



IPTV 视频个性化推荐方案

于洪涌, 邱晨旭, 闻剑峰

(中国电信股份有限公司上海研究院, 上海 200122)

摘 要: 分析了 IPTV 视频领域个性化推荐需求, 以“虚拟视频用户”为基础建立了 IPTV 视频用户画像, 构建了“离线批处理数据分析+在线流式推荐引擎”架构的个性化推荐系统, 实现了 IPTV 视频的个性化推荐。验证结果证明该方案是大数据技术在提升用户 IPTV 使用体验方面的有益尝试。

关键词: 个性化推荐; 大数据分析; IPTV

中图分类号: G203

文献标识码: A

doi: 10.11959/j.issn.1000-0801.2017318

IPTV video personalized recommendation system

YU Hongyong, QIU Chenxu, WEN Jianfeng

Shanghai Research Institute of China Telecom Co., Ltd., Shanghai 200122, China

Abstract: The personalized recommendation requirements of IPTV video were analyzed, the IPTV video user portrait with “virtual video users” was structured, a personalized recommendation system was built which included a batch data analysis system and a streaming architecture of personalized recommendation engine. Experimental results show that it is a useful attempt in enhancing the use experience of IPTV users.

Key words: personalized recommendation, big data analysis, IPTV

1 引言

2016 年是 IPTV 跨越式发展的一年, 工业和信息化部发布的数据显示, 2015 年底我国 IPTV 用户数是 4 589.5 万户, 2016 年底增长到 8 673 万户, 2017 年上半年达到 1.03 亿户。相对应我国有线电视数字电视付费用户数 2015 年底是 1.746 5 亿户, 2017 年上半年减少到 1.599 3 亿户^[1]。IPTV 相比有线电视, 具有宽带绑定和互联网视频的优势, 相比互联网视频具有带宽和直播的优势, IPTV 正以其综合优势占领着用户的客厅屏。

用户数是基础, 发展是关键。目前 IPTV 需要在丰富内容、创新业务 (高清、3D 等)、提升用户使用体验上进行加强, 进一步留住用户。现在是一个内容爆炸和突出个性的时代, 面对丰富的内容, 怎么快速找到自己喜欢的节目是用户开机后的第一选择, 于是个性化推荐系统应运而生, 个性化推荐系统能使用户在海量节目中快速定位自己喜欢的节目, 是提升用户体验、黏住用户的重要手段。

完整的个性化推荐系统至少包括数据采集、数据分析处理、得到个性化推荐列表和个性化推



荐列表展现几个部分，推荐内容展现需要和现网运营紧密结合，本文主要探讨如何通过数据采集、分析，得到满足用户需求的个性化推荐列表。

2 需求分析

IPTV 视频个性化推荐的基本需求是：结合用户以往的观看行为 and 用户信息，从节目库中挖掘用户感兴趣的节目，推荐给用户。基于这个基本需求，结合 IPTV 运营经验和用户分析，IPTV 视频个性化推荐系统还需要满足以下层次的需求。

(1) 共性和个性化

人作为社会中的一员，是社会共性和私有个性的统一。在 IPTV 视频领域，社会共性主要体现在：用户对于“好的”节目有一定程度的普遍标准，用户喜欢符合“好的”标准的节目；另一方面用户具有“从众”心理，会选择别人都说好的节目或反映当前热点和时尚的节目。

而个性化体现在：不同的人具有不同的兴趣爱好和需求，体现在观看视频节目上，有喜欢足球比赛或动作电影等的差异；同时，由于用户不同的消费习惯和家庭背景等，其 IPTV 行为也会有差异，如对于收费节目有不同的敏感度，有小孩的父母为小孩准备儿童动漫节目等。

总之，从共性和个性化角度，既要满足用户普遍审美观和从众心理，又要满足不同兴趣和背景下用户的个性化需求，避免马太效应。

(2) 不同场景下的个性化推荐

用户面对不同的页面和节目时，其心理需求也是有差异的。系统应配合不同的场景，提供有差异化的个性化推荐。

开机后面面对首页时，大多用户没有强烈的指向性时，根据用户喜好结合热点节目向用户推荐；用户在观看直播相关节目时，可以推荐相关节目的直播频道；用户观看电影或电视剧时，可推荐相关电影或电视剧的点播和回看等；当用户观看娱乐节目时，可把相似用户的观看热点推荐给用户。

总之，一套让用户满意的个性化推荐系统，应该是一套综合的、随机应变（能根据不同场景为用户提供不同推荐内容）的系统，而不是千篇一律地重复。

(3) 常规推荐和短期兴趣点推荐

在个性化推荐系统中，用户的个性化（或画像）一般是基于对用户一段时期的行为数据的分析得到的，具有一定的稳定性；但用户的实际爱好或行为具有一定的跳跃性，如世界杯期间可能热衷于比赛，同时冷落前期一直看的电视剧；一段时期内对纪实片感兴趣，把娱乐新闻抛在脑后等。

一个“好”的视频个性化推荐系统，应该能敏锐地捕捉用户“当前”的兴趣点，并基于当前兴趣点向用户推荐个性化视频内容。

(4) 直播和点播的推荐

IPTV 节目有直播和点播，直播节目还可以进行时移和回看，IPTV 个性化视频推荐系统能够根据用户观看习惯、当前节目内容，进行直播（包括时移、回看范围的直播节目）和点播节目的合理推荐。

(5) 区分家庭不同用户的推荐

IPTV 在用户模式方面，与优酷等互联网视频一个重要区别是：互联网视频在用户登录后，能明确知道用户是“谁”；IPTV 的登录更多的是合法性检验，IPTV 在家庭内部具有公共性，对于用户没有明确的指向性。

作为一个“好”的个性化推荐系统，应该能够“智能”区别用户是喜欢戏剧的 A、喜欢体育赛事的 B、喜欢动画片的 C，还是 ABC 在一起，并基于不同用户的 IPTV 视频特征推荐“恰如其分”的内容。

(6) 时限不同的需求

不同的场景对推荐列表显示时限的要求不同：在首页或点播大类首页等用户尚未开始观看的场景，应尽可能快地向用户进行推荐；在用户点播或订购时，也应快速向用户推荐相关视频；在用户观

看过程中准备推荐列表的时间要求较低等。

以上是在 IPTV 实际运营中对用户的一些分析,反映了当前阶段的 IPTV 视频个性化推荐需求,本文就如何满足这些个性化推荐需求进行方案设计和实现。

3 系统架构设计

基于以上的需求分析, IPTV 视频个性化推荐系统应该能够:面向 IPTV 家庭用户,基于 IPTV 直播和点播内容,综合考虑当前热点和用户个性化,关注用户当前兴趣点,在不同场景下向用户推荐不同的视频内容,智能化地引导用户快速定位自己“喜欢”的节目,减少节目选择的烦恼。

为此,本文进行了以下系统框架设计,该系统设计采集用户和视频节目相关数据,进行离线和在线相结合的分析,为不同场景提供不同的个性化推荐。本系统设计聚焦于如何综合不同技术实现上述需求,对于具体的算法或技术不做过多讨论。

3.1 基于 DIKW 的流程设计

DIKW (data information knowledge wisdom) 模型^[2],通常被作为知识体系的经典模型,在此作为 IPTV 视频个性化推荐系统设计的理论指导。DIKW 模型将数据、信息、知识、智慧纳入一种金字塔形的层次体系,每一层比下一层都赋予一些特质。原始观察及量度获得了数据;分析数据间的关系获得了信息,在行动上应用信息产生了知识;智慧关心未来,它含有暗示及滞后影响的意味。

基于 DIKW 的系统设计如图 1 所示,各阶段分析如下。

(1) 数据汇聚

本系统需要汇聚的数据主要包括用户基础信息、视频内容信息、用户 IPTV 业务信息、用户 IPTV 观看数据等。数据以满足需求为目的,对于历史数据(如 3 个月以前),可以直接采集已有的汇总数据或标签数据,对于近期数据,需采集详细数据。

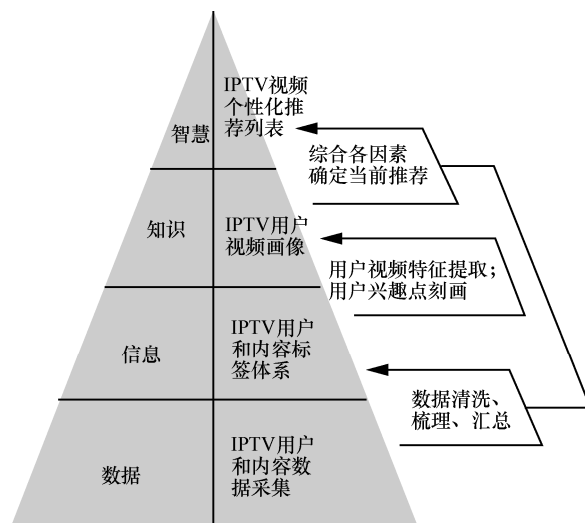


图 1 基于 DIKW 模型的 IPTV 视频个性化系统设计示意

(2) 建立 IPTV 标签体系

对汇聚的数据进行一定的筛选、计算和汇总后形成量化信息体系,并进一步提炼成用户的目标、行为特征或观点的描述,就是标签体系。由于本系统的用户标签主要来自用户 IPTV 视频观看数据,所以首先要建立内容的标签体系。

(3) 刻画 IPTV 用户视频画像

用户画像是结合具体业务,对某一用户或某一用户群体进行的用户核心目标、关键行为模式及主要观点的刻画。对于 IPTV 用户,主要刻画用户什么时间、什么场景喜欢观看什么节目等,并作为向用户进行个性化推荐的主要依据。

(4) IPTV 视频个性化推荐

结合用户画像、用户当前观看行为、当前场景和营销活动等,选择适合该场景的算法,从可选内容库中向用户提供推荐列表。

3.2 总体框架方案

基于上述需求和系统设计,结合目前主流的大数据分析体系,本系统的框架方案如图 2 所示。系统分 3 个层次。

(1) 数据采集

数据主要来自运营商 ODS、大数据平台和 IPTV 业务平台,还有网络爬取数据或第三方数据;采集的数据包括详细数据、汇总数据和标签



数据，大部分数据汇聚到“IPTV 视频个性化推荐数据处理平台”进行数据处理，实时数据传送到“IPTV 视频个性化推荐引擎”进行流式处理。

(2) 数据分析处理

汇聚的数据经过清洗、集成、变换和规约等一系列预处理，保存到数据集中；基于数据集进行分析，形成包括内容标签和用户标签的标签体系；提取用户关键标签形成用户 IPTV 视频画像。并基于用户标签画像进行离线建模分析，得到用户推荐模型，基于推荐模型计算用户、内容的邻居用户，基于各种推荐模型和邻居用户/内容的计算，得到包括不同维度和不同分析方法的多个用户推荐列表。

(3) 个性化推荐

“IPTV 视频个性化推荐引擎”实时获取 IPTV 当前数据，进行以下处理。

- 场景判断：基于用户 IPTV 实时行为判断用户场景，作为后续数据处理和个性化推荐的基础。
- 反馈处理：基于用户的浏览、观看等行为

进行隐式用户反馈分析，作为构建用户动态画像、进行个性化推荐的重要依据。

- 动态画像：基于用户本日或本次 IPTV 开机后的行为和反馈处理，形成用户的动态画像，动态画像体现了用户当前的兴趣点，是进行个性化推荐的直接依据。
- 获取列表：“IPTV 视频个性化推荐引擎”将当前场景信息、时效性、动态画像等发送给“IPTV 视频个性化推荐数据处理平台”，进行用户行为判断，并快速构建个性化推荐列表，返回给“IPTV 视频个性化推荐引擎”。
- 内容筛选：“IPTV 视频个性化推荐引擎”收到个性化推荐列表后，基于场景、动态画像、观看历史和营销策略等进行内容筛选。
- 内容排序：对筛选后的推荐内容，基于场景、用户动态画像、当前热点、营销策略等，对推荐内容进行排序，并将排序后的个性化推荐列表同步给展现系统。

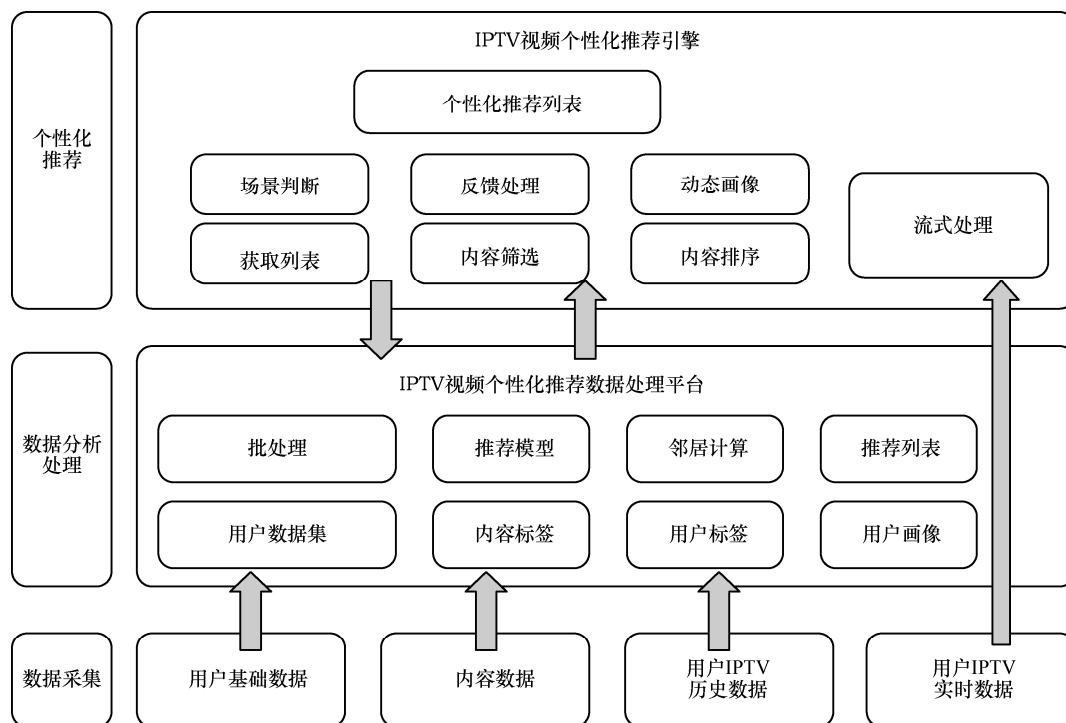


图2 IPTV 视频个性化系统框架示意

4 IPTV 标签体系建立

4.1 数据汇聚

该系统需要汇聚的数据主要包括用户基础信息、内容信息、用户 IPTV 业务信息、用户 IPTV 观看数据等，见表 1。

对数据利用的第一步是形成标签，本系统的标签体系包括内容标签和用户标签，内容标签是建立用户标签、进行个性化推荐的基础。本文所说的标签系统包括内容标签、用户标签及打好标签的内容库和用户库。

4.2 内容标签

本系统的内容标签首先来自内容提供商，然后通过网络爬虫爬取包括优酷、爱奇艺等主流视频网站的数据。本系统内容标签的基本维度主要包括以下几点。

- 内容类型维度：一级大类包括电视剧、电影、综艺、动漫、纪录片、娱乐、体育、财经、教育、少儿、游戏等，然后按影片信息、主演、地区、上映时间、子类型等进行分类，标签分类按扁平化、允许一定冗余的原则进行。
- 播放方式维度：直播和点播（点播包括回看、时移）。
- 清晰度维度：4K、高清、标清。
- 收费维度：免费、收费、促销、积分兑换。
- 推广维度：基于合作、市场预测、营销等角度为内容打上一定的推广标签，便于优先向用户推广。

对于打好标签的内容库，同时进行以下处理，便于快速进行推荐。

- 竞争力排行：对于每个节目或内容，进行内外部竞争力计算，如影片甲有 A、B、C 标签，影片甲的外部竞争力为影片甲在所有具有 A 标签的内容中的排行，影片甲的内部竞争力为其各标签的相对权重。
- 建立内容的邻居列表：对于打好标签的内容计算其相似度，结合用户观看内容的相关性，为每个内容建立基于主要标签的邻居内容列表。

4.3 用户标签

如“需求分析”中所述，IPTV 中的用户无法指向到具体的人。本文中的 IPTV 用户分析以家庭为单位，根据用户使用 IPTV 的习惯构建“虚拟”的视频用户，即通过数据分析得到用户 A 看 IPTV 时 70% 时间看体育比赛，家中的节目订购由其操作，不仅订购体育类节目，还订购了儿童类节目；用户 B 在 21 点以后通过机顶盒 2 看娱乐节目；用户 C 下午看京剧等。

基于上述分析，本文构建了“家庭—虚拟视频用户”的 IPTV 用户结构，其中家庭信息包括家庭地址、用户套餐信息、支付信息、订购退订信息、机顶盒信息、IPTV 观看汇总统计信息等，并进行标签信息选取，形成家庭的标签信息，这可以从总体上把握“这家人”的 IPTV 视频特征及可能的 IPTV 视频动作。

对于“虚拟视频用户”，需要根据用户信息从无到有构建：根据汇聚的家庭 IPTV 数据，采用分

表 1 IPTV 视频个性化推荐系统数据汇总

| 数据类别 | 标签信息 | 画像描述 |
|--------------|--|---------------------------|
| 用户基础信息 | 个人信息、住址信息、用户上网信息、家庭标签、支付标签等 | 数据主要来自运营商 ODS、DPI |
| 内容信息 | 节目直播表、点播内容表及节目信息表 | 主要来自内容提供方，部分来自网络爬取 |
| 用户 IPTV 业务信息 | 用户套餐信息、积分数据、用户 IPTV 订购退订汇总信息及近期订购行为、机顶盒数量及类型 | 数据来自运营商 ODS 系统和 IPTV 业务平台 |
| 用户 IPTV 观看数据 | 用户开机时间、观看节目信息（包括直播和点播）历史汇总数据、近期视频观看行为数据、用户实时 IPTV 行为数据 | 数据来自 IPTV 业务平台 |



类、聚类等数据分析方法，根据用户视频观看行为（包括观看时间、时长、观看内容类型、不同内容的观看相邻性等），为这个家庭构建几个 IPTV 视频行为不同的用户，构建的过程同时也是为用户打标签的过程。这些“虚拟视频用户”可能与实际的用户不是一一对应的，但其 IPTV 特征明显具有差异性。

这些“虚拟视频用户”可以被产生，也可以被改变，甚至消失；也就是说这些用户的特征，要根据不断汇聚的数据进行动态调整：例如一个幼儿特征明显的用户，随着时间推移，喜欢动漫，喜欢游戏，观看教育类节目，变成一个小学生特征明显的人；一个电视剧权重 80% 的人，慢慢体育权重上升到 50% 等。总之，用户标签特征是动态变化的，是需要去“运营”的。

用户被创建的过程也是被打标签的过程，打好标签后对用户的标签计算权重，得到用户的视频兴趣偏好，同时计算用户相似度，得到用户邻居列表。

4.4 用户画像

用户画像是用户信息由定量到定性的蜕变，是标签化的用户模型。本文所说的用户画像是根据 IPTV 标签体系构建的，可看作完整用户画像在 IPTV 视频领域的投影。

对应于标签，IPTV 用户画像也包括家庭画像和用户画像，家庭画像是对家庭标签信息的抽象概括，包括家庭支付能力及支付敏感度、套餐及业务订购特征，从标签到画像的过程见表 2。

构建“虚拟视频用户”画像的目的是为个性化推荐服务的，个性化推荐要兼顾不同场景，对应其画像也有不同形式。

（1）综合型

从较长一段时期用户数据标签分析用户总的偏向爱好，如用户 A 喜欢现代都市类电视剧（权重 30%），喜欢演员 XX 和 YY（相对权重 55%）；喜欢娱乐节目（权重 22%），喜欢选秀类娱乐节目（相对权重 75%）；用户爱好稳定度高（变异率 8%）等。

（2）场景型

从较长一段时期用户数据标签中挖掘用户某个场景的偏向爱好，如用户 B 在周末深夜喜欢看高清大片（常规权重 60%），热衷观看英超比赛权重（即时权重 80%）等。

（3）动态画像

基于用户本日或本次 IPTV 开机后的行为和反馈处理，形成用户的动态画像，动态画像基于实时数据快速构建，体现了用户当前的兴趣点，是进行个性化推荐的直接依据。

总之，本系统构建的标签和画像体系是为个性化推荐服务的，它们基于用户行为和视频内容属性构建，是进行数据分析建模、进一步得到用户推荐列表的基础。

5 个性化推荐机制

IPTV 视频个性化引擎推荐以数据处理平台为基础，针对用户的 IPTV 实时行为，向用户及时提供个性化推荐内容。

表 2 IPTV 家庭用户标签画像对照

| 类别 | 用户标签 | 用户画像 |
|--------------|---|-------------------------|
| 用户家庭财产信息 | 核心区域、三室两厅、学区房、有车、无贷款 | 家庭比较富裕 |
| 套餐信息 | 电信全家享 399 套餐、账单代扣、无欠账、两次教育产品包订购 | 支付能力强，支付敏感度不高，关注儿童教育 |
| 用户 IPTV 业务信息 | 一个 4K 机顶盒，2 个高清机顶盒 | 喜欢高质量节目 |
| 用户 IPTV 观看数据 | 每周开机 12 次，观看 24 h，娱乐、体育、教育占 78%，直播节目占 75% | IPTV 中度用户，喜欢直播，点播节目有待挖掘 |

5.1 混合的推荐机制

个性化推荐系统中,推荐算法是系统的灵魂。从个性化推荐系统发展过程来看,主要的算法包括基于人口统计学的推荐、基于内容的推荐和基于协同过滤的推荐和混合的推荐。

协同过滤(collaborative filtering, CF)^[3]又称社会过滤,是目前发展最为成熟和应用最为广泛的推荐技术。其核心思想就是根据用户对物品的偏好,发现物品本身的相关性或者用户的相关性,然后再基于这些关联性进行推荐。基于关联对象的不同,协同过滤主要分为基于用户的协同过滤(user CF)和基于物品的协同过滤(item CF)。

item CF 基于物品的内在联系进行推荐,往往在某领域中更全面,广度上有所欠缺;而 user CF 基于用户间的关联进行推荐,往往推荐各领域热门的物品,对于各领域的深度推荐上不足;同时研究发现,对相同的数据使用 user CF 和 item CF 得出的推荐结果中,只有 50% 是一样的。这些足以说明 user CF 和 item CF 是互补的关系^[4]。

本系统需要满足不同场景、不同聚焦度和不同时限的需求,以协同过滤推荐机制为主,综合使用分类、聚类等各种数据分析方法,同时采用不同策略,因地制宜,因时制宜,为用户提供适合用户喜好的个性化推荐。不同的策略分析如下。

- 加权策略:分析场景需求,为不同分析方法结果分配不同权重,结合推荐内容自身权重,进行计算确定推荐结果;特别对于内容,可能有多个标签,要进行不同标签的权重计算,这样才能突出内容的主要特点,在进行用户推荐列表选择时,才更容易筛选。
- 分区策略:在不同区域基于不同推荐机制显示不同的推荐内容,如在用户刚开机或判断用户没有聚焦点时,为用户提供当前热点、根据用户最近观看行为的推荐等;

在用户观看完一部电影时,基于该电影和用户电影喜好进行相关电影推荐;总之因地制宜是进行个性化推荐的重要原则,根据用户所在页面和状态进行推荐,才能实现场景上的个性化。

- 分层策略:在用户邻居或内容选取计算前,可采用简单的筛选、分类等方法先进行一次数据处理,减小数据计算范围;或者将一次数据分析结果作为下一次数据分析的输入。分层策略是减少计算量、及时给出推荐结果的重要手段。

5.2 在线处理和离线处理相结合的个性化推荐

个性化推荐系统除了个性化,另一特点就是实时性要求。个性化推荐系统要求系统在用户看到首页或进行内容浏览时,几乎同时向用户进行个性化推荐。同时,实时性要求算法简洁,计算快速,能够在极短时间内得到结果,这需要提前进行大量计算分析,为实时分析推荐提供基础。离线批处理分析与在线实时分析相结合,快速为用户提供符合用户口味的推荐内容。

其中批处理子系统根据汇聚的数据进行标签画像分析、邻居用户分析等,综合采用各种算法进行建模分析,并生成初步的推荐列表;实时处理子系统主要判断用户当前场景,根据用户行为获得反馈信息;基于反馈信息,结合批处理系统的成果,向用户提供个性化推荐列表。具体分析如下。

实时处理子系统根据用户所处的页面判断场景,并根据用户本次开机后行为获得隐式用户反馈信息,包括正向反馈信息:用户在浏览某类视频内容(如电影、娱乐),用户浏览时对于某类或某几类内容进行停留,用户观看/回看/时移某节目等;也包括负向反馈信息:对于某类内容一闪而过,长时间没有浏览某类内容等。并基于用户反馈信息快速构建用户动态视频画像(体现当前的兴趣点)。

批处理子系统接收实时子系统传送的用户信



息,进行用户场景、画像比较处理,若用户行为没有大的变化,基于已有分析成果提供推荐列表;否则进行快速比较,基于已有分析成果和用户画像差异,变更选择范围并进行二次筛选,快速给出个性化推荐列表,返回给实时子系统。个性化推荐列表包括基于内容的协同过滤、基于用户的协同过滤、热点内容和重点营销内容等。

实时子系统收到个性化推荐列表后,基于以下因素进行内容筛选。

- 当前场景:主要考虑内容的聚焦性和时效性,如首页聚焦性弱、时效性高、推荐内容较广泛、提供速度快;播放页面聚焦性强,时效性相对低一些,推荐内容针对性更强,时间可适当放宽。
- 营销策略:结合当前营销活动,向用户提供热销或重点营销的内容。
- 用户动态画像:批处理子系统提供的推荐内容基于用户历史信息分析提供,实时子系统判断用户当前动态画像是否和批处理子系统的用户画像契合,并基于差异性进行调整。
- 有效性:基于内容有效性、历史播放记录等进行筛选。

经过筛选后的内容还要经过排序才能提供给用户。排序时基于用户动态画像和批处理子系统提供内容的权重,进行综合处理。

总之,批处理子系统基于用户历史信息得到用户长期兴趣点并提供推荐内容,实时子系统基于当前情景随机应变,进行适当调整,并作为数据源向批处理子系统提供相关数据。两者紧密结合,功能互补,快速、智能地实现个性化推荐。

5.3 系统实现框架选择

目前以流式计算为主的实时分析系统对个性化推荐系统的实时性进行了很好的支撑,目前主流的流式计算框架包括 Storm、Spark Streaming 以及专门为实时推荐系统设计的 Kiji 等。

相比较而言,Kiji 是专为实时推荐系统设计的架构,针对性强,但目前采用的系统比较少。Storm 和 Spark Streaming 是目前比较流行的流式计算框架,对比之下,Storm 专为流式计算设计,实时性更好;Spark Streaming 用批处理的方式实现实时分析,在实时性上不如 Storm,但 Spark Streaming 继承自 Spark,可以运行上百个节点,并发性支持更好,且可以和 Spark 天然对接^[5]。

IPTV 视频个性化推荐系统的在线部分更注重用户动态画像刻画、用户场景判断,而不是数据流分析,而且 IPTV 视频个性化推荐系统用户量大,对并发性要求高,同时考虑 Spark 在离线处理上的优势,故系统采用 Spark+Spark Streaming 进行数据离线和在线分析。

6 系统试点及展望

本系统在某电信运营商省公司的 IPTV 系统中,针对电影点播业务,选取部分点播电影比较多的用户进行小规模系统试点。从试点结果来看,用户电影观看时间平均增加了 12%左右,业务订购量增加不明显。试点结果分析如下。

- 用户电影观看时间的增加说明个性化推荐系统已发挥了智能化引导的作用,能够为用户找到用户愿意看的影片。
- 业务订购量增加不明显,说明个性化推荐不是万能的,IPTV 业务发展还需要在内容质量提升、业务营销和产品设计等方面全面进行;同时,在 IPTV 点播业务中,很多用户已经购买了月包或年包。
- 个性化推荐系统应该基于全业务、全场景进行统一规划、分析,著名的在线影片租赁商 Netflix 新一代推荐系统的承载形式是“会员首页”,即每个会员登录之后的第一页,平均 2/3 的视频是从此处发起的。

个性化推荐系统起源于电子商务网站,亚马逊是个性化推荐系统成功的早期典范;后续推广

到音视频、新闻推荐、社交等众多领域,都取得了成功。目前在互联网视频领域,个性化推荐是必备工具,据 Netflix 估算,个性化推荐系统每年为它的业务节省的费用可达 10 亿美元^[6]。

然而由传统电视转化而来的 IPTV 业务,在这方面才刚起步,本文提出的 IPTV 视频个性化推荐方案基于 IPTV 运营需求,分析 IPTV 视频和用户及个性化推荐的特性,以“虚拟视频用户”解决 IPTV 中用户指向问题,以离线批处理数据分析+在线流式处理引擎的组合,解决用户画像刻画和准实时的个性化推荐。本方案是综合解决 IPTV 中个性化推荐的有益尝试,也是大数据技术在 IPTV 领域的深入应用。IPTV 以其和宽带结合的先天优势,加上大数据技术的保驾护航,必将在家庭视频领域获得更大发展。

参考文献:

- [1] 卫星参数网. 我国有线电视付费用户跌破了 1.6 亿户[EB/OL]. (2017-08-16)[2017-09-26]. http://www.sohu.com/a/165155976_283658.
Satellite Parameter com. China's cable TV subscribers fell 160 million households[EB/OL]. (2017-08-16)[2017-09-26]. http://www.sohu.com/a/165155976_283658.
- [2] 智库百科. DIKW 模型[EB/OL]. (2016-07-27)[2017-09-20]. <http://wiki.mbalib.com/wiki/DIKW%E4%BD%93%E7%B3%BB>.
MBA Lib. Data-to-information-to-knowledge-to-wisdom model[EB/OL]. (2016-07-27)[2017-09-20]. <http://wiki.mbalib.com/wiki/DIKW%E4%BD%93%E7%B3%BB>.
- [3] 尤海浪, 钱锋, 黄祥为, 等. 基于大数据挖掘构建游戏平台个性化推荐系统的研究与实践[J]. 电信科学, 2014, 30(10): 27-32.

YOU H L, QIAN F, HUANG X W, et al. Research and practice of building a personalized recommendation system for mobile game platform based on big data mining[J]. Telecommunications Science, 2014, 30(10): 27-32.

- [4] 赵晨婷, 马春娥. 探索推荐引擎内部的秘密[EB/OL]. (2011-03-16)[2017-09-22]. https://www.ibm.com/developerworks/cn/web/1103_zhaoct_recommstudy1/index.html?ca=drs-.
ZHAO C T, MA C E. Explore the secrets inside the recommendation engine[EB/OL]. (2011-03-16)[2017-09-22]. https://www.ibm.com/developerworks/cn/web/1103_zhaoct_recommstudy1/index.html?ca=drs-.
- [5] 廖建新. 大数据技术的应用现状与展望[J]. 电信科学, 2015, 31(7): 7-18.
LIAO J X. Big data technology: current applications and prospects[J]. Telecommunications Science, 2015, 31(7): 7-18.
- [6] A5 站长网. Netflix 每年靠它节省 10 亿美元 这套个性化推荐系统是怎么回事? [EB/OL]. [2017-09-20]. http://www.sohu.com/a/108901273_134135.
Admin 5 com. How does Netflix save \$1 billion a year on this personalized recommendation system? [EB/OL]. [2017-09-20]. http://www.sohu.com/a/108901273_134135.

[作者简介]



于洪涌 (1976-), 男, 中国电信股份有限公司上海研究院数据分析师, 主要从事 IPTV 数据分析、智慧家庭数据分析等方面的工作。

邱晨旭 (1973-), 男, 中国电信股份有限公司上海研究院高级工程师, 主要从事电信数据规划、项目管理等方面的工作。

闻剑峰 (1977-), 男, 中国电信股份有限公司上海研究院大数据智慧运营研发 (高级), 主要从事大数据基础架构研发方面的工作。