

基于 XDR 数据分析的 OTT 视频服务感知质量评估方法

黄鹂声¹ 冉金也¹ 罗静¹ 张翔引²

¹(电子科技大学计算机科学与工程学院 成都 611731)

²(电子科技大学航空航天学院 成都 611731)

(lsh@uestc.edu.cn)

Estimating QoE for OTT Video Service Through XDR Data Analysis

Huang Lisheng¹, Ran Jinye¹, Luo Jing¹, and Zhang Xiangyin²

¹(School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu 611731)

²(School of Aeronautics and Astronautics, University of Electronic Science and Technology of China, Chengdu 611731)

Abstract Over the top (OTT) video services have gradually become one of the most popular online services. However, due to poor network quality, overload of service platform and other reasons, OTT services often encounter quality problems such as playback failure, increased number of stuck, and long buffer time, which lead to the decline of quality of experience (QoE). Internet service providers need to accurately evaluate the QoE of OTT video services so as to identify quality problems in advance and further optimize networks and services. In this paper, a no-reference OTT video quality estimation method based on extended data record (XDR) data analysis is proposed. It extracts a small amount of information with high relevance to video quality from a large amount of XDR data, and converts large-scale, low-value XDR data into high-value and small-scale video quality feature information. This method facilitates the application of subsequent artificial intelligence algorithms for OTT video quality evaluation, reduces the cost of further data mining, and improves the accuracy of machine learning model and QoE evaluation results. The data extracted by this method is used for OTT video QoE prediction, and experimental results show that the accuracy of the QoE prediction results is significantly better than that of the current evaluation results using the original XDR data.

Key words video quality estimation; XDR data; data analysis; OTT(over the top); QoE(quality of experience)

摘要 互联网电视(over the top, OTT)视频业务逐渐成为最流行的在线业务之一,然而网络视频往往由于网络质量差、服务平台过载等原因,出现播放失败、卡顿次数增加、缓冲时间过长等质量问题,导致用户感知质量(quality of experience, QoE)下降.因此,运营商需要精确评估和掌握用户在使用网络视频业务过程中的质量体验,以便提前发现质量问题,进一步开展网络和业务优化工作.为了解决该问题,提出一种基于用户呼叫/事务/会话记录数据(extend data record, XDR)的无参考网络视频质量评估方法.该方法从大量 XDR 数据中提取出与视频质量相关性高的少量信息,将大规模、低价值的 XDR 话单数据转化为高价值、小规模的视频质量特征信息,有利于后续人工智能算法的应用和视频业务质量评价,降低进一步数据挖掘的资源成本,提升机器学习的输入样本质量和 QoE 评价结果的准确性.实验

收稿日期:2019-10-24;修回日期:2020-04-17

基金项目:国家重点研发计划项目(2018YFB0804505);国家电网公司科学技术项目(522722180007)

This work was supported by the National Key Research and Development Program of China (2018YFB0804505) and the Science and Technology Project of State Grid Corporation of China (522722180007).

通信作者:罗静(luojinn@126.com)

表明:使用该方法提取后的数据进行 QoE 预测,得到的预测结果在准确性方面明显优于目前基于原始 XDR 数据的 QoE 机器学习评估方法。

关键词 视频质量评估;用户呼叫/事务/会话记录数据;数据分析;互联网电视;用户感知质量

中图分类号 TP751

随着移动网络传输能力的快速提升和智能终端的普及,互联网电视(over the top, OTT)视频业务日趋成为网络用户中最流行的在线业务之一。以 OTT 视频技术为主导的网络电视、移动视频和多屏互动业务,通过智能手机、平板电脑和联网电视机的 OTT 视频内容消费量正以惊人的速度增长。在移动视频领域,研究表明 50% 以上的平板电脑用户会经常在线观看 OTT 视频^[1]。

有线电视长期以来的低费用和高保障,使得消费者不再满足于以往免费网络视频“尽力而为”的质量体验。用户体验质量的高低成为 OTT 视频成功的关键。一旦用户感知质量(quality of experience, QoE)没有达到用户的期望值,他们可能立刻转投竞争者。

然而网络视频往往由于网络质量差、服务平台过载等因素,出现播放失败、卡顿次数增加、缓冲时间过长等业务质量问题,这些问题会导致用户满意度下降、投诉量增加甚至退订业务等后果。因此,服务提供商需要精确评估和掌握用户在使用网络视频业务过程中的体验质量,以便于提前发现质量问题,进一步开展客户关怀、预检预修工作。

视频质量评估主要分为主观、客观 2 种方法。主观评估需要人类观察者通过视觉系统感知视频质量,双激励损伤量表(double stimulus impairment scale, DSIS)、双激励连续质量量表(double stimulus continuous quality scale, DSCQS)、绝对分级法(absolute category rating, ACR)^[2]都是典型的主观评估方法。尽管主观评估方法可以最直接地反映视频质量,但是由于成本高昂且费时,所以并不实用。

客观视频质量评估方法分为 3 类:全参考(full reference, FR)、半参考(reduce reference, RR)、无参考(no reference, NR)。FR 需要获取完整原始视频序列。在评估过程中,对比完整的原始视频图像数据和用户接收到的视频图像数据间的差异,检测用户接收到的视频质量损伤程度,根据损伤程度对视频用户体验进行评价。峰值信噪比(peak signal-to-noise ratio, PSNR)^[3]和结构相似性(structural similarity index measurement, SSIM)^[4]是常用的 FR 指标。RR 是从原始视频序列中提取和比较一些特

征,即评估时也需要参考原始视频序列,只不过使用的是基于原始视频提取的特征信息,而非全部图像数据^[5],如文献^[6],通过参考部分图像特征来对视频质量进行评估。不管是 FR 还是 RR,都需要将用户接收到的视频数据和原始视频数据传输到评估服务器中进行视频质量评估,该过程成本高昂且耗时。另一方面,运营商获取原始视频数据也相对困难。NR 估计视频质量时则不需要原始视频序列,只需要分析通过客观测量得到的与视频质量有关的指标对视频质量进行估计。因此,对于运营商来说,采用 NR 方法估计视频质量显然比 FR 和 RR 更可行。

深度报文检测(deep packet inspection, DPI)系统通过对网络关键接口的流量和报文内容进行检测分析,根据策略对流量进行过滤控制,实现信令面和用户面消息的采集,能够对用户上网行为产生的信息进行过滤、采集。该系统分为 3 层架构,其中采集层和解码层负责数据采集、流量分析、日志合成,一般以各种数据记录方式存储在解码层的数据库内,这种数据即用户话单数据(extend data record, XDR),其中 X 代表呼叫/事务/会话^[7]。因此,例如,具体地,呼叫详细记录(call detail records, CDR)包含关于提供给最终用户的网络和服务特性的大量数据,这些海量的数据被 DPI 设备采集并存储在电信运营商大数据系统中。国际电信联盟电信标准化部门(International Telecommunication Union for Telecommunication Standardization Sector, ITU-T)^[8]建议规定了 CDR 的结构和内容。应用层主要对 XDR 记录数据进行计算、整理、统计,合理组织和存储数据,并进行呈现。电信运营商对 XDR 数据源转换处理后得到可以直接识别的字段信息,内容包括时间、网络类型、国际移动用户识别码(international mobile subscriber identity, IMSI)、移动用户号码(mobile subscriber international ISDN/PSTN number, MSISDN)等 100 多个字段^[9],也正是本文所使用的原始数据。

由于 XDR 数据的庞大规模和低价值,长期以来,针对 XDR 数据分析的 OTT 视频质量评估研究工作相对欠缺且计算存储成本高昂。本文提出了一种

基于 XDR 数据的无参考的网络视频质量评估的方法,在原始 XDR 数据的基础上进行数据加工,从海量 XDR 数据中提取出与视频质量强相关性的少量信息,将大规模、低价值的 XDR 话单数据转化为高价值、小规模的视频质量特征信息,有利于后续人工智能算法的应用和视频业务质量评价,降低下一步数据挖掘的资源成本,提升下一步机器学习的输入样本质量和输出模型的准确性。

1 相关工作

目前对网络视频业务质量的 NR 评估方法主要有以下 3 类。

1.1 流量探针分析方法

在网络关键节点(如城域网路由器)部署流量探针,对网络中的视频流进行采集、分析,利用 DPI 或深度/动态流检测(deep/dynamic flow inspection, DFI)技术直接从网络报文序列中提取视频质量指标,完成质量评估,如文献[10]。但是现有的流量探针分析方法存在一些缺陷,如必须在网络中部署流量采集探针设备,成本高昂,在已经存在统一 DPI 的前提下,再次部署流量采集设备显得过于浪费;而且单台探针设备处理能力有限,在大规模流量环境下会出现丢包、资源耗尽等现象,导致分析功能失效。

1.2 APP 上报信令分析方法

各类视频终端 APP 软件在播放过程中,会向视频业务服务平台报送业务质量信息,例如播放起止时间、卡顿次数、卡顿时长占比等,对这些信令进行采集、提取、解析,可获得较为真实的视频业务质量评价结果。目前部分运营商已采取此类方法分析常见视频网站的用户体验质量。但是采用该方法时,信息的提取依赖于终端视频 APP 软件上报给视频服务平台的信令信息,各个视频服务 APP 分别设计了自己的信令格式,需要针对每一类 APP 单独定制信令解析规则,工作量大。一旦 APP 软件升级或启用数据加密传输,则原有的信令解析规则不再可用,导致对该类视频业务的质量分析全部失效。

1.3 XDR 话单大数据分析方案

在移动统一 DPI 系统中,提取与视频业务相关的 XDR 原始话单,结合已知的视频质量标签作为训练样本,采用各类机器学习算法进行训练,以获得从 XDR 原始话单数据到视频业务质量之间的映射模型。例如文献[10]给出一种基于 XDR 分析互联网业务质量的方案,即在服务器集群的架构上组建一

套系统对 XDR 数据进行采集、存储、转换处理,再输出数据到前台应用程序,做深入分析优化,能准确定位互联网业务指标异常变化的原因,提高互联网业务质量分析工作效率。然而,XDR 数据具有数据量大,数据粒度小的特点,视频服务进程通常对应于数十条 XDR 记录,且 XDR 记录中的大多数信息是统计数据,并不直接包含与视频服务质量相关的信息。所以问题的关键就是如何合理地利用这些 XDR 数据,但现有研究[11-12]仅提出了概念框架和思路,缺乏从 XDR 中挖掘视频服务质量的具体实施方法。

2 方法介绍

2.1 基本思想

视频业务是一个跨越一定时间周期的连续过程,其质量特征也表现为一个完整过程中不同时段的质量。因此,对视频业务质量的刻画,应该对视频业务的全过程进行分时段连续评价,最终将不同时段的质量评价结果进行聚合,获得完整的质量刻画指标。虽然单个 XDR 话单记录仅仅覆盖一个网络会话且不足以刻画整体业务质量,但多个 XDR 话单在时间和空间层面存在耦合,利用该耦合关系可实现 XDR 话单的聚合和关联挖掘:覆盖同一时段的多个 XDR 话单记录共同构成了该时段的视频业务行为,多个 XDR 话单的同—指标值共同构成了该时段的视频业务指标值。本文以原始 XDR 话单为输入,实现高价值质量数据的生产,主要思路可概括为:将以会话为单位的 XDR 话单以时间和空间方式进行关联,形成为以时间窗口为单位的切片记录,然后将多个切片记录进行汇聚,形成对视频业务过程的总体质量记录。

如图 1 所示,横坐标表示时间,一条水平线代表

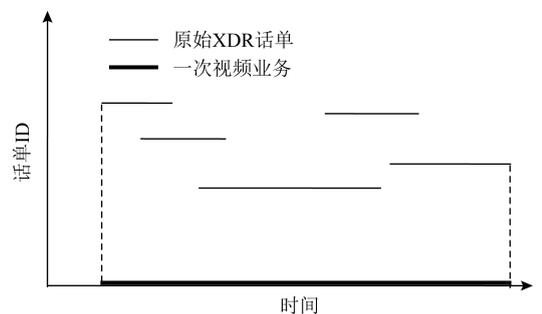


Fig. 1 Relationship between video service and XDR record

图 1 视频业务和 XDR 数据的关系

一条 XDR 记录,多个相互重叠的 XDR 记录集合表示一次视频业务。

2.2 视频质量特征信息的选择

定义网络视频质量特征信息为:用于记录某一时段内,某个移动上网用户在观看视频过程中的多个质量特征指标的数据记录.本文提出的视频质量特征信息如表 1 所示,包括但不限于 5 个基本字段,12 个统计字段,3 个计算字段,3 个推测字段。

Table 1 Filed of Feature Information

表 1 质量特征信息所有字段

字段分类	字段名称	字段含义
基本字段	BEGINTIME/s	统计开始时间,精确到秒
	ENDTIME/s	统计结束时间,精确到秒
	MSISDN	用户账号
	VT	视频应用类型
	USERIP	用户 IP 地址
统计字段	TCP_COUNT/个	往来包数
	SERVER_COUNT/个	下载包数
	SERVER_BYTES/B	下载字节数
	SERVER_LOST/个	下载丢包数
	SERVER_RETRANS/个	下载重传包
	CLIENT_COUNT/个	上传包数
	CLIENT_BYTES/B	上传字节数
	CLIENT_LOST/个	上传丢包数
	CLIENT_RETRANS/个	上传重传包数
	DELAY/ms	端到端时延
	JITTER/ms	端到端时延抖动
	MTU/B	下载最大包长
计算字段	AVG_IP_INBPS/bps	平均下载速率
	SERVER_PLR/%	下载丢包率
	CLIENT_PLR/%	上传丢包率
推测字段	MR/bps	推测码率
	FASTPUSH_FLAG	推测快推质量标记
	LABEL_SCORE	推测视频下载质量分数

基本字段表示一次视频服务的基本信息,如时间信息、用户识别信息等;统计字段可直接从原始 XDR 数据中经过累加统计获取,是一次视频服务的简单统计量,本文不过多讨论;计算字段则是对统计字段进行关联计算后得出的与视频质量强相关的信息。

计算字段能够较好地刻画视频质量,其中,丢包率(packet loss rate, PLR)被认为是刻画视频质量最重要的参数.在 IP 网络的传输中,视频文件通常被打包成固定大小的分组,不同分组的分片信息不同,丢包会直接影响其所在帧的解码^[10].为了获得更准确的结果,我们需要同时考虑客户端和服务器的双向 PLR。

另外,平均下载速率也被认为对视频质量有很大影响.一般而言,如果下载速度非常慢,视频质量将不会很好^[11],因此平均下载速度也可以用作视频质量的重要参数。

在网络视频播放中,往往采用一边下载一边播放的方式.视频服务器和客户端均会控制下载速率和缓存大小,使之与播放所需要的码率匹配,并保持适当的视频播放缓冲.本文提出一种“快推质量标记”参数,其中快推是指在视频刚刚播放的时候,为了减少缓冲时间,往往使用多线程或不限制下载速率的方式,尽快填充客户端的播放缓冲区,因此视频刚刚开始播放的时候,下载速率可能远高于观看过程中的正常码率.表 1 推测字段中的推测快推质量标记就是用于表示快推是否成功.显然,该参数对视频质量有着很大的影响。

2.3 按时间窗口切片的信息提取

图 2 是从原始 XDR 话单数据中提取出特征信息的过程,主要可分为 2 步:第 1 步是将 XDR 话单数据按时间分片得到时间窗口记录;第 2 步将时间窗口记录进行聚合,得到一条汇总记录。

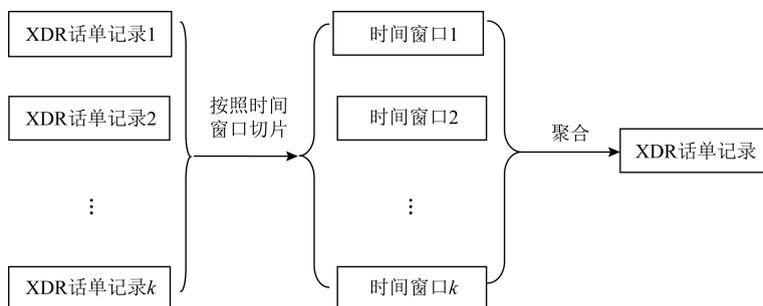


Fig. 2 Procedure of information extraction

图 2 信息提取过程

2.3.1 切片

本步骤是对以 TCP 会话为单位的视频类 XDR 话单进行切片加工,输出多条以固定时间窗口为单位的 时间窗口记录;时间窗口记录的格式如表 1 所述.

1) 将同一用户的多条视频类 XDR 话单记录进行时间聚类,形成 XDR 话单群(XDR records cluster),如图 3 所示,其中每一条线段表示一条视频类 XDR 话单记录,线段的 2 个端点分别代表该 XDR 话单的起止时间,一个 XDR 话单群应满足以下条件:话单群中任意一条 XDR 话单,均能在本群中找到至少一条起止时间与之部分或全部重合的 XDR 话单,不同 XDR 话单群之间存在明显的时间间隔.

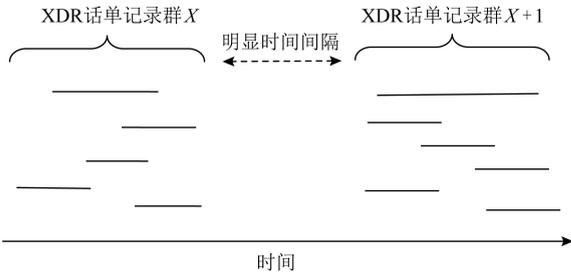


Fig. 3 XDR records cluster
图 3 XDR 话单群示意图

将 XDR 话单群 X 定义为

$$X = \{r_1, r_2, \dots, r_n\},$$

其中 r_i 表示第 i 条话单(话单按照开始时间排序),其开始时间为 b_i ,结束时间为 e_i ,对每个 XDR 话单群 X 而言,其中任意一条 XDR 记录都能在当前话单群中找到至少一个与其时间重叠的 XDR 记录,即 $\forall r_i \in X$ 满足: $\exists r_j \in X$,

$$(b_i > b_j \ \&\& \ b_i < e_j) \ \parallel \ (b_j > b_i \ \&\& \ b_j < e_i).$$

定义 XDR 话单群 X 的开始时间 bt_x :

$$bt_x = \min\{b_i | r_i \in X\};$$

结束时间 et_x :

$$et_x = \max\{e_i | r_i \in X\};$$

则对于不同的 XDR 话单群 X 和 Y , X 和 Y 之间存在明显时间间隔,即若 $bt_x < bt_y$,则有

$$bt_y - et_x > T,$$

其中 T 为给定的 2 个话单群之间最小时间间隔,在本文中,我们将其取值为 5 s.

2) 将同一 XDR 话单群 X 中的话单切片为多条时间窗口记录,形成时间窗口记录列表.

如图 4 所示,我们将一个话单群的总持续时长(最大结束时间与最小开始时间的差值)平均切分为多个固定时长的时间窗口,然后对同一 XDR 话单群中的所有话单进行数据切片,为每个时间窗口生成一条唯一的时间窗口记录;该时间窗口记录中的各个统计字段来自于与该时间窗口部分或全部重合的多条 XDR 话单的切片统计结果.

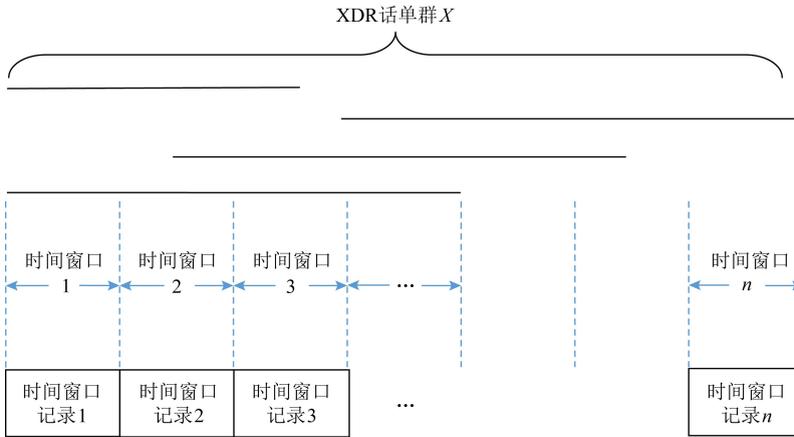


Fig. 4 Time window
图 4 时间窗口

将一个话单群的总持续时长定义为

$$dt_x = et_x - bt_x. \tag{1}$$

若将 X 平均切分为 n 个时间窗口,则第 i 个窗口 wnd_i 的开始时间 wnd_bt_i 满足

$$wnd_bt_i < bt_x + (i-1) \times \frac{dt_x}{n}, 1 \leq i \leq n,$$

结束时间 wnd_et_i 满足

$$wnd_et_i < bt_x + i \times \frac{dt_x}{n}, 1 \leq i \leq n.$$

定义每个窗口 wnd_i 的持续时间为 wnd_dt ,此时长度为 $\frac{dt_x}{n}$.

在步骤 2)中涉及到一些统计字段的计算,在此以 `SERVER_COUNT` 字段为例,说明时间窗口记录

中的各个统计字段的计算方法。

① 定义 $SERVER_COUNT_{i,k}$ 为 XDR 话单群 X 中第 i 条 XDR 话单记录 r_i 中的 $SERVER_COUNT_i$ 值在第 k 个时间窗口记录 wnd_k 中的统计分量值, 则 wnd_k 中的 $WND_SERVER_COUNT_k$ 值为该话单群中所有 XDR 话单记录的 $SERVER_COUNT_{i,k}$ 分量值之和, 即

$$WND_SERVER_COUNT_k = \sum_{r_i \in X} SERVER_COUNT_{i,k}. \quad (2)$$

② 第 i 条 XDR 话单记录 r_i 在第 k 个时间窗口记录 wnd_k 中的 $SERVER_COUNT_{i,k}$ 分量值计算方法为:

取该 r_i 的 $SERVER_COUNT_i$ 字段值, 除以该记录起止时间 ($ENDTIME - BEGINTIME$) 之差 (单位为 s), 然后乘以该 XDR 话单记录与本时间窗口的时间重合长度 $Toverlap$ (单位为 s), 表示为

$$SERVER_COUNT_{i,k} = \frac{SERVER_COUNT_i \times Toverlap}{ENDTIME_i - BEGINTIME_i}. \quad (3)$$

③ 第 i 条 XDR 话单记录 r_i 与某个时间窗口 wnd_k 的时间重合长度 $Toverlap$ 计算方法为

$$Toverlap = \begin{cases} wnd_dt, b_i \leqslant wnd_bt_k \ \&\& \ e_i \geqslant wnd_et_k; \\ \min(wnd_et_k, e_i) - b_i, b_i \geqslant wnd_bt_k \ \&\& \\ b_i \leqslant wnd_et_k; \\ e_i - wnd_bt_k, b_i \leqslant wnd_bt_k \ \&\& \\ e_i \leqslant wnd_et_k; \\ 0, \text{ other.} \end{cases} \quad (4)$$

除了统计字段外, 对 3 个计算字段的获得方法进行说明, 3 个计算字段的值为同一记录中的其他统计字段值的计算结果, 计算方法分别为

平均下载速率:

$$AVG_IP_INBPS = \frac{8 \times SERVER_BYTES}{ENDTIME - BEGINTIME}. \quad (5)$$

下载丢包率:

$$SERVER_PLR = \frac{100 \times SERVER_LOST}{SERVER_COUNT + SERVER_LOST}. \quad (6)$$

上传丢包率:

$$CLIENT_PLR = \frac{100 \times CLIENT_LOST}{CLIENT_COUNT + CLIENT_LOST}. \quad (7)$$

2.3.2 聚合

本节是将 2.3.1 节输出的时间窗口特征信息记录进行聚合, 输出一条汇总记录, 然后, 计算并回填汇总记录中的 3 个计算字段. 汇总记录的格式同样如表 1 所述. 计算方法如下:

1) 将 2.3.1 节中输出的同一 XDR 话单群的时间窗口记录列表汇聚为一条汇总记录, 即为每个 XDR 话单群生成唯一的一条汇总记录, 其中的各个统计字段来自于时间窗口记录列表中所有记录的统计结果.

2) 计算汇总记录中的推测字段.

推测码率 MR 的表达式为

$$MR = \begin{cases} AVG_IP_INBPS, SERVER_PLR = 0; \\ AVG_IP_INBPS \times \left(1 + \frac{SERVER_PLR}{200}\right), \\ SERVER_PLR > 0. \end{cases} \quad (8)$$

推测快推质量标记 $FASTPUSH_FLAG$ 字段表达式为

$$FASTPUSH_FLAG = MIN(2V, 1), \quad (9)$$

其中 V 是 $SERVER_COUNT$ 字段的离散系数, $FASTPUSH_FLAG$ 取值为 V 的 2 倍值与常数 1 的较小值.

推测视频下载质量分 $LABEL_SCORE$ 字段的表达式为

$$LABEL_SCORE = \max\left\{100 \times \frac{AVG_IP_INBPS}{MR}, 100 \times FASTPUSH_FLAG\right\}. \quad (10)$$

根据上述过程, 可将多条 XDR 数据记录融合为一条视频观看记录, 从而在大规模减少数据集规模的前提下, 实现高价值指标信息的提取, 为进一步机器学习提供优质数据集.

3 实验结果和分析

3.1 GBDT 模型

在实验中, 我们使用梯度提升迭代决策树算法 (gradient boosting decision tree, GBDT)^[13] 模型来完成后续机器学习过程, 弥补人工特征生成的局限性. GBDT 算法是一种用于数据分类和回归的集成学习^[14] 算法. 该算法是由多棵类似分类回归树 (classification and regression tree, CART)^[15] 的决策树组成, 将算法中所有决策树的输出结果累加起来就是 GBDT 的最终输出结果, 它在被提出之初就和支持

向量机(support vector machine, SVM)^[16]一起被称为泛化能力较强的算法,并都是数据分析中常用的学习算法.GBDT模型的数学表达式可表示为

$$f(x) = \sum_{m=1}^M \beta_m b(x; \gamma_m), \quad (11)$$

其中, $b(x; \gamma_m)$ 代表第 m 棵决策树, β_m 是第 m 棵决策树的权重.

我们选择GBDT算法模型的原因是,该模型中的每棵决策树都是回归决策树,因此它可以较好解决数据回归问题并具有高检测精度的特性.此外该算法可有效处理异常点,还能在一定程度上避免模型过拟合问题.

3.2 数据集

我们使用的数据集是包括8102个视频会话的XDR话单数据集,数据来自于某地区网络真实OTT业务匿名化数据集,以标准的XDR格式提供,每个视频会话有预先标记好的视频质量主观分,该主观分的表达式为

$$Lable = 100 - lag, \quad (12)$$

其中 lag 是使用APP上报信令分析方法计算出的视频卡顿时长占比,见本文1.2节.一般认为网络用户在观看视频时,卡顿时长占比越高则视频质量越差.

作为对比,本文针对原始XDR话单数据使用同样的模型进行训练和预测,该原始XDR话单数据包括了24715条XDR话单.

3.3 评估指标

我们参考每个视频会话中预先标记好的视频质量主观分(式(12)),使用3个指标来评估本文方法的可行性,分别为:

1) 平均绝对误差(mean absolute error),记为 mae .是绝对误差的平均值,能够直接反映视频质量评分误差的真实情况,定义如下:

$$mae = \frac{1}{N} \sum_{i=1}^N |Y_i - \hat{Y}_i|.$$

2) 均方差(mean squared error),记为 mse .是最简单,应用最广泛的图像评价方法之一,可以很好地衡量视频质量评分和视频质量主观分之间的偏差,定义如下:

$$mse = \frac{1}{N} \sum_{i=1}^N (Y_i - \hat{Y}_i)^2.$$

3) 平均偏差误差(mean deviation error),记为 mde .可以较好地反映视频质量评分与视频质量主观分之间的平均差异,定义如下:

$$mde = \frac{1}{N} \sum_{i=1}^N \frac{|Y_i - \hat{Y}_i|}{Y_i}.$$

其中, N 代表总样本数, Y_i 为第 i 个样本的标签值,即主观评价分, \hat{Y}_i 代表样本的估计值.

以上3个指标是常用的衡量观测值与真值之间偏差的参数,能够较好地反映出本文中视频质量评分和主观分之间的差异.本文通过比较提取后数据和原始数据在不同学习率下的 mae , mse , mde ,来说明该提取方法的高效性和有用性.除此之外,本文还选用常用性能参数 $PLCC$, $SROCC$, $KROCC$ 来辅助评价算法的好坏.这3个指标够较好地刻画数据之间的相关性.

3.4 结果及分析

本文分别使用原始数据、本文提取的特征信息进行训练和用户感知质量评估,采用0.1~0.9的不同学习率来训练模型,在每次训练过程期间训练数据是随机选择的.结果如表2、表3所示.很显然,学习率越高,得到的结果越准确,其原因是GBDT模型的估计精度取决于训练集的大小.

表2是直接以原始XDR话单数据为输入对用户感知质量评估的结果.可以看到,在学习率较低的情况下,基于原始XDR的视频质量评估结果存在较大误差;虽然随着学习率的增加, mae 和 mse 确实可以有效降低,但 mse 仍然无法降低至可以接受的范围.基于 mse 对异常值敏感,而 mae 对异常值不敏感这一特性,基本可以得出结论:该方法对一些异常值的预测能力较差.而事实上,卡顿导致的视频质量下降恰好是属于这些异常值,因此使用原始XDR话单数据对视频质量进行评估,很难取得良好效果.

Table 2 Experimental Result of Initial Data at Different Learning Rates

表2 使用原始数据在不同学习率下的实验结果

学习率	mae	mse	mde
0.1	14.45	2601.02	0.221
0.2	12.42	2012.34	0.211
0.3	9.224	1601.12	0.205
0.4	6.231	1389.45	0.193
0.5	5.678	1261.07	0.186
0.6	5.124	823.46	0.182
0.7	4.52	470.24	0.177
0.8	4.321	314.97	0.172
0.9	4.32	245.2	0.17

表 3 是以本文提取的特征信息为输入、使用同样的算法模型和参数的用户感知质量评估结果,可以看到,在不同的学习率中, mae , mse , mde 均明显低于原始数据实验结果;其中 mse 虽然随着学习率的变化而显著变化,但总体保持在较低水平。

Table 3 Experimental Result of Extraction Data at Different Learning Rates

表 3 使用本文提取的特征信息在不同学习率下的实验结果

学习率	mae	mse	mde
0.1	1.923	51.005	0.018 9
0.2	1.689	47.921	0.017 6
0.3	1.687	44.209	0.017 2
0.4	1.557	40.142	0.017 1
0.5	1.558	38.109	0.016 9
0.6	1.55	35.218	0.016 9
0.7	1.52	33.983	0.016 8
0.8	1.522	31.236	0.016 6
0.9	1.492	29.974	0.016 4

进一步将基于两类数据的评估结果进行对比,如图 5~7 所示,可以看出,采用本文方法提取的质量特征信息,只需要很少的学习率即可实现较高的评估准确性.本文提取的质量特征数量仅为原始 XDR 字段数量的 6.7%,可以看出原始 XDR 数据不仅数据量大,而且直接利用价值较低.而在本文方法进行信息加工后,不仅能显著减小数据规模,还能大幅度降低学习率并显著提升质量评估结果的准确性。

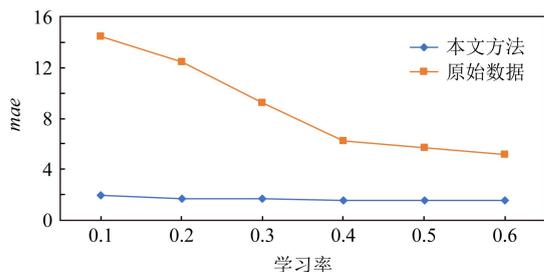


Fig. 5 Comparison of mae between two data sets

图 5 2 个数据集平均绝对误差对比

进一步,选用图像质量分析领域常用的评价参数 $PLCC$, $SROCC$, $KROCC$ 来辅助评价本文方法的效果,这 3 个参数均为相关系数,取值范围是 $[-1, 1]$,越接近 1 表示相关性越强.由于目前已知的文献中,尚缺乏可验证的从 XDR 数据中提取特征的方法介绍,因此本文不做类似方法的比对.分别

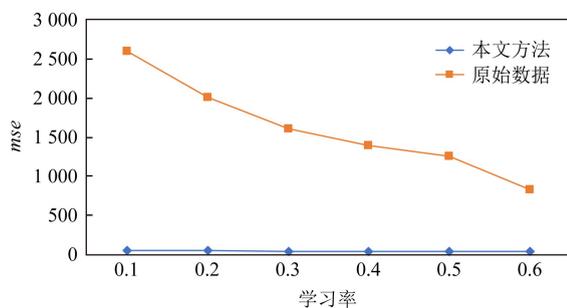


Fig. 6 Comparison of mse between two data sets

图 6 2 个数据集均方差对比

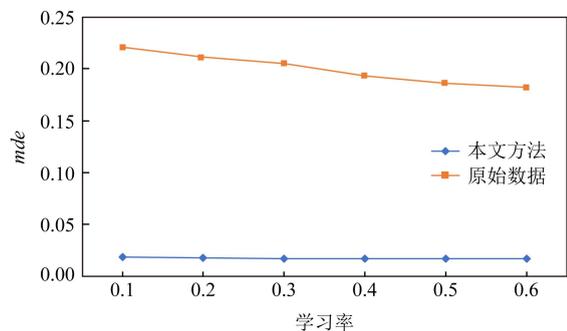


Fig. 7 Comparison of mde between two data sets

图 7 2 个数据集平均偏差误差对比

将基于本文提取数据、原始 XDR 数据的视频质量评估结果与主观评分进行相关性分析,获得的相关系数数值如表 4 所示:

Table 4 Performance of the Proposed Method

表 4 本文算法性能参数

方法	$PLCC$	$SROCC$	$KROCC$
本文方法	0.611 9	0.791 3	0.691 0
原始数据	0.316 1	0.255 1	0.233 7

通过表 4 可以看出,使用本文方法的视频质量评价结果表现出了与主观评价结果的明显相关性,其 $PLCC$, $SROCC$, $KROCC$ 指标均明显优于基于原始 XDR 话单的评价结果。

4 结 论

针对 OTT 视频业务质量评估,本文提出了一种基于 XDR 的信息提取方法.首先将多条原始 XDR 话单进行聚合,将以会话为单位的 XDR 话单以时间和空间方式进行关联,形成为以时间窗口为单位的切片记录,然后将多个切片记录进行汇聚,形成对视频过程的总体质量记录。

以本文提取的信息作为机器学习算法输入,能够得到较为精确的视频质量评估结果,与原始 XDR 数据相比,本文生成的特征数据与真实视频质量具有更高的相关性,数据量更小,所需学习率更低,评估结果也更为准确。

参 考 文 献

- [1] Zhao Gang, Yang Congcong. Research on quality assurance of OTT video [J]. Information & Communications, 2016 (8): 246-247 (in Chinese)
(赵纲, 杨聪聪. OTT 视频业务质量保障研究[J]. 信息通信, 2016 (8): 246-247)
- [2] ITU-R. Recommendation BT.500-11: Methodology for the Subjective Assessment of the Quality of Television Pictures [S/OL]. ITU Telecom Standardization Sector of ITU, 2002 [2020-03-10]. <https://www.itu.int/rec/R-REC-BT.500-11-200206-S/en>
- [3] VQEG. Final Report From the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment [EB/OL]. 2000 [2020-03-10]. <http://www.vqeg.org>
- [4] Zhou Wang, Bovik A C, Sheikh H R, et al. Image quality assessment: From error visibility to structural similarity [J]. IEEE Transactions on Image Processing, 2004, 13(4): 600-612
- [5] Goldstein E B. Cognitive Psychology: Connecting Mind, Research and Everyday Experience [M]. Wadsworth: Cengage Learning, 2014: 100-121
- [6] Li Wengjing, Luo Qian, Yu Peng, et al. Reduced-reference video QoE assessment method based on image feature information [C] //Proc of the 17th Asia-Pacific Network Operations and Management Symp APNOMS. Piscataway, NJ: IEEE, 2015: 519-522
- [7] Ye Shuxiang. High-speed rail user precision recognition algorithm based on XDR big data [J]. Information & Communications, 2018 (6): 226-227 (in Chinese)
(叶树祥. 基于 XDR 大数据的高铁用户精准识别算法[J]. 信息通信, 2018 (6): 226-227)
- [8] ITU-T. Recommendation Q. 825: Specification of TMN Applications at the Q3 Interface: Call Detail Recording [S]. ITU Telecom Standardization Sector of ITU, 1998 [2020-03-10]. <http://www.docin.com/p-587150115.html>
- [9] Zheng Yinghang, Zhong Qizhu, Luo Yaoman. Estimating Internet quality through XDR analysis [J]. Telecommunications Information, 2017 (7): 16-19 (in Chinese)
(郑颖航, 钟其柱, 罗耀满. 基于 XDR 分析互联网业务质量的方案[J]. 电信快报, 2017 (7): 16-19)
- [10] Wu Di, Jin Yuehui, Yang Tan. A probe system for analyzing transmission and playback quality of online videos [C] //Proc of IEEE Int Conf on Cloud Computing & Intelligence Systems. Piscataway, NJ: IEEE, 2014: 475-480
- [11] Cheng Jie, Geng Yan. Research and evolution of pause phenomena in mobile video services [J]. Designing Techniques of Posts and Telecommunications, 2017 (12): 11-15
- [12] Zhao Gang. Design and implementation of mobile video service quality analysis system [J]. Telecommunications Science, 2017, 33(7): 159-166
- [13] Friedman J H. Greedy function approximation: A gradient boosting machine [J]. Annals of Statistics, 2000, 29(5): 1189-1232
- [14] Polikar R. Ensemble learning [J]. Scholarpedia, 2009, 4(1): 2776
- [15] Breiman L I, Friedman J H, Olshen R A, et al. Classification and regression trees (CART) [J]. Encyclopedia of Ecology, 1984, 57(1): 582-588
- [16] Joachims T. Making large-scale SVM learning practical [J/OL]. Advances in Kernel Methods Support Vector Learning, 1998 [2020-03-10]. http://www.kernel-machines.org/publications/Joachims99/bibliography_exportForm



Huang Lisheng, born in 1975. PhD. Associate professor. His main research interests include network measurement, network security, network management.

黄鹏声, 1975 年生. 博士, 副研究员. 主要研究方向为网络测量、网络安全、网络管理等。



Ran Jinye, born in 1994. Received his master's degree from Computer Science at University of Electronic Science and Technology of China (UESTC) in 2020. His main research interests include network security and network anomaly detection.

冉金也, 1994 年生. 硕士. 主要研究方向为网络安全和网络异常检测。



Luo Jing, born in 1994. Master candidate in computer science at University of Electronic Science and Technology of China (UESTC). Her main research interests include network security and big data.

罗静, 1994 年生. 硕士研究生. 主要研究方向为网络安全和大数据。



Zhang Xiangyin, born in 1983, PhD. His main research interests include signal processing and distributed network system.

张翔引, 1983 年生. 博士. 主要研究方向为信号处理和分布式网络系统。