



(12) 发明专利

(10) 授权公告号 CN 103002329 B

(45) 授权公告日 2015. 06. 17

(21) 申请号 201210510024. 2

CN 102223358 A, 2011. 10. 19, 全文.

(22) 申请日 2012. 12. 03

CN 102708495 A, 2012. 10. 03, 全文.

CN 102592235 A, 2012. 07. 18, 全文.

(73) 专利权人 合一网络技术(北京)有限公司

地址 100080 北京市海淀区海淀大街 8 号中
钢国际广场 A 座 5 层 A、C 区

审查员 芦伟

(72) 发明人 白雪 付鹏 卢学裕 姚健
潘柏宇 卢述奇

(51) Int. Cl.

H04N 21/258(2011. 01)

H04N 21/2668(2011. 01)

(56) 对比文件

CN 1889679 A, 2007. 01. 03, 全文.

CN 101673385 A, 2010. 03. 17, 全文.

CN 102665118 A, 2012. 09. 12, 全文.

US 2010/0299209 A1, 2010. 11. 25, 全文.

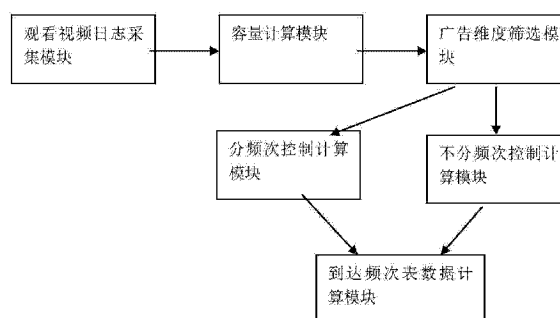
权利要求书2页 说明书10页 附图1页

(54) 发明名称

一种预测网络视频广告的到达频次表数据的
装置及方法

(57) 摘要

本发明涉及一种预测网络视频广告的到达频次表数据的装置及方法,该装置包括:用于记录并产生观看视频日志的视频日志采集模块、用于计算满足投放维度的在历史日期范围内每天的容量计算模块、用于根据随机数与概率的比较结果来排除日志的广告维度筛选模块、用于计算访问 1 次以上到 n 次以上的用户数的不分频次控制计算模块、用于统计到达频次表数据的到达频次表数据计算模块。该方法包括:计算满足投放维度的在历史日期范围内每天的容量,对日志进行筛选,统计出每天每个 cookie 的访问次数;统计每天的累积访问次数;统计访问过不同次数的用户数;计算从访问过 1 次以上到 n 次以上的用户数。



1. 一种预测网络视频广告的到达频次表数据的装置,其特征不在于该装置包括:
观看视频日志采集模块,用于记录用户观看视频的状态为一个观看视频日志;
容量计算模块,选择邻接预测日期范围之前的历史日期范围,在不进行频次维度控制时,计算满足投放维度的在历史日期范围内每天的容量;

广告维度筛选模块,根据预期的投放量与计算出的容量计算概率 C,用随机类对每天中每个日志产生一个随机数,如果该随机数小于概率 C 则保留该日志,否则排除;

不分频次控制计算模块,根据以上保留的日志数据统计出每天每个 cookie 的访问次数,用日期和 cookie 作为 key,统计访问次数;用日期和 cookie 作为 key,统计每天的累积访问次数;用日期和访问次数作为 key,统计访问过不同次数的用户数;计算从访问过 1 次以上的用户数到 n 次以上的用户数;

到达频次表数据计算模块,根据上述计算结果统计预测日期内每天的 1 次以上的用户数到 n 次以上的用户数;

其中,该装置还包括分频次控制计算模块,在进行频次维度控制时,将符合投放维度条件的日志记录筛选出来;计算概率 $E = \text{频次维度数值} / \text{用户实际观看次数}$,用生成随机数取概率的方式,如果随机数小于此概率 E,则保留观看日志,同时记录总保留日志不超过频次控制次数;计算每天的容量;每个日期点都要进行随机数取概率,计算概率 $C = \text{预设的每天投放量} A / \text{实际的每天容量} B$,小于此概率 C 的日期点日志保留,大于概率 C 的日期点日志丢弃。

2. 根据权利要求 1 所述的装置,其特征不在于:

所述对每天的容量进行计算是按以下方式进行计算:对每天的日志进行分析,符合投放维度的即记为 1,不断累加,累加值为容量。

3. 根据权利要求 1 所述的装置,其特征不在于:

所述投放维度包括地域维度、频道维度、版权维度、剧目维度、小时维度。

4. 一种预测网络视频广告的到达频次表数据的方法,其特征不在于该方法包括:

步骤 (1) 选择邻接预测日期范围之前的历史日期范围,在不进行频次维度控制时,计算满足投放维度的在历史日期范围内每天的容量;

步骤 (2) 根据预期的投放量与计算出的容量计算概率 C,用随机类对每天中每个日志产生一个随机数,如果该随机数小于概率 C 则保留该日志,否则排除;

步骤 (3) 根据以上保留的日志数据统计出每天每个 cookie 的访问次数,用日期和 cookie 作为 key,统计访问次数;

步骤 (4) 用日期和 cookie 作为 key,统计每天的累积访问次数;

步骤 (5) 用日期和访问次数作为 key,统计访问过不同次数的用户数;

步骤 (6) 计算从访问过 1 次以上的用户数到 n 次以上的用户数;

在进行频次维度控制时,该方法包括以下步骤:

步骤 (7) 将符合投放维度条件的日志记录筛选出来;

步骤 (8) 计算概率 $E = \text{频次维度数值} / \text{用户实际观看次数}$,用生成随机数取概率的方式,如果随机数小于此概率 E,则保留观看日志,同时记录总保留日志不超过频次控制次数;

步骤 (9) 计算每天的容量;

步骤(10) 每个日期点都要进行随机数取概率, 计算概率 $C = \text{预设的每天投放量 } A / \text{实际的每天容量 } B$, 小于此概率 C 的日期点日志保留, 大于概率 C 的日期点日志丢弃。

5. 根据权利要求 4 所述的方法, 其特征在于步骤(1) 中的容量按如下方式进行计算:
对每天的日志进行分析, 符合投放维度的即记为 1, 不断累加, 累加值为容量。

6. 根据权利要求 4 所述的方法, 其特征在于:
所述投放维度包括地域维度、频道维度、版权维度、剧目维度、小时维度。

一种预测网络视频广告的到达频次表数据的装置及方法

技术领域

[0001] 本发明涉及互联网领域,尤其是涉及一种预测网络视频广告的到达频次表数据的装置及方法。

背景技术

[0002] 目前视频广告的主要形式有如下方式,前贴片:在一段视频播放开始之前出现广告;后贴片:在一段视频播放完成之后出现广告;暂停:在一段视频点击暂停后浮现在页面播放器中的广告;以上不同的广告形式对于此专利都适用。广告投放量计算单位为 cpm(每千人成本):广告被播放 1000 次即称为投放了 1 个 cpm。视频广告的常用投放维度有,时间:从起始日到结束日;地域:是否针对某一个省份或者城市做定向投放(比如北京或者上海),还是全国普投;频道:是否针对某一个特定的频道投放(比如电视剧或者电影),还是针对所有的频道;内容:是否针对某一个特定的内容投放(比如只针对《还珠格格》这部剧做投放);其他:更细粒度的维度,比如小时维度(比如只投放在每天的晚 8 点到 10 点),版权维度(比如只投放在电视剧版权下,不投放在自制剧上);频次:是否针对某一个用户做频次控制(比如在某一单的广告投放中,规定同一用户在一定的时间周期内,最多只能看到不超过一定次数的这个广告)。

[0003] 举例说明:客户在 6.1 到 6.21(时间维度)确定了一轮前贴片投放(广告形式),投放内容为某某活动,定投 4 个城市,分别是北京,上海,深圳,天津(地域维度),电视剧频道(频道维度),每天投放量为 1000 个 cpm,投放周期内每个用户观看此广告不超过 2 次(频次维度)。用户观看视频日志为当用户点开一个视频的时候,服务器会发送一条请求,记录当前的用户观看的状态,记录的信息如下表所示(这里只列出和本发明有关系的字段,其它字段没有列出):

[0004]

时间戳	IP 地址	频道	二级频道	视频识别	关键字	Cookie	剧目识别
1341504000314	218.107.16.182	D	2071	105027107	天涯明月刀	1339936540824eot	196106

[0005] 此条记录表示 cookie 为 1339936540824eot 的用户, 在 2012 年 7 月 6 日 0:00:00 (1341504000314 是毫秒数, 换算成时间) 时, 观看了视频标识为 105027107 的视频, 此视频关键字是“天涯明月刀”, 此视频所属频道是 d (d 表示电视剧频道), 二级频道是 2071 (电视剧中的武侠类), 剧目标识为 196106。用户每观看一次视频, 则记录一条这样的日志, 目前一天记录的日志经过压缩大小为 8 到 10 个 GB。广告容量: 在一定时期内, 限定一些广告投放的维度, 最多能够观看多少次广告。比如, 在 6.9 这天, 广东省电视剧频道的前贴片一共投放了 10189.428CPM 广告。那么 10189.428CPM 就是广东省电视剧频道前贴片在 6.9 这天的容量。广告效果描述:

[0006] 目前常用的是 Reach_Frequency (到达 / 频次) 表数据。如下表所示:

[0007]

日期	展示 次数	1+	2+	3+	4+	5+	6+	7+	8+	9+	10+	11+	12+	13+	14+	15+	16+	17+	18+	19+	20+
2012-06-09	3566300	1346851	769317	480592	309011	202309	133822	89970	61003	42081	29598	21144	15356	11410	8610	6654	5142	4095	3305	2706	2246
2012-06-10	7395414	2269871	1380965	949746	679355	497194	368900	276843	209516	159709	121561	93164	71791	55410	42908	33564	26452	21001	16813	13632	11088
2012-06-11	10419249	2932741	1810897	1272932	936735	708763	545880	426082	335218	265966	211733	169076	135614	108772	87359	70505	57178	46408	37976	31148	25801

[0008] 上表中第一行表示 2012.6.9 这天此广告一共展示了 3566300 次, 看了至少 1 次的用户数有 1346851, 至少 2 次的用户数有 769317, 依次类推到至少看过 20 次的用户数有 2246。2012.6.10 的数据是表示从 2012.6.9 到 2012.6.10 这两天的展示次数和不同频次的用户数。

[0009] 根据现有技术, 用户无法预测要投放的广告的效果, 即无法预测到达频次表数据, 然而这正是用户预先想要获得数据, 因此, 如何预测网络视频广告的到达频次表数据就是现在亟待解决的技术问题。

发明内容

[0010] 本发明针对现有技术存在的不足, 提供了一种预测网络视频广告的到达频次表数据的装置及方法, 其中该装置包括: 观看视频日志采集模块, 用于记录用户观看视频的状态为一个观看视频日志; 容量计算模块, 选择邻接预测日期范围之前的历史日期范围, 在不进行频次维度控制时, 计算满足投放维度的在历史日期范围内每天的容量; 广告维度筛选模块, 根据预期的投放量与计算出的容量计算概率 C, 用随机类对每天中每个日志产生一个随机数, 如果该随机数小于概率 C 则保留该日志, 否则排除; 不分频次控制计算模块, 根据以上保留的日志数据统计出每天每个 cookie 的访问次数, 用日期和 cookie 作为 key, 统计访问次数; 用日期和 cookie 作为 key, 统计每天的累积访问次数; 用日期和访问次数作为 key, 统计访问过不同次数的用户数; 计算从访问过 1 次以上的用户数到 n 次以上的用户数。到达频次表数据计算模块, 根据上述计算结果统计预测日期内每天的 1 次以上的用户数到 n 次以上的用户数。

[0011] 进一步,在同时进行频次控制时该装置中的不分频次控制计算模块替换为:分频次控制计算模块,将符合投放维度条件的日志记录筛选出来;计算概率 $E = \text{频次维度数值} / \text{用户实际观看次数}$,用生成随机数取概率的方式,如果随机数小于此概率 E ,则保留观看日志,同时记录总保留日志不超过频次控制次数;计算每天的容量;每个日期点都要进行随机数取概率,计算概率 $C = \text{预设的每天投放量} A / \text{实际的每天容量} B$,小于此概率 C 的日期点日志保留,大于概率 C 的日期点日志丢弃。

[0012] 进一步,所述对每天的容量进行计算是按以下方式进行计算:对每天的日志进行分析,符合投放维度的即记为1,不断累加,累加值为容量。

[0013] 进一步,所述根据预期的投放量与计算出的容量计算概率 C 是指订单的每天预期投放量除以计算出的每天的容量。

[0014] 进一步,所述投放维度包括地域维度、频道维度、版权维度、剧目维度、小时维度。

[0015] 此外,本发明还提供了一种预测网络视频广告的到达频次表数据的方法,该方法包括:步骤(1)选择邻接预测日期范围之前的历史日期范围,在不进行频次维度控制时,计算满足投放维度的在历史日期范围内每天的容量;步骤(2)根据预期的投放量与计算出的容量计算概率 C ,用随机类对每天中每个日志产生一个随机数,如果该随机数小于概率 C 则保留该日志,否则排除;步骤(3)根据以上保留的日志数据统计出每天每个cookie的访问次数,用日期和cookie作为key,统计访问次数;步骤(4)用日期和cookie作为key,统计每天的累积访问次数;步骤(5)用日期和访问次数作为key,统计访问过不同次数的用户数;步骤(6)计算从访问过1次以上的用户数到 n 次以上的用户数。

[0016] 进一步,步骤(1)中的容量按如下方式进行计算:对每天的日志进行分析,符合投放维度的即记为1,不断累加,累加值为容量。

[0017] 进一步,如果进行频次维度控制时,包括以下步骤:步骤(7)将符合投放维度条件的日志记录筛选出来;步骤(8)计算概率 $E = \text{频次维度数值} / \text{用户实际观看次数}$,用生成随机数取概率的方式,如果随机数小于此概率 E ,则保留观看日志,同时记录总保留日志不超过频次控制次数;步骤(9)计算每天的容量;步骤(10)每个日期点都要进行随机数取概率,计算 $C = \text{预设的每天投放量} A / \text{实际的每天容量} B$,小于此概率 C 的日期点日志保留,大于概率 C 的日期点日志丢弃。

[0018] 进一步,所述投放维度包括地域维度、频道维度、版权维度、剧目维度、小时维度。

附图说明

[0019] 图1是根据本发明装置的模块图。

具体实施方式

[0020] 为使本发明的上述目的、特征和优点更加明显易懂,下面结合附图和具体实施方式对本发明作进一步详细的说明:

[0021] 客户在5.26到6.15(时间维度)确定了一轮前贴片投放(广告形式),投放内容为某某活动,定投4个城市,分别是北京,上海,深圳,天津(地域维度),电视剧频道(频道维度),每天投放量为1000个cpm,投放周期内每个用户观看此广告不超过2次(频次维度)。比如客户在5月底谈这个合同,想预测此次投放的结果如何,也即预测到达频次表数

据。因为频次维度的有无决定了算法稍有不同,所以此处分两类说明。

[0022] 首先,根据每天记录的用户观看行为日志,选择邻接 5.26 到 6.15 之前的日期范围,比如选择 5.5 到 5.25 的周期。广告形式为前贴片。(越接近的历史周期效果越好)

[0023] 不分频次控制时,即如果不控制每个用户观看此广告不超过 2 次(频次维度),此投放的维度有 2 个,分别是地域维度和频道维度。

[0024] 第一步:计算次投放维度每天的容量:

[0025] 根据 5.5 到 5.25 的视频观看日志,从 5.5 到 5.25 共 21 天,对每天此 4 个城市电视剧频道的容量进行计算,分别为(下划线分隔)6310124_6668956_5125200_5145431_5150704_5294901_5807339_6721603_7299027_5665583_5493653_5660381_5722277_6192772_7302331_7528980_5844377_5836514_5924083_6125748_6392961

[0026] 计算方式为:凡是符合此投放维度(即 4 个城市的电视剧频道,城市根据 ip 解析),即记为 1,不断累加。

[0027] 第二步:广告维度的筛选

[0028] 声明一个产生随机数的类,再次读取 5.5 到 5.25 的视频观看日志,根据投放维度进行筛选。不符合投放维度的记录排除,找到投放记录的记录时开始判断(或者将第一步符合条件的日志筛选出来后,可以读取这个筛选后的日志,并且读取过程中根据随机数筛选出比例)。在读到 5.5 的日志的时候,刚才计算出的容量为 6310124(记为 A),这里的订单需要每天投放 1000 个 cpm(即为 1000*1000 次广告)(记为 B),那么得到概率 $C = B/A = 1000*1000/6310124 = 0.1585$,用随机类生成一个小于 1.0 的随机数,如果小于概率 C,则将此记录保留下来,如果大于此概率,则将此记录排除。通过这样的方式,可以保证每天投放的 cpm 均稳定在 1000 左右。

[0029] 第三步:广告效果的计算

[0030] 根据第二步保留下来的日志记录,即可看作预测投放出去的基础数据,在此数据的基础上,可以进行广告效果的计算。因为视频的容量在一定的周期内都会保持一个稳定的趋势,所以用此结果即可以知道 5.26 到 6.15 的投放结果也大致如此。计算的关键在于容量的计算和比例的计算。为了得到表中的预测结果数据,将筛选后的日志进行统计。假设,筛选出的日志如下,这里为了简便,省略了其他维度。

[0031]

日期	访问 cookie
----	-----------

20120505	cookie_A
----------	----------

20120505	cookie_A
----------	----------

20120505	cookie_B
----------	----------

20120505	cookie_C
----------	----------

20120506	cookie_A
----------	----------

20120506	cookie_C
----------	----------

20120507	cookie_B
----------	----------

20120507	cookie_C
----------	----------

20120508	cookie_A
----------	----------

20120509	cookie_B
----------	----------

[0032] 这里可以看出:20120505日,cookie_A访问了2次,cookie_B访问了1次,cookie_C访问了1次

[0033] 假设根据这些日志计算 20120505 到 20120509 的投放效果

[0034] 3.1:统计出每天每个 cookie 的访问次数,用日期和 cookie 作为 key, 统计访问次数,得到结果如下

[0035]

20120505	cookie_A	2
----------	----------	---

20120505	cookie_B	1
----------	----------	---

20120505	cookie_C	1
----------	----------	---

[0036]

20120506 cookie_A 1

20120506 cookie_C 1

20120507 cookie_B 1

20120507 cookie_C 1

20120508 cookie_A 1

20120509 cookie_B 1

[0037] 数据说明举例:20120505日,cookie_A访问了2次,cookie_B访问了1次,cookie_C访问了1次。

[0038] 3.2:因为广告效果是按照累计天统计(如最上面表格的说明2012.6.10的数据是表示从2012.6.9到2012.6.10这两天的展示次数和不同频次的用户数。)

[0039] 统计从20120505到20120509累计的访问次数,用日期和cookie作为key,比如cookie_A,为了得到20120505到20120509的访问效果,逻辑如下

[0040]

startdate=日志中的访问日期

enddate=20120509

tempdate=startdate

while(tempdate<=enddate){

key=tempdate+cookie;

key的计数+1;

tempdate=tempdate+1天;

[0041]

}

[0042] 针对cookie_A,计算出的效果(累计访问次数)如下

[0043] 20120505 cookie_A 2(20120505访问了2次)

[0044] 20120506 cookie_A 3(20120506访问了1次,所以从20120505到20120506共访问了3次)

[0045] 20120507 cookie_A 3(20120507没有访问,所以从20120505到20120507还是共访问了3次)

[0046] 20120508 cookie_A 4(20120508访问了1次,所以从20120505到20120508共访问了4次)

[0047] 20120509 cookie_A 4(20120509 没有访问,所以从 20120505 到 20120509 还是共访问了 4 次)。

[0048] 同样,得到 cookie_B 的累计访问次数如下:

[0049]

20120505 cookie_B 1

20120506 cookie_B 1

20120507 cookie_B 2

20120508 cookie_B 2

20120509 cookie_B 3

[0050] 得到 cookie_C 的累计访问次数如下:

[0051]

20120505 cookie_C 1

20120506 cookie_C 2

20120507 cookie_C 3

20120508 cookie_C 3

[0052]

20120509 cookie_C 3

[0053] 3.3:统计访问过不同次数的用户有几个,用日期和访问次数作为 key

[0054] 20120505 1 2(20120505 访问过 1 次的有 cookie_B 和 cookie_C, 共 2 人)

[0055] 20120505 2 1(20120505 访问过 2 次的有 cookie_A, 共 1 人)

[0056] 20120506 1 1(20120505 到 20120506 访问过 1 次的有 cookie_B, 共 1 人)

[0057] 20120506 2 1(20120505 到 20120506 访问过 2 次的有 cookie_C, 共 1 人)

[0058] 20120506 3 1(20120505 到 20120506 访问过 3 次的有 cookie_A, 共 1 人)

[0059] 20120507 2 1(20120505 到 20120507 访问过 2 次的有 cookie_B 共 1 人)

[0060] 20120507 3 2(20120505 到 20120507 访问过 3 次的有 cookie_A 和 cookie_C 共 2 人)

[0061] 20120508 2 1(20120505 到 20120508 访问过 2 次的有 cookie_B 共 2 人)

[0062] 20120508 3 1(20120505 到 20120508 访问过 3 次的有 cookieC 共 1 人)

[0063] 20120508 4 1(20120505 到 20120508 访问过 4 次的有 cookie_A 共 1 人)

[0064] 20120509 3 2(20120505 到 20120509 访问过 3 次的有 cookie_B 和 cookie_C 共 2 人)

[0065] 20120509 4 1(20120505 到 20120509 访问过 4 次的有 cookie_A 共 1 人)

[0066] 3.4:计算从访问过 1 次以上到 20 次以上的人有多少,计算逻辑如下:

[0067] int n = 访问次数 ;

[0068] n = n > 5 ? 5 : n ; (该命令表示如果 n > 5 则按照 5 计算, 因为只计算到 5+, 即最大只统计到访问超过 5 次以上的用户有多少. 最上面的表格是按照统计到 20 次计算, 这里因为数据较少, 统计到 10 次即可)

[0069] for (int i = 1 ; i <= n ; i++) {

[0070] key = 日期 + i ;

[0071] key 的次数 +1 ;

[0072] }

[0073] 20120505 1 3 (20120505 访问过 1 次及以上的有 3 人, 包括访问过 1 次的有 2 人, 访问过 2 次的有 1 人)

[0074] 20120505 2 1 (20120505 访问过 2 次及以上的有 1 人)

[0075] 同样,

[0076] 20120506 1 3 (20120505 到 20120506 访问过 1 次及以上的有 3 人)

[0077] 20120506 2 2 (20120505 到 20120506 访问过 2 次及以上的有 2 人)

[0078] 20120506 3 1 (20120505 到 20120506 访问过 3 次及以上的有 1 人)

[0079] 20120507 1 3 (20120505 到 20120507 访问过 1 次及以上的有 3 人)

[0080] 20120507 2 3 (20120505 到 20120507 访问过 2 次及以上的有 3 人)

[0081] 20120507 3 2 (20120505 到 20120507 访问过 3 次及以上的有 2 人)

[0082] 20120508 1 3 (20120505 到 20120508 访问过 1 次及以上的有 3 人)

[0083] 20120508 2 3 (20120505 到 20120508 访问过 2 次及以上的有 3 人)

[0084] 20120508 3 2 (20120505 到 20120508 访问过 3 次及以上的有 2 人)

[0085] 20120508 4 1 (20120505 到 20120508 访问过 4 次及以上的有 1 人)

[0086] 20120509 1 3 (20120505 到 20120509 访问过 1 次及以上的有 3 人)

[0087] 20120509 2 3 (20120505 到 20120509 访问过 2 次及以上的有 3 人)

[0088] 20120509 3 3 (20120505 到 20120509 访问过 3 次及以上的有 3 人)

[0089] 20120509 4 1 (20120505 到 20120509 访问过 4 次及以上的 1 人)

[0090] 最终得到表格如下 :

[0091]

日期	展示次数 (sum(1+ 到 5+))	1+	2+	3+	4+	5+
20120505	4	3	1	0	0	0
20120506	6	3	2	1	0	0
20120507	8	3	3	2	0	0
20120508	9	3	3	2	1	0
20120509	10	3	3	3	1	0

[0092] 数据说明 : 最后一行表示从 20120505 到 20120509 这段期间, 共展示过广告 10 次,

其中看过 1 次及以上的有 3 人,看过 2 次及以上的有 3 人,看过 3 次及以上的有 3 人,4 次及以上的有 1 人;这样就得到了从 20120505 到 20120509 的广告效果预测结果:这里所有的实现算法均用 hadoop 集群处理和 HDFS 分布式存储实现。

[0093] 如果将频次维度的条件加入,算法如下:

[0094] 第一步:将符合投放条件的日志记录筛选出来,因为涉及到每个用户观看的频次,所以保留的形式如下:

[0095] cookie: 观看日期序列(按照从早到晚排序)

[0096] 比如用户 A:20120505, 20120505, 20120510

[0097] 表示 A 用户共看了 3 次符合投放维度的视频,分别是在 20120505 日看了 2 次, 20120510 看了 1 次

[0098] 第二步:

[0099] 在第一步的基础上进行频次控制。比如用户 A 看过 3 次,而客户要求控制在 2 次。那么还是用生成随机数取概率的方式,计算概率 $E = 2/3 = 0.667$, 如果随机数小于此概率,则保留两次观看点。同时记录总保留观看点不超过频次控制次数。通过第二步得到的结果比如为用户 A:20120505, 20120510。

[0100] 第三步:

[0101] 计算容量。在第二步的结果基础上,计算每天的容量。比如用户 A 为 20120505 观看一次,20120510 观看一次,那么 20120505 的容量累加 1,20120510 的容量也累加 1。计算出在频次控制下每天的容量大小。比如 20120505 的容量数据在频次控制下只能达到 2310124(记为 B)。

[0102] 第四步:

[0103] 根据第三步计算出的容量数据,再次读取第二步的结果数据,每天的投放量 1000 个 cpm(记为 A),同样用概率取值,为了均匀得到日志,比较概率的时候均是取随机数。这里频次控制每个日期点都要进行随机数取概率,计算 $C = A/B = 1000*1000/2310124 = 0.4329$, 比较随机数概率与 C,小于此概率 C 的话这个观看日期点保留,大于概率 C 的丢弃。

[0104] 第五步:

[0105] 根据第四步计算出的结果,即可计算广告效果。(统计效果的计算方法与上述内容相同)

[0106] 使用本专利能够根据客户的投放要求,准确量化的预测投放的效果。目前大多数广告预测都是凭借经验,甚至不做预测。而本专利根据历史的真实数据,结合客户的需求,经过设计算法运算,将准确的计算结果数据提供给客户,用以指导未来的投放行为。

[0107] 如图 1 所示,根据本发明提供了一种预测网络视频广告的到达频次表数据的装置,其中该装置包括:观看视频日志采集模块,用于记录用户观看视频的状态为一个用户观看视频日志;容量计算模块,选择距离预测日期范围较近的历史日期范围,在不进行频次维度控制时,计算满足投放维度的在历史日期范围内每天的容量,投放维度包括地域维度和频道维度;广告维度筛选模块,根据预期的投放量与计算出的容量计算概率 C,用随机类对每天中每个日志产生一个随机数,如果该随机数小于概率 C 则保留该日志,否则排除;不分频次控制计算模块,根据以上保留的日志数据统计出每天每个 cookie 的访问次数,用日期和 cookie 作为 key, 统计访问次数;用日期和 cookie 作为 key,统计每天的累积访问次数;

用日期和访问次数作为 key,统计访问过不同次数的用户数;计算从访问过 1 次以上到 n 次以上的用户数。到达频次表数据计算模块,根据上述计算结果统计预测日期内每天的 1 次以上到 n 次以上的用户数。

[0108] 进一步,该装置还包括:分频次控制计算模块,将符合投放维度条件的日志记录筛选出来;计算概率 $E = \text{频次维度数值} / \text{用户实际观看次数}$,用生成随机数取概率的方式,如果随机数小于此概率 E,则保留观看日志,同时记录总保留日志不超过频次控制次数;计算每天的容量;每个日期点都要进行随机数取概率,计算 $C = \text{预设的每天投放量} A / \text{实际的每天容量} B$,小于此概率 C 的日期点日志保留,大于概率 C 的日期点日志丢弃。

[0109] 本专利可以根据投放维度不断细分,未来如果新增了广告投放形式或者其他广告投放维度,本专利都能根据实际需求做相应变化。

[0110] 使用本专利实现的广告预测系统,用客户实际投放的结果和预测的结果做对比,误差率平均在 5% 左右。

[0111] 以上是对本发明的优选实施例进行的详细描述,但本领域的普通技术人员应该意识到,在本发明的范围内和精神指导下,各种改进、添加和替换都是可能的。这些都在本发明的权利要求所限定的保护范围内。

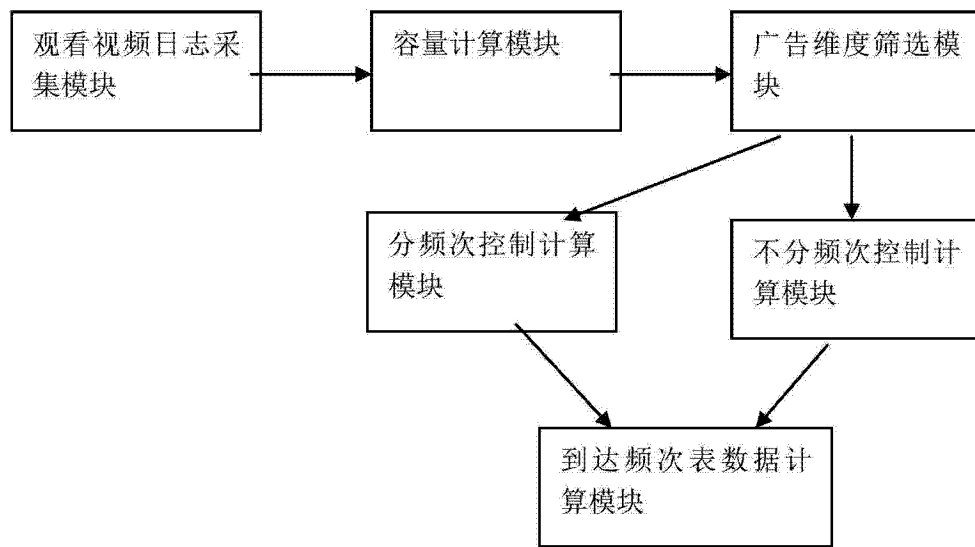


图 1