择最适合的距离函数。将查准率和查全率与验证集进行比较,以学习距离函数或其参数。

[0205] 假设目标是基于已知高收入人群的种子节点集的属性来预测高收入人群的相似受众群。计算不同距离函数或不同的特定距离函数的种子节点与所有其他节点之间的相似度。距离函数通过使用节点属性(如,在线和电视收视率)来计算相似度。

[0206] 例如,如果距离函数是具有下列参数s的径向基函数:

[0207] $f(N_i, N_j) = \exp(-||A_i - A_j||^2 / s^2)$,

[0208] 则利用相同的距离阈值T计算不同参数值s的种子节点与其他所有节点之间的成对距离,以生成相似人群节点集。对于不同参数值s(需要学习的参数值),该计算可生成不同的相似人群节点集,相似人群节点集以 $N_S(s)$ 表示。

[0209] 对于节点集 $N_S(s)$,可计算真正类(TP)率、假正类(FP)率和假负类(FN)率。TP是 $N_S(s)$ 中的所有节点,也是验证集中的目标集。在本示例中,所有节点也是(标注数据集内的高收入人群)。FP是 $N_S(s)$ 中的所有节点,但不属于目标集(不是高收入人群)。FN是 $N_S(s)$ 中的所有节点,但不属于目标集(不是高收入人群)。FN是属于验证集的所有节点(是高收入人群),但是不属于 $N_S(s)$ 。

[0210] 根据应用程序,查准率与查全率之间可能需要存在不同的交替损益。在向受众定

向投放广告时,需要较高查全率,这是因为曝光(广告)成本低,而如果疏漏了目标受众成员,则付出的成本高。

[0211] 本文的示例旨在从可能的s值中选择查准率和查全率均较高的s值。而其他类型的距离函数,则可尝试采用其他参数来最大化查准率和查全率。

[0212] 相似人群模型的准确性只能用于定义目标受众群。例如,可利用电视收看和在线行为数据集根据高收入人群的种子集来预测相似人群中是否包含有高收入人群。可利用预测的节点集的真实收入水平集来验证预测结果。这将使预测结果具备准确性。然而,一个细分群体的准确预测对于新的目标受众群而言是没有意义的,比如,相同的用户是否也驾驶豪车。

[0213] 非重复到达率的计算

[0214] 消费者图将节点(消费者)连接至节点(消费者)所使用的所有设备上。因此,消费者图中删除了个体的广告总曝光次数中的重复曝光次数。例如,如果用户abc 123已经在其电视、台式设备和移动设备上的每一个观看过指定广告,则总的非重复曝光次数计为1次。这使得下列度量可通过直接测量进行计算。

[0215] 非重复性曝光受众是指属于消费者图中的已删除重复曝光的目标受众人群的用户人数。则非重复直接到达率的计算公式如下:

[0216] $\text{非重复到达率} = \text{非重复曝光受众人数} / \text{总受众人数}$

[0217] 对于样本测量,可将非重复曝光的样本受众人数作为已删除广告重复曝光的目标受众群中的样本用户人数进行计算。则样本到达率的计算公式为:

[0218] $\text{非重复样本到达率} = \text{非重复曝光样本受众人数} / \text{总样本受众人数}$

[0219] 在建模测量数据中,消费者图中的采集数据的用户的ID是未知的。因此,无法一对一地删除到达数据中的重复数据。

[0220] 如果广告主想设置一个消费者频次上限(例如,如果广告主不想同一广告向同一用户或在同一电视上展示两次以上),则计算非重复到达率对电视内容的定向投放是非常有用的。因此,非重复到达率还为优化广告活动效果提供了一种便捷的度量方法,例如,可通过随时间计算非重复到达率来调整广告活动,从而通过改变活动参数提升活动效果,如,消费者人口统计或电视内容的播放时间及频道。

[0221] 此外,随着消费者数据的不断接收和利用,广告主不需要再等待数周或数月才能获知活动成果,从而使广告主可对后续活动的改进做出决定。例如,广告主可在活动仍在进行的同时,通过撤销已向接触过广告多次的消费者投放的内容,或者撤销已向无响应的消费者投放的内容来优化活动。

[0222] 增量到达率的计算

[0223] 在第t天,将非重复到达率(直接或样本到达率)设为x。增量到达率是投放广告活动后的额外非重复到达率。在跨屏环境中,增量到达率是用于计算广告主是否需要通过扩充移动平台将30%的电视到达率扩增至35%的重要参数。但需要注意的是,在直接测量如电视数据时,由于目前智能电视在整个人群中的使用比例相对较少,因此,针对智能电视获得的部分样本数据仅仅是总数据中的一部分。

[0224] 在建模测量数据中,如,样本属性必须经由推断得出的面板数据,消费者图中的采集数据的用户的ID是未知的。因此,无法辨别同一用户是否曾经看过该广告。由于设备无法

与特定用户相关联,因此,不能通过建模数据来计算非重复增量到达率。如上所述,非重复的增量到达率可通过样本测量数据进行计算,因此,本文的方法优于基于面板数据的方法。

[0225] 散列函数

[0226] 图6是用于识别可能成为广告主目标受众的电视受众群的各种数据的示图。如下所述,广告主可基于特定目标市场的预期受众数据选择购买的广告库存。

[0227] 实时电视数据201由有限电视公司或TiVo等OEM供应商储存。优选地,电视数据包括来自电视、OTT设备和VOD(如,HBO Go)等任意组合资源的数据。该数据包括收看数据205,收看数据205包括在特定电视或设备(观看电视内容节目的设备)观看过的节目集以及电视的IP地址或观看电视内容的设备的IP地址。在使用程序化电视或机顶盒的情况下,收看数据205可以与电视ID 203相关联,电视ID 203是特定设备的专属ID,与IP地址不同。IP地址可用于关联同一家庭内的多台设备。它可能包含表示设备品牌或其他属性的信息。电视ID 203和观看数据205将被提供给需求方平台等实体,以供进一步定位广告内容。这些数据可以用任何适合的格式(如,API)提供。

[0228] 电视ID、IP地址和收看数据将被转换成包含与指定电视相关联的收看数据的电视散列ID 207。散列处理对于保护消费者隐私而言非常重要,消费者的主要可识别个人信息将被掩盖,以避免进一步操作,如,检索和直接定位内容。可识别的个人数据包括特定设备(智能电视等)的IP地址。因此,系统一旦从智能电视接收到数据就会对接收到的数据进行混淆编译,使任何人都无法通过IP地址等进行检索。因此,在散列函数的一个实施例中,IP地址中的标识被去除,以散列ID表示,从而可通过散列ID进行查询,而无需通过个人电视进行检索。IP地址数据在传送给DSP之前可由OEM供应商或DSP进行散列处理。优选地,散列函数为本领域技术人员熟知的任何单向散列函数。散列处理过程可由DSP执行,也可由OEM供应商完成。

[0229] 通过在线数据源与DMP等第三方数据合作伙伴交换数据可获得更新的消费者群体行为数据211,如,第三方社交媒体数据、第三方在线购买历史记录、第三方返回链接历史记录、第三方链接历史记录、第一方获得的cookie数据及其他数据(如,包括浏览、社交媒体内容和购买等在内的互联网行为数据)。因此,消费者行为数据包括消费者访问过的网站的信息和消费者的人口统计学数据(如,性别、年龄、位置和以前浏览过的广告)。

[0230] 消费者行为数据通常可从多个其他资源(如,DMP)汇聚而成,或者可由DSP随着时间推移通过学习获得。消费者行为数据可包含或转化成含有消费者散列ID标签221的数据形式,以便移除数据中可识别的个人信息。数据交换可以是连续的或间歇的。

[0231] 与其他第三方数据合作伙伴(如,Experian)进行额外的或可替代的数据交换可更新消费者(如,近期从特定商店购买特定类别商品的每个消费者)的细分数据。该细分数据可包含或转化成含有消费者散列ID标签221的数据形式,以便移除数据中可识别的个人信息。与各数据合作伙伴之间的数据交换可以是连续的或间歇的。

[0232] 图8中示出了数据交换的另一个方面。DSP820可接收可选数据馈送,用于更新第一方信标事件830,如,特定事件(如,消费者点击广告或者最小化其设备802上含有广告的窗口)发生时触发的VAST或VPAID标签804。信标事件中含有消费者散列ID标签221,用于去除数据中可识别的个人信息。数据馈送可以是连续的。因此,第一方数据可包含已知特定消费者ID的信息。在808中,SSP向设备投放广告。在809中,如果DSP决定对展示次数投标,则提交

竞价;如果决定不投标,则做出“不投标”响应。来自设备的Cookie ID(如,来自消费者台式计算机)或设备ID数据806通过SSP传送至DSP,然后在DSP与消费者ID进行匹配。

[0233] 现在,散列用户ID包含与指定电视相关联的用户组,如,家庭中可访问这台电视的用户。

[0234] 然后,使电视ID与用户ID之间的散列ID同步223。散列ID同步可在单个服务器或多个服务器上进行。鉴于会不断接收到消费者的电视收看习惯数据以及消费者本身的数据,因而应经常进行同步或匹配更新。因此,该过程会定期或持续进行。

[0235] 然后,匹配电视收看数据205与散列用户ID,在此期间,电视ID中的标识将被去除。

[0236] 匹配操作可在“沙箱环境”下进行,例如,在隔离的安全服务器上进行,此时,服务器不可由第三方查询或以其他方式访问。删除识别信息(如,IP地址、姓名和联系信息)的操作也可在沙箱环境下进行。

[0237] 该过程的结果是生成一个可实时更新并包含聚集的散列用户ID的数据库,其中,每队散列用户ID与多个行为数据、细分数据及电视收看数据相匹配。DSP可将该整合数据提供给广告主使用。该整合数据可使广告主识别并由此定位由人口统计学数据及收看习惯混合定义的人群子集。例如,广告主可将目标定位为生活在西北太平地区的过去两周内观看过辛普森一家的年龄在18至30周岁的男性。

[0238] 优选地,将经全面整合的散列ID数据存储在各种用户类别下,以供广告主搜索和选择。

[0239] 广告主可利用散列用户ID搜索到观看特定电视节目的用户信息。广告主可选定消费者子集,然后搜索其观看过的电视内容和与其相关的电视节目。广告主通过与程序化电视分离的独立频道在模拟电视上购买时段。

[0240] 此外,广告主可基于特定目标市场的历史收看数据预测需要购买的广告库存。这对预测那些已有观看数据的季播或停播电视节目而言是非常有用且很重要的。广告主可通过识别收看数据中的收看模式(如,特定节目的特定观众倾向于收看的其他特定节目)在其他特定节目播放期间向该节目的观众重新定向投放广告内容。

[0241] 程序化电视竞价

[0242] SSP(如,WideOrbit、Videa和Clypd)将来自多个地方电视台的电视资源汇聚到一个公共平台上。DSP对单个或多个电视广告点位出价。与实时竞价(RTB:采用《实时竞价协议》(2.0-2.4版),请参阅互联网广告局:www.iab.com/guidelines/real-time-bidding-rtb-project/)不同,实时竞价可由本文的技术所用,且适用于数字广告库存,其确认中标与否的响应时间通常只需几分之一秒(通常以毫秒计),但电视竞价的反馈时间则需要一天至数周不等。模拟电视竞价的速度比程序化电视竞价慢,这是因为模拟电视竞价并非实时竞价,不受算法实现的影响。

[0243] 此外,程序化电视竞价的反馈响应模式是接受、保留或拒绝中的一种,而不仅仅只有中标/失标。这就给当前的数字竞价方案带来了一个无法解决的难题。与本文的技术及方法一致的竞价架构及方法如下。

[0244] 在如图10所示的PTV平台中,一个或多个SSP(如,WideOrbit、Videa和Clypd等SSP1和SSP2)拥有连接所有电视资源的程序化界面,并聚集这些电视资源,其中,电视资源来自电视台TV1、TV2.....TV9。需求方平台DSP1~DSP3(如,可执行本文方法的实体)以前只使

用过实时竞价方法 (RTB) 对在线广告库存出价, 并且已经接受电视广告库存竞价过程耗时较长的状况。

[0245] 根据本文的方法, 一条广告库存的竞价决策及相应的投标价基于某些关键绩效指标 (KPI) 的目标绩效确定。在购买PTC广告库存时, 有两个主要的KPI, 即“受众到达率”和“直接反应”。KPI可用于了解相关的优化指标。另一个KPI是总收视率 (GRP), 它是衡量活动是否成功的标准。还有一个KPI与成本有关。由于活动的预算金额是已知的, 因此, 使用本文的方法可计算出每个受众成员接触指定活动所需的成本。广告主可通过这种方式基于定向投放的成效比较各个活动的经济效益, 并且也可使用该KPI来修改或改进广告活动。

[0246] 在受众到达率KPI中, KPI与通过到达率或非重复到达率测量广告在目标受众人群中曝光率有关。

[0247] 相反地, 直接反应KPI与消费者接触广告后的直接行为有关, 如, 网站访问量 (以用户接触广告后访问广告主网站的平均次数表示)、在线购买 (用户在接触广告后在亚马逊等在线商店中的购买行为)、线下购买 (用户在接触广告后在零售店或杂货店等实体店中的购买行为) 以及位置事件 (用户在观看广告后到访特定位置的平均次数)。

[0248] 基于程序化电视数据进行的广告时段竞价取决于广告主设定的目标, 如, 某个最低GRP以及一个或多个优化度量, 其最终变成时段和投标价格的选择。

[0249] 可对未来的电视广告点位 (如, 节目时间表中的时段) 出价, 并且通常可在该点位广告播出前两周提前出价。对于事先已知日期的体育赛事等特殊事件, 竞价开始时间可能甚至更早。如果优选日期的竞价未成功, 为了以防万一, 可在不同的日期对同一个广告点位进行多次竞价。广告点位由几个参数定义, 包括: 节目名称 (如, 辛普森一家)、时段 (节目播出的某一天的某个时间段, 如, 黄金时段和深夜, 这可能需要判断是播放新内容, 还是重播) 以及节目播放的地域 (如, 纽约指定市场区域“DMA”)。

[0250] 在竞价和反馈过程中存在两种不确定性: 竞价是否成功, 如果成功, 怎样展示广告。这又将导致出现两种竞价方法: “探索”和“开发”。这种竞价类型的独特之处在于两种竞价方法之间存在交替损益。

[0251] 竞价结果可能有三种: 中标 (成功: 卖方已接受报价); 失标 (失败: 报价被卖方已拒绝); 以及“保留” (卖方已接受将此次投标作为广告点位/广告库存的时段中的一部分播放或“轮流”播放)。

[0252] 对于广告主而言, “保留”是一个中性结果。例如, 如果广告主对21:00~24:00之间播放的8个广告点位中的2个投标, 而卖方承诺在8个广告点位的2个中播放广告, 但却没有指定在哪个特定点位中播放广告, 这就是“保留”。

[0253] 失标结果中将附带其他反馈信息, 如, 修改和重新提交投标的可能性, 但广告主通常可从“保留”投标结果中获得更多参考信息。

[0254] 这就是程序化电视内容竞价的独特一面。如果竞标价格胜出, 则可能是因为出价过高。相反地, 在实际使用中, “保留”标是指广告点位被锁定并与其他具有相似特征的广告点位合二为一。例如, 如果有10个可用时段, 而广告主可对其中2个时段出价, 那么当整个时段被出清时, 两个被当做“保留”返回的广告点位也将被出清。这使广告主更好地了解应该怎样出价, 并且更加自信自己不会失标。为此, 对于广告主而言, “保留”结果比中标结果提供的信息多 (因而可极大地减少不确定性), 这是因为当反馈结果是“保留”时, 广告主可对

整个广告点位时段的清算价做推断。因此,在程序化电视竞价中,最佳的启发算法是针对“保留”结果,而不是“中标”结果,这也是程序化电视竞价区别于数字RTB的一个因素。

[0255] 在竞价过程中,针对特定广告点位(由节目名称、时段和DMA等定义)的出价是否会成功、失败或保留是不确定的。在指定投标价格(P)下,设参数(用 θ 表示)的竞标结果(中标、失标或保留)的概率分布为 Π :

[0256] $\Pi(\text{结果}=\text{中标/失标/保留}|\theta, P)$

[0257] DSP获得越多针对特定参数的不同投标价格下的竞标结果的样本数据时,结果的不确定性越小。在观察每个新的数据点之后,可通过贝叶斯更新来细化概率分布 Π 。

[0258] 另一种不确定性是通过受众到达率或直接反应KPI衡量指定电视广告点位的绩效。在广告点位上投放广告后,可通过测量到达率(如,GRP)或直接反应(如,网站访问量)来量化广告在某些方面的表现。由于广告点位的绩效是不可重复的(但可随时间发生变化),因此,广告点位绩效的不确定性可用下式表示,例如: $\Pi(KPI=50GRP|\theta)$ 。

[0259] 可用贝叶斯更新来描述这种不确定性,随着观察到的数据点增多,这种不确定将降低。

[0260] 在对电视广告点位出价时,广告主希望设定一个可使其中标或获得保留标资格的投标价格,并且希望以特定的价格对特定广告点位出价,以达到预期绩效(受众到达率或直接反应KPI)。可以通过一个中标/保留概率高的价格来确定可达到预期绩效、不确定性低的广告点位。这适用于有很多可用数据且处于“探索”情况下的广告点位。或者,也可选择样本数据较少或没有样本数据的广告点位及价格;通过找到绩效高且可以以较低投标价格中标或获得保留资格的广告点位来极大地降低未知广告点位的不确定性。这就是“探索”竞价。

[0261] 根据之前的不断增长的已中标(失标)的信息来投标使得指定广告主的整个竞价过程更有效率。图7就示出了这样一种情况。根据大量样本假设广告主在特定投标价格(本示例中为22美元)及其一组其他变量的特定值下可获得特定的GPR。其中,投标价格表示多因素空间位置,用图7中的黑框表示。在该示例中,为了表示方便,仅用了3个变量(如正交轴上所示的位置、时段和节目名称),但在实际中可用3个以上变量。灰框表示未成功的投标的信息(反馈值或结果概率)。如果广告主的报价背离最优报价,则获得的结果相近,且广告的绩效类型也相似。该信息可使竞价结果中的不确定性降低,但需基于迭代学习过程。网格中的框的颜色越黑,结果越确定。广告主可基于参数相似的广告库存(如,时段或相同的地域位置)利用其他框的数值来对相似时段的广告库存的值进行推断。

[0262] 竞价的目的是旨在以指定的价格最大化绩效指标(受众到达率或直接反应KPI)或者以最低成本实现绩效水平。由于绩效是不确定的,因此,另一个没有可用数据的广告点位可能已经以较低成本取得了不俗的成绩。因此,从长远来看,对于竞标者而言,通过探索不确定性高的广告点位来采集数据是有价值的。该原理可通过探索收益参数U实现。目标函数的一般形式由公式(1)给出:

[0263] [广告点位]值=[广告点位]预期值+U×[广告点位]不确定性降低值(1)

[0264] 广告点位的预期值通过预期绩效与不确定值的积分获得。因此,预期值的示例性定义由公式(2)给出:

[0265] [广告点位]预期值= $\sum_{\theta} P(KPI=x|\theta)\sigma(KPI|\theta)$ (2)

[0266] 式中, $\sigma(KPI|\theta)$ 是在给定 θ 的情况下,KPI等于x的值的的不确定性。和用所有 θ 值代

替。

[0267] 广告点位的不确定性的降低值可根据本领域中的公式通过熵减少、信息值或信息增益等准则得出。

[0268] 不确定性的变化的示例性定义可由公式 (3) 给出：

$$[0269] \quad \Delta \text{不确定性}[\theta] = \sigma(\text{KPI} | \theta) - \sum_x \sigma(\text{KPI} = x | \theta) \quad (3)$$

[0270] 式中,和用KPI的全部离散值代替。

[0271] 竞价方法

[0272] 程序化电视时段的示例性竞价方法可从以下几点进行说明。首先,广告主设定目标预算B。然后,将所有可用的广告点位集设为S。

[0273] i. 列出S中的全部广告点位集s。

[0274] ii. 用公式 (1) 计算集s的值。

[0275] iii. 利用先验分布数据为每个s分配一个投标价格P(s)。

[0276] P(结果=中标/失标/保留 | s, P(s))。

[0277] 先验分布数据可通过SQAD(纽约州柏油村,网址:sqad.com)等企业提供的刊例价格数据进行估算。某些数据供应商专门提供电视时段购买信息。这些供应商基于与电视库存广告点位的关系提供广告时段的价格范围。

[0278] iv. 列举所有组合的s(非重复),并用 θ 表示每种组合。

[0279] v. 通过对 θ 中的所有值s求和来计算出一个分数。

$$[0280] \quad v^i. \text{分数}[\theta] = \sum_s P(s) P(\text{结果} = \text{中标/保留} | s, P(s)) P(\text{KPI} = x | s)$$

[0281] vii. 选择具有最高分数的 θ ,使得预期目标预算 $\sum_s P(s) P(\text{结果} = \text{中标/保留} | s, P(s)) \leq B$ 。

[0282] 准备投标的广告点位是组合 θ 中可使分数最大化并且预期目标预算低于B的广告点位。

[0283] 预期目标预算通过考虑投标价格P(s)可能带来的结果(中标或保留)的概率来计算。基本上,如果中标或保留的平均概率为0.10,则预计广告主可通过预算B对总预算为10B的广告点位进行投标。

[0284] PTV竞价与数字实时竞价(RTB)不同,PTV竞价可至少通过两种方式用于在线及其他数字应用程序(如,由其他DSP处理)。

[0285] 在RTB中,由供应商针对即刻可用的广告展示次数提出竞价请求。竞标者基于竞价请求参数报一个最大投标价格。其接收投标的时间窗口通常很短(小于50ms)。平台在接收所有投标后将开始第二次竞拍。中标者会收到中标通知,所有其他竞标者将收到未中标通知。广告主的不确定性大于其以指定价格中标的可能性。在一些实施例中,可使用机器学习系统构建中标率与价格之间的概率分布,从而学习最优投标价格。

[0286] 在程序化电视竞价中,可提前14天接收未来广告点位的投标。除中标/失标结果外,竞标者还会接收到中标、失标或保留中的某一个结果。失标信号中还会提供其他反馈信息,使竞标者可根据反馈修改和重新提交报价。机器学习系统将构建中标/失标的概率分布、不同投标价格的失败率以及广告点位的定义参数。需要学习的概率分布比RTB更复杂(且具有更多维度)。

[0287] 计算实现

[0288] 计算机用于处理以比特串等形式存在的广告活动数据、广告库存、消费者图和设备图的功能可由本领域的程序人员或程序员团队进行开发。这些功能可以用多种编程语言实现,在某些情况下,可通过混合编程实现。例如,函数及脚本函数可用函数式编程语言编程,如:Scala、Golang及R。部分函数可用其他编程语言实现,如,Prolog、Pascal、C、C++、Java、Python、VisualBasic和Perl,但C#等.Net语言及其他同等语言不在此列。技术能力不受限于或取决于用于实现或控制基本功能访问的底层编程语言。或者,此项功能可从更高级别的功能开始实现,如,依赖于之前开发的数学表达式(如,比特串和稀疏向量)处理功能的工具包。

[0289] 本文中的技术可开发用于与目前使用的任何公知的计算机操作系统以及本说明书中未列出的其他系统一起运行。这些操作系统包括但不限于:Windows(包括可从微软公司获得的Windows XP、Windows95、Windows2000、Windows Vista、Windows7、Windows 8、Windows Mobile、Windows 10及其中级版本更新版本等变体);Apple iOS(包括iOS3、iOS4、iOS5、iOS6、iOS7、iOS8、iOS9及其中级版本更新版本等变体);OS9、OS 10.x等Apple Mac操作系统(包括被称为“Leopard”、“Snow Leopard”、“Mountain Lion”和“Lion”的变体);UNIX操作系统(如,Berkeley标准版);以及Linux操作系统(例如,可从许多分销商或“开源”软件免费获得)。

[0290] 对于依赖于已经实现的其他软件组件(处理稀疏向量和计算相似性度量向量的功能)的给定实现而言,假设这些功能可由本领域程序技术人员提供。

[0291] 此外,应当理解的是,指示适当编程的计算机执行本说明书中所述的方法的可执行指令可用任何适当的计算机可读格式存储及发送。这可包括但不限于:大容量“硬盘驱动器”等便携式可读驱动器、连接至计算机USB端口的笔式驱动器、计算机内部驱动器、CD-ROM或光盘。还应进一步理解的是,虽然可将可执行指令存储在便携式计算机可读介质上并以有形形式发送给购买者或用户,但是,可执行指令还可通过如互联网等远程下载至用户计算机,其中,互联网连接可部分地依赖于无线技术,如,WiFi。本技术的这一方面并非意味着可执行指令是以信号或其他非有形实施例的形式实现。可执行指令也可以作为“虚拟机”实现的一部分来执行。

[0292] 本文的计算并不局限于特定版本或类型的网页浏览器。可设想该技术可通过以下一种或多种方式来实现:Safari、Internet Explorer、Edge、FireFox、Chrome、Opera及其任何版本。

[0293] 计算设备

[0294] 适合用于实施本说明书中所述方法的示例性通用计算设备900的示意图,如图9中所示。

[0295] 计算机系统900,包括:至少一个数据处理单元(CPU)922;存储器938,通常包括高速随机存取存储器和非易失性存储器(如,一个或多个磁盘驱动器);用户界面924;一个或多个磁盘934;以及至少一个用于通过网络(包括互联网)与其他计算机和其他设备(如通过高速网络连接或无线连接)通信的网络或其他通信界面接口936。可选地,计算机与互联网之间可设置防火墙952。至少在CPU 922、存储器938、用户界面924、磁盘934和网络接口936之间通过至少一根通信总线933相互通信。

[0296] 可选地,CPU 922可包括用于优化大向量数据处理的向量处理器。

[0297] 存储器938用于存储程序和数据,通常包括以下程序和数据中的一些或全部:提供基本系统服务的操作系统940;一个或多个应用程序,如,解析程序950和编译程序(图9中未示出);文件系统942;一个或多个用于存储广告库存946、活动描述948及其他信息的数据库944;以及可选地,执行高阶数学运算所需的浮点运算处理器。本发明的方法还可调用包含在一个或多个动态链接库(图9中未示出)中、但存储于存储器938内或磁盘934上的函数。

[0298] 可选地,在数据库中的数据太大而无法在存储器938上有效存储时,可将图9中所示存储在存储器938内的数据库及其他程序存储于磁盘934上。数据库也可替代地或部分地存储在通过网络接口936与计算机系统900通信的一个或多个远程计算机上。

[0299] 存储器938包括用于接收一个或多个广告主的输入以及计算消费者之间的相似度分数的编码指令。指令进一步包括用于执行一个或多个解析程序、计算度量以及执行统计分析的程序指令。在一些实施例中,稀疏向量本身并不是在计算机900上计算的,而是在不同计算机上计算的,如,通过网络接口936传递至计算机900。

[0300] 可设想的是本文技术的实施方式有多种,尤其是在计算机设备上进行不同复杂度的计算时,其包括但不限于:工作站、个人计算机、笔记本电脑、平板电脑及其他移动计算设备(包括手机、可穿戴设备及个人数码助手等)。该计算设备可适当配置用于运行执行本文方法的软件的处理器,包括但不限于:图形处理器、向量处理器和数学协处理器。此外,某些计算功能通常分布于多台计算机上,例如,一台计算机用于接收输入和指令,而另一台或其他计算机用于通过互联网连接接收指令并执行远程处理,或者可选地,将结果或输出送回第一台计算机。

[0301] 计算装置可通过用户界面924进行控制,其中,用户界面924可包括显示器、鼠标926、键盘930和/或图9中未示出的其他组件(如,触控板、控制球、触屏、手写笔、语音识别及手势识别技术,或者基于用户眼球运动的其他输入或其输入次组合或组合)。此外,这些实施方式的配置还允许广告库存的购买者通过网络连接远程访问计算机900,并通过一个属性与界面924相当的界面浏览广告库存。

[0302] 在一个实施例中,计算设备可配置成通过二维码扫描、手势识别以及输入生物特征数据或密码来限制用户的访问。

[0303] 将该技术还原至实施例中时,可像一个或多个软件模块、函数或子程序处理存储的广告库存及消费者数据的数据库一样批处理运行该技术,或者与输入单个广告活动的特定指令的用户交互。

[0304] 本文中的技术匹配广告库存与广告活动标准的结果可以以有形形式显示,如,在一个或多个计算机显示设备(如,监视器、笔记本电脑显示器、平板电脑显示器、上网本显示器或手机显示屏)上显示。匹配结果还可被进一步打印成纸质文件,存储为计算机可读介质的电子文件,可在计算机之间传输或共享的电子文件,或者投影到放映厅屏幕上(如,展示期间)。

[0305] 工具包:本文的技术可通过允许用户(如,广告库存购买者)访问和控制那些可提供广告活动的关键管理元件的基本函数来实现。在计算机实现中,内置某些默认设置,但同时也为用户提供足够多的用于选择广告库存的分配的功能,使用户可根据自身考量移除某些功能或调整其权重(若适用)。

[0306] 工具包可以通过脚本工具、图形用户界面(通过触屏选择)和/或下拉菜单实现,以

适应用户的技术水平。用户访问基础工具的方式并非旨在以任何方式对本技术的创新性或实用性进行限制。

[0307] 因此,本文的方法可在一台或多台具有处理器(配置成执行该方法)的计算设备上或者跨一台或多台具有处理器的计算设备来实现,并且可编码为计算机可读介质内的可执行指令。

[0308] 例如,本文的技术包括具有指令编码的计算机可读介质,其中,指令编码用于执行向电视受众群定向投放广告内容的方法。这些指令包括:接收第一消费者群体的程序化电视数据;去除每台电视的IP地址标识,为每台电视创建与第一消费者群体中的每个消费者相关联的散列ID;将每台电视的散列ID与消费者数据图中的消费者标识符相关联;根据第二消费者群体的在线行为和人口统计学数据构建消费者数据图;基于第二消费者群体中的消费者的一个或多个人口统计因素以及与该消费者相关联的电视收视数据列表,识别电视广告库存的一个或多个时段;购买与第二消费者群体相关的电视广告库存时段;以及与媒体渠道通信,指示其将广告内容投放给库存的一个或多个时段中的第二消费者群体。

[0309] 因此,本文的技术进一步包括具有至少一个处理器的计算设备,其中,处理器配置用于执行向电视受众群定向投放广告内容的方法。这些指令包括:接收第一消费者群体的程序化电视数据;去除每台电视的IP地址标识,为每台电视创建一个与第一消费者群体中的每个消费者相关联的散列ID;将每台电视的散列ID与消费者数据图中的消费者标识符相关联;根据第二消费者群体的在线行为和人口统计学数据构建消费者数据图;基于第二消费者群体中的消费者的一个或多个人口统计学因素以及与该消费者相关联的电视收视数据列表,识别电视广告库存的一个或多个时段;购买与第二消费者群体相关的电视广告库存时段;以及与媒体渠道通信,指示其将广告内容投放给库存的一个或多个时段中的第二消费者群体。

[0310] 此外,本文的技术包括:计算设备,配置用于执行程序化电视内容的竞价指令;以及计算机可读介质,具有执行程序化电视竞价的编码指令。

[0311] 云计算

[0312] 本文的方法可通过在“云”中运行实现。因此,由一个或多个计算机处理器执行的用于基于计算机执行本文方法的过程不需要由单个计算机或计算设备来执行。过程和计算可分布于一个或多个数据中心中物理位置彼此不同的多个处理器。数据通过互联网等网络连接与各处理器交换。优选地,可使用加密等安全协议来使消费者数据被盗用的可能性达到最小化。远离实体(如,DSP)进行的一个或多个远程计算包括消费者图和设备图及其更新。

[0313] 示例

[0314] 示例1:电视广告细分市场的自动索引标记的散列处理

[0315] 已利用本说明书中其他地方所述的设备和过程实现了一个实施例。

[0316] 向广告库存购买者提供了一个在线用户界面。购买者可通过该用户界面从提供的目标细分市场列表中选择,如,年龄在17至28周岁想要购买汽车的女性。细分市场的选择范围可宽泛或局限,使库存购买者可从多个可能的组合中进行选择。

[0317] 目标细分受众是各种在线消费行为的聚合视图。目标细分市场通过在线数据源进行分类,如,第三方社交媒体数据、第三方在线购买历史记录、第三方返回链接历史记录、第

三方链接历史记录、第一方获得的cookie数据及其他数据(如,包括浏览、社交媒体内容和购买等在内的互联网行为数据)。终端用户平台从多个第三方资源(包括多个独立收集和整合数据的DMP合作伙伴)获取数据。该平台提供服务器端到服务器端的同步和映射服务,并将数据(包括多个第三方数据资源和第一方采集到的信标数据)映射到平台服务器中的现有数据中。数据整合后将被映射到电视/OTT/VOD收看数据中,这些数据中可能包含IP识别地址。在沙箱环境中,电视/OTT/VOD收看数据将被映射到各种散列ID数据中,并立即通过IP地址自动删除过程删除数据中的标记。

[0318] 在一个示例性应用中,平台用户通过下列标准来选择细分市场:年龄在18至34周岁之间对宝马汽车感兴趣、吃麦圈(Cheerios)的男性;年龄在18至34周岁之间且吃麦圈的人;年龄在18至34周岁之间且不吃麦圈的人或任何合理组合。系统进行搜索,并返回与多个第三方和第一方消费者行为数据相关联的消费者散列ID列表,从而保护消费者识别信息(如,IP地址、姓名或联系信息),避免向终端用户暴露。

[0319] 用户提交细分市场选择之后将进入后台流程。将选定的细分市场与满足用户选择标准的实际在线用户的对应散列ID进行匹配。散列ID代表平台收集和标记的IP地址、移动设备、社交媒体账号以及其他唯一识别的媒体媒介。

[0320] 终端用户系统包含细分市场类别库,该类别库随着不断实时整合第三方数据形成的细分市场新信息而不断更新。实时整合数据可使终端用户准确选择细分市场,并衡量和预测未来的消费者行为。在整合新的第三方数据时,终端用户系统为消费者个人提供一个唯一的匿名终端用户ID,例如,首次添加入终端用户库的媒体消费者将获得一个唯一ID:#WHFKCNJ164829DHAHH17。

[0321] 在并行过程中,平台继续跨设备聚合用户的媒体消费行为,并为该行为编译一个唯一的消费者ID。例如,平台从尼尔森和Exelate等社交媒体供应商收集第三方数据,并将其与从其他第三方平台(如,VOD电视和电子商务网站等)获得的识别信息进行匹配。然后,平台进行归化运算,汇聚唯一消费者ID下的数据,将在多个第三方平台上有在线行为的单个用户ID标记为该唯一ID。

[0322] 平台使用唯一消费者ID继续归化输入数据。现在,可将这些被标记有唯一消费者ID的聚合数据提供给终端用户广告库存系统使用。

[0323] 终端用户广告库存系统通过消费者ID收集新的消费者数据,并不断将消费者ID与内部散列ID进行匹配。终端用户广告库存购买者和平台供应商在选择目标市场时将被屏蔽,无法使用这些识别信息。

[0324] 平台供应商可通过唯一登录、用户管理(user-dashboards)、用户控制和隐私设置等为终端用户提供各种账户。

[0325] 本文中引用的所有参考文献均引入本文作为参考。

[0326] 以上描述旨在说明本技术的各个方面,并非旨在通过本文中所述的示例限制所附权利要求的范围。现已对本发明做了充分描述,但对本领域技术人员而言显而易见的是,可在不脱离所附权利要求的精神或范围的前提下对其做多种修改和变更。

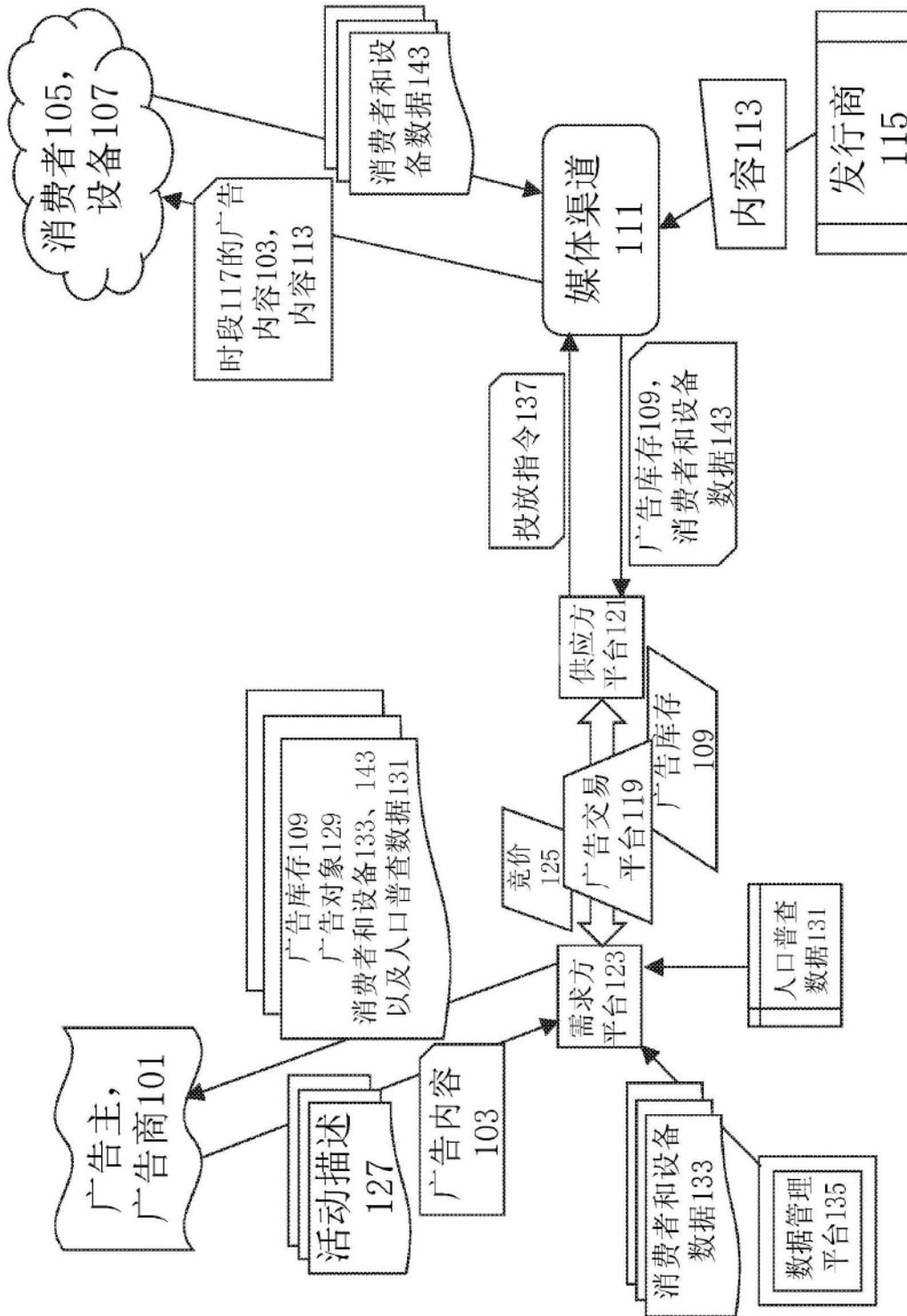


图1

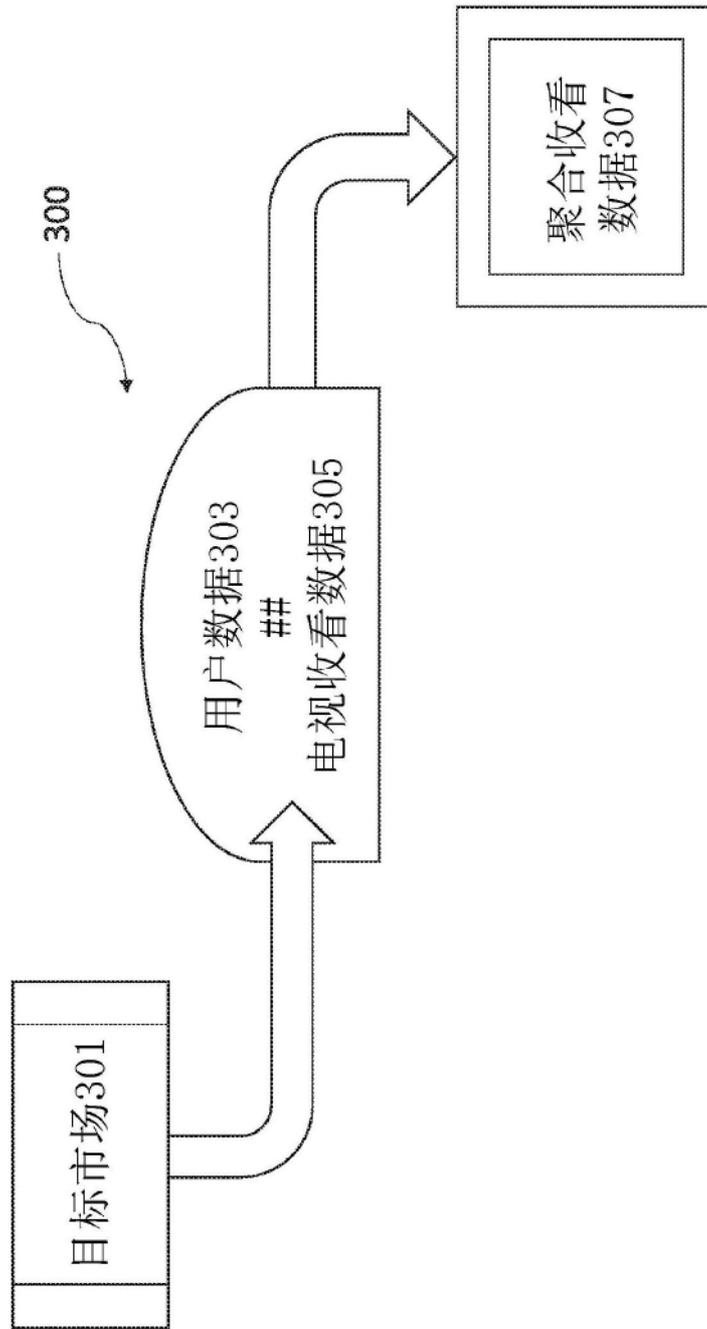


图2

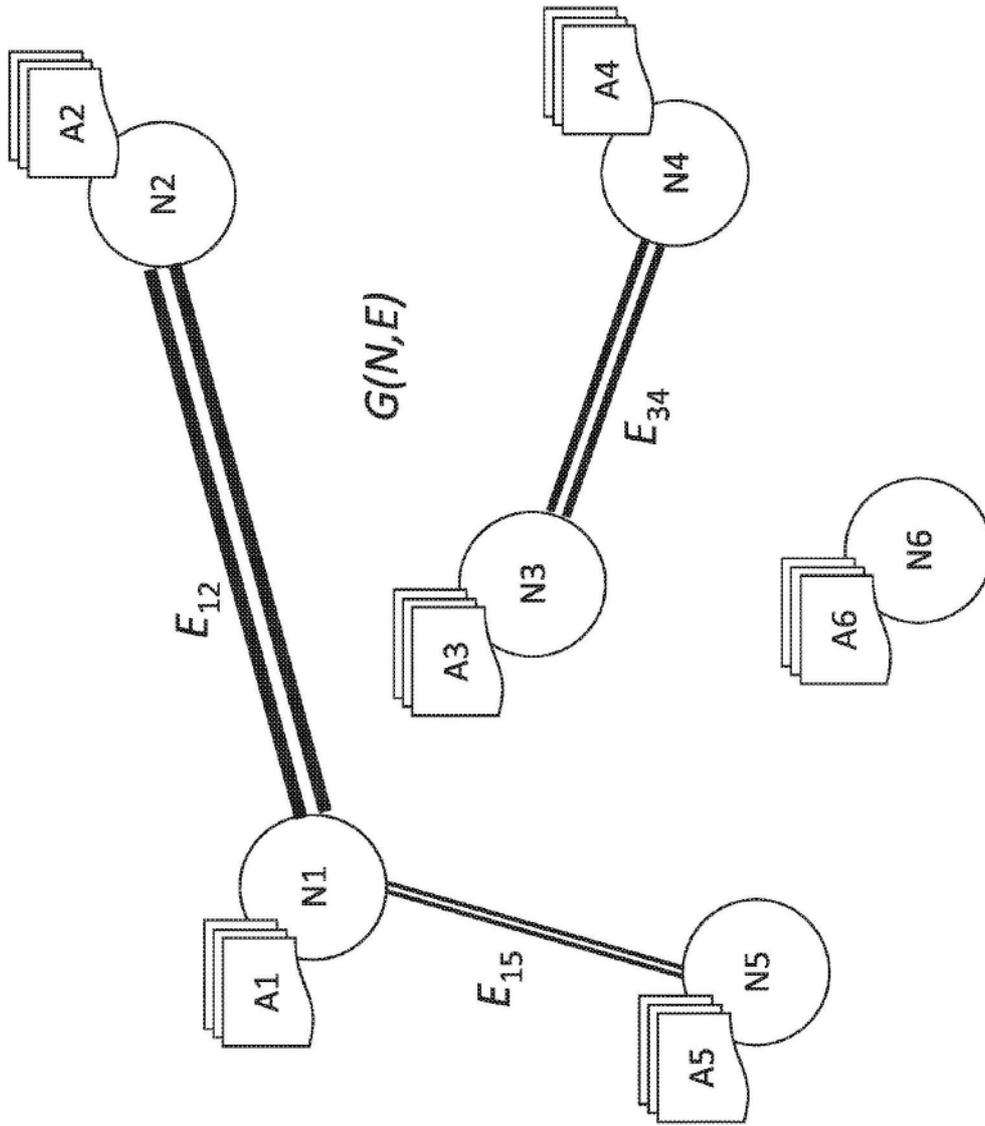


图3

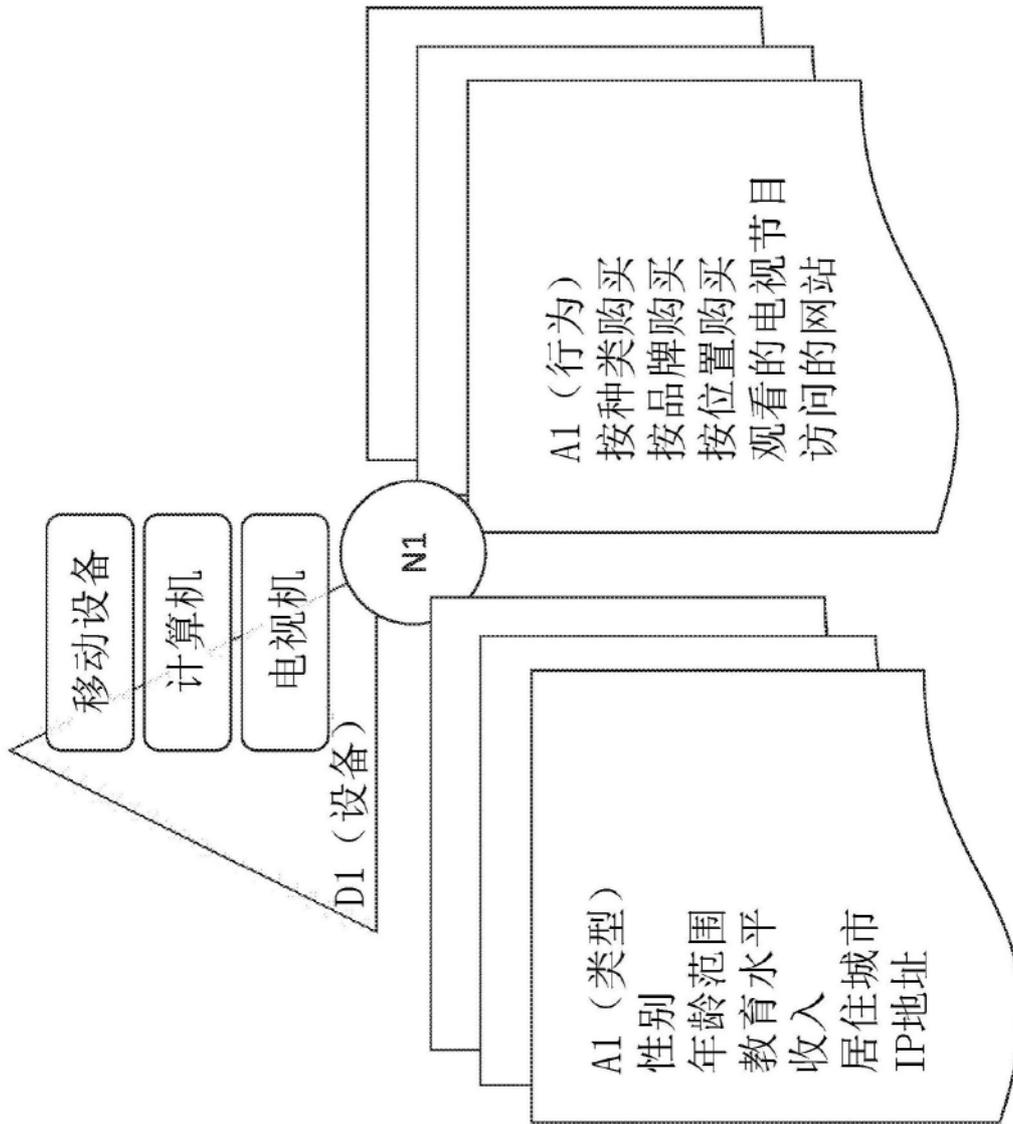


图4

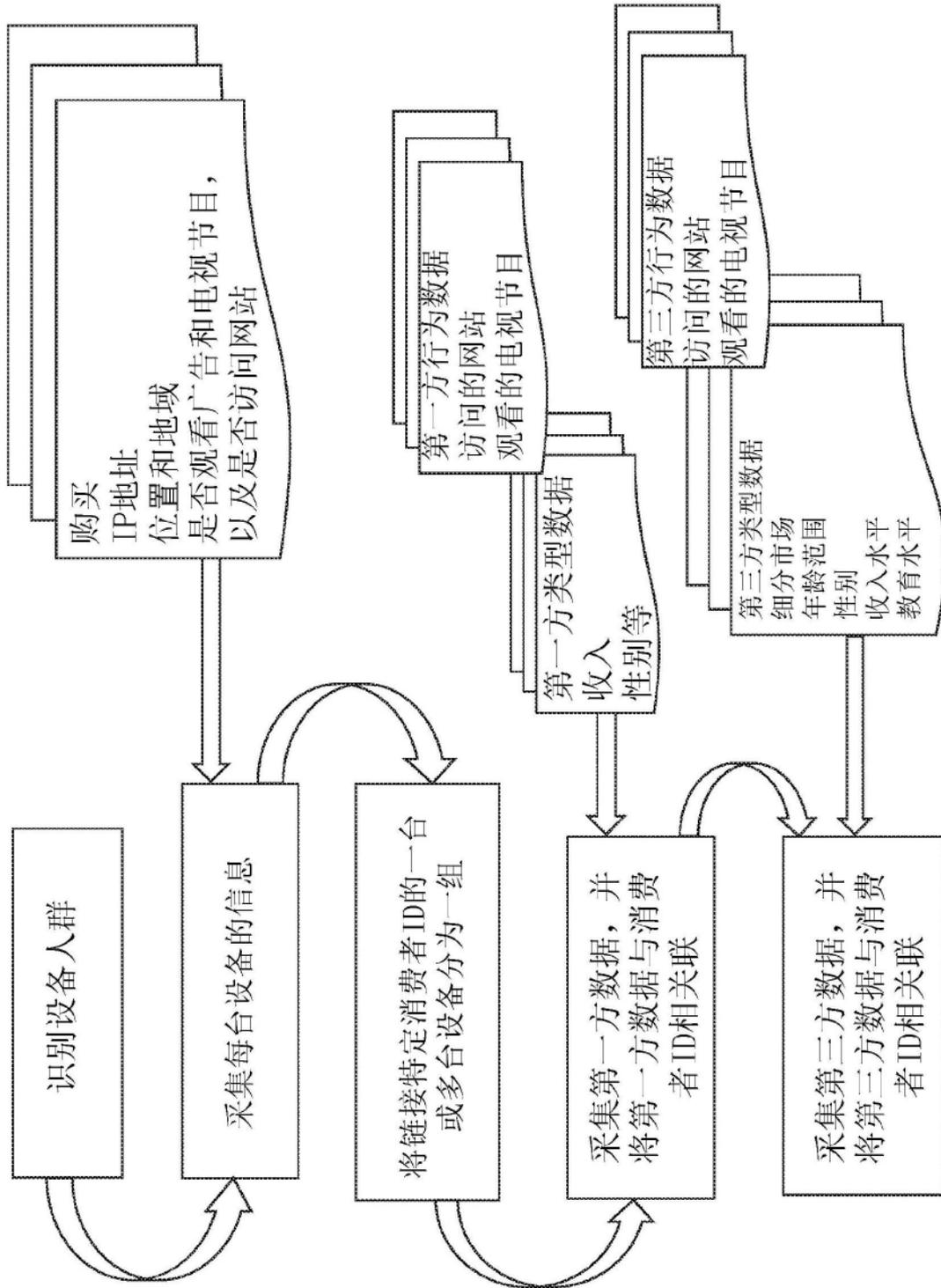


图5A

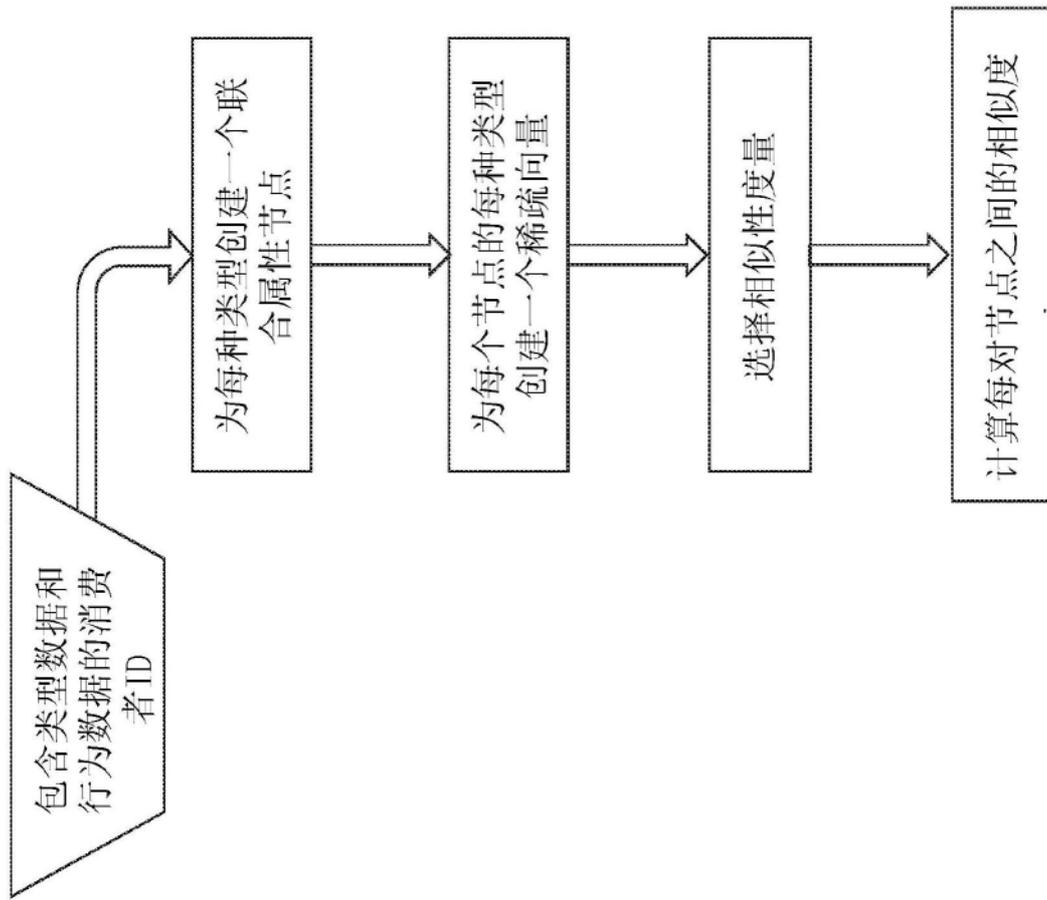


图5B

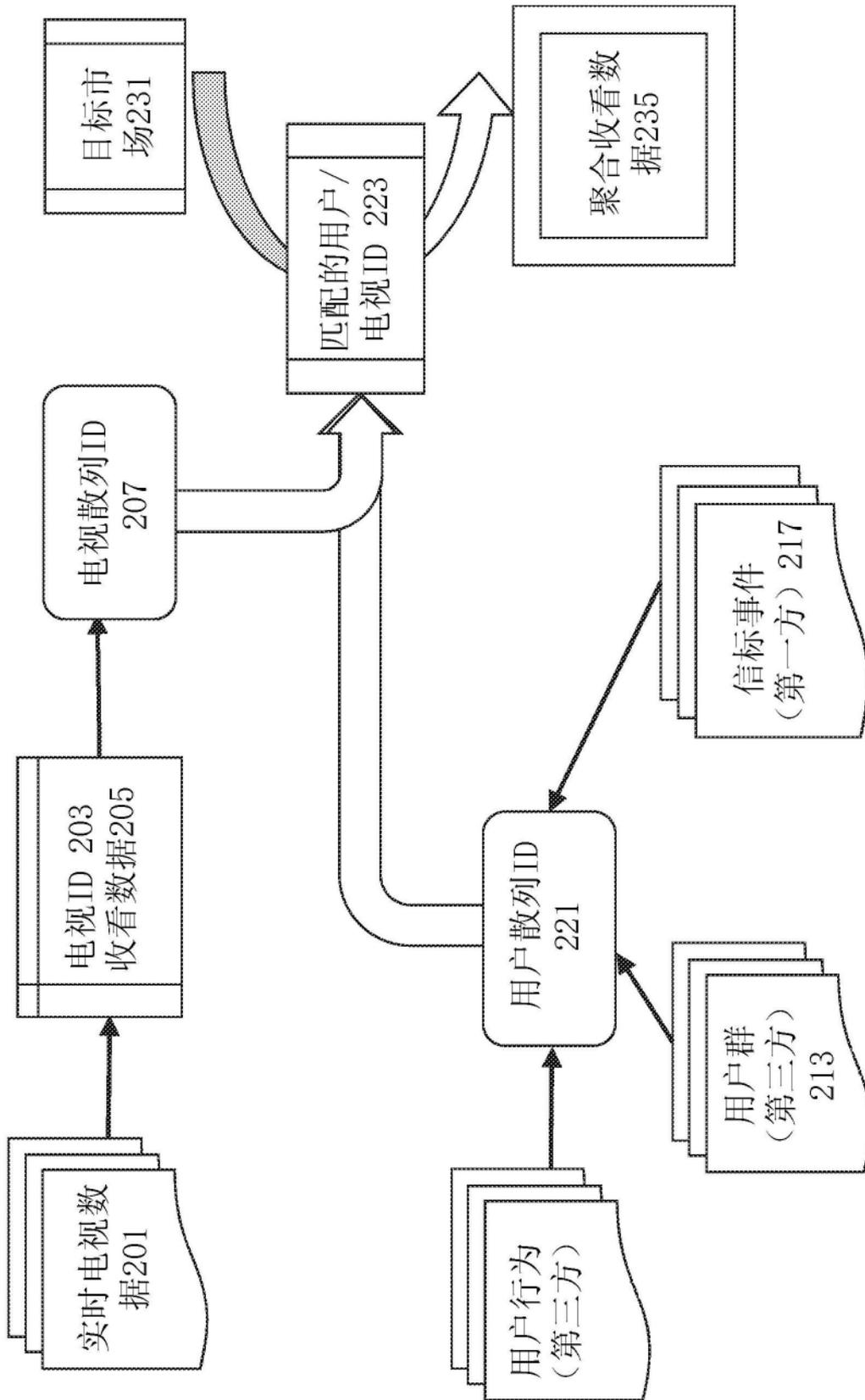


图6

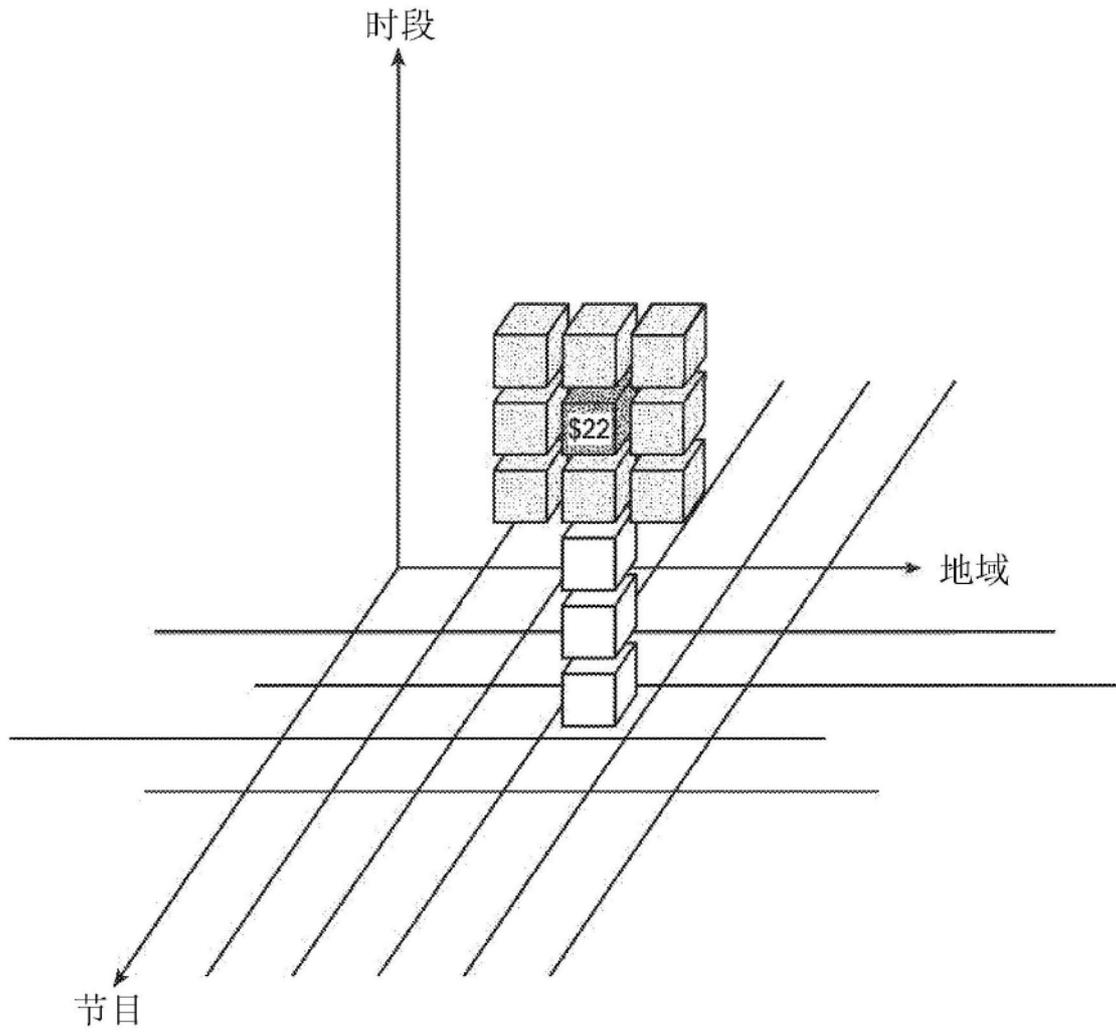


图7

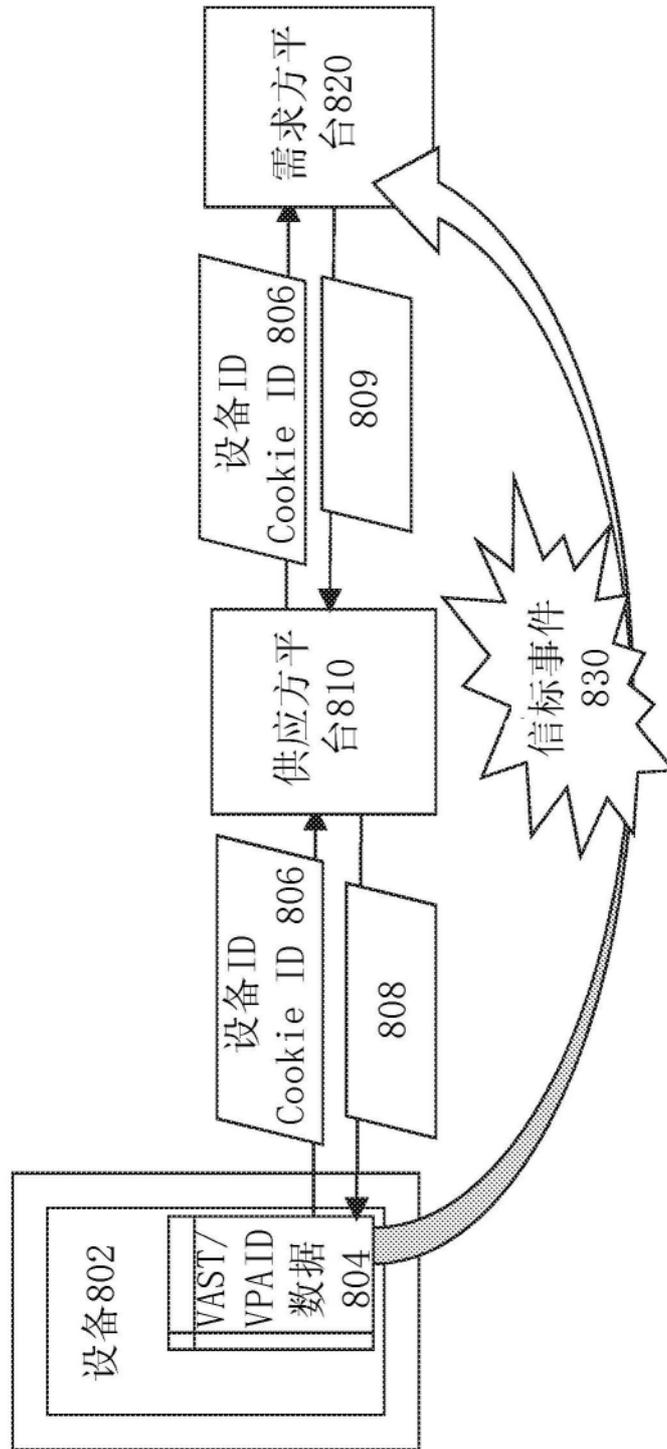


图8

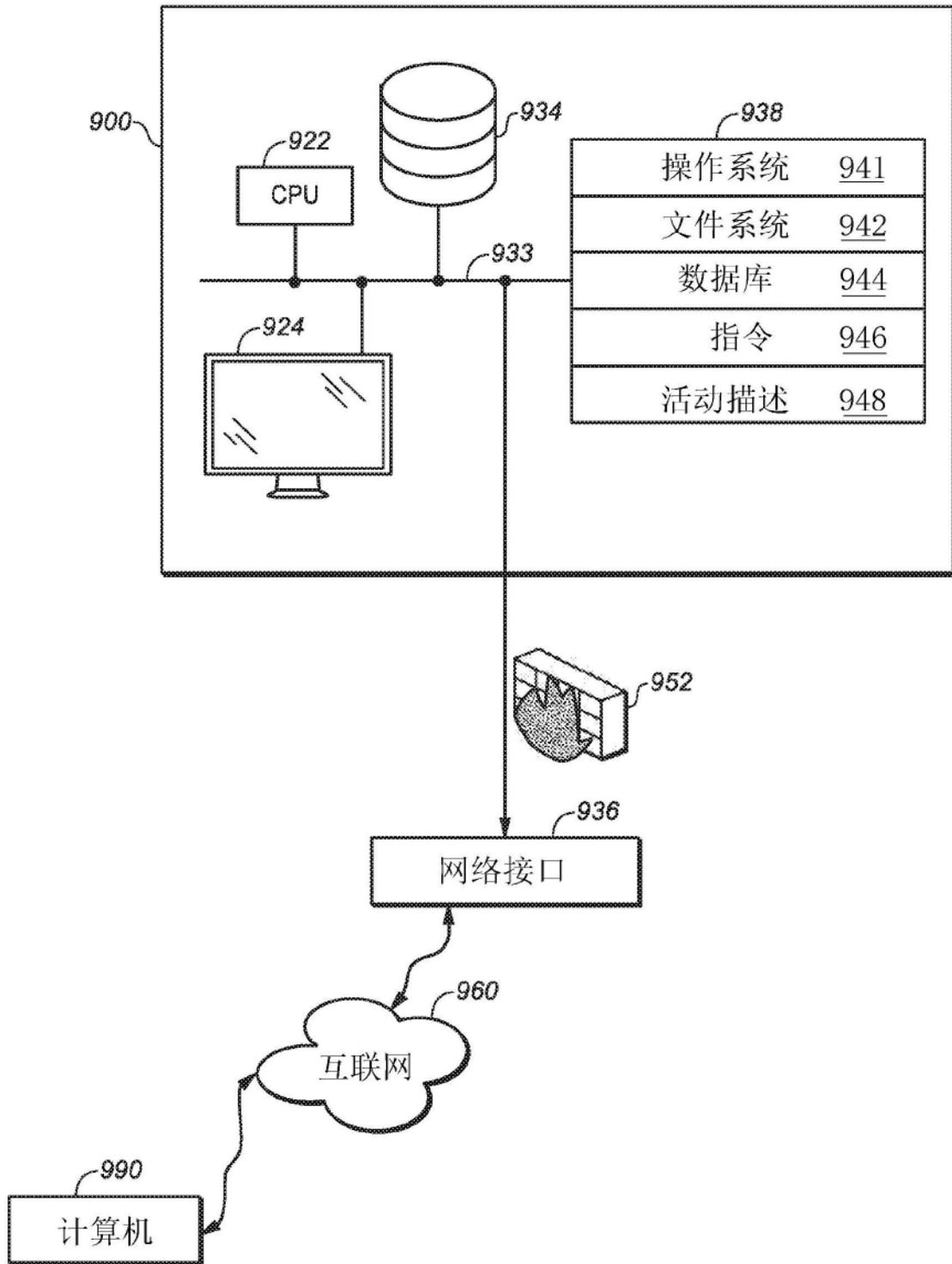


图9

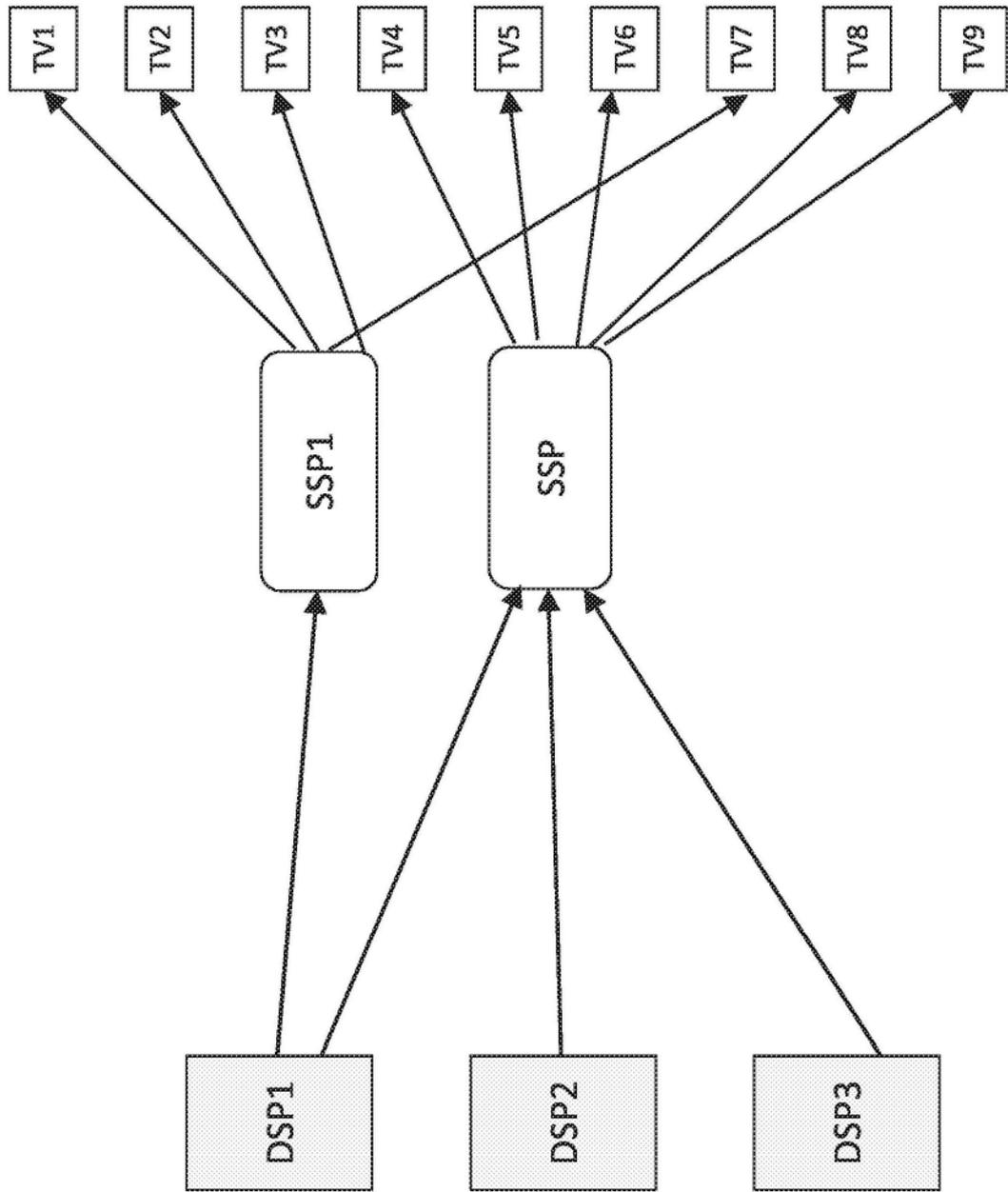


图10