

众包直播中精彩片段的自动化识别^①



兰荣亨¹, 胡雨晗², 朱格², 田野², 朱明¹

¹(中国科学技术大学 信息科学技术学院, 合肥 230027)

²(中国科学技术大学 计算机科学与技术学院, 合肥 230027)

通讯作者: 田野, E-mail: yetian@ustc.edu.cn

摘要: 近年来, 基于众包的视频直播平台逐渐兴起, 以其丰富的观众-主播交互机制吸引广大用户观看. 针对直播平台的分析也随之成为流媒体服务领域的一个研究热点. 直播过程中精彩片段的自动提取对于标签生成、视频分类和内容推荐等方面而言至关重要, 然而现有的精彩片段检测大多围绕音频、视频数据本身展开, 如视频语义分析、音频情感感知等, 缺乏对用户交互属性的合理利用. 本文以斗鱼直播平台为例, 通过分析观众的发弹幕与送礼物行为, 提出了基于直播间弹幕数量时间序列和礼物价值时间序列的精彩片段自动化检测方法. 首先利用 z-score 方法检测序列高潮, 然后对高潮做样本标注和特征构建, 最后采用随机森林对序列高潮分类并识别内容高潮, 即精彩片段. 结果表明, 模型能够以较高的准确率完成精彩片段的自动化识别任务.

关键词: 众包直播; 精彩片段检测; 特征挖掘; 监督学习; 数据挖掘

引用格式: 兰荣亨, 胡雨晗, 朱格, 田野, 朱明. 众包直播中精彩片段的自动化识别. 计算机系统应用, 2019, 28(9): 219-224. <http://www.c-s-a.org.cn/1003-3254/7082.html>

Automatic Content Highlight Detection in Crowdsourced Live Video Streaming

LAN Rong-Heng¹, HU Yu-Han², ZHU Ge², TIAN Ye², ZHU Ming¹

¹(School of Information Science and Technology, University of Science and Technology of China, Hefei 230027, China)

²(School of Computer Science and Technology, University of Science and Technology of China, Hefei 230027, China)

Abstract: Crowdsourced live video streaming, which attracts vast number of users by its rich viewer-broadcaster interaction mechanism, has flourished and expanded over the past few years. The analysis of live video streaming platform has become a research hotspot in the field of streaming media services. Automatic extraction of highlights in live video streaming is crucial for tag generation, video classification and content recommendation. However, the existing highlight detection analysis mostly focuses on audio or video data itself, such as video semantic analysis, audio emotional perception, etc., lacking the rational use of user interaction attributes. In this study, we take Douyu live video streaming platform as a case study. Through analysis of the viewer's danmu posting and virtual gift donating behavior, we propose an automatic content highlight detection method based on the time series of danmu quantity and virtual gift value in the broadcasting. Firstly, we use z-score method to detect the sequence highlights, then we conduct highlight sample labeling and feature constructing. Finally, Random Forest is used to classify sequence highlights and identify the content highlights. The results show that the model we proposed can accomplish the task of automatic content highlight detection with high accuracy.

Key words: crowdsourced live video broadcasting; highlight detection; feature mining; supervised learning; data mining

① 基金项目: 国家自然科学基金 (61672486) 国家科技重大专项 (2017ZX03001019-004)

Foundation item: National Natural Science Foundation of China (61672486); National Science and Technology Major Program (2017ZX03001019-004)

收稿时间: 2019-03-09; 修改时间: 2019-04-04; 采用时间: 2019-04-10; csa 在线出版时间: 2019-09-05

作为互联网时代的杀手级应用,视频流媒体服务在过去的20多年正不断驱动着互联网技术的发展与进步^[1]。近年来,基于众包的视频直播逐渐兴起,吸引了大量用户,相应的催生了全民直播这个巨大的娱乐产业。国外知名的早期直播平台有Twitch.tv, Youtube Live等,国内也有如斗鱼、虎牙、战旗等大量直播平台,在2016年甚至出现了千播大战的局面^[2]。

直播过程中精彩片段(Highlight)的自动提取对于标签生成、视频分类和内容推荐等方面而言至关重要,然而现有的精彩片段检测大多围绕音频、视频数据本身展开^[3-7]。Yao T^[3]等人根据视频Highlight片段和非Highlight片段的视频帧差异,使用深度学习方法学习了两种片段特征;Wang J^[5]等人利用原始视频和音频生成关键字序列,并使用HMM模型捕捉了视频Highlight。也有一些工作结合了用户行为做Highlight检测,如Zhao Y^[8]等人根据用户观看视频时的跳转行为,利用CUSUM和MB-GT算法检测了视频点播中的精彩片段。针对众包直播系统,也有一些文献进行了相关研究。Li^[9]通过研究网络直播系统的访问日志,提出了一系列用于直播持续时间,用户活动,用户的到来与离开时间建模的模型。此外还有一系列研究直播系统生态、架构设计、用户行为等的工作^[10-12]。

国内的直播平台,基本都引入了送礼机制,即观众可以通过购买平台提供的虚拟礼物来打赏自己喜欢的主播,而平台则以抽成的方式分享观众打赏的礼物。观众的打赏,成为了主播和平台的主要收入来源之一。所以,在海量带宽、技术开发、商业竞争等各种成本的巨大压力下,提升服务竞争力和用户变现能力成为了平台发展的重中之重。

本项研究对当前众包直播生态做了大量工作,尤其对于虚拟礼物赠送机制,观众行为,频道流行度等问题进行了一系列的建模分析^[13,14]。作为研究后续,本文从识别精彩片段的视角,提出了基于直播间弹幕数量时间序列和礼物价值时间序列的精彩片段自动化检测方法,并讨论了如何将其应用于优化内容推荐和优质主播发掘,在提升用户体验的同时提升用户变现能力。

1 平台简介与研究背景

1.1 平台简介

作为国内最大的直播平台之一,斗鱼(Douyu.com)^[15]已经从最初的游戏直播,发展成为涵盖游戏、娱乐、

户外、体育等各个方面的泛娱乐直播平台。利用平台提供的服务,用户可以随时随地在网络上直播并分享自己的生活。

和大多数国内平台类似,斗鱼提供了丰富的交互机制,来拉近主播与观众的距离。观众在观看直播的同时,可以实时在直播间内发送评论,评论会在屏幕上划过,称之为弹幕;对于喜欢的主播,观众也可以花钱购买平台提供的虚拟礼物来打赏主播。平台的虚拟礼物从价值0.1元到上千元不等,满足了观众的各种打赏需求。当观众在直播间送出了价值500元及500元以上的礼物时,平台会对这条送礼消息在平台所有直播间进行广播并派送虚拟道具(鱼丸),观众看到广播消息可以点击进入该直播间抢夺鱼丸。根据平台机制,观众可以通过发送弹幕消息来提高抢到鱼丸的概率。作为虚拟礼物的一种,鱼丸也可以用来打赏主播。所以,在观众送出高价值礼物后,往往会导致该直播间的弹幕数量激增。

利用斗鱼官方提供的API^[16],本研究采集了自2016/11/22至2016/12/19连续四周的数据。数据包含近750万个观众发送的2.5亿条弹幕数据和送出的689万个礼物数据,以及24万个主播产生的近179万条开播记录。经过简单统计,在这四周内斗鱼观众总共送出了价值近4700万元的礼物。

1.2 研究背景

在短时间内,如果直播间出现礼物价值和弹幕数量的激增,则称之为一个高潮。一个主播有了精彩的表演或者游戏操作,直播间的观众往往会发送大量的弹幕夸赞主播,也可能会送出高价值礼物来打赏主播,这样产生的高潮称为内容引起的高潮,简称内容高潮,也即本研究需要识别的精彩片段。另一方面,如上文所述,高价值礼物的送出会导致弹幕数量的激增,一些特殊的人群,如主播的经纪人,主播的忠实观众,可能会通过送出高价值礼物来帮助主播吸引人气。所以从弹幕数量上看,对于一些直播间可能会产生直播气氛热烈高涨的假象,这样产生的高潮称为礼物引起的高潮,简称礼物高潮。

本研究的目标,是从所有高潮中识别内容高潮,即精彩片段。

2 高潮检测与特征构建

2.1 高潮检测

根据从斗鱼直播平台采集到的弹幕数据和送礼数

据,对于一次开播 (session) s , 可获得其弹幕数量秒量级时间序列, 记为:

$$D_s = \{d_1, d_2, \dots, d_{l_s}\} \quad (1)$$

其中 d_i 为所有用户在此次开播的第 i 秒发出的总弹幕数量.

使用 z-score^[17]方法, 计算得到弹幕数量序列的 z-score 序列:

$$Z_s^d = \{z_1^d, z_2^d, \dots, z_{l_s}^d\} \quad (2)$$

其中 $z_i^d = \frac{d_i - \mu_s^d}{\sigma_s^d}$, μ_s^d 和 σ_s^d 分别为弹幕数量时间序列 D_s 的均值和标准差.

由 Z_s^d 序列进一步筛选出离散时刻序列 $T_s^d = \{i \mid |z_i^d| > \theta_p\}$, 其中阈值 $\theta_p > 0$, 可根据置信系数 p 动态调节; 对于离散序列 T_s^d 中的每一个时刻 i , 前后各扩充 τ 秒, 即得到一个小时间区间 $i_\tau = [i - \tau, i + \tau]$; 合并所有有交集的 i_τ , 得到若干大的连续时间区间. 每一个大的时间区间内都包含大量弹幕, 定义为弹幕激增 (burst) 时间区间集, 记为 DB_s .

为了便于理解, 图 1 和图 2 展示了离散的点通过扩展、合并得到大时间区间的过程.

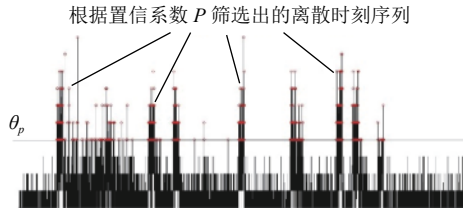


图 1 筛选获得离散时刻序列 T_s

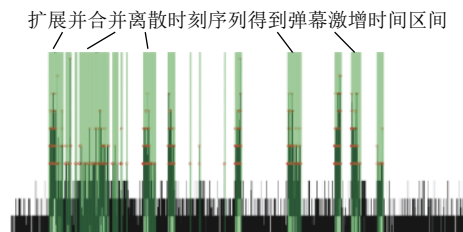


图 2 离散时刻序列扩展与合并

弹幕激增时间区间集 DB_s 即由图 2 中用阴影标识的一些连续的小时间区间组成.

对于礼物价值序列 $G_s = \{g_1, g_2, \dots, g_{l_s}\}$ 作同样的处理, 可以得到礼物价值激增时间区间集 GB_s . 同样的, GB_s 中的每一个小时间区间内, 都包含了高额的礼物价值.

最后, 对于礼物激增时间区间集 GB_s 中的每一个区间 a_i , 若在弹幕激增时间区间集 DB_s 中存在某个区间

b_j 与之有交集, 即

$$a_i \cap b_j \neq \emptyset \quad (3)$$

则合并这两个时间区间, 得到的新区间 h_i :

$$h_i = a_i \cup b_j \quad (4)$$

新的时间区间 h_i 中既有弹幕激增又有礼物价值激增, 即为前文所定义的高潮, h_i 定义为高潮时间区间. 开播 s 所有高潮时间区间构成的集合定义为高潮时间区间集 H_s :

$$H_s = \{h_i\} \quad (5)$$

图 3 展示的是一个包含弹幕激增和礼物价值激增的高潮示例样本.

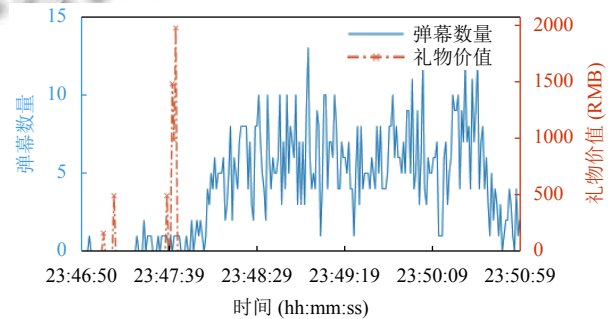


图 3 高潮示例样本

算法 1 描述了上述高潮检测过程.

算法 1. 高潮检测算法

- 1) 获取开播 s 的弹幕数量时间序列 D_s 和礼物价值时间序列 G_s ;
- 2) 计算 D_s 和 G_s 的 z-score 序列 Z_s^d 和 Z_s^g ;
- 3) 根据置信系数 p 筛选得到离散时刻序列 T_s^d 和 T_s^g ;
- 4) 分别对 T_s^d 和 T_s^g 中的每一时刻前后扩充 τ 秒得到连续时间区间并合并有交集的区间, 得到弹幕激增时间区间集 DB_s 和礼物价值激增时间区间集 GB_s ;
- 5) 合并 DB_s 和 GB_s 中有交集的时间区间, 即获得高潮时间区间集 H_s .

2.2 训练样本构建

从数据库中随机选择一定数量开播, 获得其弹幕数量和礼物价值时间序列数据, 应用上述高潮检测算法, 得到了由 801 个高潮构成的高潮时间区间集 H (简称高潮集), 作为后续模型的训练样本.

根据前文论述, 高潮有两种, 一是由于一些观众纯粹为了通过给主播送高价值礼物 (如“火箭”), 来提高直播间人气引起的, 即由礼物引起的高潮, 这种高潮里的弹幕内容基本上是为了提高获得虚拟道具 (鱼丸) 而发送的与直播内容无关的弹幕, 如“鱼丸大军驾到, 都闪开”、“大家好, 我是新来的喷子, 是直接喷还是走程

序”等;二是由于主播的精彩表演或游戏里的精彩操作引起,即由内容引起的高潮,这种高潮里的弹幕消息虽然也有和前者一样和内容无关,而只是为了抢夺虚拟道具而发的弹幕,但却有着大量和内容相关的诸如称赞主播、和主播积极交互的有意义的弹幕。

基于此观察,本研究采取人工查看高潮持续时间内的弹幕内容的方式,来标注高潮是由礼物引起,还是由内容引起,从而构建一个有监督的训练集。为了降低主观误差,本研究请了3个志愿者观看所有样本的弹幕内容并独立作出标注,最后综合3位志愿者的标注结果,以少数服从多数的原则作出最后标注。最终样本构成如表1所示。

表1 样本构成

	数量
内容高潮	348
礼物高潮	453
总样本	801

2.3 特征构建

为了更好的训练模型,本研究先后为每一个高潮构建了多达20个特征。经过模型测试,得到了如下7个对模型贡献最大的特征:

AvgGiftByUser: 观众在高潮持续时间内送出的人均礼物价值,单位:元/人;

AvgGiftByTime: 观众在高潮持续时间内送出的时间平均礼物价值,单位:元/秒;

AvgDanmuByUser: 观众在高潮持续时间发送的人均弹幕数量,单位:条/人;

AvgDanmuByTime: 观众在高潮持续时间发送的时间平均弹幕数量,单位:条/秒;

Hour: 代表高潮发生的时刻,精确到小时,取值0-23;

RatioOfBurstTime: 根据高潮检测算法,每一个高潮由一对有交集的弹幕激增时间区间和礼物价值激增时间区间合并得到,此特征描述的是弹幕激增与礼物价值激增的时间区间长度比值;

RatioOfCumArea: 此特征描述的是在高潮持续时间内,礼物价值序列与弹幕数量序列累积分布曲线的线下面积比。

前6个特征的计算方法与含义是显而易见的,然而最后一个特征构建却不是平凡的。为了更好的理解这个特征,下面给出进一步的解释。

直观上理解,对于由礼物引起的高潮,礼物激增的

开始时刻要领先于弹幕激增。极端情况下,在高潮开始的短时间内,观众的送礼行为就已经结束,接下来产生由高价值礼物引来的观众为了抢夺虚拟道具,不断发送内容无关弹幕而引起的弹幕激增。相反,对于内容引起的高潮,由于这是因主播的精彩直播而产生的小高潮,观众们自发的发送弹幕,或赞美或鼓励的与主播积极互动,期间穿插着礼物送出事件。所以,礼物价值激增开始时刻往往滞后于弹幕激增,并且礼物送出是一个持续的过程。

图3所示为礼物高潮示例。为了便于对比,图4给出了一个内容高潮示例。

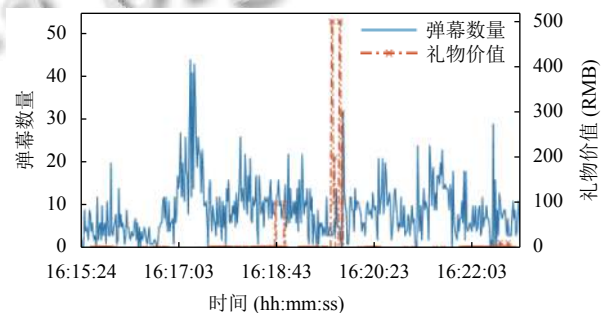


图4 内容高潮示例

为了定量描述这种特征,本研究提出了累积分布曲线线下面积比。

假设在高潮持续时间内,用户发送弹幕的时刻是一个随机事件 E ,则高潮持续时间区间构成了随机事件的样本空间 Ω 。从概率论角度看,一个高潮内观测到的弹幕数量时间序列就构成了一个经验分布函数 $f_d(t)$ 。根据弹幕数量时间序列可进一步计算得到弹幕经验累积分布函数 $F_d(t)$ 。类似的,可计算得到礼物价值经验累积分布函数 $F_g(t)$ 。

对于图3和图4展示的两个高潮,分别作出弹幕数量-礼物价值累积经验分布图,如图5所示。可以看出,对于由礼物引起的高潮,其礼物价值经验累积分布函数迅速攀升到最大值1,这和之前的讨论,即观众的送礼行为在高潮开始的短时间内结束是一致的。而对于由内容引起的高潮,其礼物价值经验累积分布函数是随着时间持续上升,且其上升滞后于弹幕数量经验累积分布函数,这也和之前讨论一致。

下面给出线下面积比来定量描述此特征的方法:

$$RatioOfCumArea = \frac{S_g}{S_d} \quad (6)$$

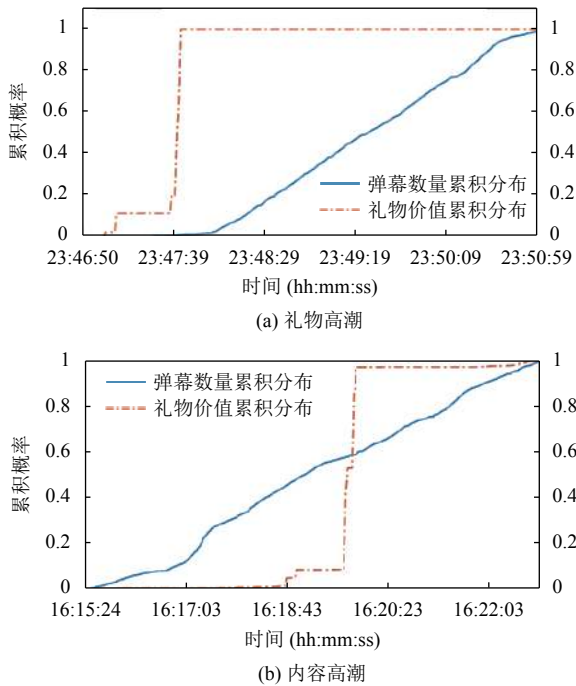


图5 弹幕数量-礼物价值累积分布图

其中, S_g 为礼物价值经验累积分布曲线的线下面积, 由 $F_g(t)$ 积分得到:

$$S_g = \int F_g(t) dt \quad (7)$$

同理可得 S_d :

$$S_d = \int F_d(t) dt \quad (8)$$

从直观上可以理解, 在典型情况下, 对于由礼物引起的高潮, 其 $RatioOfCumArea$ 值会大于 1, 而由内容引起的高潮, 其 $RatioOfCumArea$ 值因会小于 1. 对于图 5 中的两个样本, 由礼物引起的高潮和由内容引起的高潮的 $RatioOfCumArea$ 值分别为 2.05 和 0.87.

表 2 总结了上述 7 个特征的含义. 特征构建完毕后, 研究样本中的每个高潮被映射为一个 8 维的特征向量, 其中最后 1 维为样本标记. 最终得到 801×8 维的高潮特征数据.

3 模型设计及结果分析与应用

3.1 模型设计

本研究的目标是对高潮进行分类, 找出内容高潮, 所以这是一个有监督学习中的分类问题.

随机森林 (Random Forest, RF) 作为一种集成方法, 其具有强大的拟合能力和泛化性能, 可以处理非线性

数据, 训练速度快, 且训练过程无需对数据进行规范化. 基于随机森林的众多优点, 本研究采用随机森林作为高潮分类模型.

另一方面, 由于样本标注的人工成本较大, 所以为了充分利用已有标注数据, 采用交叉验证 (Cross-validation) 的方法来完成模型训练. 具体而言, 采用十重交叉验证.

影响随机森林模型拟合能力的一个重要参数为子树的数量. 一般而言, 较多的子树可以让模型的拟合能力更强, 性能也更加稳定, 但同时也会让训练过程变得缓慢. 本研究从较少的子树数量开始, 逐步增加子树数量, 观察模型精度和训练速度, 兼顾模型性能和效能, 最后将子树个数定为 200.

表 2 特征含义

特征名称	特征含义
<i>AvgGiftByUser</i>	人均礼物价值
<i>AvgGiftByTime</i>	时间平均礼物价值
<i>AvgDanmuByUser</i>	人均弹幕数量
<i>AvgDanmuByTime</i>	时间平均弹幕数量
<i>Hour</i>	高潮发生的时刻
<i>RatioOfBurstTime</i>	弹幕-礼物激增时间区间长度比值
<i>RatioOfCumArea</i>	弹幕-礼物经验累积分布曲线线下面积比

3.2 结果分析与应用

模型的分类结果如表 3 混淆矩阵所示.

表 3 混淆矩阵

		真实	
		1	0
预测	1	303	49
	0	45	404

注: 1 代表由内容引起的高潮, 0 代表由礼物引起的高潮

进一步计算, 可得到模型的准确率 (Accuracy), 精确率 (Precision), 召回率 (Recall), F1 分数 (F1 Score), 如表 4 所示.

表 4 分类结果

参数	Accuracy	Precision	Recall	F1 Score
结果	88.3%	86.1%	87.1%	86.6%

可以看出, 评估模型性能的各项指标都达到了令人满意的结果, 模型能够以较高的精确率识别出由内容引起的高潮. 这反映了本研究特征构建以及模型设计的有效性.

检测出由内容引起的高潮, 即精彩片段, 有许多应用, 下面作简单讨论.

1) 用于优化推荐. 当直播平台检测到某个直播间内产生了一定量的精彩片段, 则可在首业推荐版块实时推荐该直播间. 能够产生精彩片段的直播间, 首先其直播一般比较精彩, 其次观众们愿意在该直播间送礼物. 所以推荐此类直播间既能提升用户体验, 又能吸引更多的潜在的送礼观众进入直播间, 进而提高平台的虚拟礼物收入;

2) 用于发掘潜在的优质主播. 直播平台在评估每个主播时, 可以考虑其直播历史中的精彩片段数量. 主播的历史精彩片段数量从一定程度上反映了其优质内容生产能力以及观众变现能力. 所以, 将历史精彩片段数量作为主播的评估因素之一, 有助于及时发掘既有才能, 又能吸金的主播.

4 结论与展望

本文以斗鱼直播平台为例, 通过观众的发弹幕行为和送礼行为, 研究了众包直播系统中精彩片段的自动化检测方法. 首先, 根据主播开播的弹幕数量和礼物价值时间序列, 给出了直播高潮的检测算法; 其次, 将直播高潮分为由礼物引起和由内容引起, 通过人工查看弹幕内容的方式标注训练样本, 并构建了高潮特征; 最后, 利用随机森林方法对高潮进行分类, 得到了令人满意的结果. 对于精彩片段检测的应用场景, 本文也做了简单的讨论.

接下来的工作, 将采集一些直播视频数据, 结合视频内容标注做样本标注, 来降低标注误差, 并尝试结合视频语义理解技术, 进一步提高模型性能.

参考文献

- 1 Li BC, Wang Z, Liu JC, *et al.* Two decades of internet video streaming: A retrospective view. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 2013, 9(1s): 33.
- 2 Mihawk. 千播大战过去后, 直播还是风口么? <https://36kr.com/p/5061698.html>, [2017-01-12].
- 3 Yao T, Mei T, Rui Y. Highlight detection with pairwise deep ranking for first-person video summarization. *Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas, NV, USA. 2016. 982–990.
- 4 Lin YL, Morariu VI, Hsu W. Summarizing while recording: Context-based highlight detection for egocentric videos. *Proceedings of 2015 IEEE International Conference on Computer Vision Workshops*. Santiago, Chile. 2015. 443–451.
- 5 Wang JJ, Xu CS, Chng E, *et al.* Sports highlight detection from keyword sequences using HMM. *Proceedings of 2004 IEEE International Conference on Multimedia and Expo (ICME)(IEEE Cat. No.04TH8763)*. Taipei, China. 2004. 599–602.
- 6 Gong YH, Han M, Hua W, *et al.* Maximum entropy model-based baseball highlight detection and classification. *Computer Vision and Image Understanding*, 2004, 96(2): 181–199. [doi: 10.1016/j.cviu.2004.02.002]
- 7 Otsuka I, Nakane K, Divakaran A, *et al.* A highlight scene detection and video summarization system using audio feature for a personal video recorder. *IEEE Transactions on Consumer Electronics*, 2005, 51(1): 112–116. [doi: 10.1109/TCE.2005.1405707]
- 8 Zhao Y, Tian Y, Liu Y. Extracting viewer interests for automated bookmarking in video-on-demand services. *Frontiers of Computer Science*, 2015, 9(3): 415–430. [doi: 10.1007/s11704-014-3490-2]
- 9 Li ZY, Kaafar MA, Salamati K, *et al.* Characterizing and modeling user behavior in a large-scale mobile live streaming system. *IEEE Transactions on Circuits and Systems for Video Technology*, 2017, 27(12): 2675–2686. [doi: 10.1109/TCSVT.2016.2595325]
- 10 Kaytoue M, Silva A, Cerf L, *et al.* Watch me playing, I am a professional: A first study on video game live streaming. *Proceedings of the 21st International Conference on World Wide Web*. Lyon, France. 2012. 1181–1188.
- 11 Zhang C, Liu JC. On crowdsourced interactive live streaming: A Twitch. tv-based measurement study. *Proceedings of the 25th ACM Workshop on Network and Operating Systems Support for Digital Audio and Video*. Portland, OR, USA. 2015. 55–60.
- 12 Wang BL, Zhang XY, Wang G, *et al.* Anatomy of a personalized livestreaming system. *Proceedings of the 2016 Internet Measurement Conference*. Santa Monica, CA, USA. 2016. 485–498.
- 13 Wang XD, Tian Y, Lan RH, *et al.* Beyond the watching: Understanding viewer interactions in crowdsourced live video broadcasting services. *IEEE Transactions on Circuits and Systems for Video Technology*, 2018. [doi: 10.1109/TCSVT.2018.2877694]
- 14 兰荣亨, 朱格, 杨文, 等. 基于聚类的网络直播群体行为建模分析. *计算机系统应用*, 2019, 28(1): 69–74. [doi: 10.15888/j.cnki.csa.006728]
- 15 斗鱼. <https://www.douyu.com>. 2016.
- 16 斗鱼 API 文档. <https://open.douyu.com/>.
- 17 Z-Score. https://en.wikipedia.org/wiki/Standard_score.