

阿里文娱 **技术**

阿里云 开发者社区

5G+AI

阿里文娱技术实践

文娱大脑
互动式视频 大数据 5G
优酷 5G 优酷 6DoF 机器智能
智能 5G 优酷 大数据
灯塔 5G 大麦 5G 机器智能
窄带高清 文娱大脑 6DoF 灯塔 大麦 窄带高清
互动式 5G 服务 大数据 机器智能 5G
现场服务 机器智能
互动式 6DoF 优酷 互动式视频
优酷

关注我们



(阿里文娱技术公众号)

关注阿里技术



扫码关注「阿里技术」获取更多资讯

加入交流群



- 1) 添加“文娱技术小助手”微信
 - 2) 注明您的手机号 / 公司 / 职位
 - 3) 小助手会拉您进群
- By 阿里文娱技术品牌

更多电子书



扫码获取更多技术电子书

目录

优酷发布窄带高清 3.0，用户看片流量再省 30%	4
内容全生命周期里的文娱大脑	9
5G 下的 6DoF 视频技术，到底有何不同？	21
大麦物联网技术实践：挑战、解法、趋势	27
大数据赋能电影数智化宣发	37

优酷发布窄带高清 3.0，用户看片流量再省 30%

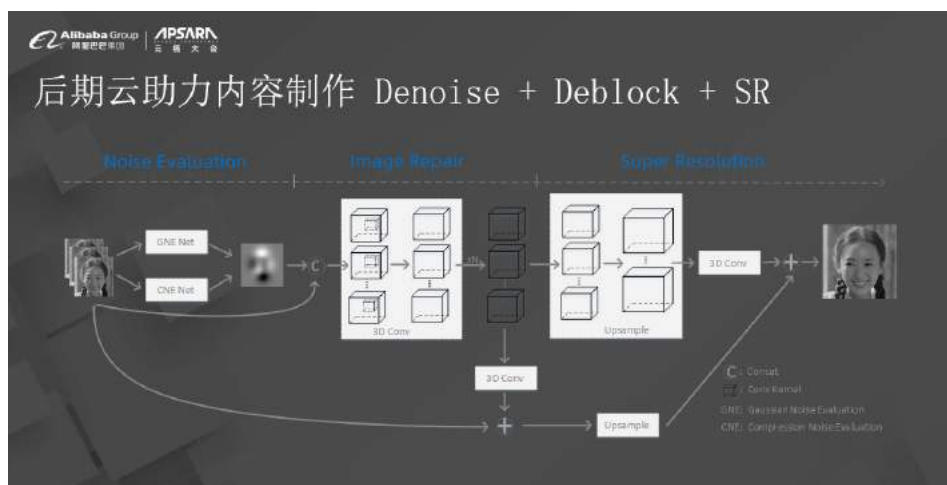
作者：江文斐 阿里巴巴文娱事业群资深算法专家

9月27日，在2019杭州·云栖大会【智慧文娱技术】分论坛上，优酷带来了行业关注已久的窄带高清3.0。在5G大趋势下，优酷加速布局智能高清，意在通过技术创新，让更多用户享受普惠高清服务。

阿里巴巴文娱事业群资深算法专家江文斐表示：除了延续前序的转码技术方案的优点，窄带高清3.0的核心改进在于将播放渲染环节纳入到其技术体系中。窄带高清3.0是一套“基于标准编码格式的私有视频传输播放协议”，即窄带高清3.0的视频流完全遵循国内外视频编解码标准，在此基础上增加一套加密的私有信令来指导播放器渲染出最佳的画面。这套云+端的播放协议，通过对全链路的通盘优化，再一次突破视频传输效率瓶颈，再一次降低超高清观影门槛。

从1.0到3.0，窄带高清的技术演进路线

早在2015年，阿里巴巴就提出窄带高清概念。这个概念一经提出，就被行业及视频厂商引用，并且类似服务开始涌现。本质上，窄带高清代表的是成本与体验相调和的视频服务理念。在体验方面，观众在观看视频时会受到片源清晰度、分辨率、码率、帧率、编码算法、播放策略等影响，所谓的窄带高清就是希望从这些角度提升观看体验；而影响成本的因素包括码率、分发策略和播放策略，从1.0到2.0，窄带高清一直从这三个层面出发，实现降本提体验。今日发布的窄带高清3.0，除了继续在2.0基础上再节省约30%的宽带码率。更重要的是，打破了终端设备和内容平台之间的数据转码的信息屏障，让超高清观影惠及更多的普通用户。



窄带高清不仅仅是节省码率。1.0 是实现最佳分辨率——码率的匹配关系，平均节省 30% 带宽，在用户低带宽条件下也可看高清。2.0 实现云端渲染，迎合人眼主观视觉，让 720P 达到 1080P 观感。紧接着，优酷技术团队花了整整 2 年来突破 2.0 的效率，从云 + 端联动切入，窄带高清 3.0 再一次降低约 30% 的码率。在即将爆发 5G 时代，让使用移动网络、使用千元机的用户同样享受超高清视频服务，实现真正的普惠。

窄带高清 3.0 的两大特征

优酷的窄带高清 3.0 具有两大鲜明特征：一是完全遵循通用视频编解码标准（例如 H.264 / HEVC），标准码流是建设良好生态的基础，更具普适性。二是从云到端的私有通信链路。除视频流外，一些根据终端性能定制的 side information 通过 SEI 等标准封装格式传输，根据预设协议在播放端做视频渲染。综合看，窄带高清 3.0 是一套完全符合通用标准的私有视频传输播放协议，可直接与现有的协议和设备接入，输出给行业的上下游。

一般而言，普通流媒体分为 3 个环节，原始介质、码流传输、终端呈现。在云端将视频压缩并传输，在播放侧解码并渲染到屏幕显示。



窄带高清 3.0 在对视频进行编码压缩的同时，通过在云端对终端设备的渲染特征进行数据化建模，对视频的各个场景进行 AI 智能分析，让每一帧画面、每一台设备都能获得最合适的渲染处理，达到最佳的观影效果。这种云端基于 AI 内容分析生成渲染信令，客户端依照信令进行渲染的机制，既享受了深度网络带来的收益，又免除了客户端实时分析的计算压力和电量消耗。

目前，窄带高清 3.0 已应用于优酷部分视频节目中，以下是对比画面案例。





端到端的高清技术

除了窄带高清，老片修复和超高清重制也作为优酷的超高清战略融入平台全链路中。在内容制作环节，对标清老片采用工业化 + 算法修复的方式批量高清化。对高清新片采用深度学习算法实现超高清化（4K / 60 帧 / HDR / 环绕音效）。在生产环节，有了窄带高清转码的加持，让广大用户得以用最少的流量看高清。在播放环节，通过

云 + 端联动渲染，让普通设备看超高清成为可能。

从业务模式出发，优酷更聚焦窄带高清的软件解决方案。从行业视角看，5G 时代下的窄带高清云 + 端联动趋势亦然来到。所以，优酷正在积极与硬件终端厂商合作，将软 + 硬件能力打通，低成本、低能耗地实现各种视频的高清化，也希望更多上下游合作伙伴一起，共同引领超高清的发展。

同时，在智慧文娱技术分论坛，优酷同步发布优酷工业级超分辨率及增强视频数据集，这也是目前业界最大高清版权多类别数据集，将应用实践反哺科研学术，并再应用于实践，助力行业人才的培养。

内容全生命周期里的文娱大脑

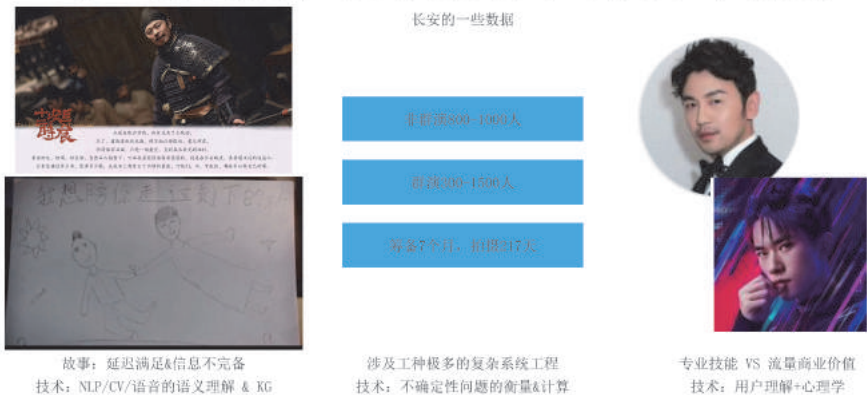
作者：蔡龙军 阿里巴巴文娱事业群资深算法专家

文娱内容不同于其他商品，很难有完整的量化指标体系，而内容的复杂性决定了文娱产品的不确定性。我们如何利用 AI+ 大数据能力，建造文娱的内容认知大脑？串联内容全生命周期，实现对内容、流量、宣推等的“未卜先知”，提升对爆款的预测和生产能力？

一、文娱产业趋势及技术挑战



行业技术挑战：内容的复杂性决定了不确定性



文娱内容不像商品有完整的量化指标体系，它是一个复杂的实体，它跟意识形态以及用户体验强相关，对内容进行量化评估和衡量是非常困难的。

比如，选角儿。我们不能通过单一指标去衡量一个演员，我们需要综合考量演员的演技、气质、颜值、潜力等与是否与某一个角色匹配，并且能生成数据指标，以实现纵横向的对比。另外，导演、主演组盘是否为最优组合，能否成为爆款？这是更加复

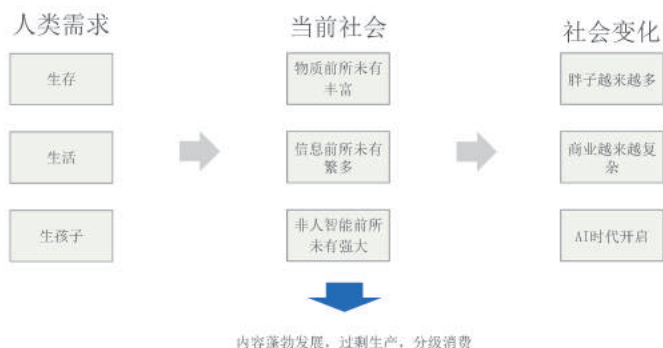
杂的选择模式问题。今天面临的技术挑战是如何进行知识的抽取、挖掘以及推理，确定什么样的组合是最优解。

除上述两个问题，影片的拍摄过程更是一个庞大的系统工程和艺术创作过程。以《长安十二时辰》为例，该片非群演有约 1000 人，群演有 300 到 1500 人，历时 7 个月拍摄 217 天。我们参考软件工程行业，软件工程发展了 70 年，主要研究三个层面：方法论、过程以及工具，然后是如何将三者组合。软件行业的敏捷开发对于软件工程的质量和效率都有非常大的提升，如何将这些理论应用到内容制作产业，让内容制作敏捷起来？

内容敏捷即知晓过程对结果造成的影响是什么，并快速地调整内容创作过程，让它更敏捷。但内容行业面临的独有特点“延迟满足”，让用户在内容的某一分钟特别嗨，可能来自于前面的 30 分钟铺垫在那一分钟爆发了，针对内容的这个特点，我们除了要做基本的知识图谱语义的理解之外，还要考虑如何做有效的对应分析，如何做对应的知识抽取等问题。



行业趋势及挑战：多，从商业驱动到消费驱动

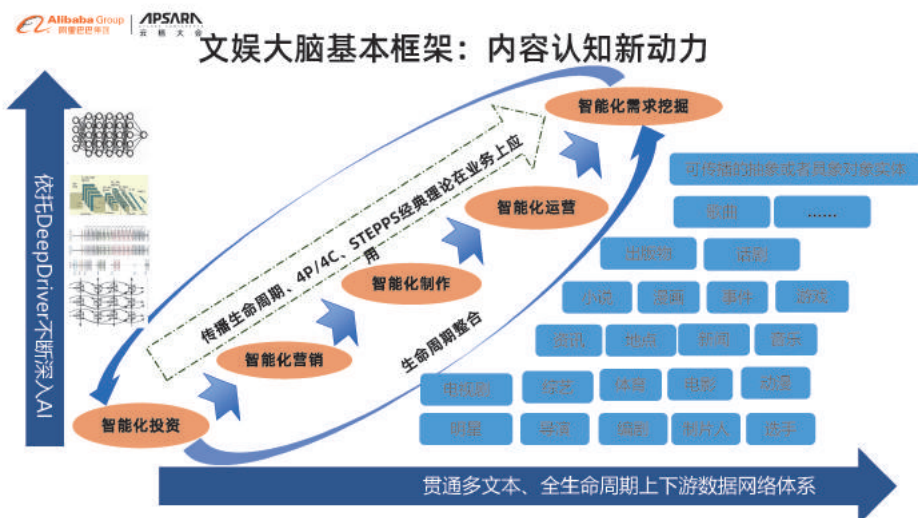


行业趋势：分层、分级消费加剧挑战



今天这个问题加剧了，比过去还要复杂。在过去的 5 到 10 年里，UPGC 加上整个内容的生产量极大的发展，用户的消费分层化、多样化。全民爆款越来越少，用户对内容的需求更加个性化。相应于内容生产端，就需要考虑不同用户群的个性化需求。

二、文娱大脑基本框架：内容认知新动力



针对上面几大困难，我们今天在做文娱大脑——优酷北斗星智库来解决。我们将所有的内容形式和用户消费的数据都采集下来，将人工智能的技术手段、业务领域的细分理论做整合融合，构建内容认知框架。



内容认知框架 (Cognitive Framework)



内容认知框架分为两部分，内容和用户。其思路就是心理学发展的基本的思路。

- 1) 内容侧：对内容进行理解，包括外延和内涵。外延就是内容的各种基本属性，比如主创阵容、题材类型等；内涵主要研究内容的戏剧理论和视听语言，围绕制作内容的支撑要素，我们用传统的机器学习方式对内容进行理解，再基于戏剧理论和视听语言构造内容的衡量要素。
- 2) 用户侧：分析用户的观看行为。用户行为来自于用户的心理偏好、心理情绪。用户心理偏好、心理情绪来自于生理构造，基于心理学的五大人格理论和用户的观看行为，构建模型建立左边和右边的连接，从而知道创造什么样的内容，用户会有什么样的感受。

三、贯穿全生命周期的文娱大脑生产力

Alibaba Group 阿里巴巴集团 | APSARA 数据大脑

贯穿全生命周期的文娱大脑生产力

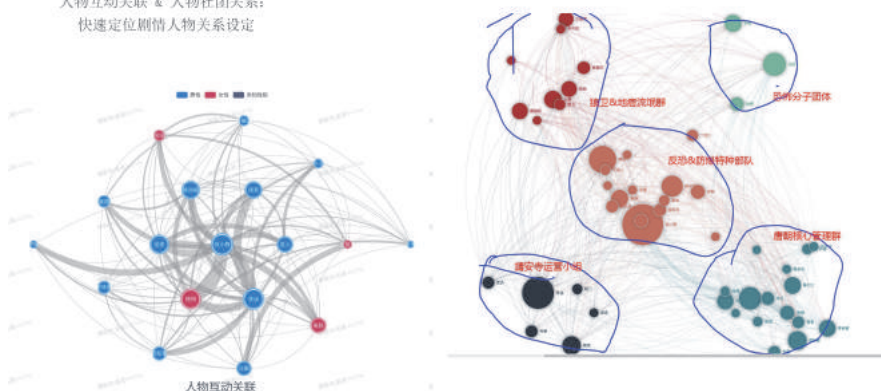


基于内容认知框架，我们在内容生命周期的每个阶段都做了具体工作：开播前提供内容评估、艺人挖掘和内容情绪挖掘等能力；在早期为内容评估提供有效的数据支撑；在制作阶段提供现场解决方案，比之前更敏捷的反馈机制；同样在播出后也提供数据支持，实现更好的宣发。

Alibaba Group 阿里巴巴集团 | APSARA 数据大脑

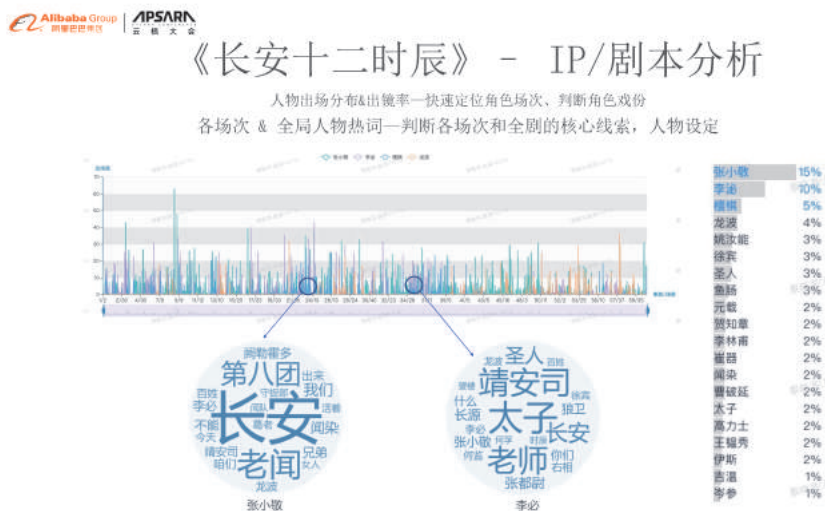
《长安十二时辰》- IP/剧本分析

人物互动关联 & 人物社群关系：
快速定位剧情人物关系设定



1. IP/ 剧本分析

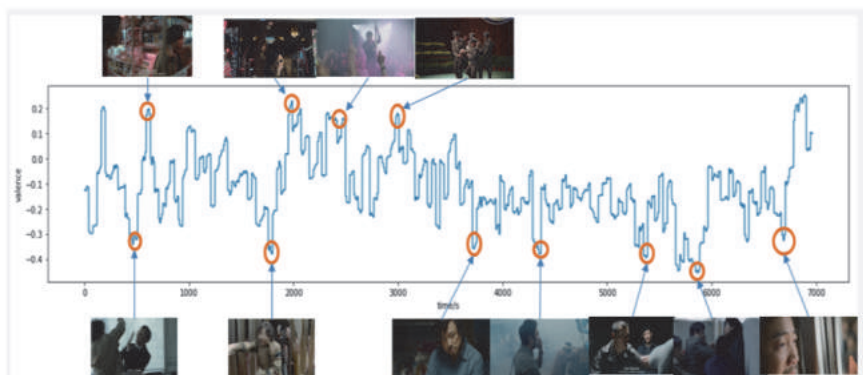
上图是《长安十二时辰》的分析示例，我们把已有的剧本作为样本，让机器去学习，识别出剧本的所有角色，把角色直接交互的对白、行为识别出来，再进行社团的划分。《长安》剧本最终划分出来几个群体：反恐防暴小分队以张小敬为中心，唐朝核心管理团队以皇上为中心。通过这种方式快速定位整个剧本的人物和人物关系的展开。



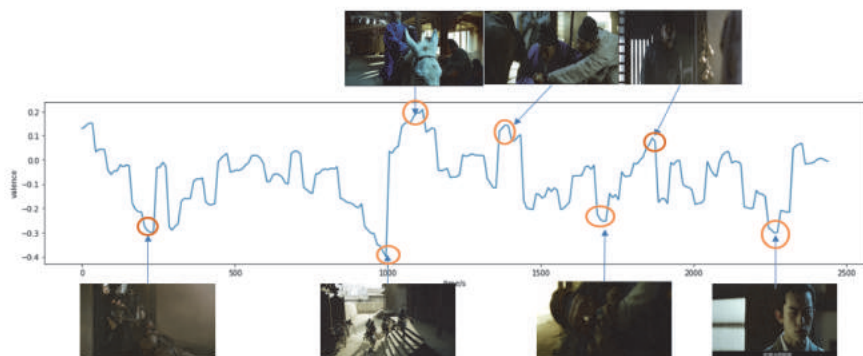
2. 用户情绪识别与成片情绪挖掘

围绕角色关系，将整个剧本的角色情绪也识别出来，构造成如上的曲线。基于对海量剧本的分析曲线，抽取出各个指标（出镜率、戏份、情绪值等）并形成 benchmark，对于之后的每一个剧本进行衡量，相当于对剧本进行一个“体检”。

《药神》用户情绪VA



《长安十二时辰》成片情绪挖掘

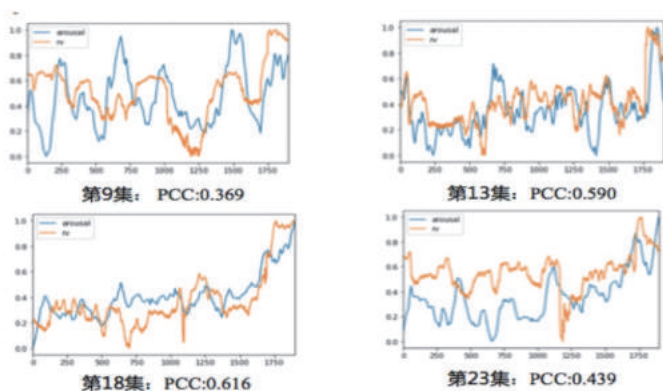


同样是“体检”的方法，对于《药神》和《长安十二时辰》，我们做了用户情绪的识别、体检的扫描，参考零线的位置。我们发现《药神》几乎都是正向和负向级的，直到最后出现一个正向区间，基本上后期都是以眼泪为主。而《长安十二时辰》的情绪状态比较稳定。对照情绪高低点的具体情节，我们发现，曲线表达的情绪和具体的故事情节是非常相符的。

3. 情绪强度预测与网络收视率

Alibaba Group | APSARA
阿里巴巴集团 | 云栖大会

《长安十二时辰》情绪强度预测 VS 网络收视率

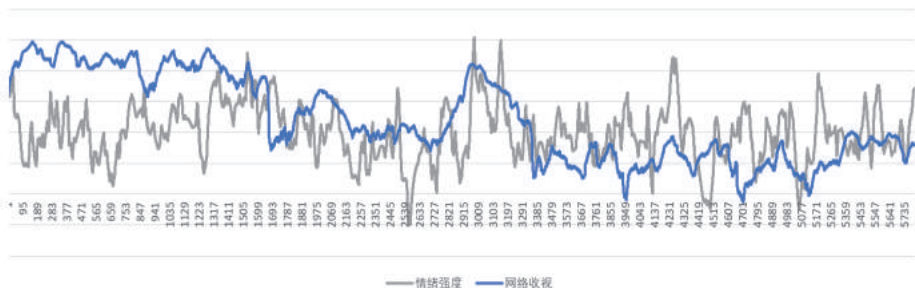


然后我们拿更多的方式去验证它的合理性，上图抽取《长安十二时辰》的剧集，每集有两条曲线，蓝线是刚才预测的情绪曲线，黄线是播放指数（表示每一秒钟有多少用户在观看），通过两条曲线对比，我们可以发现，两条曲线的相关性比较高的将近60%，情绪的高峰、低谷和用户的观看行为状态是吻合的，由此我们就提供了一种能力，基于这种能力对剧本或影片做情绪扫描，实现对影片热度的未播先知，再对比benchmark，帮助制作者更高效的完成制作。

Alibaba Group | APSARA
阿里巴巴集团 | 云栖大会

用户情绪强度预测 VS 网络收视率：综艺/电影

《一起乐队吧》用户情绪强度预测 VS 网络收视



4. 用户情感曲线在技术上是如何实现的？

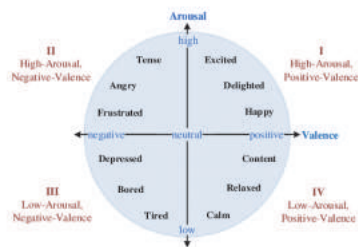


基于VA的情感模型

业内通用的情感模型：

Valence: 情绪正负向。-1 到 +1 之间，-1 表示负向情感，如悲伤，+1 表示正向情感，如高兴

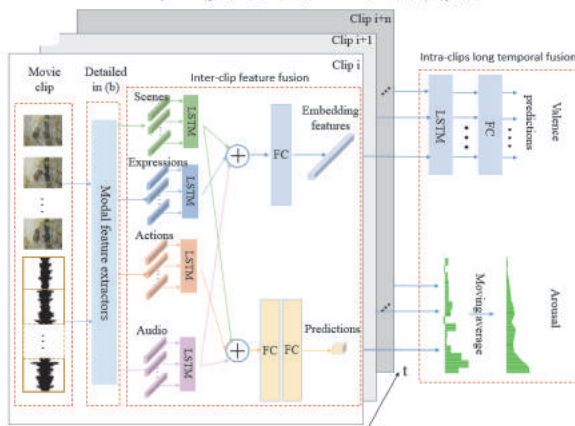
Arousal: 情绪的强烈程度。-1 表示情绪最不强烈，如困乏平静，+1 表示轻度最强烈，如激动兴奋。



首先，我们把用户观影情绪的表述，映射到认知计算中常用的二维空间表示，也就是 Valence 和 Arousal。Valence 表示情绪正负极性，Arousal 表示情感激烈程度。

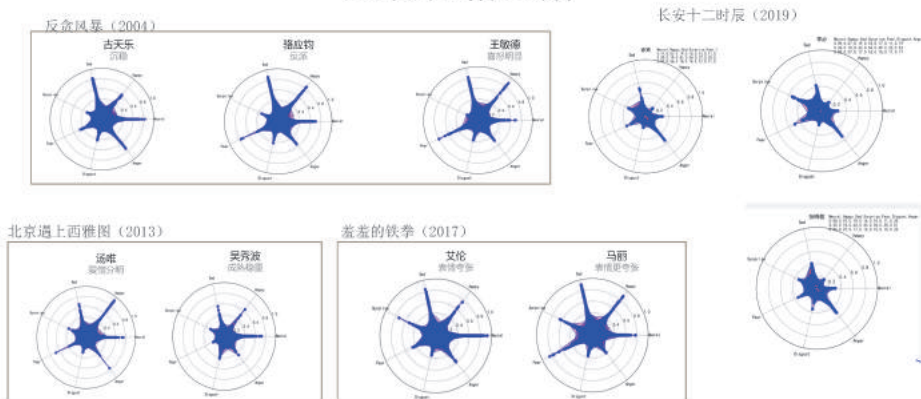


多模态的VA识别模型



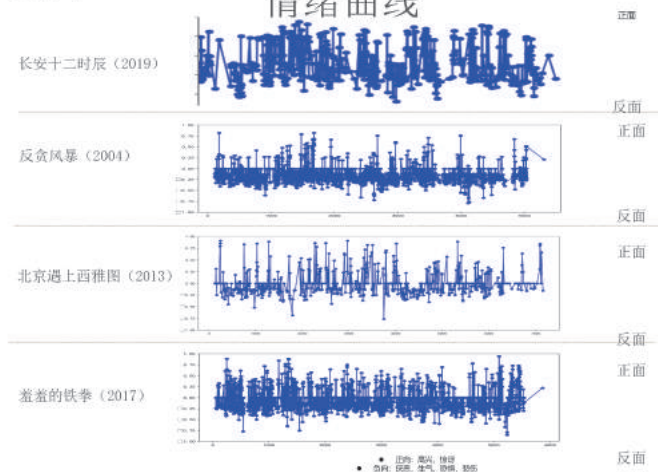
其次，基于情绪极性跟强度提供一个预测，这个是我们今年产出的论文，（论文地址：<https://arxiv.org/abs/1909.01763>）。近两年，心理学研究的核心观点是为什么用户会感同身受？这来自于前两年的一个理论——静向神经元，所以我们选择场景、表情、动作以及声音作为基本的模型的输入，对模型参数进行学习。

人物性格理解



如上所讲，内容产业有强延迟满足的问题，我们通过两层分析来解决长短期满足的问题，除用户情绪分析，我们也做内容角色的情绪识别。通过图片表情识别模型，识别不同题材类型的影片，可以获得不同角色刻画的人物性格。如 2004 年的《反贪风暴》，时隔十多年，主创人物形象的脸谱还是正向的。上图显示的负面角色情绪以开心、害怕为主，正面形象以悲伤、生气为主，与负面反派的开心正好相对，正面的人一直很沮丧，是一个有些压抑角色形象。

情绪曲线

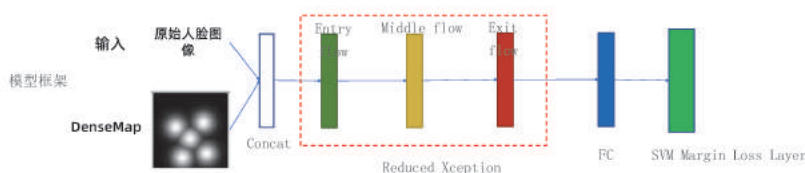


同样，我们分析角色的每秒情绪，形成角色的正负情绪曲线，部分影片的分析结果曲线如上图，不同题材类型的节目会有不同的情绪密度。所以，你想放松的时候，要看的不一定是喜剧，喜剧其实不一定会放松，因为角色的正负向情绪不停交替，由于延迟满足，大脑负荷非常大，需要做长短记忆，反而很多爱情片对大脑的占用相对低。



情感识别：图片表情识别

改进模型 (Reduced Xception with Margin Loss)



输入：引入人脸关键点densemap

原理：精确判断人脸表情需重点关注五官如眼睛，鼻子，嘴的区域信息

关键点检测模型 (MTCNN) Densemap 计算

$$I(i, j) = \max_{k=1}^K e^{-\frac{(i-x_k)^2 + (j-y_k)^2}{2\sigma^2}}$$

$I(i, j)$ 为 densemap 在像素 i, j 处的强度， x_k, y_k 关键点 k 的坐标， $k \in \{1, 2, \dots, K\}$



情感识别：图片表情识别

模型提升 (Reduced Xception with Margin Loss)

Reduced Xception*

使用可分离卷积 (deepwise 卷积 + pointwise 卷积)

Entry flow, middle flow, exit flow 各缩减至2层卷积

$$\min_w \frac{1}{2} w^T w + C \sum_{n=1}^N \max(1 - w^T x_n t_n, 0)$$

$x_n \in R^D$ 表示上一层输出， $t_n \in [-1, 1]$ 为分类标签， w 表示本层参数

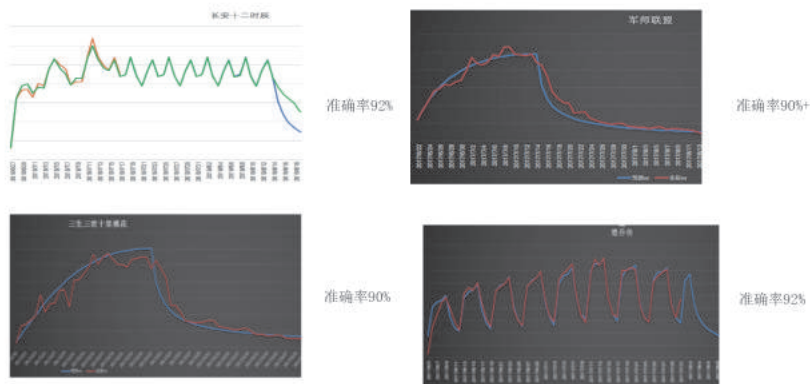
$$\begin{aligned} & \min_{w, \xi_n} \frac{1}{2} w^T w + C \sum_{n=1}^N \xi_n, \\ & \text{s.t. } w^T x_n t_n \geq 1 - \xi_n \quad \forall n \\ & \xi_n \geq 0 \quad \forall n \end{aligned}$$

*Chollet F. Xception: Deep Learning with Depthwise Separable Convolutions[J]. 2016:1800-1807
** Yichuan Tang, Deep Learning using Linear Support Vector Machines, ICML 2013

角色情绪检测是一个分类问题，所以利用人脸 landmark 对初始图像做识别，生成 densemap 作为附加通道，和原始图片 RGB 三通道拼接合并后作为模型输入，这样可以使 densemap 对应的关键区域权重更大，更容易让模型捕捉关键区域特征；合成的输入送入到 Reduced Xception 网络进行特征提取；在 loss 方面，我们引入了基于 SVM 的 marge loss，提升各情绪类别的类间差距，提升情绪识别的效果，具体如上图。

Alibaba Group 阿里巴巴集团 | APSARA 文娱大脑

采制阶段：预测能力建设



基于前面对内容的各种理解产生的各种纬度的内容的量化纬度，我们构建了预测模型，可以提前预测出节目的流量走势，如内容认知框架中所讲的，首先对内容进行量化，然后对内容相应的量化纬度进行提前的预测，为业务决策提供辅助支撑。

最后，分享我对未来趋势的一些见解。在强人工智能尚遥远的情形下，如何结合机器 AI 和人工经验将是个永恒主题。一是结合符号学派智能和链接学派智能，建设和完善决策引擎，包括结合人工逻辑规则和可学习数据 AI，不确定性分析框架和经久不衰的贝叶斯因果决策，以及神经元化的混合智能计算框架。二是量化的心理学研究也越来越重要，如何结合大数据应用价值非常大。这也是阿里文娱大脑探索的方向。

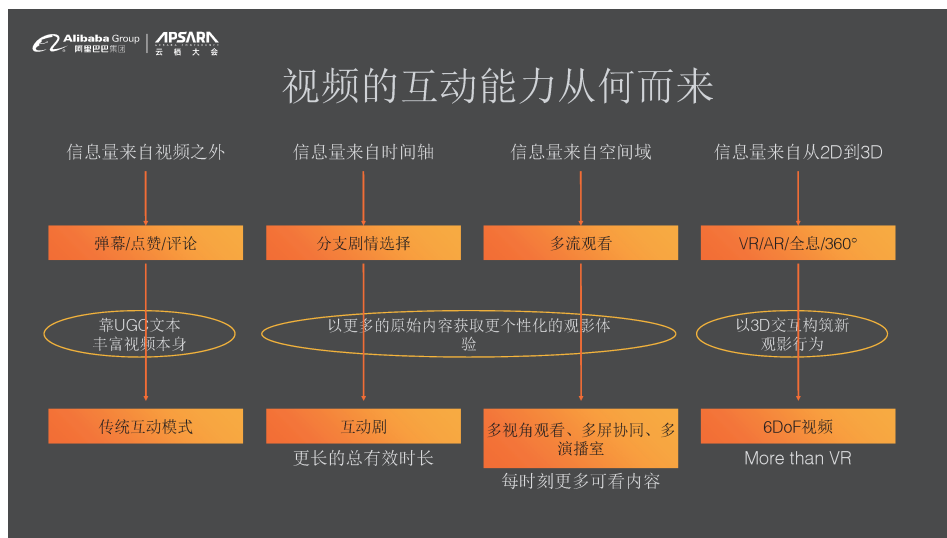
5G 下的 6DoF 视频技术，到底有何不同？

作者：张天若 阿里巴巴大文娱事业群高级算法专家

5G 时代，视频内容形态将迎来三个趋势：一是依托 5G 的大带宽，超高清视频将真正落地；二是沉浸式视频迎来发展，但如同《头号玩家》所描述的，沉浸式视频体验是人类感官消费的最终形态，还有很长的路要走；三是交互式视频，包括 3D 视频、AR 形态视频、游戏化视频，以及借由 AI 赋能的新视频消费方式。优酷在这些领域都在布局，希望借助技术创新，引领用户观看体验的提升。

在 2019 杭州云栖大会【智慧文娱技术】分论坛上，阿里文娱高级算法专家张天若重点分享 5G 下，交互式视频中的 3D 视频技术及应用趋势。以下为分享精选。

一、视频的交互能力从何而来？四种典型方式



交互式视频与单向接收式视频的最大区别在于，交互要有额外的信息反馈，目前四种交互能力：一是最传统的方式，如评论、点赞等，依靠视频之外的内容提供

交互的可能性；二是信息量来自时间轴。视频形态以互动剧为主，让用户选择情节走势，在时间轴上创造更多信息量；三是信息量来自空间域。同一时刻在空间上为用户提供多样化的观看选择。如《这！就是街舞》中的多视角直拍，世界杯中不同演播室和不同风格的解说，都是让用户在同一时刻有更多可看内容。

四是信息量来自从 2D 到 3D 的视频形态变化。当视频形态从 2D 变成 3D 时，视频内容的整体组织形式就会有更丰富的结构化信息。基于立体视觉的 6DoF 视频就是方式之一。

二、交互式视频形态：准 6DoF 视频，子弹时间

首先介绍下 6DoF，6 degrees of freedom，其概念来自于物体在空间中运动时有六个不同的自由度——前后、左右、上下三个平移自由度，以及三个旋转的自由度，对应的就是点头、摇头、歪头。从自由度视角重新审视 VR，会明显发现，基于固定的 VR 相机拍摄的 VR 视频只有三个自由度：可转头但不能移动，代入感非常差。

而 6DoF 视频，是一种更自由的视频形态。首先“脚可以动”，用户可以虚拟和移动观看位置；其次“手可以动”，通过一定的手势操作来影响视频内容画面本身。优酷目前正在做的 6DoF 视频，通过相机阵列来拍更大范围，获得更多的观看自由度。

其实，类似的 6DoF 技术已有很长历史，最经典的是《黑客帝国》中男主角 NEO 仰身躲子弹的慢动作镜头，当时依靠一整排照相机拍摄，将每个相机拍好的图片叠加在一起，再生成视频，在上个世纪看起来非常具有视觉冲击力。

20 年后，我们依然在做类似的视觉呈现，技术上有何不同？其变化主要源自三方面，1) 从电影特效到工业化的制作，不在使用原始拍摄图片，拍摄成本降低，展示自由度提升。2) 自由的特效创作，更大的发挥空间。3) 使用摄像机，支持视频形态。

今年，优酷将 6DoF 视频技术首次应用到国内体育赛事（如 CUBA）直播中，我们将投篮画面做定格，将运动员相对的位置关系、动作，通过多视角呈现，带来现场感的观看体验。

三、交互式视频形态：6DoF 视频

严格意义上，子弹时间是 2D 视频，更侧重后期制作。真正的 6DoF 视频形态是在观看过程中，用户可以通过手指拖拽操作，选择观看角度和位置。虽然摄像机是线状的，但可以不依赖原始摄像机位置，实现上下和前后位置的调整，包括近景的人物特写、远景的全景画面。

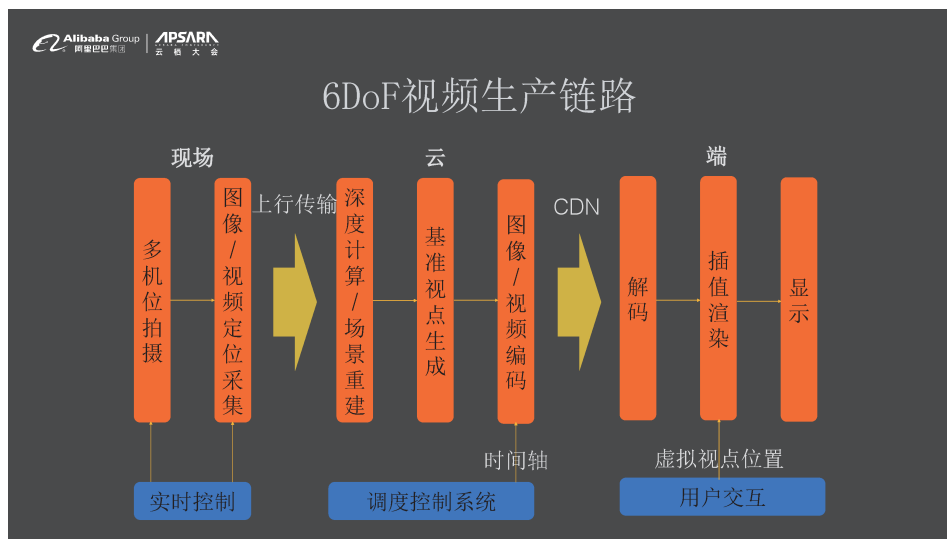
很多人会联想到 VR 视频，但 6DoF 和 VR 视频有明显区别。比如，在一场体育赛事中，VR 视频可以向前看球员、向左右看场地，向后看观众，听起来视觉上更自由，但实际画面是散点的，并不跟随“踢球”这一主线来组织画面内容，用户看到的越多，越难以抓住视觉重点。而 6DoF 视频，摄像机阵列都在拍摄画面中的同一个位置，呈现的永远是兴趣度最高、最有趣的内容，用户只是选择观看角度和方式，实际上比传统 VR 视频有更好的视觉体验。



上图是《这！就是灌篮》、双 11 天猫潮流盛典等的场景，无论俯拍还是仰拍，无论相机安装是弧形还是直线，都可以实现。

四、交互式视频的技术链路

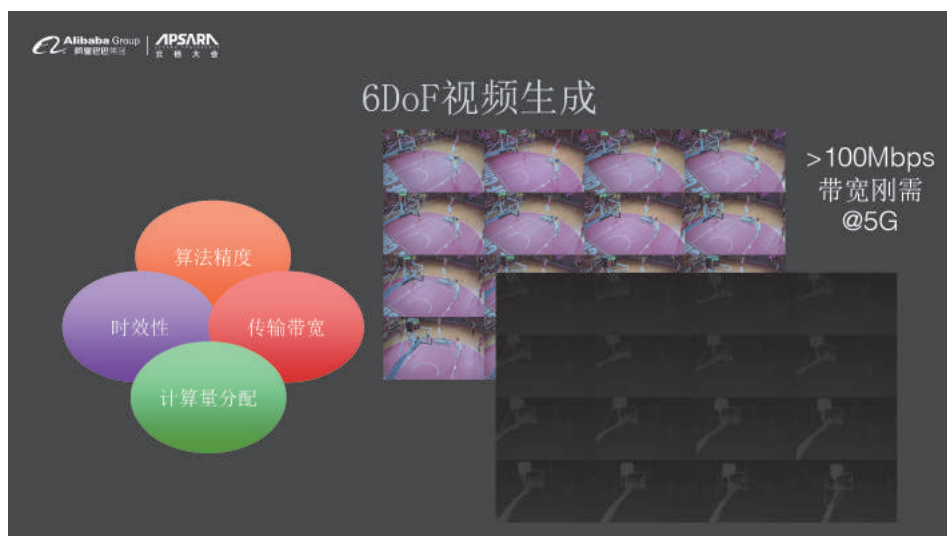
交互式视频的技术链路，以 6DoF 视频为例，其拍摄跟普通视频拍摄的链路是一样的，包括三个环节：



第一，现场拍摄环节。多机位同步控制与拍摄，将采集好的视频和图像上传到云端做处理。

第二，云端处理环节。在云上计算画面中物体的景深关系，生成基准视角的视频图像载体，并将视频图片编码，通过 CDN 传到用户手机上。

第三，端侧播放环节。在用户端做解码，根据用户的交互操作，计算视点位置，对虚拟视点做渲染，最后显示到屏幕上。传输视频格式是通用的 H.264/H.265，播放设备本身具备解码能力，保证覆盖几乎全部机型。端到端的实时画面生成效率，同时兼顾手感，从技术角度来就是低延迟。



技术链路中的关键点，是如何生成 6DoF 视频的处理环节。

首先，传输到用户端的画面，由色彩图和深度图构成。深度图用灰度来表示物体之间的距离，其复杂性在于如何实时、准确地推算出深度并生成基准视频。深度评估与影像生成之间是反比关系，需要技术权衡与取舍。

其次，传输带宽也是影响因素之一。深度图和色彩图被压缩后会模糊，基于模糊的图像所进行的新视角生成会把瑕疵放大。目前，从优酷的实践看，100Mbps 以上的带宽，在 4G 移动网络下是很难实现的，必须依赖 5G 视频技术。

6DoF 视频节省宽带传输的第二个解决方案类似云游戏，不需要将拼接视频传到用户端，只需将用户视角的信息传到云端。云端按照用户希望看到的画面做处理和渲染，生成 2D 视频，这样所占用的带宽就很小了。但为了观看流畅，整体链路的延迟需要非常低，这部分又依赖 5G 的低延迟特性。

五、交互式视频应用前景

从行业角度出发，视频标准化更倾向于从传统的视频压缩到 3D 相关的视频处理。可以看到 MPEG 的 Roadmap 中，从 6DoF 到光场到点云的压缩处理，都是围

绕 3D 和立体视频的技术进行的。优酷在国内的视频标准 AVS 部分，从 VR 标准就开始深度参与。目前我们与北大一起制定基于深度的 6DoF 视频标准，预计今年会公布初步标准。

综上所述，6DoF 视频、3D 建模视频以及 VR 视频，在业务模式、技术路径和应用场景方面都有区别。6DoF 视频的特点是自由视角的播放以及事后的内容组织重建能力，它在记录真实世界、在以体育比赛和舞台表演为核心的内容上有很明显优势。落到优酷的内容场景上，适用于信息密度大、需要有临场感的内容，所以演出、团体组合、体育和舞蹈教学类视频，6DoF 视频都可以发挥作用。

另外，在交互式视频的趋势下，视频和游戏的边界正在模糊。传统游戏最核心的特点，包括情节代入感强、用户可通过操作影响内容走势，都在逐渐纳入互动视频的制作中。另一方面，传统视频的特质，包括非常有艺术感的剧本和情节，真实画面和真人演出，在游戏中也越来越常见。二者之间的融合源于两方面：一是随着算力的大幅提升，提供了将真实世界数据化的能力，其核心是将画面进行结构化的提取，通过 3D 建模去理解场景内容。二是连接能力的提升，将用户和生产者的距离拉近，在消费内容时的交互能力越来越高。

随着 5G 带来的连接能力和云带来的算力的双重提升，交互式视频将成为视频内容革新浪潮中那一朵最闪亮的浪花。

大麦物联网技术实践：挑战、解法、趋势

作者：楚佩斯 阿里巴巴文娱事业群高级无线开发专家

这是一个万物智联的时代，是 AI 技术和物联网相结合的时代，今天我主要介绍 IoT 技术在大麦现场服务部分的应用。

大麦现场服务平台，分为设备端和服务端两部分。端上设备的一大特点就是碎片化非常严重，因此如何标准化是一大挑战。另外端基础能力，如物联网及安全等技术能力的建设也是重点，除此之外，5G 时代的来临，未来的现场娱乐设备也值得期待。

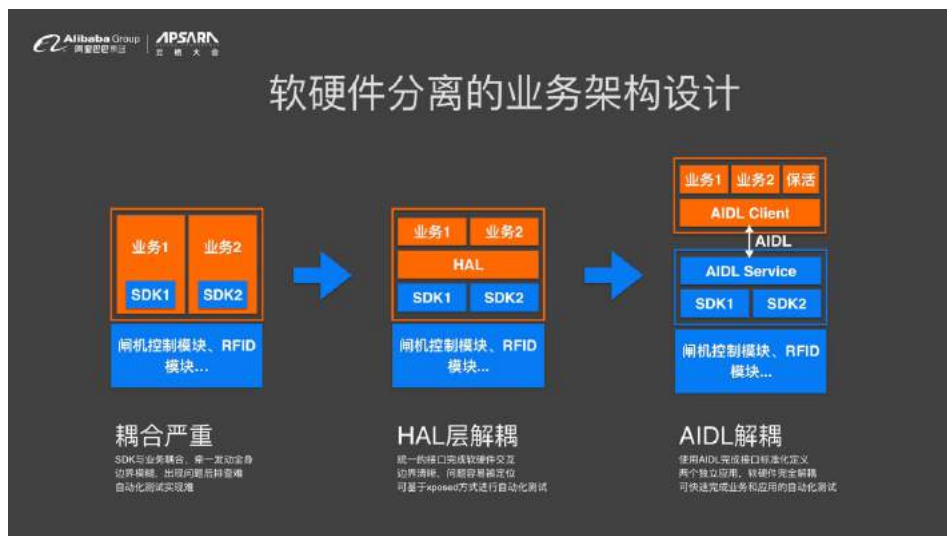
在 2019 杭州云栖大会【智慧文娱技术】分论坛上，阿里巴巴大文娱事业群高级无线开发专家楚佩斯重点分享 5G 下，大麦现场服务解决方案实践。以下为分享精选。

一、硬件设备标准化之路



原来的大麦在现场服务这一块，大麦和麦座都有相应的现场业务，存在重复建设问题。大家可以看图中的硬件设备，种类繁多，但是现场对稳定性的要求是异常严苛的。试想一下，一个十万人的大型活动，现场服务中断一分钟，将导致非常严重的事故发生。我们在全中国运用的非常多的自助的设备如果全部要依赖于人工来运维的话，对于人力的消耗也是非常大的问题，因此如何提效降本、提升运营效率是一个非常现实的话题。

1. 软硬件分离架构的演进的过程



这是软硬件分离架构的演进的过程，经历了三个不同的阶段，第一个阶段其实就像原始社会，SDK 与业务耦合严重，每一次改动都牵一发而动全身，自动化测试无法实现。

第二个阶段做了硬件抽象层，通过在业务层和硬件 SDK 层中间增加一层硬件抽象层，这个事情在 Android 的 Framework 底层也是这么做的，这样确实一定程度上解决了硬件适配难和自动化测试的问题，但是仍然不够，问题是当你需要集成的硬件的设备越来越多的时候，整个应用程序包的大小会急剧增长，甚至会出现不同的供

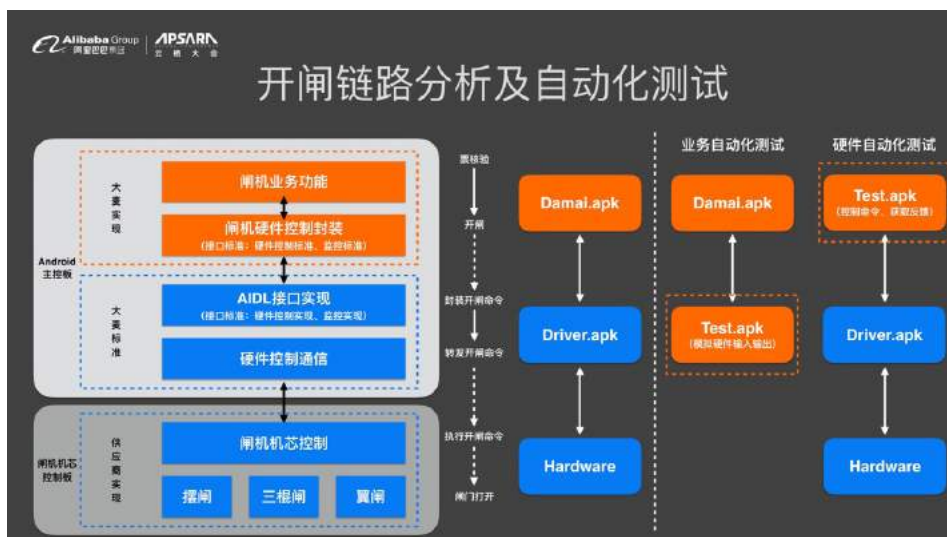
应商提供的 SDK 是冲突的，最终会导致整个应用包的打包失败。

第三个阶段，完全将业务逻辑和驱动程序完全的解耦，分别在两个不同的进程中来实现，进程间使用 Android 的 AIDL 来实现，定义标准的 Service 接口。这样带来的好处就是驱动程序可以随着 ROM 一起打包发布，A 供应商的自助机和 B 供应商的自助机，我们使用的是完全无差别的业务 apk。C 供应商和 D 供应商只需要按照我们的标准去完成相应的驱动 apk 的开发，我们的业务就可以在上面跑起来。

当然也有两个比较重要的点值得我们注意，第一个是需要注意进程的保活，因为在不同的进程中，可能会出现驱动程序因为各种原因出现闪退，这时业务程序需要有保活机制重新拉起驱动程序以免影响业务运行。第二个是安全问题，不能让所有的应用程序都可以通过 AIDL 接口来操作硬件，驱动程序需要检查调用者的签名和包名。

2. 开闸链路分析及自动化测试

通过标准化的建设是非常有利于整个的自动化测试的，整个的流程分为三部分，业务程序、驱动程序、硬件模块，左边这个是我们从一个票的核验到完成的开闸的整个调用链路。



自动化测试在这个场景下有两个诉求，第一个是业务的自动化测试，业务的自动化测试主要检验的是业务逻辑，比如说刷二维码入场，但是刷二维码后除了校验二维码有效性之外，还有各种复杂的权限规则的校验。第二个是硬件的自动化测试，我们可以通过模拟一个业务程序，通过 AIDL 标准接口调用完成硬件的压测。在实际测试中，我们发现了一个硬件模块在万分之一概率的情况下出现了控制失败的问题，最终推动供应商完成了问题修复，通过这两个各有侧重点的测试，保障整个系统的稳定性。

3. 故障监控与告警

虽然做了自动化测试，但是仍然不能保证线上设备不会出现任何问题，出现问题其实不可怕，可怕的是没有办法发现，然后去解决。这是我们的某一个硬件打印机里面有卡纸的异常，故障上报到运营平台，运营平台结合设备的信息，将这个故障通过钉钉告警对应的设备维护的同学，在完成设备维护后，整个监控告警与故障恢复可以完成业务闭环。



通过这些建设我们也取得了一些成果，实现了硬件的标准化和运维的智能化，同时我们也愿意把能力开放出来赋能行业，目前已经有一些系统在对接中。

二、物联网技术应用实践

有很多硬件设备虽然联网提供服务，但是运维管控能力都很弱、甚至没有，因此很难称之为一个物联网设备。大麦有非常多的硬件设备，在物联网场景下也进行了不少探索。



在现场面临的场景是很复杂的，比如大型活动，人流聚集，线上体验会很差，大家都会共享基站资源，并不是你的信号不好，而是系统已经无法满足业务需求。

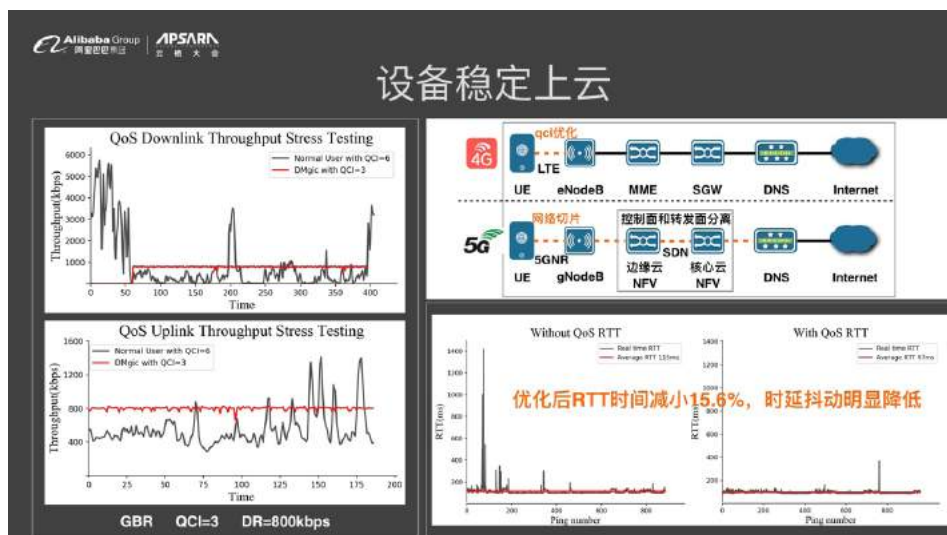
在 4G 时代，20MHz 的带宽下提供 100Mbps 的速率，但是这个速率是大家共享的。5G 时代这个问题会好很多，但还是要理性看到问题的本质，5G 时代大家基本上使用的是 100MHz 的带宽，本身就要比现有 4G 大 5 倍，5G 的频率利用效率是 4G 的 3 ~ 5 倍。容量大了之后，需求也增大了，也许到时候还是供不应求的。高密度区域还是需要做专项优化的。我们创新性的使用一些物联网技术来解决现场挑战，但物联网技术的安全性和稳定性成为比较大的挑战。

设备上云肯定需要考虑安全问题，阿里的物联网平台提供了整个设备上云解决方案，设备安全的认证通过一个三元组的设备密钥去完成设备的激活入网，而且一

机一密，即使攻破一个设备也无法将其应用到其他设备，整个网络的安全性可以得到保障。



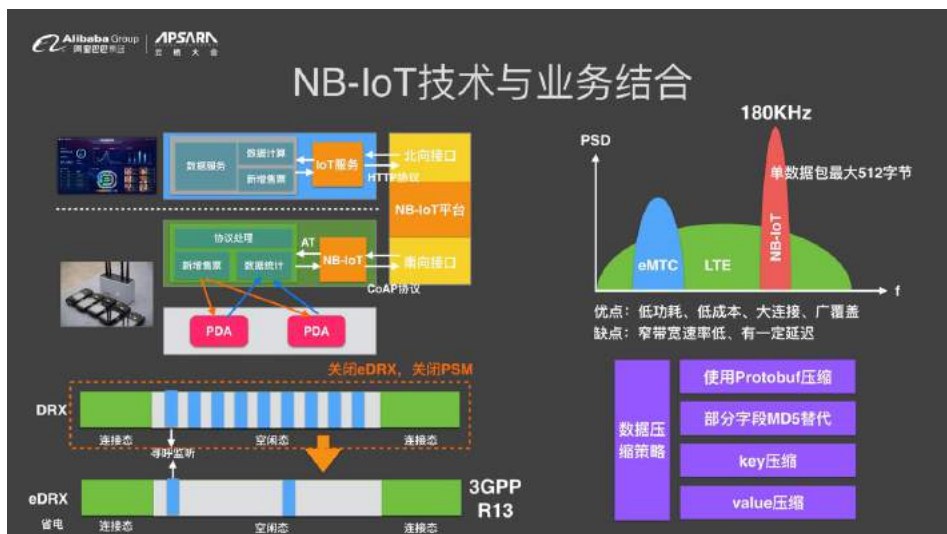
另外物联网设备的数据如果需要存储，阿里云也提供了表格存储和 DataHub 等，让设备上云及数据处理变的非常的方便。



设备不仅需要安全的上云还需要稳定的上云，在 4G 时代我们前面已经提到了，带宽是有限的，在密度区域是供不应求的。那么如何解决不同业务对于带宽、时延不同的需求呢？

在 4G 时代其实是有 QCI(QoS Class Identifier) 技术，QCI 是 EPS 承载最重要的 QoS 参数，它是一个数量等级，3gpp 一共定义了 9 个等级，语音电话是工作在 QCI1 等级，优先级最高。普通联网 tcp 数据流是工作在 QCI6 等级。从上面图中可以看到，在网络压力比较大的情况下，QCI3 等级的业务相对 QCI6 等级的业务要稳定很多，RTT 延迟和时延抖动都有明显的降低。

这个技术在 5G 时代其实粒度会更细，也有了一个新的命名叫网络切片。通过网络切片可以为不同的业务提供不同的带宽、时延。5G 有 SDN(软件定义网络)，通过这些技术可以灵活的分配核心网的资源。因此在 5G 里面的网络切片是端到端的，而在 4G 时代只是基于 QCI 完成的终端设备和基站之间这一部分的 QoS 保障。



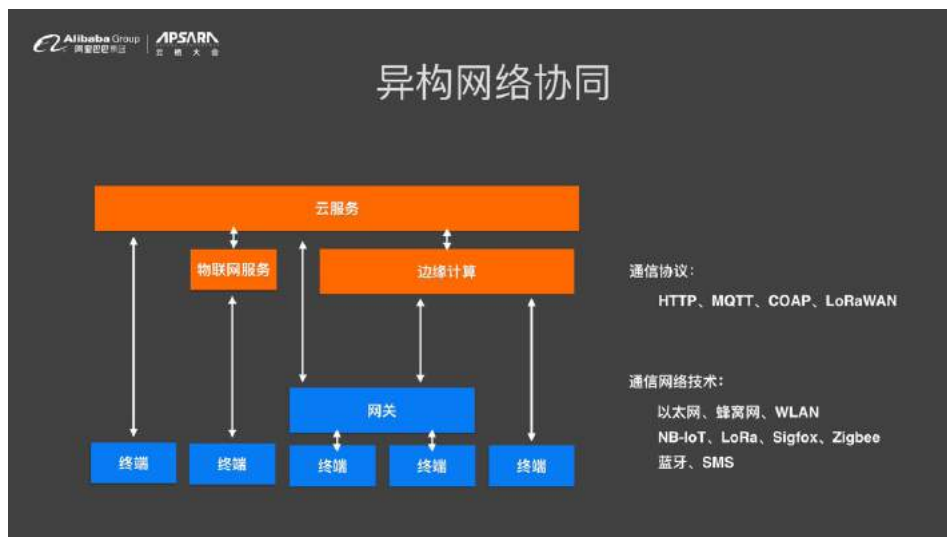
在设备上云这条路上，除了传统的蜂窝网络，其实还有新的物联网技术如 NB-IoT 技术，这个是我国运营商正在大力推广的，也是 5G 的物联网候选标准。

NB-IoT 是窄带物联网，只有 180kHz 带宽，每个数据包可以传输 512 个字

节的内容。NB-IoT 是运营商部署在独立频段上的，因此在现场人流量集聚拥塞的情况下，仍然能稳定的运行。

为了满足低功耗的特性，在 R13 版本的时候加入的 eDRX 和 PSM 的模式，目的就是节省电量，通信完了就开始休息了，有利于设备的低功耗和基站的大容量。在我们的场景中，功耗不是最主要的考量因素，但是数据尽可能快的到达设备是一个重要考量因素，因此关闭设备的 PSM 和 eDRX 模式成为了一个选项。

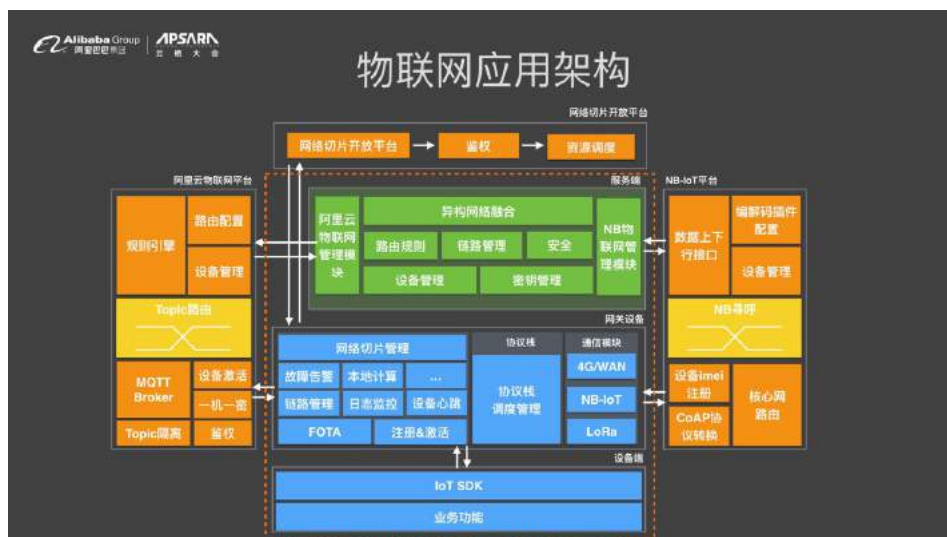
同时因为窄带物联网带宽是非常低的，我们的业务数据是需要经过一些处理再运行到上面。比如新增售票，某一个用户购票后，数据如何进入设备中。由于 NB-IoT 的带宽非常小，因此需要将数据包进行压缩，以满足在 NB-IoT 网络中运行。NB-IoT 网络使用 CoAP 协议并不保证数据包可靠到达，因此如果业务需要可靠到达，需要设计应用层的应答机制，来保证整个数据的可靠、稳定到达。



在物联网中也很复杂，协议是没有既定的，比如通信协议有 HTTP、MQTT、COAP 等，通信网络有熟悉的以太网和 Wi-Fi，也有新兴的 NB-IoT/LoRa 等物联网技术。如何在异构网络中保证数据的可靠到达成为一个挑战。

我们有一个场景，比如新增售票，其实可以通过 4G 的通道下发到设备中，但现

场网络变差，数据包到达能力也会变差，这时候可以通过 NB-IoT 网络进行补偿，以保证在异构网络下数据的可靠和实时的到达。



这是大麦的物联网架构，中间绿色和蓝色部分是大麦系统，周围黄色部分是外围系统，比如阿里云的物联网平台，以及 NB-IoT 的平台以及网络切片开放平台。

三、数字化现场娱乐的趋势



无纸化是一个很重要趋势，现场娱乐票务行业相对来说还是比较传统的行业，通过撕票根的方式来核验，电子票的占比还有待进一步提升。12306 也在推动票务及核验的无纸化，我们也希望通过无纸化的方式打造数字化的现场娱乐。



5G 和现场会有什么结合？VR / 全息等技术将带来真正的沉浸式体验，改变现场的观看习惯。虽然目前技术还是无法达到这种效果，但未来可以期待。在阿里巴巴 20 周年晚会时，也会有互动手环，通过 IoT 设备通信完成很好的互动体验，对超大规模物联网节点设备进行实时控制还是存在不小的技术挑战，未来 5G 也会支持高密度的物联网设备通信，新的创新应用将会层出不穷，给用户更好的体验。

大数据赋能电影数智化宣发

作者：冯新平 阿里影业高级数据技术专家

电影宣发正在从传统线下走向互联网，传统线下和互联网的宣发预算投入，由原来的以线下为主，转变为线下和互联网预算比例为 1: 2。好的宣发能为电影票房打下非常好的基础，助力电影口碑快速发酵，并下沉到三四线的市场。

由于电影宣发的长链路和环节的复杂性，在数据化的进程中，我们遇到不少难题。在 2019 杭州云栖大会【智慧文娱技术】分论坛上，阿里影业高级数据技术专家冯新平，从技术层面分享了阿里影业的数据化宣发的解决方案，希望对大家有借鉴。

一、电影行业的互联网化



电影行业的互联网化程度怎么样呢？从专资办 2018 年的数据看，每年有 2.5 亿用户在线上购票，85% 的票房来自线上，90% 影院都在网上进行售票，线上化程度非常高。



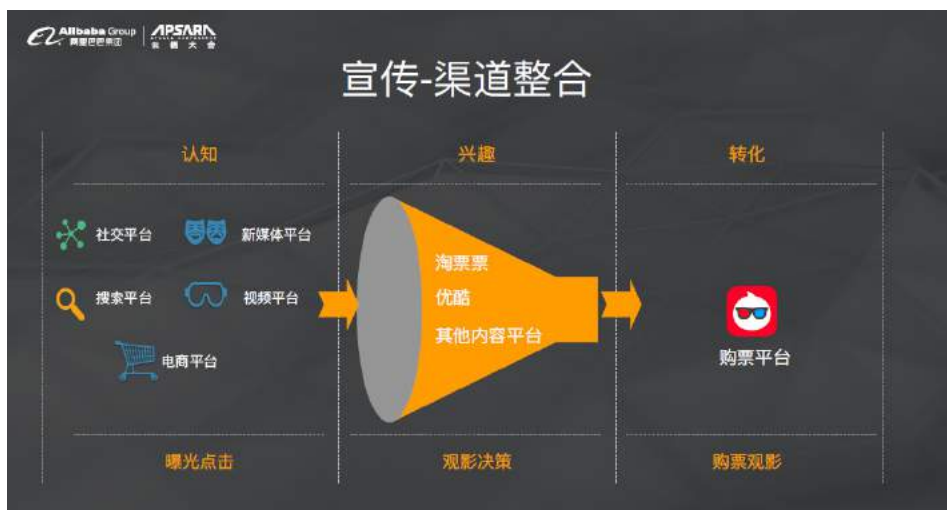
在线上化的基础上，我们实现了数据积累，包括用户基础数据、偏好数据，线上用户的观影路径、决策路径、看预告片、最终购票；还有影片和影院的基本信息，这是货和场的基础信息，对后续的票房预测有比较大的帮助。



线上化和数据化到达一定程度后，是不是所有的一切都顺理成章？其实困难还非常多。首先，由于影视宣发的渠道是分散的，宣传成本居高不下。其次，全链路购票数据分析存在难度，宣传效果难跟踪。再次，全国影院系统分散在各个地方，数据上报不及时，票房数据不能及时同步。

所以，接下来我将从宣传和发行这两个点进行分享。

二、宣传：渠道整合，快速宣推能力



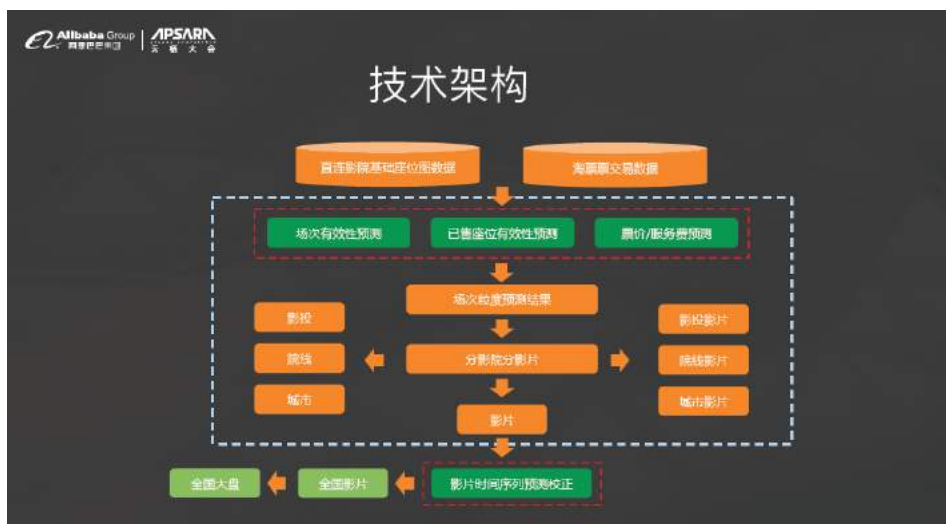
宣传部分，主要是整合宣发渠道，比如社交平台、新媒体平台、抖音号，在非常快的时间内把所要的渠道做整合。另外无论渠道多么繁多，85% 的电影票都在线上购票，在渠道触达后，用户最终都是要回到购票平台的，例如通过淘票票购票，即使数据渠道比较分散，但是我們也能做一定程度的链路分析。

例如《碟6》。第一波主打存量观众，比如《碟5》评分较高的用户，引导其参加《碟6》的点映礼，容易形成好口碑；第二是引导核心观众去看，核心观众就是电影拍摄前所确定的受众人群。第三是分析潜在观众在哪里？灯塔中有两个比较好的功能，大V宝和渠道通，能够在一小时之内精准触达到上亿用户，同时对宣发公司提供相对系统性的数据报告。

三、发行阶段：票房数据的实时预测



映前票房预测的难度特别大，因为影响因素特别多，与当时的舆情、热点相关性比较大。我们在支持发行阶段的数据指导，所以做的是实时票房预测，包括有效场次、已售座位数和平均票价都要计算出来。另外全国 1.1 万个影院中约 90% 我们有直联（指通过淘票票可以直接售票的影院），如何基于直联影院推测全国的票房情况？



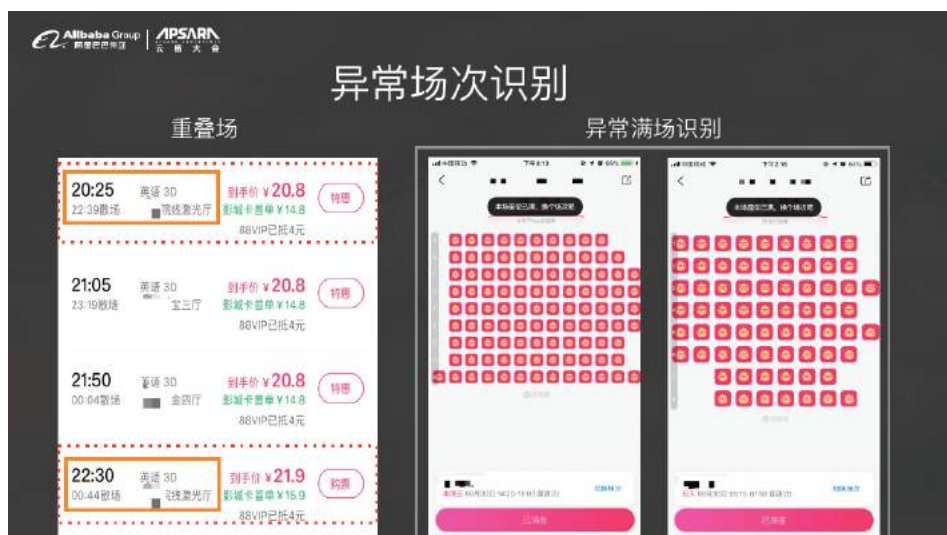
数据上，我们有两个输入，直联影院的实时的排片信息和淘票票的订单情况。技术上需要解决四个难点，如何识别场次的有效性、识别真实售卖情况，还有预算平均票价；把每一个场次的票房全部计算出来了，然后进行累加。这样看起来是不是很简单，其实第一版，包括同行都在解决这个问题，都不是这样做的，是怎么样来做的呢？

行业全部影院中，有的影院上报专资办票房中规中矩的，和真实数据基本一致，有的和真实数据有较大出处，所以要识别出可信影院，标准就是计算影院排座之后，上映的最后时间节点的不可售状态和真实上报的数据差距，越小越可信，然后根据这些可信影院的票房来预测大盘票房。

第一个版本我们线上运行半年多，运维成本特别高，小影片体量小，误差大；更重要的是预测误差没有合理解释，因为基于小样本推测全局。还有不可抗力出现，如2018年广州台风时，广州整个影院都不能售票了，导致预测误差很大。

所以我们推出优化版本，主要解决两个技术难点，一个是异常场次的识别，二是已售座位的识别。

第一个，异常场次的识别。





主要是重叠场和异常满场。重叠场，比如两个场次同一个厅，但是上一场的散场时间为 22:39，下一场的开场时间为 22:30，明显时间重叠。全国大档期每天有 50 万场，平时也有 30 万场，都要识别出重叠场。另外是异常满场，一个就是“幽灵场”，比如 8 月 30 号凌晨是满场。我们将潜在异常满场识别出来，再基于历史数据，是不是有惯例？还有一种是包场，也要识别出来。

第二个，有效座位的识别。



例如一部影片，在同一个电影院同一个厅的两个场次，开场时座位数一模一样，就会引起我们的怀疑。因为售票平台的系统商不一样，每个系统商的座位数状态都不太一样，总的来说有四种，不可用、已售、预留、锁座，有的影院经理喜欢把预留标成不可用，这个也要识别出来。基于国家权威数据，与阿里影业所监测到的状态进行对比，每一个影院都需要一套算法和模型来进行识别。

第三个，平均票价的识别。

平均票价&服务费预测

票价

服务费

分账票价 = 票价 - 服务费

非分区定价模型

$$f(x_1, x_2) \rightarrow y$$

$$\min_{i, \delta_i} \|x_i + \delta_i - y\|_1$$

分区定价模型



$$f(x_1, x_2, w) \rightarrow y$$

$$\min_{i, \delta_i} \left\| \frac{\sum_j \omega_j x_{i,j}}{\sum_j \omega_j} + \delta_i - y \right\|_1$$

有效已售座位 w_i
排期票价 x_1
订单票价 x_2
历史专资办票价 y

平均票价，由于每一个影院基础设施不同，活动情况和票补力度不同，所以在影院维度的平均票价都不一样。我们能够获取到的信息主要是有效订单票价、排期票价、历史上的专资办的票价数据，通过三个票价去拟合真实票价。

第四个，片票预测的修正。

全国影片票房预测修正

直连影院影片票房 $x(t)$ → 大盘影片票房 $y(t)$

$y(t) = r(t)x(t)$ $r(t)$ 有周期性

$$r(t) = g(t) + s(t) + h(t) + e(t)$$

$$g(t) \text{ 是趋势项, 采用线性趋势}$$


$$g(t) = (k + \mathbf{A}(t)\delta)t + (b + \mathbf{A}(t)^T \eta)$$

$$h(t) \text{ 是节假日项, 节日 } i = \text{元旦, 春节, 清明节, 劳动节} \dots$$

$$h(t) = \sum_i k_i \delta_{D_i}(t)$$

$$s(t) \text{ 是周周期项, 周期性周数 } P = 7, 30$$

$$s(i) = \sum_{n(i)=1-N}^N c_n \psi^{\frac{n(i)}{P}}$$



我们能够获取部分影院的直联数据 (90% 到 95%), 剩下的 5% 到 10% 数据获取不到, 如何将直连影院的票房校准到全国呢? 从数据分析上发现, 校准系数与档期、周期性有关。比如周一到周四, 与周五、六、日的校准系数是不一样的, 我们基于时间序列分析模型, 考虑了自然趋势分析, 大档期因素, 还有周期性影响等进行建模。

效果如何? 2019 年春节档, 我们的预测票房与大盘票房的误差约 1%, Top10 影片平均误差在 3% 之内, 体量越大误差越小, 体量越小误差越大, 50% 的院线的平均票价是 5% 之内。

电影宣发的数据化过程中, 数据积累的完整度还有用户的链路分析还有很长的路要走, 我们的目标是打造电影行业数智化宣发的新基础设施, 让天下没有难做的宣发。

关注我们



(阿里文娱技术公众号)

关注阿里技术



扫码关注「阿里技术」获取更多资讯

加入交流群



- 1) 添加“文娱技术小助手”微信
 - 2) 注明您的手机号 / 公司 / 职位
 - 3) 小助手会拉您进群
- By 阿里文娱技术品牌

更多电子书



扫码获取更多技术电子书