

文章编号:1005-9679(2019)01-0113-05

关于 IPTV 业务中广告库存分配算法的研究

邵进明¹ 蒋 炜¹ 李艳婷²

(1. 上海交通大学 安泰经济与管理学院, 上海 200030; 2. 上海交通大学 机械与动力工程学院, 上海 200240)

摘要: 目前国内关于 IPTV 广告投放的算法研究还未完全成熟, 很多企业在实践中更多依赖于人工经验分配。主要研究 IPTV 业务中广告库存匹配问题, 建立了广告流量分配算法模型, 并根据问题特性优化精简了模型约束条件, 同时利用稀疏矩阵的便利性求解, 最后通过某新媒体公司的 IPTV 广告投放案例对算法模型的可行性进行了验证。

关键词: IPTV; 新媒体广告; 库存; 分配算法

中图分类号: F 22 **文献标志码:** A

Study on the advertising inventory allocation algorithm of IPTV business

SHAO Jinming¹ JIANG Wei¹ LI Yanting²(1. Antai College of Economics and Management, Shanghai Jiao Tong University, Shanghai 200030, China;
2. School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China)

Abstract: Allocation algorithms for the current domestic IPTV advertising allocation has not yet fully mature, many enterprises depends more on human experience in practice. This paper mainly studies the advertising inventory allocation and matching problems on IPTV business, and an advertising traffic allocation model is established, of which the constraints are simplified according to the characteristics of the problem, while taking advantage of the convenience of the sparse matrix. Finally, we verify the feasibility of the algorithm model through a new media company's IPTV advertising allocation case.

Key words: IPTV; new media advertising; inventory; allocation algorithm

0 引言

IPTV(Internet Protocol Television)是宽带电视的一种,它利用广电网络提供的丰富的节目资源,采用数据传输速率更高、覆盖范围更广、运行状况良好的宽带通信网作为基础网络设施。除了计算机和移动设备终端,用户使用网络机顶盒和普通电视机即可通过互联网接入享受个性化、交互式、可定制的 TV 视听业务及其他增值服务。

从 2001 年 3 月 15 日通过的“十五规划纲要”第一次明确提出“三网融合:促进电信、电视、互联网三

网融合”开始,到近日国务院办公厅印发《三网融合推广方案》,明确在总结试点经验的基础上,加快在全国全面推进三网融合,推动信息网络基础设施互联互通和资源共享,将广电、电信业务双向进一步扩大到全国范围,并开展实质性工作。在中央推进三网融合的重点工作中,IPTV 融合了电信、广电、互联网三方面的技术和优势并创造了新的价值领域。广电运营商拥有丰富的节目内容资源和稳定的用户群体,电信和互联网运营商则拥有交互技术和网络,在很大程度上,IPTV 将会促进三网融合的进程,推进电信、广电、互联网利用各自的特点更好地为人们

收稿日期:2018-03-14

作者简介:邵进明,男,硕士研究生,主要研究方向为基于数据的商业运营分析,E-mail: yuanqiao90@yeah.net;蒋炜,教授,博士生导师,主要研究方向为质量管理、大数据与商务智能;李艳婷,副教授,主要研究方向为质量管理、数据挖掘。

提供服务。2012 年,中国 IPTV 用户规模占全球的 1/4,根据工信部官网 2015 年 2 月的数据,2014 年我国三家基础电信企业 IPTV 用户达到了 3363.6 万户。不管是从用户规模、政策扶持,还是从与之相关的互联网行业发展状况来看,IPTV 在实践中的发展前景相当可观。

国内外学者关于 IPTV 广告投放算法的研究,并不能很好地解决我国目前 IPTV 行业的实际问题。我国的 IPTV 业务采取的是广电与电信合作的模式,双方掌握的用户数据是不共享的,IPTV 广告运营方无法获取准确的用户行为数据来定制用户标签,也就无法使用竞拍投放定向广告的方式将用户出售给广告主,因而目前使用的依然是基于节目内容而非用户的售卖模式,在面对众多广告主时,如何优化流量分配方案、提高广告流量利用效率显得尤为重要。

1 问题描述

在 IPTV 广告库存的分配问题中,涉及节目内容所有者、运营媒体、广告主(需要投放产品广告商家)、广告代理商、收视用户等多个主体,在不影响问题本质的前提下,这里简化成媒体、广告主、用户三者之间的关系(如图 1 所示)。在这三者之中,媒体拥有节目资源版权,并通过 IPTV 渠道将这些节目内容以直播、点播、回看等形式供用户自行免费或付费收看;用户根据个人兴趣喜好及空闲时间,通过遥控器向媒体的服务器发出收视请求,然后媒体服务器根据接收到的用户请求将对应的节目投播到用户终端;节目资源在用户终端的每一次播放都意味着一次投放广告的机会,或称之为广告流量,媒体将拥有的广告流量出售给不同的广告主,并在协商过程中根据广告主提出的投放诉求制定具体的投放分配计划。

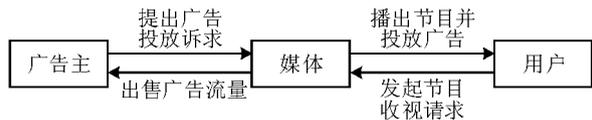


图 1 广告业务流程

一般而言,IPTV 运营方采用预售方式(例如出售未来三个月的广告流量):广告主可以对节目内容、播出时间、曝光次数提出要求,形成意向订单,媒体结合自身广告流量的预期库存水平与广告主协商后决定是否签订正式售卖合同。

广告主的需求可以用播放时段(播放广告的时间段)、节目路径(播放广告的节目载体)和投放数量

(广告曝光次数)三个维度来描述。例如:广告主希望于 2013 年 4 月 1 日至 4 月 3 日及 4 月 11 日至 4 月 17 日这 10 天的 19:00—21:00 在《中国达人秀》及《港台电影》将广告素材播放 50 万次。广告素材由广告主提供,可以是图片、文字或短视频等。有别于传统电视直播或重播均由电视台控制的特点,IPTV 用户收看节目有点播、回看、时移、直播等多种形式且可自由选择,因而在广告主的广告投放订单描述中,对投放时间及投放节目的描述也有相应的特殊之处。

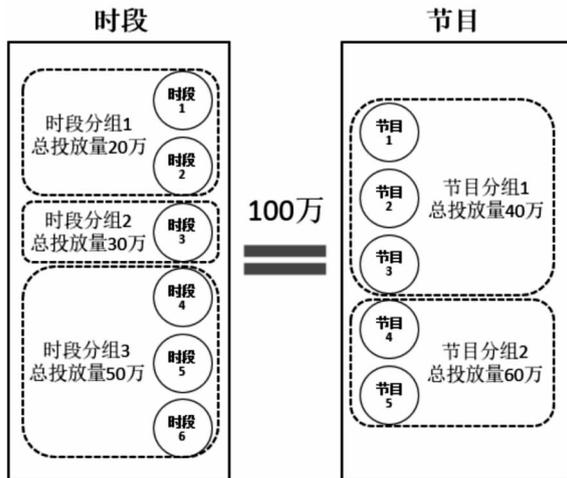


图 2 IPTV 广告投放要求图示

如图 2 所示,假设某订单一共需要投放 100 万广告流量。投放协议约定,广告必须在指定的 6 个时段内播出,且这 6 个时段被拆分成 3 个时段分组,各组投放流量分别为 20 万、30 万、50 万,合计 100 万;同时,广告必须在指定的 5 个节目中播出,且这 5 个节目被拆分成 2 个节目分组,各组投放流量分别为 40 万、60 万,合计 100 万。订单在时段和节目两个维度对流量的需求是独立的,不要求某个节目与某个时段一一对应,但由于广告主并不会为超额投放的广告流量买单,却会对未完成投放的流量索取违约金(惩罚成本),因而媒体只能在对节目流量预测的基础上,完成订单的同时充分利用节目流量,尽可能让广告主需求的时段和节目匹配投放。

在实践中,对于 IPTV 媒体运营者而言,由于采用预售方式,需要对未来一段时间的广告流量进行预测,终端用户作为广告流量的创造者,其收视行为的偶然性造成流量预测必然存在不确定性;与此同时,广告主的订单到达时间及订单需求的数量也是不确定的。在这样一个广告流量供给和需求均存在不确定性的售卖模式中,如何制定广告投放计划成为亟待解决的问题。

本文主要基于媒体的视角,研究如何在广告流量库存有限的情形下快速有效且合理地对其进行分配以满足不同广告主的投放需求,简而言之就是库存与需求的匹配问题。为了简化考虑,本文假设广告流量和广告主订单都是已知的,广告流量单价在售卖合同签订时约定,在此基础上设计以月为周期的广告流量分配算法。

2 建立模型

经过对问题的分析,本文选用线性规划模型给出数学表述形式。首先对模型中涉及的变量进行定义:

i, j, k : 订单编号, 投放时段(以小时为单位), 节目编号;

I, J, K : 所有订单集合, 所有时段集合, 所有节目集合;

m, n, M_i, N_i : 时段分组编号, 节目分组编号, 订单 i 指定的时段分组集合, 订单 i 指定的节目分组集合;

p_i : 订单的合同约定单价(常用每千次展示成本 CPM, 这里处理成单次播放售价);

Φ_m^i : 订单 i 要求的编号为 m 的时段分组中包含的所有时段集合, 其中 $m \in M_i$;

Ψ_n^i : 订单 i 要求的编号为 n 的节目分组中包含的所有节目集合, 其中 $n \in N_i$;

V_m^i : 订单 i 在编号为 m 的时段分组要求投放的广告流量, 其中 $m \in M_i$;

H_n^i : 订单 i 在编号为 n 的节目分组要求投放的广告流量, 其中 $n \in N_i$;

z_{jk} : 节目 k 在时段 j 的实际流量, 则预期流量 $Z_{jk} = E(z_{jk})$, 销售取期望值;

X_{ijk} : 节目 k 在时段 j 实际分配给订单 i 的广告流量;

c_i : 订单完全违约的惩罚成本, 则该订单实际发生的惩罚成本为 $c_i \left[1 - \frac{\sum_{j \in \Phi_i} \sum_{k \in \Psi_i} X_{ijk}}{V_i} \right]$ 。

显然, 满足订单要求的所有时段可以表示为集合:

$$T_i @ \bigcup_{m \in M_i} \Phi_m^i \quad (1)$$

满足订单要求的所有节目可以表示为集合:

$$P_i @ \bigcup_{n \in N_i} \Psi_n^i \quad (2)$$

对于订单, 从时段和节目两个维度独立定义的广告流量总需求量应当一致, 从而有

$$D_i @ \sum_{m \in M_i} V_m^i = \sum_{n \in N_i} H_n^i \quad (3)$$

对于订单 i , 若时段或节目不在指定的集合内, 则默认分配的广告流量为零:

$$X_{ijk} \equiv 0 \quad \forall i \in I, j \notin T_i, \text{ or } k \notin P_i \quad (4)$$

此外, 预期广告流量 Z_{jk} 使用预测值, 视为已知量, X_{ijk} 为决策变量, 最终目标是使得收益 R 最大化。具体模型如下:

$$\text{Max } R = \sum_{i \in I} p_i \sum_{j \in T_i} \sum_{k \in P_i} X_{ijk} - \sum_{i \in I} c_i \left[1 - \frac{\sum_{j \in T_i} \sum_{k \in P_i} X_{ijk}}{D_i} \right] \quad (5)$$

s. t.

$$\sum_{j \in \Phi_m^i} \sum_{k \in P_i} X_{ijk} = V_m^i \quad \forall i \in I, m \in M_i \quad (6)$$

$$\sum_{j \in T_i} \sum_{k \in \Psi_n^i} X_{ijk} = H_n^i \quad \forall i \in I, n \in N_i \quad (7)$$

$$\sum_{i \in I} X_{ijk} \leq Z_{jk} \quad \forall j \in T_i, k \in P_i \quad (8)$$

$$X_{ijk} \geq 0 \quad \forall j \in I, j \in T_i, k \in P_i \quad (9)$$

在上述模型中, 式(5)作为目标函数, 优化目标是使得总收益 R 最大化, 式(6)、(7)表示实际分配的流量在节目和时间两个维度都应当与订单需求量相等, 式(8)表示实际分配流量不应当超过流量的预期库存水平, 式(9)是非负约束。若式(6)、(7)不成立则直接导致客户订单未满足或因过分配造成流量浪费, 若式(8)不成立则表示存在超卖现象(按超过预期的库存水平出售), 这两者都会带来违约风险和惩罚成本, 严重者甚至会逐渐失去客户的信任, 引发企业信誉危机。

故而, 企业在实际决策时需要考虑以下条件和目标:

① 应尽可能达到并超过计划收益 \bar{R} (\bar{R} 为给定值);

② 应尽可能满足已形成正式售卖合同的所有客户订单, 但不要超过需求量;

③ 应尽可能充分售出广告流量, 但不要过度超卖。

这样, 流量分配决策转化成多目标决策问题, 从而可以将以上线性规划转化为目标规划问题。现假定以上三个目标对应的优先因子分别为 P_1, P_2, P_3 并规定 $P_1 \gg P_2 \gg P_3$, 表示首先保证第一个目标的实现, 这时可以不考虑第二个目标, P_2 级目标是在实现 P_1 级目标的基础上考虑的, 以此类推。此外, 对于同一个优先因子的两个目标差别, 可以赋予不同的权重系数 w 。优先因子 P 和权重系数 w 都由决策者按照具体情况确定。

原问题转化为目标规划数学模型如下:

$$\begin{aligned} \text{Min } C = & P_1 d_{11}^- + P_2 \sum_{i \in I} (w_{2i}^- d_{2i}^- + w_{2i}^+ d_{2i}^+) + \\ & P_3 \sum_{n=1}^N (w_{3n}^- d_{3n}^- + w_{3n}^+ d_{3n}^+) \end{aligned} \quad (10)$$

$$\sum_{i \in I} p_i \sum_{j \in \Phi_i} \sum_{k \in \Psi_i} X_{ijk} - \sum_{i \in I} c_i \left[1 - \frac{\sum_{j \in \Phi_i} \sum_{k \in \Psi_i} X_{ijk}}{V_i} \right] +$$

$$d_{11}^- - d_{11}^+ = \bar{R} \tag{11}$$

$$\sum_{j \in \Phi_i} \sum_{k \in \Psi_i} X_{ijk} + d_{2i}^- - d_{2i}^+ = V_i \quad \forall i \in I$$

$$\sum_{i \in I} X_{ijk} + d_{3r}^- - d_{3r}^+ = Z_{jk} \quad \forall j \in J \quad \forall k \in K$$

$$r = 1, 2, \dots, Q \quad Q = j \cdot k \tag{12}$$

$$X_{ijk}, d^-, d^+ \geq 0$$

$$\forall i \in I \quad \forall j \in J \quad \forall k \in K \tag{13}$$

在对以上模型求解时可以按照线性规划的常见算法进行,例如单纯形法、内点法、外点法等。当然,在具体的实践案例中,要根据案例自身的特殊性 & 数据的特征灵活处理。

3 实证分析

本文的研究案例来自一家新媒体 IPTV 公司,所使用的数据在不影响业务描述和学术研究实证分析的前提下进行了抽样、缩放等处理,仅作为学习研究使用,不对外公开。需要解决的问题正如前文中的问题描述所示:基于媒体的视角,研究如何在广告流量库存有限的情形下快速有效且合理地对其进行分配,以满足不同广告主的投放需求。

本文选择了一个月的节目播放历史数据及对应时段的广告投放订单,共计包含 215 个节目、264 个需求订单,其中需求细分包括 371 个时段分组及 446 个节目分组。媒体服务器在播出节目时系统会自动记录收看用户的操作信息,包括用户 ID、节目路径 ID 和点播时间。由于广告主订单一般对时间的要求都是精确到小时为单位,故而首先需要对数据进行预处理,将原先精确到秒级别的播放记录统计成小时级别的流量。此外,假定所有的流量售卖单价均为 1,惩罚成本设定为零。

若直接按照模型中的订单编号、投放时段、节目编号三个维度来定义,将会产生超过 4000 万个 $(264 \times (30 \times 24) \times 215)$ 决策变量及 15.6 万条 $(371 + 446 + 30 \times 24 \times 215)$ 约束条件。在硬件及软件处理能力的双重限制下,如此庞大的线性规划模型无法快速有效求解。但是经过观察,对于同一个订单,在时段维度上的持续时间一般为 3 至 7 天中的部分时段,而非完整的 1 个月,并且在节目维度上也不会同时覆盖全部 215 个节目。剔除每个订单中不涉及的节目及对应播放时段后,留下决策变量约 70 万个,不足原先个数的 2%。同时,由于广告主对节目内容及播放时段有自主选择性,使得部分节目的部分点击量超出所有广告主的需求范围,从而在考虑库存约束条件时

可以剔除这一部分。于是,约束条件剩下约 9.5 万条,仅为原来个数的 60%。经此简化,线性规划模型中的系数矩阵 A 缩减为约 10 万行 70 万列。同时,观察到该矩阵中存在大量的零元素,故而选择使用稀疏矩阵形式对其进行存储和相关运算。

对于本案例的实际运算,硬件设施是普通的台式机 (RAM-12G, CPU-i7-4770),软件为 MATLAB 及 CPLEX,最终在 10 秒之内输出最优解。

图 3 所示为最终对每个订单计算实际投放的流量达到需求数量的比例累积频率分布图。从图 3 中可知,约有五分之一的订单满足率不足 10%,约有 40% 的订单被完全满足。从实践来看,当切实考虑未完成投放的惩罚成本时,这部分满足率特别低的意向订单很可能被拒绝接受,即不会形成正式合同。换言之,本文的投放分配模型可以给企业销售部门在考虑是否接受某意向订单时提供决策支持。

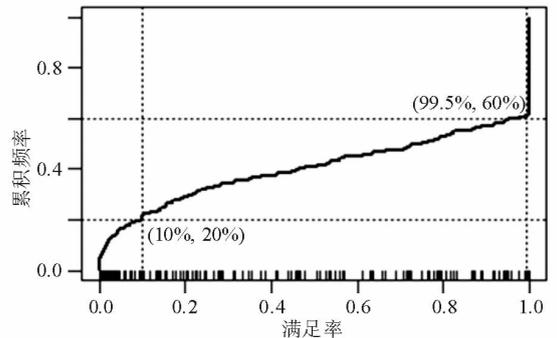


图 3 订单满足率分布图

图 4 则是从投放结果中随机抽取的 8 个订单获得的节目投放分布图,可以看出这 8 个订单一共涉及 24 个节目,其中编号为 5122 的订单只投放了两个节目,而编号为 5120 的订单投放了 8 个节目。总体而言,全部 264 个订单的需求总量约 8200 万,总库存约 1 亿,最终实际投放数量约 5800 万,订单满足率约为 71%。在总库存超过总需求的情形下却没有完全满足所有订单的原因主要有三点:第一,部分节目收视的点击量在时间上较为分散,超出广告主选择的投放时段,造成广告流量闲置;第二,部分节目点击量很高,但其收视观众不被广告主青睐,广告流量难以全部售出变现;第三,部分节目的观众群体受到较多广告主的争夺,但点击量有限造成广告流量严重短缺,导致一些对节目定位范围较窄、出价又缺乏竞争力的广告主未如愿分配到合适流量,最终以低满足率甚至零满足率收场。从实践来看,可考虑结合节目本身内容及实际投放情况,对销售不太理想的节目和时段进行打包,设计成新的产品,并通过适当的营销手段与定价方式,主动寻找并出售

给合适的广告主。

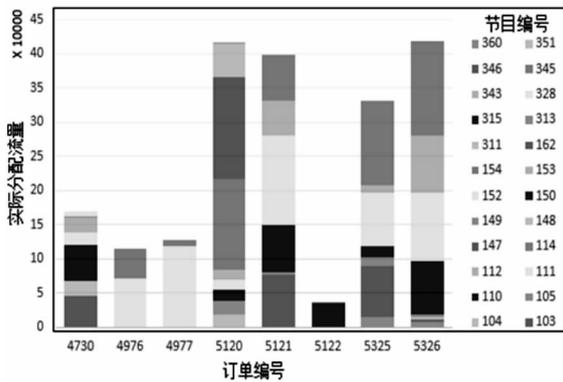


图 4 订单实际投放节目分布(取样)

4 结论建议

本文主要从 IPTV 业务中广告库存售卖的实践问题出发,建立了广告流量分配算法模型,并通过实践案例对模型的可行性加以检验。在求解过程中,针对问题具体特性进行的简化手段,以及后续为了提高运算效率采取稀疏矩阵的处理防水,都在很大程度上提高了模型的实用性。本文的研究结果对于当前更多依赖人工经验进行广告流量分配的 IPTV 行业实践具有一定的借鉴意义。对于未来三个月预售模式,在流量预测的基础上通过本文提供的模型完成分配方案;在分配完毕之后,对于各个节目剩余未能售出的部分流量,考虑其分散化、碎片化的特性,可以考虑重新打包设计成新的产品,继续寻找合适的新广告主,从而充分提高流量变现效率。

后续研究主要有三方面:首先,提高前期流量预测的准确性,这是售卖的基础;其次,将广告流量的具体售价及订单惩罚成本具体化并纳入实证求解过程,尝试研究广告定价策略及违约成本的确定;最后,考虑分配模型对流量预测不确定性的适应性,以及实际投放过程与计划方案出现偏离时如何快速及时地进行动态调整的问题。希望这一系列的后续研究能与实践的联系更加密切,并进一步提高模型的实用性。

参考文献:

- [1] 王梦琳. 三网融合背景下中国 IPTV 版权保护及发展建议[J]. 今传媒, 2013(8):31-33.
- [2] 彭劲, 戴秋华, 何青. 解析 IPTV 及其所面临的关键技术问题[C]. 2005 国际有线电视技术研讨会, 中国杭州, 2005.
- [3] 杨状振. IPTV 广告的传播特点与产业运营[J]. 现代视听, 2008(6):44-46.
- [4] 方乐莺. IPTV 广告盈利新模式[J]. 消费导刊, 2010

(6):17.

- [5] 康丽. 浅析数字媒体时代下的数字电视广告的发展趋势[J]. 电子世界, 2014(16):472-473.
- [6] YAMAGISHI K, HAYASHI T. Parametric packet-layer model for monitoring video quality of IPTV services; Communications, 2008. ICC'08. IEEE International Conference on, 2008[C]. IEEE.
- [7] López V, García-Dorado J L, Hernández J A, et al. Performance comparison of scheduling algorithms for IPTV traffic over polymorphous OBS routers[C]. ICTON Mediterranean Winter Conference, 2007. ICTON-MW 2007.
- [8] HEI X, LIU Y, ROSS K W. IPTV over P2P streaming networks; The mesh-pull approach[J]. Communications Magazine, IEEE, 2008,46(2):86-92.
- [9] YU H, YAN Y, BERGER M S. IPTV traffic management using topology-based hierarchical scheduling in carrier ethernet transport networks[C]. Communications and networking in China, 2009. ChinaCOM 2009. Fourth International Conference on, 2009.
- [10] YU G, WESTHOLM T, KIHLM M, et al. Analysis and characterization of IPTV user behavior [C]. Broadband Multimedia Systems and Broadcasting, 2009. BMSB' 09. IEEE International Symposium on, 2009.
- [11] QIU T, GE Z, LEE S, et al. Modeling user activities in a large IPTV system[C]. Proceedings of the 9th ACM SIGCOMM conference on Internet measurement conference, 2009.
- [12] ABDOLLAHPOURI A, WOLFINGER B E, LAI J, et al. Elaboration and formal description of IPTV user models and their application to IPTV system analysis [C]. Proceedings of MMBnet 2011 Workshop, 2011.
- [13] KODIALAM M, LAKSHMAN T V, MUKHERJEE S, et al. Online scheduling of targeted advertisements for IPTV [C]. INFOCOM, 2010 Proceedings IEEE, 2010.
- [14] KONOW R, TAN W, LOYOLA L, et al. Recommender System for contextual advertising in IPTV scenarios[C]. Computer Supported Cooperative Work in Design (CSCWD), 2010 14th International Conference on, 2010.
- [15] 林宏伟, 邵培基. 基于混合多属性决策的网络广告投放媒体的评价与应用[J]. 管理评论, 2012(12):53-63.
- [16] 彭敦陆, 姚楠, 徐文杰. 基于页面模糊聚类的网络广告投放策略[J]. 华东师范大学学报(自然科学版), 2013(3):54-59.