arXiv:2601.07850v1 [cs.MM] 8 Jan 2026

MLLM-VADStory：领域知识驱动的多模态大语言模型助力视频广告故事线洞察

Jasmine Yang Meta 美
国纽约州纽约
yumingyang@meta.com

Poppy Zhang Meta 美
国纽约州纽约
poppyzhang
@meta.com

Shawndra Hill Meta
美国纽约州纽约
sha
wndrahill@meta.com

摘要

我们提出了 MLLM-VADStory，这是一种新颖的领域知识引导型多模态大语言模型（MLLM）框架，旨在大规模、系统地量化并生成视频广告故事线理解方面的洞察。该框架的核心理念是：广告叙事由功能意图构建，每个场景单元都承担独特的交际功能，在数秒内传递产品和品牌相关信息。MLLM-VADStory 将广告划分为多个功能单元，利用一种新颖的广告专用功能角色分类体系对各单元的功能进行分类，然后汇总不同广告中的功能序列，以还原数据驱动的故事线结构。我们将该框架应用于四个行业细分领域的 5 万条社交媒体视频广告，发现故事型广告创意能够提高视频留存率，并建议采用表现最佳的故事弧线，以指导广告主进行广告创意设计。我们的框架展示了利用领域知识引导多模态大语言模型生成可扩展的视频广告故事线洞察的价值，可作为理解视频广告创意的通用工具。

CCS 概念：• **通用与参考** → **实证研究**；**测量**；• **应用计算** → **媒体艺术**；• **信息系统** → **内容分析与特征选择**。

关键词：视频广告创意、故事线设计、多模态大语言模型、功能角色、广告效果

1 引言

视频广告（“广告创意素材”）的叙事结构处于一个连续谱上，[1]，一端是非叙事广告，通常采用视觉拼接，着重展示产品促销信息和功能；另一端则是叙事广告，通过产品体验或使用产品带来的个人转变来讲述故事。尽管故事已被证明是能在不同媒体中娱乐消费者的有力载体，[例如3, 4, 7, 8]，人们对故事如何影响广告效果，尤其是哪些因素能让某些广告故事比其他故事更有效，仍知之甚少。了解这一设计空间，有助于寻求可大规模应用的行动洞见的广告主和平台优化其创意策略。

本文介绍了*MLLM-VADStory*，这是一个由广告领域专业知识驱动、基于多模态大语言模型的框架，旨在系统地大规模量化和衡量视频广告故事及其结构弧线。与以往通过情绪和信息流来构建故事结构的研究不同 [4, 8]，我们的方法基于一个新颖的理念：短视频广告创意中的故事讲述在功能上不同于传统的故事结构（例如弗莱塔格金字塔 [例如 10]）：广告故事必须在数秒内传达信息，并融合产品与品牌导向信息，以实现特定的营销目标。因此，广告故事可以视为由一系列功能意图构成的结构化序列，其中每个场景单元都承担特定的传播目的。

我们的框架整合视觉和音频信号，将视频分解为语义连贯的功能单元，然后基于我们构建的、以广告领域专业知识为基础的新型功能角色分类体系，引导多模态大语言模型描述每个单元的功能意图。通过总结视频广告创意中的功能序列，我们揭示了数据驱动的叙事结构；这些结构可通过建模与广告效果指标相关联，从而大规模生成洞见。

我们将此框架应用于四个行业细分领域中的 5 万条活跃视频广告，以归纳创意故事设计的最佳实践。我们发现，故事有助于提升视频留存率，而表现最佳的故事结构弧线因行业细分领域而异。对广告的分析揭示了一些早期规律：以吸引点开篇、采用问题驱动弧线可能有助于提高停留率；展示社会认同并提供明确解决方案，则可能有助于促进产品转化。

我们的贡献主要有三点：

- 据我们所知，鉴于分析视频广告创意所需的可扩展性和计算复杂度，本研究是首个利用多模态大语言模型，基于大规模广告创意素材理解视频创意故事线的工作。
- 我们提出了*MLLM-VADStory*，这是一种基于多模态大语言模型的新型视频故事框架，以广告领域专业知识为基础，通过构建独特的功能角色分类体系，并依据视频广告背后的传播中的功能意图对故事线进行结构化，从而捕捉数据驱动的故事线结构。
- 我们大规模提炼出可解释、可执行的故事洞察，并展示了该框架的价值

它可用于视频创意理解，是视频洞察生成的多功能工具。

2 方法：MLLM-VADStory

我们提出了*MLLM-VADStory*，这是一个领域知识驱动框架，整合视觉和音频模态，借助多模态大语言模型系统地量化嵌入广告创意素材中的故事线结构。图1展示了该框架。该框架首先利用基于视觉和语音的信号，将视频分解为语义连贯的片段，这些片段对应于叙事中的功能转变（例如问题构建、产品展示或功能说明）。随后，我们采用三步法进行叙事理解：第一步，将各单元中的视觉和语音线索输入多模态大语言模型，以检测视频创意中是否存在故事（在我们的案例中，使用 Llama 多模态大语言模型视频理解模型[1]），这一过程基于由领域知识充分支持的故事定义。第二步，使用 Llama 多模态大语言模型处理每个独立单元，对视觉和文本线索进行联合推理，并在叙事学原理和广告领域专业知识的指导下生成功能角色预测。最后，在第三步中，我们总结检测到的叙事功能序列，以获得可扩展的数据驱动的故事结构。我们将故事和故事结构与广告效果指标关联起来，利用机器学习模型进行预测分析，从而获得可扩展的洞察（见第3节）。

2.1 将视频分割为功能单元

我们提出了一种内容感知多模态分割方法，整合视觉和音频语言线索。该方法并非仅依赖基于视觉的分割（PySceneDetect 的内容感知检测器[2] 及自适应阈值）。如果视频创意由承担相同功能的场景组成视觉拼接，这种方法可能会对其过度分割；因此，我们还使用 Whisper[3]，这一稳健的语音识别模型提取自然语音边界。我们根据时间线索（例如停顿、静音）和语言转换（例如“于是”和“然后”等连词）推导语音分割时间戳。随后，我们对两种分割流进行同步和细化，生成能够展现视频广告创意中叙事转换方式的功能单元。

2.2 叙事测量

2.2.1（S1）故事检测。我们首先根据领域知识驱动的故事定义，引导 Llama 多模态大语言模型结合视频创意在所有单元中的视觉场景和语音文字稿，对广告进行评估，以检测其中是否包含故事。在既有研究对叙事的支持下[例如2, 5, 6, 9]，我们将故事定义为“对某一事件或一系列相互关联的事件的叙述，这些事件促成了从初始状态向后续阶段或结果的转变”。

[1] https://www.llama.com/ [2] https://www.scenedetect.com/docs/latest/api/detectors.html [3] https://github.com/openai/whisper

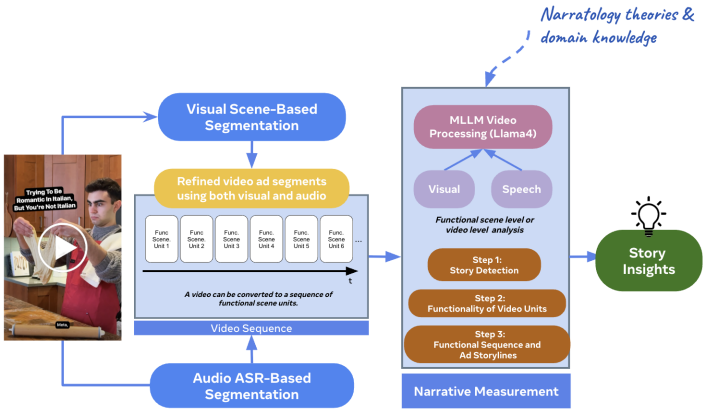

图 1. 故事理解视频分析流程

一系列相互关联的事件，这些事件促成了从初始状态向后续阶段或结果的转变”。

2.2.2 (S2) 视频单元的功能。

功能角色分类体系。 为了理解故事如何通过功能序列展开，我们随后基于广告领域知识，提出了一套细粒度、互不重叠且灵活的功能角色分类体系（见表 1），用于刻画每个单元所承担的功能。该分类体系涵盖六个核心类别，共同覆盖视频的完整范围。(1)**开场与背景铺陈**：在广告开头吸引观众注意，并确立场景。(2) **问题、需求与挑战**：呈现促使叙事展开的矛盾、需求或痛点。(3)**产品介绍与说明**：展示产品或服务如何解决问题，并传达其特性或益处。(4)**说服性框架**：通过情绪、促销激励、时间线索及质量相关线索等不同提示，凸显产品的吸引力。(5) **收尾与品牌身份**：通过传达结果、行动号召或品牌身份来提供解决方案或强化品牌。(6) **其他**：涵盖维持节奏或基调、但不推进核心叙事的过渡性或辅助性元素。

功能角色检测。 我们使用 Llama 多模态大语言模型，根据功能单元中的视觉场景和语音转录文本，并在提示中提供完整的分类体系及各角色的描述，自动推断每个功能单元所承担的主要功能。图2 展示了一个示例。

2.2.3（S3）功能序列与广告故事线。我们使用 Llama-4-Scout-17B-16E Instruct[4] 总结视频广告创意中的功能序列，并通过人工参与命名和合并语义相似的结构，以获得宏观故事结构

[4] https://ai.meta.com/blog/llama-4-multimodal-intelligence/

表1. 视频广告创意的功能角色分类体系

开场	**吸引点：** 吸引观众的注意力或兴趣，但不一定与产品相关；通常出现在视频开头的几秒钟内。**建立背景：** 在故事推进之前，交代故事的现状——人物、地点或时间。
问题、需求与挑战	**问题设定：** 首次提出待解决的问题、需求或痛点 **问题强化：** 强化问题，使其更容易引起共鸣，或显得更加严重。
产品介绍与说明	**功能说明：** 解释产品特征及其带来益处的原因；不只是展示。 **产品亮点展示：** 展示产品的关键属性或益处（仅作表层展示，不深入说明）。 **演示/试用：** 展示产品被使用或测试以完成某项任务**对比：** 将产品与竞品或先前状态进行对比。 **社交证明：** 展示其他人的群众评价或推荐语。 **解决方案揭示：** 将产品呈现为某个问题的解决方案
说服性框架	**情感诉求：** 运用情绪与受众建立联系并吸引受众。 **幽默：** 运用喜剧元素进行娱乐，并使信息更易引发共鸣。**愿景描绘：** 描绘由产品带来的理想生活方式或未来。 **促销：** 传达优惠机制：折扣、捆绑套餐、优惠码、定价条款或价值主张细节（不使用紧迫性语言）。 **紧迫性触发：** 增加时间压力以加速行动。 **稀缺性触发：** 强调供应有限，以制造错失恐惧并促使行动；不同于一般促销。**警示：** 提醒观众注意潜在风险或免责声明。
收束与身份认同	**行动号召：** 引导观众采取行动；促使其立即行动。**结果：** 展现干预后的收益或转变。**品牌塑造时刻：** 展现品牌身份（例如标志、标语、口号）。**洞察／理念：** 表达品牌或产品理念；引导观众发现关于生活、人类行为或世界运作方式的新事物。**预告／悬念：** 让叙事保持开放式结尾，以吸引观众再次参与。
其他	**视觉填充：** 提供过渡性的节奏，不对叙事作出贡献。

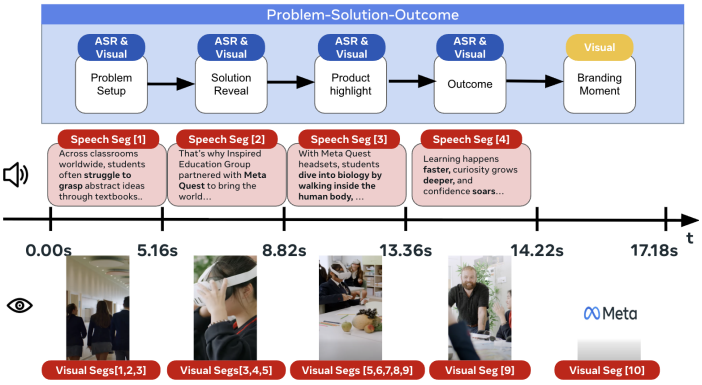

图 2. 视频创意的故事线示例

以图 3 展示了该方法以及一些数据驱动的故事线示例。

创意策略师对多模态大语言模型完成的所有步骤分类进行了人工检查，并凭借领域专业知识确保其准确性。

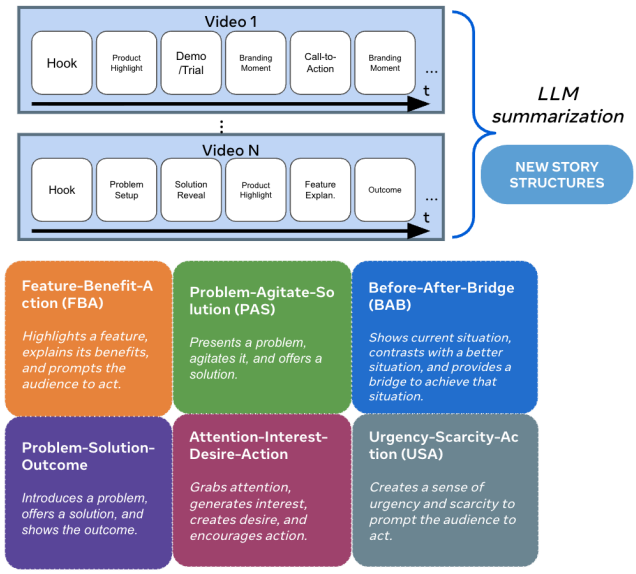

图 3。 功能序列摘要及故事线结构示例图解

3 应用

数据与设置。 我们将该框架应用于 50,000 条活跃的社交媒体视频广告。我们聚焦于英语广告，这些广告均包含单条视频，时长为 15 至 60 秒，来自四个行业细分领域（服饰与配饰、美容、食品和饮料）。此外，我们将分析范围限定为美国广告主投放的广告，这些广告在上线首日达到一定的每日展示量和预算门槛，以确保广告主为这些广告创意素材投入了精力，并使研究结果能够反映故事设计的最佳实践。

广告故事洞察。 在各类广告创意素材中，我们发现：（1）平均有 48.2% 的广告创意素材包含一个故事，其中饮料（65%）和美容（53.7%）在故事讲述上的投入要高得多；(2) 规模较大的广告主会更多地投入故事内容，而故事内容的集中度会随着广告主规模的增大而降低；(3) 当广告主选择的广告系列目标是提升品牌认知和促进销售时，故事内容的使用会更多；(4) 故事内容在提升视频开头几秒的整体视频留存率方面最为有效。图 4 展示了视频前 10 秒内的故事内容对平均停留时间百分比变化的影响（定义为观看一定秒数的观众比例在多个视频上的平均值），该结果通过普通最小二乘法（OLS）计算，并控制了视频相关特征（例如是否包含语音、视频时长）、广告及广告系列相关特征（例如广告系列目标、受众规模）。我们发现，绝对百分比变化在 2 秒时达到最高。以基准值为

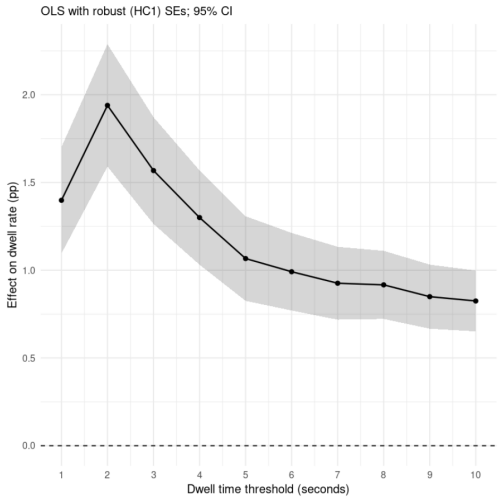

图 4. 广告包含故事时，前 10 秒停留率的变化百分比

表2. 按广告指标和细分垂直领域划分的主要故事线模式

按广告指标划分 细分垂直领域	表现最佳的故事线	提升幅度
2秒停留率		
服装与配饰	钩子–特性–优势–行动（HFBA）	+5.8%
美容	钩子–问题–演示–解决方案（HPDS）	+4.9%
Food	问题-激化-解决方案（PAS）	+6.3%
饮料	钩子–问题–解决方案（HPS）	+4.1%
点击率（CTR）		
服装与配饰	注意–兴趣–欲望–行动（AIDA）	+8.9%
美容	钩子–特性–优势–行动（HFBA）	+2.8%
Food	注意–兴趣–欲望–行动（AIDA）	+4.8%
饮料	特征-利益-行动（FBA）	+3.9%
转化率（CVR）		
服装与配饰	社交证明-行动（SPA）	+4.6%
美容	社交证明-行动（SPA）	+3.7%
Food	问题-激化-解决方案（PAS）	+8.5%
饮料	特征–亮点–解释–行动（FHEA）	+5.1%

不含故事内容的广告的平均停留率计算，相对停留率约提升 5% 至 10%。

我们还要求多模态大语言模型解释广告中为何存在故事，并总结不同广告创意素材中的共性，以提取故事信号。与以往的营销见解一致 [7],，我们发现，对话、分享个人经历、呈现挑战/冲突/问题解决方案等，通常意味着故事的存在。另一方面，出现播报员、使用宣传用语，以及着重强调产品特征或益处，则意味着故事缺失。

广告故事线结构洞察。。跨不同子垂类，我们发现：（1）最常用的故事线结构是特征–益处–行动和问题–解决方案类型的故事线（问题–激化–解决方案、吸引点–问题–解决方案，以及问题–解决方案–结果）；（2）与其他子垂类相比，之前–之后–桥梁在美容领域的使用频率要高得多，其次是服饰领域；（3）与其他子垂类相比，美容领域的视频中故事线结构的使用方式更多样；（4）注意–兴趣–欲望–行动出现在食品领域采用率最高的结构之列，这可能是因为从心理层面来说，食欲和购买动机受欲望驱动。

为预测故事结构对广告效果（前两秒后的停留率、点击率、转化率）的提升幅度，我们针对每个细分垂直领域拟合 XGBoost 模型。这类机器学习模型利用梯度提升决策树，同时将视频特征（例如是否包含音频、是否有语音、视频时长、宽高比）和其他元数据（例如受众、优势定向）纳入考量。表 2 通过部分依赖关系，展示了对各项广告指标提升贡献最大的故事结构。我们发现，虽然最有效的故事结构因细分垂直领域而异，但四个细分垂直领域也呈现出一些共同趋势。总体而言，吸引点优先、以问题驱动的弧线可能有助于提高前两秒后的停留率。清晰地呈现从吸引注意到激发兴趣/展示益处的弧线，可能有助于提升点击率；展示社会认同和明确的解决方案，则可能提高转化率。广告主在优化或设计新的广告创意素材以开展进一步实验时，可以参考这些见解。

4 结论

我们提出了 MLLM-VADStory，这是一个由广告领域知识驱动的框架，利用多模态大语言模型系统地量化广告创意素材中蕴含的故事线结构。我们将该框架大规模应用于视频广告创意，从而为广告主提炼出可解释的故事洞见。我们希望这项工作能够激发更多研究，利用领域知识指导多模态大语言模型进行视频创意理解。

参考文献

[1] Jennifer Edson Escalas。2003。广告叙事：它们是什么，又如何发挥作用？载于*Representing consumers*。Routledge，283–305。[2] Gérard Genette。1980。叙事话语：方法论论文。第3卷。康奈尔大学出版社。[3] Wesley R Hartmann 和 Daniel Klapper。2018。超级碗广告。营销科学37，1（2018），78–96。[4] Samsun Knight、Liu Liu 和 Laura J Kornish。2025。通过情感序列构建有说服力的故事。可见于 *SSRN 5489708*（2025）。[5] Robert McKee。1997。内容、结构、风格与编剧原理。*Alba Editorial*（1997）。[6] Andrew Piper、Richard Jean So 和 David Bamman。2021。面向计算叙事理解的叙事理论。载于*2021*年自然语言处理实证方法会议论文集。298–311。[7] Ron Shachar、Lev Muchnik 和 Oded Netzer。2025。给我讲个故事。可见于 *SSRN 5236601*（2025）。[8] Olivier Toubia、Jonah Berger 和 Jehoshua Eliashberg。2021。量化故事形态如何预测其成功。《美国国家科学院院刊》118，26（2021），e2011695118。[9] Tom Van Laer、Jennifer Edson Escalas、Stephan Ludwig 和 Ellis A Van Den Hende。2019。拉斯维加斯发生的事会留在 TripAdvisor 上吗？理解消费者评论叙事性的理论与方法。《消费者研究杂志》46，2（2019），267–285。[10] Leni Yang、Xian Xu、XingYu Lan、Ziyan Liu、Shunan Guo、Yang Shi、Huamin Qu 和 Nan Cao。2021。将弗莱塔格金字塔结构应用于数据故事的设计空间。《*IEEE*可视化与计算机图形学汇刊》28，1（2021），922–932。