

策略算法工程师-音视频

留用实习面试速背版：CAE / Ladder / ABR / QoE / AB 实验 / ML 八股

面试定位：不要装成音视频专家，要表现成“懂音视频核心链路、会做策略建模、会做实验验证、能快速落地”的算法候选人。

0. 岗位一句话理解

这个岗位是在音视频播放链路里做策略优化：上游通过内容感知编码和 Ladder 设计决定视频有哪些清晰度和码率档位，下游通过 ABR、缓存和后处理策略决定用户实际播放哪个档位。最终目标是在清晰度、流畅度、成本和时长/营收之间做最优权衡。

1. 自我介绍模板

面试官您好，我叫 XXX，目前主要方向是机器学习/算法工程。过去项目主要集中在数据建模、策略优化和工程落地上，比较关注如何从业务目标出发，通过特征分析、模型训练、策略设计和实验验证提升核心指标。

我看到这个岗位主要做音视频点播和直播场景下的策略算法，包括内容感知编码、转码 Ladder 设计、ABR 播放策略，以及基于用户、视频、主播画像做个性化优化。我理解它的核心不是单一模型，而是围绕清晰度、流畅度、成本、时长/营收之间做多目标权衡。

虽然我在音视频领域的经验还在补充，但我对机器学习建模、数据分析、AB 实验和策略迭代比较熟悉，也提前系统了解了 CAE、VMAF、ABR、QoE 等基础概念，希望能把已有的算法和工程能力迁移到音视频策略场景中。

2. 必背主线：CAE → Ladder → ABR → QoE → AB 实验

- **CAE：**根据视频内容复杂度自适应选择编码参数、分辨率、码率，目标是在质量不下降的前提下降低成本。
- **Ladder：**多档分辨率/码率组合，本质是在质量和成本之间找 Pareto 最优点。
- **ABR：**播放器根据带宽、buffer、设备和用户场景动态选择码率档位。
- **QoE：**综合清晰度、卡顿、首帧、切档、成本的体验目标函数。
- **AB 实验：**所有策略最终都要用线上实验看主指标、护栏指标和诊断指标。

3. 音视频基础速记

概念	速背答案
编码	把原始视频压缩成可存储/传输的码流，如 H.264/H.265/AV1。
转码	把已有视频重新编码成不同格式、分辨率、码率或档位。
封装	把音频、视频、字幕和元数据放入容器，如 MP4、FLV、TS。
码率	单位时间的数据量，通常越高越清晰但成本越高。
分辨率	画面像素尺寸，提升分辨率不一定提升感知质量。
GOP	两个关键帧之间的帧组；GOP 长压缩效率高但 seek、抗错、低延迟较差。
I/P/B 帧	I 帧独立编码，P 帧参考过去，B 帧双向参考，压缩效率更高但延迟更高。
CBR/VBR/CRF	CBR 稳定码率，VBR 按复杂度变码率，CRF 按目标质量控制码率。

4. 高频问答：CAE / Ladder / VMAF

Q: 什么是 CAE?

CAE 是 Content Adaptive Encoding, 内容感知编码。不同视频内容复杂度不同, 所以不应该用统一固定码率模板。动画、PPT、访谈类视频低码率也清晰; 体育、游戏、舞蹈类视频运动剧烈、纹理复杂, 需要更高码率。CAE 会分析空间复杂度、时间复杂度、运动强度、纹理、场景类型, 也可以结合抽样编码得到码率-质量曲线, 为每个视频选择合适的分辨率、码率和编码参数。目标是在主观质量基本不下降的情况下节省带宽、存储和转码成本。

Q: 如何设计转码 Ladder?

我会把 Ladder 设计看成多目标优化问题。第一, 分析视频内容, 包括空间复杂度、时间复杂度、运动强度、纹理复杂度和场景类型。第二, 生成候选编码点, 如不同分辨率、码率、QP/CRF 参数。第三, 用 VMAF、SSIM、PSNR 和主观抽检得到码率-质量曲线。第四, 去掉冗余档位, 例如某个 1080p 档位比 720p 高很多码率但 VMAF 提升很小, 就不值得保留。第五, 根据业务目标选择最终档位, 高价值内容偏质量, 长尾内容偏成本, 直播考虑实时性, 点播可用更复杂离线分析。最后通过 AB 实验验证清晰度、卡顿、播放时长和成本。

Q: VMAF 是什么?

VMAF 是 Netflix 提出的感知视频质量评价指标, 它融合多个视觉特征, 用机器学习拟合人的主观观看评分。相比 PSNR 只看像素误差、SSIM 主要看结构相似度, VMAF 更接近人的主观感受。但 VMAF 不是万能的, 可能不完全覆盖二次元、游戏、直播、美颜、人脸、文字等场景, 所以线上还要结合播放时长、卡顿率、用户反馈和主观抽检。

Q: 如何判断视频内容复杂度?

可以看空间复杂度、时间复杂度和语义复杂度。空间复杂度包括纹理、边缘、细节; 时间复杂度包括运动强度、镜头切换、运动矢量; 编码复杂度包括 QP、残差、预测误差、码率-质量曲线; 语义复杂度包括人脸、文字、游戏、体育赛事等。

5. 高频问答：ABR / QoE / 缓存

Q: 什么是 ABR?

ABR 是 Adaptive Bitrate, 自适应码率播放。它的目标是在网络波动下动态选择合适播放档位, 尽量做到清晰、流畅、不卡顿。ABR 通常根据历史下载吞吐量、当前 buffer 水位、可选码率档位、设备性能、网络类型等信息决策。简单策略基于带宽估计选择低于带宽的最高码率, 更稳健的策略会同时考虑 buffer, buffer 低时保守降档, buffer 高时再尝试升档, 避免频繁切换。

Q: 为什么不能只按带宽选码率?

因为带宽估计有噪声，网络短时波动大。如果只根据瞬时带宽升档，下一段网络下降时容易导致 buffer 下降甚至卡顿。播放体验里卡顿的负向影响通常大于清晰度提升的正向收益，所以 ABR 要综合考虑 buffer、带宽趋势、切档稳定性、首帧时延和用户场景。

Q: QoE 公式怎么讲?

$QoE = \alpha \cdot \text{清晰度} - \beta \cdot \text{卡顿时长} - \gamma \cdot \text{清晰度切换} - \delta \cdot \text{首帧时长} - \lambda \cdot \text{成本}$ 。清晰度可用平均码率、VMAF 或用户感知质量表示；卡顿和首帧是强负向指标；频繁切档影响观看稳定性；成本包括带宽、存储和转码资源。短视频重首帧，长视频重稳定清晰度，直播重低延迟和少卡顿。

Q: 如何设计 ABR 策略?

先做规则 baseline，再逐步模型优化。baseline 可以基于带宽估计和 buffer 水位：用过去几个分片下载速度估计吞吐量，选择低于安全带宽的最高码率；同时引入 buffer 保护，buffer 低于阈值就保守降档，高于阈值再允许升档。进一步加入用户画像、设备、网络类型、视频类型和历史卡顿风险。蜂窝网络或弱网用户更保守，高端设备和 WiFi 用户更积极。最终优化 QoE，而不是单纯最大化平均码率。上线通过 AB 实验看播放时长、卡顿率、首帧、平均清晰度、切档次数和带宽成本。

6. AB 实验与因果推断

Q: AB 实验怎么设计?

先明确主指标、护栏指标和诊断指标。转码策略实验中，主指标可以是播放时长、完播率、用户清晰度感知或成本节省；护栏指标包括卡顿率、首帧时长、退出率、投诉率；诊断指标包括平均码率、VMAF 分布、各档位命中率、不同设备和网络下的表现。实验设计要用用户级随机分流，保证实验组和对照组在设备、网络、地区、内容类型上分布一致。先小流量灰度，护栏指标不劣化再逐步放量。整体不显著时做分层分析。

Q: 高清用户时长更高，能说明提高清晰度能涨时长吗？

不能直接说明因果关系。高清用户本身可能网络更好、设备更好、用户价值更高、观看意愿更强，时长更高可能由这些混杂因素导致，而不是清晰度本身导致。要验证清晰度提升是否真的带来时长提升，最好做随机 AB 实验。如果不能实验，可用倾向得分匹配、DID、因果森林等方法控制网络、设备、活跃度、内容类型等混杂变量。

Q: 离线指标涨，线上没涨怎么办？

可能原因：第一，离线指标和真实体验不一致，例如 VMAF 提升但用户不敏感。第二，样本分布不一致，离线集没覆盖线上设备、网络和内容类型。第三，策略副作用，例如清晰度提升导致码率升高，弱网用户卡顿增加。第四，实验设计问题，如样本量不足、分流不均、周期太短。我会先做分层分析，再看护栏和诊断指标，定位是哪类人群或场景没有收益。

Q: 多目标冲突怎么处理？

先区分硬约束和优化目标。卡顿率、首帧、投诉率可以作为护栏指标，不能明显劣化；在满足护栏的前提下，再优化清晰度、时长或成本。策略上可以写成加权 QoE，也可以按业务场景设置不同权重：弱网用户优先流畅，高价值内容优先质量，长尾内容更关注成本。

7. 策略设计题模板

题 1: 设计内容感知转码系统

- 目标：降成本、保质量。
- 输入：视频内容特征、历史播放、业务场景。
- 分析：空间复杂度、时间复杂度、场景类型。
- 候选：多分辨率、多码率、多编码参数。
- 评估：VMAF、码率-质量曲线、主观抽检。
- 决策：选择 Pareto 最优 Ladder。

- 上线：AB 实验验证质量、时长、卡顿、成本。

题 2：设计 ABR 播放策略

- 输入：带宽、buffer、档位、设备、网络、用户。
- 目标：最大化 QoE。
- baseline：吞吐量 + buffer。
- 优化：个性化阈值、卡顿风险预测。
- 护栏：首帧、卡顿、切档次数、成本。
- 上线：分层 AB，弱网/强网/设备分别看。

题 3：做用户/视频/主播画像

用户画像：设备、网络、地区、历史卡顿、清晰度偏好、观看时长、付费/活跃度。

视频画像：内容类型、运动强度、纹理复杂度、时长、热度、历史码率-质量收益。

主播画像：开播网络稳定性、内容类型、互动强度、历史卡顿和观看指标。

画像最终不是为了描述，而是作为策略输入，用来决定初始档位、升降档阈值、转码 Ladder 或后处理策略。

8. 机器学习八股速背

Q: LR 和 GBDT 区别?

LR 是线性模型，可解释性强、训练快，适合高维稀疏特征，但表达能力有限。GBDT 是树模型集成，可以自动学习非线性和特征交叉，对表格数据效果通常更强，但训练和调参更复杂，也更容易过拟合。

Q: XGBoost 和 LightGBM 区别?

XGBoost 是经典 GBDT 实现，支持二阶梯度、正则化和列采样，稳定性强。LightGBM 主要优化训练效率，用 histogram 算法、leaf-wise 生长和 GOSS/EFB 等方法，在大规模数据上更快，但 leaf-wise 如果不限深度更容易过拟合。

Q: AUC 是什么?

AUC 可以理解为随机抽一个正样本和一个负样本，模型把正样本排在负样本前面的概率。它关注排序能力，对分类阈值不敏感，适合点击率、转化率这类排序场景。

Q: 过拟合怎么办?

从数据、特征、模型和验证四方面处理：增加数据和数据增强；减少泄漏和噪声特征；加正则、限制树深、early stopping、dropout；使用交叉验证，并确保训练集和验证集按时间或用户合理切分。

Q: 样本不均衡怎么办?

可以重采样、调整 class weight、使用 focal loss、优化排序指标或 PR-AUC，也可以调整阈值。业务上还要关注负采样方式是否和线上分布一致。

Q: 特征穿越是什么?

训练时使用了线上预测时不可获得的信息，导致离线效果虚高、线上失效。例如预测用户是否卡顿时使用了播放结束后才知道的总卡顿时长。处理方式是严格按时间切分特征，确保所有特征在决策时刻可获得。

9. 不会细节时的保命回答

这个细节我之前没有在工程里深度实践过，但我的理解是它背后仍然是一个多目标策略优化问题，需要在质量、流畅度、延迟和成本之间权衡。

如果让我做，我会先建立可解释 baseline，明确输入特征、目标函数和护栏指标；然后用离线数据验证，再通过小流量 AB 实验迭代。音视频的领域特征，比如 VMAF、buffer、码率、GOP、设备解码能力，我会作为关键约束纳入策略。

10. 反问问题

- 这个岗位更偏上游转码策略，还是下游播放器 ABR 策略？
- 团队当前优化的核心指标更偏 QoE、时长，还是成本？
- CAE 和 ABR 策略目前是规则为主，还是模型化/学习化为主？
- 线上 AB 实验一般会重点看哪些护栏指标？
- 实习生进去后更可能负责离线分析、策略建模，还是线上实验落地？

11. 明天面试心法

核心人设：我不一定已经是音视频专家，但我理解这个岗位的本质是策略算法和多目标优化；我能快速学习领域知识，能基于数据建模，能设计实验验证，也能推动策略上线。

只要能把 CAE + Ladder + ABR + QoE + AB 实验这条链路讲顺，就已经超过很多只背机器学习八股的人。