

人工智能赋能的设计评价方法研究与应用

杨智渊^{1,2}, 杨文波^{1,2}, 杨光^{1,2}, 杨昌源^{1,2}

(1.阿里巴巴集团, 杭州 311121; 2.阿里巴巴—浙江大学前沿技术联合研究中心, 浙江大学, 杭州 310027)

摘要: **目的** 人工智能赋能的设计评价方法正成为智能设计领域的新关注点, 也带来了全新的挑战。分析智能设计评价方法的发展趋势, 总结评价方法及其模型构建过程中的关键问题, 能为设计评价研究和应用提供参考, 也对智能设计的发展提供支持。**方法** 以人工智能赋能的设计评价方法在阿里巴巴的应用实践为例, 重点描述面向电商场景的短视频智能评价方法的构建, 深入分析短视频设计特征对业务效果的影响, 探讨智能评价方法在短视频智能设计中的实际应用。**结果** 短视频智能评价方法可以通过预测短视频广告效果, 形成数据驱动的设计优化链路, 为阿里巴巴的短视频智能设计提供支持。**结论** 人工智能赋能的设计评价方法是基于大数据分析和挖掘设计对象与特征之间的影响关系, 可以有效地支持设计过程及结果的优化。智能评价方法是智能设计体系的延伸, 对智能设计的进一步发展有着巨大的价值。

关键词: 设计评价; 人工智能; 计算机辅助设计; 智能设计

中图分类号: TB472 **文献标识码:** A **文章编号:** 1001-3563(2021)18-0024-11

DOI: 10.19554/j.cnki.1001-3563.2021.18.004

Research and Application of Artificial Intelligent Empowered Design Evaluation Method

YANG Zhi-yuan^{1,2}, YANG Wen-bo^{1,2}, YANG Guang^{1,2}, YANG Chang-yuan^{1,2}

(1.Alibaba Group, Hangzhou 311121, China; 2.Alibaba-Zhejiang University Joint Institute of Frontier Technologies, Hangzhou 310027, China)

ABSTRACT: The design evaluation method enabled by artificial intelligence is becoming a new focus in the field of intelligent design, and it also brings new challenges. The work aims to analyze the development of intelligent design evaluation methods, summarize the key issues in the evaluation method and its model construction process, which can provide references for design evaluation research and application, and also provide support for the development of intelligent design. Taking the application of artificial intelligence empowered design evaluation methods in Alibaba as example, this work focuses on describing the construction of intelligent evaluation methods for e-commerce business scenarios, deeply analyzing the impact of video design features on advertising effects, and exploring the practical application of intelligent evaluation methods in video intelligent design. The video intelligent evaluation method can form a data-driven design optimization link by predicting the advertising effect of video, and provide support for the video intelligent design in Alibaba. The design evaluation method enabled by artificial intelligence is based on big data analysis and mining the impact relationship between design objects and features based on big data, which can effectively support the optimization of the design process and results. The intelligent evaluation method is an extension of the intelligent design, and has great value to the further development of intelligent design.

KEY WORDS: design evaluation; artificial intelligence; computer aided design; intelligent design

收稿日期: 2021-04-09

作者简介: 杨智渊(1990—), 男, 浙江人, 博士, 阿里巴巴集团高级体验设计师, 主要研究方向为智能设计、用户体验等。

通信作者: 杨昌源(1979—), 男, 江西人, 硕士, 阿里巴巴集团资深设计专家, 主要研究方向为人机交互、智能设计等。

人工智能技术的迅速发展给创意设计领域带来了巨大的影响,推动着智能设计的发展。人工智能技术被不断地用于设计活动中,参与设计的各个环节^[1]。设计评价是设计的重要环节之一,贯穿整个设计过程。人工智能赋能的设计评价可以从用户感知、美学等多个维度对设计结果进行量化评估,为设计师提供更加客观有效的参考建议,也为设计过程中的决策和优化提供重要依据。而存在于人工智能和创意设计间的特征差异,导致人工智能在面对设计评价问题时仍存在诸多挑战。在智能设计背景下,如何将主观的、不明确的开放性评价问题抽象为数据驱动的封闭问题是设计评价研究和应用的重点。本文以人工智能赋能的评价方法在阿里巴巴智能设计体系中的应用实践为例,探究数据驱动的智能评价方法的构建过程,揭示设计评价研究的发展趋势。

1 人工智能赋能的设计评价方法

设计评价是指对设计对象价值的确定,是设计对象与其对应的价值标准相比较,从而明确价值的过程^[2]。创意设计过程伴随着大量的评价与决策。一方面,设计评价可以对设计结果的质量进行评估,为设计决策提供指导;另一方面,设计评价可以干预设计流程,为设计的迭代优化指明方向。

传统的设计评价方法常由人来完成,如外包评价方法等^[3-4]。这类方法主要依靠人的经验和直觉来完成,评价结果受人的知识等多方面因素影响,表现出一定的主观性,并且随着智能设计的爆炸式增长,通过人工评价大规模的设计结果变得不太现实,无法及时有效地支持设计优化。人工智能赋能的设计评价方法可以对设计知识、用户感知等多维内容进行量化计算,建立统一的评价指标,实现更加精细的设计分析与评估,提升评价结果的准确性和客观性。此外,人工智能还可以有效弥补传统方法的效率问题,可以让智能系统或设计师在短时间内获取客观的评价结果,实现设计的快速迭代。

人工智能赋能的设计评价方法在设计领域应用广泛,可以用来衡量设计结果在不同用户体验维度上的质量好坏,如信任度、吸引力、生动性、流行度等^[5-7]。人工智能赋能的设计评价方法可以分为人工定义特征和计算机自动提取特征两种方式。采用人工定义特征的评价方法从人的感知层面出发,将抽象的设计先验知识和经验转化成合适的可量化特征,针对设计的不同维度进行建模,再基于分类、回归等机器学习算法预测设计评分^[8]。例如在美学评价领域,评价方法主要集中于通过建立审美计算模型,让计算机可以自动地评估不同类型的设计,包括图像、视频、用户界面等^[9]。Aliaksei Miniukovich 等人从界面的整体感知出发,通过量化视觉复杂度对界面审美进行评估^[10];Reinecke 等人构建了一种基于网站色彩丰富度和视

觉复杂性的感知计算模型来预测用户对网站的初始印象^[11];Li 等人基于在艺术方面的先验知识,通过机器学习对颜色和构图等特征进行训练,从而评估数字绘画的视觉质量^[12];Jia 等人为了评估服装设计的美学程度,建立了一个包含视觉特征—图像空间—审美词汇的 3 级模型,可以将美学词汇映射到服装外观设计上,从而捕捉两者之间的关系^[13];Cutumisu 等人则提出了一种海报的构图规则分析与评估方法,并基于评估结果为设计师提供相应的设计反馈^[14]。这类方法的优点在于其构建的评价特征体系符合人类的思维模式,具有良好的可解释性,然而由于其评价特征体系的定义需要很高的专业知识,并且很难准确地量化描述全部设计因素,从而影响了评价结果的准确性^[15]。

随着卷积神经网络等深度学习算法的不断成熟,采用计算机自动提取特征的评价方法研究与应用逐渐完善。这类方法依靠不同的算法从数据集中提取多维特征,包括计算机视觉特征、深度学习特征等,并建立预测模型,而无需设计专业人员的介入。例如 Dou 等人提出了一种使用深度学习技术自动计算网页美观度的方法,通过深度神经网络从网页中抽取代表性特征从而量化其美观度,为设计师提供一种客观的美学评估维度^[16];Hammad 等人利用基于卷积神经网络提取的视觉和语义特征来预测图像和视频等多媒体内容的难忘度^[17];Bosse 等人提出了一种基于深度神经网络的图像质量评估方法,该方法以大数据为驱动,不依赖于人类视觉系统的先验领域知识^[18];Ma Ning 等人则使用一个成对神经网络来对图像的吸引力进行评估与排序^[19]。这类方法的主要问题在于其特征与评价结果的可解释性较差,无法为智能系统或设计师提供有针对性的优化建议。

上述两类人工智能赋能的设计评价方法各有优劣,也有许多学者通过结合这两种方式来构建计算模型,以达到更好的评价效果。如 Khosla 等人通过定义图像的颜色特征、社交属性特征、内容语义特征,并结合提取的低维计算机视觉特征、深度学习特征建立图像评价体系,共同预测图像的流行度^[20]。总之,人工智能赋能的设计评价方法以数据驱动为基础,有着多样的实现方式,需要根据实际的应用场景和评价对象,选择相对应的量化特征,建立合适的评价方法。

2 智能评价方法构建

人工智能赋能的设计评价方法见图 1,主要包含 3 个步骤。

1) 数据收集。基于评价对象和目标,收集和标注相对应的设计数据,构建数据集。数据集是人工智能赋能的设计评价方法的基础。

2) 评价体系构建。设计评价体系包括评价标准和评价特征两部分。基于用户视角,可以将设计评价标准简单分为两类:用户体验标准和用户行为标准^[21-22]。

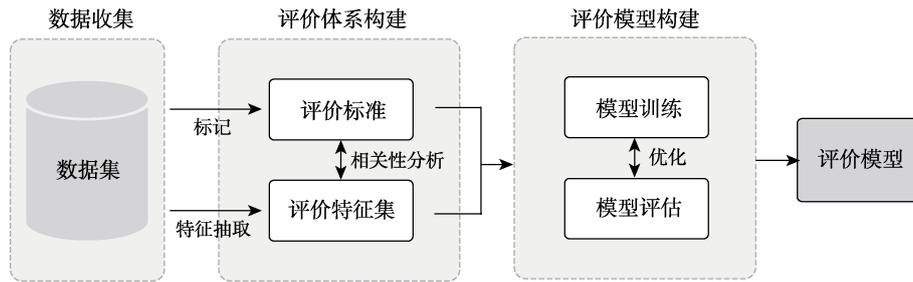


图1 人工智能赋能的设计评价方法

Fig.1 The design evaluation method enabled by artificial intelligence

用户体验标准是指从人的情感、愉悦度、惊奇度、美感、注意力等主观感受出发的评价标准。用户体验标准源于人的感知,通常采用问卷的形式进行指标的量化。用户行为标准则是指基于人的客观行为反馈的评价标准,例如点击、收藏、分享等社交行为反馈。用户行为反馈是用户心理变化的外在表现,具有客观、可量化的特点,其相关指标常用于数据驱动的设计评价研究中。用户体验标准和用户行为标准是用户心智模型在不同层面上的表现,存在着一定的间接映射关系,因此为了提高评价方法的客观性和准确性,经常会使用客观的行为指标来指代主观的体验指标,例如,用眼动数据来代表用户的注意力^[23]。

评价特征是对评价对象在多个维度上的抽象表述。设计评价研究中的评价特征种类繁多,常见的包括视觉特征、音频特征、高维语义特征、情感特征、深度学习特征等^[24-26]。评价特征有人工定义和计算机提取两种方式,不同研究中评价特征的选择和组合也不尽相同,需要根据具体的评价对象和标准进行定义。如何选择合适的特征特征是人工智能赋能的设计评价方法中的重难点。

3) 评价模型构建。构建评价模型的根本目的是探究评价标准与评价特征间的相互影响关系。常见的算法模型包括逻辑回归(LR, Logic Regression)、随机森林(RF, Random Forest)、GBDT(Gradient Boosting Decision Tree)等。LR模型作为一个经典的线性模型,具有使用简单、易理解、易解释等特点,因此,该基础模型及其相关优化模型常在设计评价研究中被使用。随着人工智能技术的发展,使用深度神经网络(DNN)进行模型训练是近年来的发展趋势。此外,通过模型集成的方法来降低预测偏差也是机器学习中的通用做法,例如Facebook提出了一种将GBDT与LR相结合的模型构建方法,极大地提升了模型性能^[27]。

人机交互(HCI)领域和机器学习(ML, Machine Learning)领域的设计评价研究在模型构建上的侧重点有所不同^[28]。HCI研究中构建模型的主要目的是揭示评价对象的评价标准与特征之间的数学关系,结果预测是评价模型的附属产物;而ML研究中构建模型的主要目的则是结果预测。即HCI领域的设计评价研究关注于关系描述,而ML领域的研究则更侧重于

结果预测。因此,在HCI研究中,通常在模型构建中引入传统的算法模型,如LR等;而在ML研究中则通常引入学习能力以提高模型的泛化能力。相较而言,HCI研究方法会对领域知识进行深入分析,其模型更具有可解释性。事实上,当前智能评价领域的前沿研究通常会结合CHI和ML方法,以达到更好的评价效果。

3 短视频智能评价方法与应用

近年来,短视频行业的蓬勃发展使人们的视听内容发生了翻天覆地的变化。短视频不仅成为人们信息发布和社交分享的重要载体,也成为了广告营销的一种强有力形式。在阿里巴巴的电商生态中,短视频广告增长迅速,并且带来了更好的业务效果。相较于平面广告,短视频广告存在制作难度大、投入成本高等问题。为了更好地解决短视频广告需求和生产之间的矛盾,鹿班短视频智能生成工具应运而生。鹿班短视频智能生成工具基于人工智能技术,可以根据商家提供的素材,自动完成视频分析、剪辑、合成等工作,满足短视频广告的批量化生成需求^[29]。

在阿里巴巴鹿班短视频的智能生产链路中,仍面临着重要问题。首先,虽然短视频智能生成工具可以实现批量化生成,但缺乏对这些大规模生成结果的快速评价,无法对结果进行优化筛选以及制定相应的广告投放建议。其次,目前智能短视频的生成与制作主要依靠专业人员总结的经验、知识,生成规则固化,缺乏对生成结果的评价和反馈机制,无法为短视频生产提供有效的优化策略。短视频智能评价是当前智能生产链路中的关注重点。具体到阿里巴巴生态中,如何利用业务场景中生产和沉淀的大量视频数据对生成结果进行评价,以及如何将短视频智能设计与业务场景、用户等相关联,提高视频广告的业务效果,仍没有得到有效解决。

本文以面向业务场景的短视频智能评价方法为例,介绍其在阿里巴巴生态中的应用,展现数据驱动的评价方法在智能设计体系中的价值。该方法面向淘宝“猜你喜欢”场景,建立短视频吸引力评价体系,基于真实的短视频及业务数据,分析和挖掘影响视频

效果的关键设计特征，构建短视频评价模型，实现对短视频的业务效果预测，从而提出相应的优化策略和投放建议，为短视频智能设计提供支持。

3.1 短视频评价体系

“猜你喜欢”是淘宝中的一个重要商品推荐场景，可以根据用户的长短期行为、偏好等个性化地推荐和展示不同的商品。短视频已经成为该场景下商品的重要展示载体。“猜你喜欢”有着独特的场景特性，在用户行为上，用户多以无目的闲逛为主；在视频形式上，以 5 s 的商品摘要视频为主。因此，如何在短时间内快速吸引用户注意力，如何更有效地通过短视频展现商品信息，引导用户转化是该场景下短视频的首要问题。“吸引力”是“猜你喜欢”场景下短视频的重要设计和评价标准。围绕“猜你喜欢”场景下的商品摘要视频，通过建立短视频在“吸引力”维度上的评价体系，挖掘短视频吸引力和设计特征间的影响关系。

短视频吸引力评价体系主要由评价标准和设计特征两部分构成。评价标准是短视频吸引力的量化表征。在设计研究中，通常采用主观评价量表的方式来收集用户对视频在吸引力维度的体验反馈，获得短视频吸引力评分代表短视频的吸引程度^[30-31]。然而“吸引力”是一个抽象、综合的概念，为了更好地衡量视频广告的吸引程度，建立客观可靠的评价标准，采用点击率（CTR, Click-Through-Rate）作为评估“猜你喜欢”场景下短视频吸引力的关键指标。CTR 指广告被点击次数与其展示次数之比，是对用户的行为反馈的量化统计，是电商广告业务的重要衡量指标，反映广告的关注程度，可以视为短视频吸引力的客观表现。

设计特征是对短视频本身及其设计过程的客观理解。基于专家知识和经验，确定影响短视频吸引力的设计特征集，并定义相应的特征计算方法。虽然已有的视频智能评价研究中定义了大量视频评价特征，如计算机视觉特征、时序特征、情感特征等^[26, 32]，但这些评价特征无法与短视频设计过程相联系，形成对智能生产链路的有效支持。因此，本文从设计视角出发，基于短视频设计过程中色彩、构图和内容（视觉）这 3 个维度，整理短视频的视觉影响因素，定义计算特征，建立短视频设计特征集，见表 1。

3.1.1 构图特征

构图特征表示的是画面元素的空间关系，通过调整画面主体的大小和位置、景深等方式来强调不同的画面信息，起着突出主体、吸引用户注意力的作用，可以使视频画面更加富有故事性。

主体占比和位置特征依赖于显著性物体检测算法。首先，对视频等距提取若干采样帧 $F = \{f_1, f_2, \dots, f_n\}$, ($n < N_s$)，基于 Hou 等人提出的显著性物体检测算法^[33]对采样帧进行主体区域提取，获得视频的主体区域序列 $O = \{O_1, O_2, \dots, O_n\}$ 。对于主体占比特征，基于提取的主体区域 O_i ，计算主体区域面积与视频画面面积的比值：

$$P = \frac{sq(O_i)}{(L \times H)}$$

其中 $sq(O_i)$ 为主体区域的像素面积， L 为视频画面的宽度， H 为视频画面的高度。所有采样帧所得比值的平均值即为主体占比特征值，其取值范围为 (0,1)。

表 1 短视频设计特征集
Tab.1 Video design feature set

| 设计目标 | 设计维度 | 设计特征 | 特征描述 | |
|--------|--------|---------|--------------------|--------------------------|
| | 构图 | 主体占比 | 主体区域面积与画面面积的平均比值 | |
| | | 主体位置 | 主体重心相对坐标的平均值 | |
| | | 三分法则 | 主体重心与画面分割线和交叉点的距离 | |
| | | 景深 | 主体区域清晰度与背景区域清晰度的比值 | |
| | | 景别 | 景别识别算法：近景、中景、全景、远景 | |
| | 吸引力 | 色彩 | 色相、饱和度、亮度 | 基于 HSV 色彩空间的色相、饱和度、亮度平均值 |
| | | | 色彩冷暖 | 基于 LCH 色彩空间的色温值 |
| | | | 色彩柔和度 | 基于 LCH 色彩空间的明度和纯度值 |
| | | | 色彩丰富度 | 二维色彩空间的色彩均值和标准差的加和 |
| | | | 影调 亮度偏向 | 基于画面平均亮度值的分类 |
| 亮度分布 | | | 画面亮色比例和暗色比例特征值的组合 | |
| 色调 | | | 画面饱和度和亮度特征值的组合 | |
| 内容（视觉） | 内容（视觉） | 前后景明暗对比 | 主体区域和背景区域的亮度比值 | |
| | | 前后景色相对比 | 主体区域和背景区域的色相比值 | |
| | | 视觉复杂度 | 灰度共生矩阵（GLMC）同质性计算值 | |
| | | 画面运动幅度 | 相邻画面帧之间的运动矢量均值 | |
| | | 画面连贯度 | 相邻画面帧之间的汉明距离 | |

对于主体位置特征,则先计算上述每个采样帧的主体区域 O_i 的重心坐标 (x_i, y_i) , 然后分别与视频画面长宽相比得到主体重心的相对坐标 $(\frac{x_i}{L}, \frac{y_i}{H})$ 。对所有采样帧的重心相对坐标取平均值即为主体位置特征

$$\text{值} \left(\frac{\sum_i^n (x_i / L)}{n}, \frac{\sum_i^n (y_i / H)}{n} \right)$$

三分构图法则,是指将画面根据两条垂直线和两条水平线分成9个相等的部分,画面主体重心位于这些线条或交叉点之上,通常认为该画面在构图上会更为美观。因此,基于视频主体区域 O_i 的重心坐标 (x_i, y_i) , 分别计算其到交叉点和分割线间的最小距离 d_i 和 d'_i , 通过使用这两个距离的乘积值, 即 $D = d_i \times d'_i$, 来预测视频画面是否遵循三分法则。

景深是指画面焦点前后方位内所呈现清晰图像的范围,通常表现为视频画面前景与背景的清晰对比程度。因此,景深特征的计算方法为,对于视频抽样帧的主体区域 O_i , 根据灰度方差乘积算法,得到画面主体区域的清晰度值 df_i ; 类似的,计算背景区域,即非主体区域的清晰度值 df'_i 。最终,把主体区域清晰度与背景区域清晰度的比值 df_i / df'_i 作为视频的景

深特征值,可表示为: $\frac{\sum_i^n (df_i / df'_i)}{n}$ 景别是指由镜头与被拍摄主体的不同位置距离而呈现的不同视觉画面效果,可以通过不同的景别表达不同的情绪、氛围等。景别由近至远可以分为以下5种:特写、近景、中景、全景和远景。本文使用了基于公开数据集训练的景别识别算法对短视频进行分类,其计算方法为对短视频每隔5帧进行一次识别测试,统计返回结果取众数作为该短视频景别类别,在人工校验下,该场景识别方法的准确率为95.6%。由于近景和特写具有一定相似性,为了保证识别结果的准确率,将特写和近景合并为近景类别,最终识别结果为近景、中景、全景、远景4个类别,短视频景别特征计算结果示例见图2。景别特征与主体占比特征在设计表现上相互影响,具有很强的关联性。

3.1.2 色彩特征

画面色彩是引起用户情绪反应的显著特征之一,色彩会影响用户的兴趣度、信任感、购买意向等,是构成短视频吸引力的重要设计维度^[34-35]。

色相、饱和度、亮度是衡量色彩的基础特征,基于HSV色彩感知模型,分别计算所有采样帧的色相、饱和度、亮度值,其平均值即为视频的色相特征值

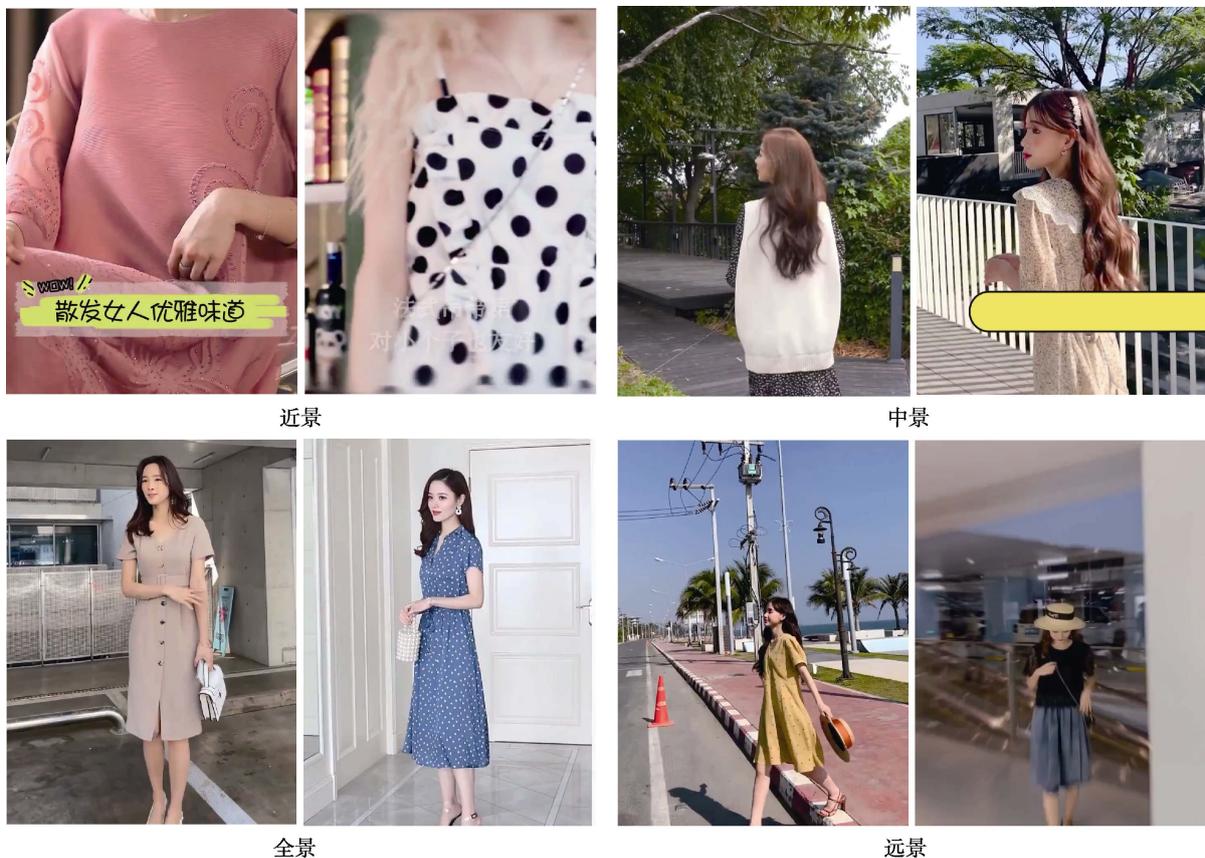


图2 短视频景别特征计算结果示例

Fig.2 Examples of field feature calculation results

H_i 、饱和度特征值 S_i 、亮度特征值 V_i 。色彩冷暖特征是指色彩带给用户的心理感觉，其冷暖性是相对的，主要通过色彩间的相互对比体现。基于 Niu 等人提出的色彩冷暖计算方法^[36]，提取视频采样帧的色温特征，其平均值代表视频画面的色彩冷暖度。类似的，基于 LCH 色彩空间可以计算得到色彩柔和度特征值。柔和度与色彩的明度和纯度相关，明度高、纯度低则视频画面色彩柔和。色彩丰富度是对画面色彩多样性程度的量化，采用 Hasler 提出的色彩丰富度计算方法^[37]，将采样帧的三维 RGB 色彩空间转换为包含 rg 、 yb 的二维色彩空间，通过该二维色彩空间的均值和标准差的加和来表示色彩丰富度。

为了更好地衡量视频画面的色彩关系及其变化，进一步定义了影调和色调两个设计特征。影调主要是指画面中的明暗关系，通过解决画面上的明暗层次等关系，使用户感受到光影构成的视觉美感。影调特征可以具体细分为两类：按照画面的亮度偏向，分为低调、中调和高调，代表不同的画面明亮程度；按照画面的亮度分布，分为短调、中调和长调，代表不同程度的画面明暗层次变化。

影调与亮度特征息息相关，根据摄影大师亚当斯的分区曝光法，将亮度值 V_i 从暗（0）到亮（10）划分为 11 个区域，其中，0—3 区为低调区，4—6 区为

中调区，7—10 区为高调区。结合亮度特征的计算取值范围 (0,1)，可以对短视频的亮度偏向特征进行计算，即，画面平均亮度值 V_i 低于 0.3，标记为低调；画面平均亮度值 V_i 在 0.3~0.7，标记为中调；画面平均亮度值 V_i 高于 0.7，标记为高调。而对于亮度分布特征，定义视频画面的亮度分布在 0—10 区中任意 3 区作为短调，任意 5 区作为中调，0—10 区全分布作为长调。亮度分布特征的计算结果基于视频画面的亮色比例、暗色比例这两个色彩特征的计算值排列组合得出。亮色（或暗色）比例特征是指画面中亮度值 V_i 高于（或低于）某个阈值的像素点占整体画面的比例。短视频影调特征计算结果示例见图 3，在亮度偏向特征中，低调的短视频表现为画面昏暗，而高调的短视频则表现为画面明亮；而在亮度分布特征中，短调短视频的画面明暗层次变化少，整体画面视觉平滑、简单，中调短视频画面明暗层次小幅度变化，画面明暗对比和谐，呈现一定的朦胧感，而长调短视频则画面明暗层次多变，细节丰富，画面呈现一定复杂度。

色调则指的是画面的色彩关系，是画面呈现出的色彩倾向。与影调特征类似，基于饱和度 S_i 和亮度 V_i 两个色彩特征计算值间的组合关系，色调特征可细分为 5 种类型，分别是艳彩色调、浓彩色调、暗彩色调、

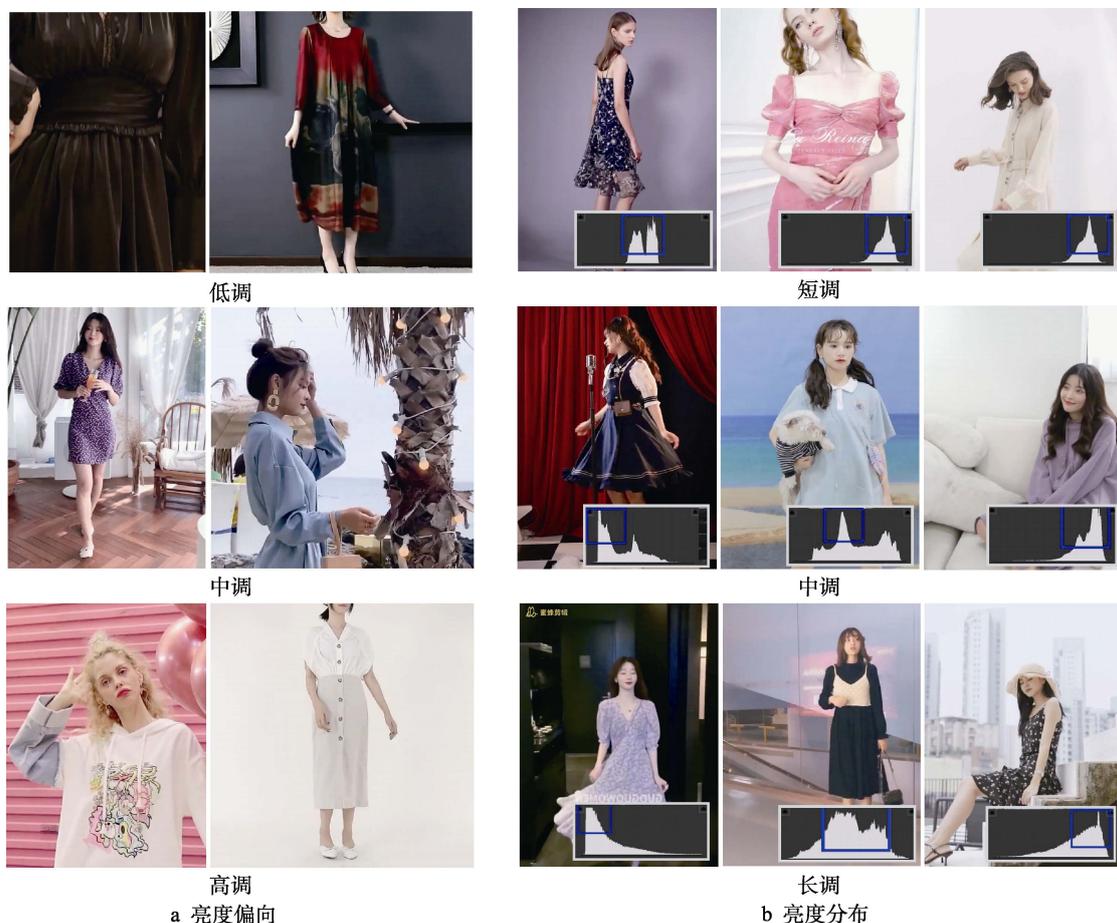


图 3 短视频影调特征计算结果示例
Fig.3 Examples of luminance feature calculation results

亮彩色调和灰彩色调。

此外，画面主体（前景）和背景的色彩对比度也会影响视频的吸引力。基于上述显著性物体检测算法提取的视频主体区域 O_i ，计算前后景在亮度特征 V_i 和色相特征 H_i 上的差异。其具体计算方法为：获取视频所有采样帧主体区域 O_i 的亮度值 fV_i 和背景区域 bV_i ，分别取平均值后计算两者间的比值可得到前后景明

暗对比特征值 $\frac{\sum_i^n (V_i / bV_i)}{n}$ 。同理，可计算得到前后

景色相对比特征值 $\frac{\sum_i^n (fH_i / bH_i)}{n}$ 。短视频色调特征

计算结果示例见图 4。

3.1.3 内容特征

内容特征主要表现为短视频内容在视觉上的呈现效果。视觉复杂度是影响短视频用户体验的重要因素，代表画面元素在用户视觉感知上的复杂性。视频画面的视觉复杂度特征可以通过纹理特征计算来量化。纹理特征计算是对图像灰度空间分布模式的提取和分析，用来描述图像元素组成成分的排列情况，从而可以反映画面的视觉复杂度。灰度共生矩阵（GLMC）算法是纹理特征分析中的常用方法，其中 GLMC 同质性特征值是通过度量图像的灰度级在行或列方向上的相似程度，研究图像中灰度的空间相关性，反映图像纹理的局部变化。其同质性特征值越高，表示画面纹理变化均匀，区域相似度高，则整体画面简单，视觉复杂度低。

画面运动幅度表征视频内容的移动剧烈程度，其特征计算通过视频光流法实现。光流法是利用图像序列中像素在时间域上的变化以及相邻帧之间的相关性来找到上一帧跟当前帧之间存在的对应关系，从而

计算出相邻帧之间物体的运动信息的一种方法。基于 Gunner Farneback 提出的光流场算法，计算相邻采样帧的光流信息，得到像素点的运动矢量 (ρ, θ) 。对于每一个采样帧 f_i ，该帧的运动幅度是该帧所有运动矢量的均值，即 $m_i = \text{mean}(\rho)$ ；则视频的画面运动幅度特征值为所有帧运动幅度序列 $M = \{m_1, m_2, \dots, m_n\}$ ， $(n < N_s)$ 的均值。其特征值越高，表示画面主体或镜头运动速度越快。

画面连贯度特征描述了视频相邻画面间的相似程度。不同的画面连接方式会对用户的观看体验产生不同的效果。其具体计算方法为，使用 Hash 算法^[38]计算 N_s 帧视频每相邻两帧画面之间的汉明距离 $D = \{d_1, d_2, \dots, d_n\}$ ，其均值即为画面连贯度特征值。其值越大说明两帧内容差异性越大，视频内容越丰富。

3.2 短视频评价模型

首先，收集“猜你喜欢”场景下的短视频数据，基于上述的短视频评价体系，通过抽取设计特征与评价指标，建立短视频数据集；其次，基于 XGBoost 模型，挖掘和学习短视频设计特征和 CTR 指标间的映射关系，短视频评价模型构建流程见图 5。

3.2.1 数据集构建

为了创建短视频数据集，收集了淘宝“猜你喜欢”场景在 2020 年 4 月 1 日—6 月 30 日投放的女装类目短视频，累计共 31 289 条。为了提升模型及其分析结果的准确性，筛选后去除了视频曝光量过低 ($PV < 400$)、透出时间过短 ($\text{day} < 7$) 等不具有分析参考价值的商品短视频，最终得到有效短视频数据共 17 378 条。

基于业务数据计算每条短视频的平均 CTR 值，作为短视频的评价指标数据集。由于短视频的 CTR 数据整体偏小，对其进行放大 100 倍并做对数变化处理进行校正，使 CTR 数据更符合正态分布（由于 CTR 数



图 4 短视频色调特征计算结果示例
Fig.4 Examples of tone feature calculation results

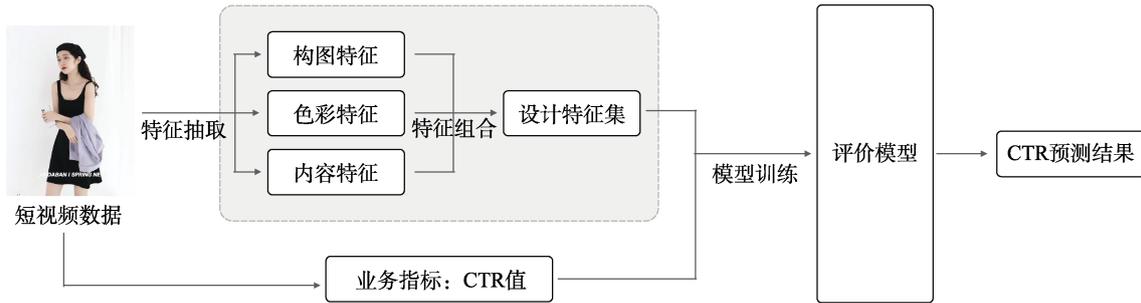


图 5 短视频评价模型构建流程
Fig.5 Video evaluation model construction process

据为公司业务敏感数据,受数据安全规定限制,本文仅展示 CTR 数据的变化趋势,而不描述具体数值)。

基于上述章节描述的设计特征计算方法,抽取短视频的设计特征,作为设计特征数据集。为了挖掘这些设计特征之间可能存在的相互关系,探索组合特征对短视频 CTR 的影响,将构图、色彩、内容特征分别进行特征交叉组合,共获得 268 个组合设计特征。通过特征组合可以扩展设计特征种类,有利于评价模型的构建。最终,获得了包含 287 维设计特征的短视频数据集。同时,对提取的设计特征数据进行连续特征离散化和归一化处理,滤除噪点数据,并且删除取值变化小等冗余特征,加强模型防滑能力,减小过拟合风险。

3.2.2 评价模型构建

在短视频数据集的基础上,采用机器学习方法中的 XGBoost 模型对短视频的 CTR 指标进行建模,拟合设计特征 X 和 CTR 指标 Y 间的映射关系。XGBoost 模型是一种结合了监督学习和集成学习的方法,通过不断迭代输入输出对来调整模型的参数,最终获得模型最优解。该模型算法基于 GBDT 框架,构建 K 个回归子树,通过迭代得到一个最强回归树来近似输出预测值,可表示为:

$$Y_i = \sum_{k=1}^K f_k(X_i), f_k \in F$$

其中, f_k 是 K 个回归树中的每一个独立子树,它将输入变量设计特征 X_i 映射到 CTR 指标 Y_i , 并拟合上次预测的残差。

将短视频数据按 7 : 3 的比例随机分成训练集和测试集,对训练集进行 5 折交叉验证,以确定训练模型的最佳参数设置。为了衡量基于 XGBoost 的评价模型在结果预测上的准确性,选择逻辑回归 (Logic Regression)、随机森林 (Random Forest)、GBDT 这 3 个经典模型作为参照,使用相同的训练集和测试集进行模型训练。模型训练实验结果见表 2。AUC 值越

高,表示训练模型分类效果越好。结果表明,通过这 4 种算法模型训练得到的短视频评价模型都有着较好的 CTR 预测效果,但基于 XGBoost 的评价模型表现最好,其 AUC 值最高。

3.3 设计特性影响分析

为了进一步探究视频设计特征与业务指标间的影响关系,为该场景下的短视频设计和优化提供方向,根据模型训练结果计算各个设计特征的贡献值并排序,使用最大互信息系数 (MIC) 对排名前十的重要设计特征和 CTR 指标进行相关性分析,相关性分析结果见图 6。与 Spearman 等常用线性相关分析算法相比, MIC 有着更好的普适性,能够均衡覆盖数据可能的函数关系,包括线性关系、非线性关系等。由于无法确定设计特征与 CTR 指标间的相关关系类型,采用 MIC 在分析结果上有着更高的准确性。

分析结果表明,短视频的设计特征对 CTR 指标存在一定的影响,但其影响存在上限。这是由于短视频 CTR 指标受多种外部因素共同影响,包括促销、商品品牌、价格、平台推荐机制等。设计是影响短视频业务效果 (CTR 指标) 的因素之一,但非唯一决定性因素。

进一步地引入价格因素探索重要设计特征对短视频 CTR 指标的影响程度。将短视频按商品价格划分成 4 个区间: 0—100、100—300、300—500、500 以

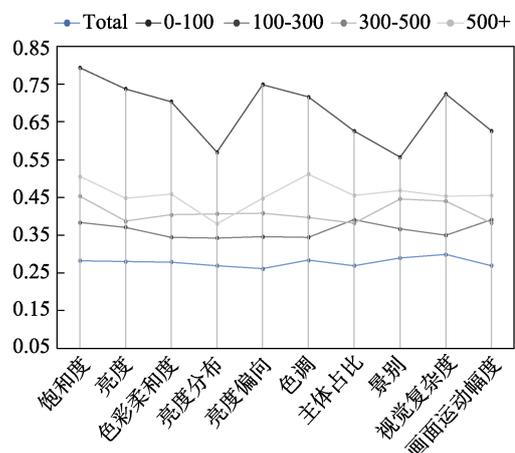


图 6 相关性分析结果
Fig.6 The results of correlation analysis

表 2 模型训练实验结果
Tab.2 Model training results

| 算法模型 | XGBoost | LR | RF | GBDT |
|-------|---------|------|------|------|
| AUC 值 | 0.88 | 0.78 | 0.83 | 0.85 |

上,并针对每类价格区间内短视频的设计特征与 CTR 指标分别进行相关系分析。结果表明,在不同价格区间,设计特征对 CTR 指标的影响程度并不相同。0—100 价格区间的短视频设计特征与 CTR 指标间的相关性明显高于其他价格区间,色彩特征分析见图 7。即相比于高价格商品,低价格商品的短视频业务效果 (CTR 指标) 更容易受到设计因素的影响。

短视频色彩维度的各个设计特征均对 CTR 有影响,且影响各不相同。其中,画面色彩的亮度和饱和度特征对 CTR 的影响最为明显。具体到影调的亮度偏向特征中,“猜你喜欢”场景下的短视频画面亮度以中高调为主,且 CTR 值随着视频画面亮度升高而呈现上升趋势(如图 7a)。而对于亮度分布特征,短视频画面的明暗关系以中长调为主,短调视频的平均 CTR 值相对更高,CTR 值随着视频画面的明暗分布跨度升高而逐渐降低(如图 7b)。分析结果表明,整体画面明亮、明暗对比层次简单、关系和谐的短视频在业务效果上表现更好,尤其是对于低价商品。

在色调特征中,“猜你喜欢”场景的短视频以灰彩色调和亮彩色调为主,且灰彩色调的短视频 CTR 值远低于其他类型色调的短视频(如图 7c)。色调特征受画面亮度和饱和度影响,其影响关系具体表现为,在中低画面亮度的情况下,相较于低饱和度(灰彩色调),中高饱和度(浓彩色调)的短视频平均 CTR 值更高;在高画面亮度的情况下,饱和度对 CTR 值

的影响不大。结果表明,画面明亮自然、色彩强烈浓郁的短视频更吸引用户,业务效果表现更好。

构图特征分析见图 8。在构图特征中,随着短视频画面主体占比的增加,平均 CTR 值逐渐降低(如图 8a)。基于景别特征计算结果进一步分析发现,全景和中景的短视频平均 CTR 值要明显高于其他景别类型(如图 8b)。构图维度的两个特征分析结果在视频设计上表现出一定程度的一致性,即画面主体(模特或商品)能完整展示的短视频更吸引用户注意力。

内容特征分析见图 9。在内容(视觉)维度的设计特征中,视觉复杂度对视频 CTR 值有明显影响,表现为在当前数据分布范围内,随着视觉复杂度的升高,商品短视频的平均 CTR 值明显提升(如图 9a)。即画面纹理简单、区域变化均匀、视觉复杂度低的短视频有着更好的广告业务效果表现,例如纯色背景短视频等。这一结果可以与色彩特征中影调的分析结果相印证,共同表明了画面明亮、主体突出的短视频会更吸引用户点击。对于画面运动幅度,该设计特征对短视频 CTR 值有微弱影响,具体表现为在一定范围的运动幅度区间内,短视频的 CTR 值随运动幅度的增加而有所上升(如图 9b)。

3.4 方法应用

基于上述设计特征的分析结果,优化鹿班短视频的智能生产链路,为短视频智能设计提供支持。其应

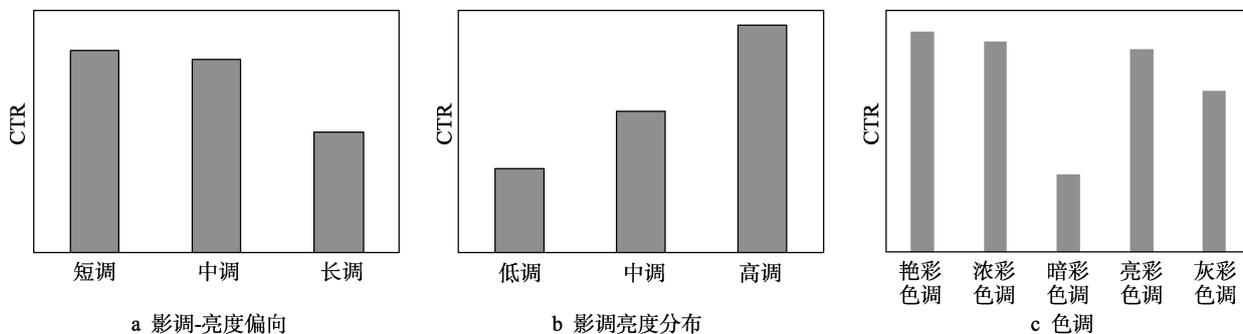


图 7 色彩特征分析
Fig.7 Analysis of color features

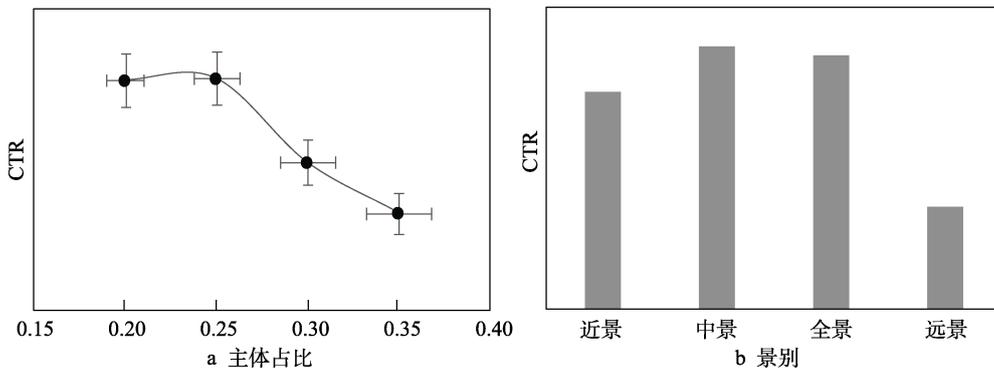


图 8 构图特征分析
Fig.8 Analysis of layout features

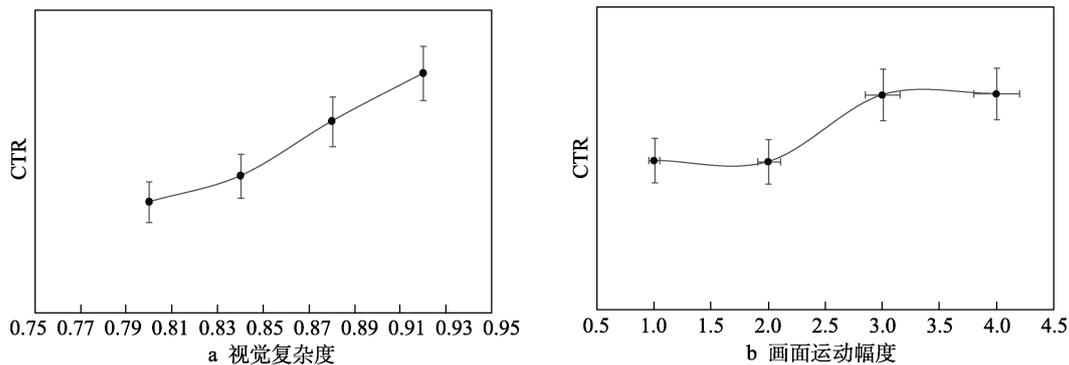


图 9 内容特征分析
Fig.9 Analysis of content features



图 10 短视频优化案例
Fig.10 Video optimization cases

用主要包括以下两个方面，一是短视频剪辑优化；二是短视频美化。短视频剪辑优化是在原有视频特征的基础上，加入构图和色彩特征，基于转场点识别对切分后的视频镜头进行优化筛选，组合成新的摘要视频。短视频优化案例见图 10。如图 10a 所示，其优化结果直观体现在视频景别和画面亮度的变化上。短视频美化则是在原有短视频生成结果的基础上，基于设计特征计算，实现自动化的背景替换、视频裁切、亮度调节等，从而对视频画面进行美化。其应用案例如图 10b。

4 结语

人工智能技术的迅速发展，使得设计评价方法从规则驱动向数据驱动转变。人工智能技术为设计评价研究与应用带来了更广阔的发展空间，也面临着诸多挑战。本文首先介绍了人工智能赋能的设计评价方法的发展趋势，然后重点描述了面向电商业务场景的短视频智能评价方法及其模型的构建，分析短视频设计

特征对业务效果的影响，展示了数据驱动的智能评价方法在阿里巴巴智能设计体系中的应用实践。人工智能赋能的设计评价方法是对智能设计方法的进一步补充，对智能设计研究的发展有着巨大作用。人工智能赋能的设计评价方法进一步提升了设计评价研究的应用范围和价值，并在以互联网企业为代表的产业界中逐步推广。

参考文献：

[1] 谭浩, 尤作, 彭盛兰. 大数据驱动的用户体验设计综述[J]. 包装工程艺术版, 2020, 41(2): 7-12.
TAN Hao, YOU Zuo, PENG Sheng-lan. Big Data-Driven User Experience Design[J]. Packaging Engineering, 2020, 41(2): 7-12.

[2] 周济, 查建中, 肖人彬. 智能设计[M]. 北京: 高等教育出版社, 1998.
ZHOU Ji, ZHA Jian-zhong, XIAO Ren-bin. Intelligent Design[M]. Beijing: Higer Education Press, 1998.

- [3] XU A, RAO H, DOW S P, et al. A Classroom Study of Using Crowd Feedback in the Iterative Design Process[C]. Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing, 2015.
- [4] KRAUSE M, GARNCARZ T, SONG J, et al. Critique Style Guide: Improving Crowdsourced Design Feedback with a Natural Language Model[C]. Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, 2017.
- [5] ORÚS C, GURREA R, FLAVIÁN C. Facilitating Imaginations through Online Product Presentation Videos: Effects on Imagery Fluency, Product Attitude and Purchase Intention[J]. Electronic Commerce Research, 2017, 17(4): 661-700.
- [6] WU Z, KIM T, LI Q, et al. Understanding and Modeling User-Perceived Brand Personality from Mobile Application Uis[C]. Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, 2019.
- [7] 周蕾, 薛澄岐, 汤文成, 等. 界面元素布局设计的美度评价方法[J]. 计算机辅助设计与图形学学报, 2013, 25(5): 758-766.
ZHOU Lei, XUE Cheng-qi, TANG Wen-cheng, et al. Aesthetic Evaluation Method of Interface Elements Layout Design[J]. Journal of Computer-Aided Design & Computer Graphics, 2013, 25(5): 758-766.
- [8] WANG W, CAI D, WANG L, et al. Synthesized Computational Aesthetic Evaluation of Photos[J]. Neurocomputing, 2016(172): 244-252.
- [9] IQBAL A, VAN DER HEIJDEN H, GUID M, et al. Evaluating the Aesthetics of Endgame Studies: a Computational Model of Human Aesthetic Perception[J]. IEEE Transactions on Computational Intelligence and AI in Games, 2012, 4(3): 178-191.
- [10] MINIUKOVICH A, DE ANGELI A. Computation of Interface Aesthetics[C]. Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, 2015.
- [11] REINECKE K, YEH T, MIRATRIX L, et al. Predicting Users' First Impressions of Website Aesthetics with a Quantification of Perceived Visual Complexity and Colorfulness[C]. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 2013.
- [12] LI C, CHEN T. Aesthetic Visual Quality Assessment of Paintings[J]. IEEE Journal of Selected Topics in Signal Processing, 2009, 3(2): 236-252.
- [13] JIA J, HUANG J, SHEN G, et al. Learning to Appreciate the Aesthetic Effects of Clothing[C]. Thirtieth AAAI Conference on Artificial Intelligence, 2016, 30(1): 1-7.
- [14] CUTUMISU M, BLAIR K P, CHIN D B, et al. Assessing Whether Students Seek Constructive Criticism: the Design of an Automated Feedback System for a Graphic Design Task[J]. International Journal of Artificial Intelligence in Education, 2017, 27(3): 419-447.
- [15] 周子洪, 周志斌, 张于扬, 等. 人工智能赋能数字创意设计: 进展与趋势[J]. 计算机集成制造系统, 2020, 26(10): 2603-2614.
ZHOU Zi-hong, ZHOU Zhi-bin, ZHANG Yu-yang, et al. Artificial Intelligence Empowered Digital Creative Design: Today and Its Future[J]. Computer Integrated Manufacturing Systems, 2020, 26(10): 2603-2614.
- [16] DOU Q, ZHENG X S, SUN T, et al. Webthetics: Quantifying Webpage Aesthetics with Deep Learning[J]. International Journal of Human-Computer Studies, 2019(124): 56-66.
- [17] SQUALLI HOUSSAINI H, DUONG N Q, GWENAËLLE M, et al. Deep Learning for Predicting Image Memorability[C]. 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2018.
- [18] BOSSE S, MANIRY D, MÜLLER K-R, et al. Deep Neural Networks for No-Reference and Full-Reference Image Quality Assessment[J]. IEEE Transactions on Image Processing, 2017, 27(1): 206-219.
- [19] MA N, VOLKOV A, LIVSHITS A, et al. An Universal Image Attractiveness Ranking Framework[C]. 2019 IEEE Winter Conference on Applications of Computer Vision (WACV), 2019.
- [20] KHOSLA A, DAS SARMA A, HAMID R. What Makes an Image Popular?[C]. Proceedings of the 23rd International Conference on World Wide Web, 2014.
- [21] KRISHNAN S S, SITARAMAN R K. Understanding the Effectiveness of Video Ads: a Measurement Study[C]. Proceedings of the 2013 Conference on Internet Measurement Conference, 2013.
- [22] AMINI F, RICHE N H, LEE B, et al. Hooked on Data Videos: Assessing the Effect of Animation and Pictographs on Viewer Engagement[C]. Proceedings of the 2018 International Conference on Advanced Visual Interfaces, 2018.
- [23] KRAFKA K, KHOSLA A, KELLNHOFER P, et al. Eye Tracking for Everyone[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016.
- [24] DEMARTY C H, SJÖBERG M, CONSTANTIN M G, et al. Predicting Interestingness of Visual Content[M]. Springer: Visual Content Indexing and Retrieval with Psycho-Visual Models, 2017.
- [25] 刘丽娟, 杨佳玉, 肖颖喆. 基于情感交互的包装设计评价方法研究[J]. 包装工程, 2020, 41(9): 181-185.
LIU Li-juan, YANG Jia-yu, XIAO Ying-zhe. Evaluation Method of Packaging Design Based on Emotional Interaction[J]. Packaging Engineering, 2020, 41(9): 181-185.
- [26] YOON S, PAVLOVIC V. Sentiment Flow for Video Interestingness Prediction[C]. Proceedings of the 1st ACM International Workshop on Human Centered Event Understanding from Multimedia, 2014.