



中国管理科学  
*Chinese Journal of Management Science*  
ISSN 1003-207X, CN 11-2835/G3

## 《中国管理科学》网络首发论文

题目：可读性和吸引力对商品销量的影响——基于电影简介的文本分析  
作者：高维和，刘德文，闵凉宇  
DOI：10.16381/j.cnki.issn1003-207x.2019.1776  
收稿日期：2019-11-07  
网络首发日期：2020-09-21  
引用格式：高维和，刘德文，闵凉宇. 可读性和吸引力对商品销量的影响——基于电影简介的文本分析. 中国管理科学.  
<https://doi.org/10.16381/j.cnki.issn1003-207x.2019.1776>



**网络首发：**在编辑部工作流程中，稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定，且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式（包括网络呈现版式）排版后的稿件，可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定；学术研究成果具有创新性、科学性和先进性，符合编辑部对刊文的录用要求，不存在学术不端行为及其他侵权行为；稿件内容应基本符合国家有关书刊编辑、出版的技术标准，正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性，录用定稿一经发布，不得修改论文题目、作者、机构名称和学术内容，只可基于编辑规范进行少量文字的修改。

**出版确认：**纸质期刊编辑部通过与《中国学术期刊（光盘版）》电子杂志社有限公司签约，在《中国学术期刊（网络版）》出版传播平台上创办与纸质期刊内容一致的网络版，以单篇或整期出版形式，在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊（网络版）》是国家新闻出版广电总局批准的网络连续型出版物（ISSN 2096-4188，CN 11-6037/Z），所以签约期刊的网络版上网络首发论文视为正式出版。

DOI: 10.16381/j.cnki.issn1003-207x.2019.1776

# 可读性和吸引力对商品销量的影响

## ——基于电影简介的文本分析

高维和<sup>1</sup>，刘德文<sup>1</sup>，闵凉宇<sup>2</sup>

(1. 上海财经大学 商学院，上海 200433；

2. 上海财经大学 信息管理与工程学院，上海 200433)

**摘 要：**商品简介为企业提供着向消费者传递商品信息，并激起消费者购买欲的重要作用。但是鲜有研究对商品简介文本的营销效果进行分析。为了弥补上述研究缺陷，本文基于营销实践的特征和相关文献梳理，并采用机器学习的方式度量了中国大陆 2018 年上映 264 部电影简介文本的可读性和吸引力水平，并检验其对首周票房和总票房的影响。研究发现，简介文本的可读性和吸引力对首周票房和总票房有差异化的影响，可读性对于首周票房的影响较大，而吸引力对总票房的影响较大。此外，明星的存在削弱了文本可读性和吸引力对票房的影响；而且，在高电影评分组中，可读性和吸引力对首周票房及总票房的负向交互作用显著。本文基于启发式-系统式模型凝练出商品简介文本的可读性和吸引力两个特征并研究其对票房的影响，对丰富中国电影研究，指导中国电影营销实践具有重要参考价值。

**关键词：**电影简介；可读性；吸引力；机器学习

**中图分类号：**F274

**文献标识码：**A

### 1 引言

商品简介是企业与消费者进行沟通和交流的基石。在消费者有限注意的限制条件下，企业须精心地设计简介文本，以达到传递产品信息并唤起消费者购买欲的目的。尤其在网络环境中，消费者面临着信息过载。大量商品在同一时间呈现，消费者不可能仔细阅读每个商品的具体介绍，而商品简介提供给消费者类似于论文摘要的功能。在选择商品时，消费者会根据企业提供的简介文本形成第一印象，继而对其后续行为产生影响。现有研究从多个视角出发，指出文本特征的重要作用。例如，文本的可读性<sup>[1,2]</sup>、说服力<sup>[3]</sup>、情感倾向<sup>[4]</sup>等。现有的研究基于英语语境探讨了商品评论相关特征对产品绩效的影响<sup>[1,5]</sup>，忽略了企业提供的自身商品介绍的相关文本特征的结果影响。但商品评论是消费者自发形成的，属于企业所不能控制的范畴。企业自身提供的文本信息能够对产

品表现产生何种影响还缺乏探索。

作为一种典型的体验商品<sup>[6]</sup>，电影消费者在实际购买前难以直接获得消费经验。影评、电影介绍等外部信息成为消费者的重要决策参考。与此同时，各大网络购票渠道的涌现与高概念电影的盛行为本文提供了良好的研究情景。在电影商业的实践中，发行公司采用不同的文本形式来描述和推介电影<sup>[7]</sup>。作为影响消费者购买决策的重要信息来源，电影简介承担着向消费者梗概电影情节，介绍主要人物和突出电影亮点的任务；同时，电影简介也必须适当给消费者营造悬念，并激发其好奇心和观影欲。以往文本研究的结果变量仅包括消费者分享、评论和点赞等非购买变量。由于难以在总体层面上匹配产品销量，用实际的二手数据来量化文本结果影响的研究不多。作为一种典型的体验商品和享乐商品<sup>[6]</sup>，电影简介的什么特征能够在什么程度上裨益其票房收入的提升，文献并没有给出解答。因此，本文欲以 2018 年上映的电影为实证对象，利用机器学习的方法对电影简介的文本进行建模，并从文本的可读性和吸引力两个角度出发来研究其对票房的影响。

本文的边际贡献如下：第一，从文本研究领域出发，本文聚焦于电影简介文本的可读性和吸引力对短期票房和长期票房的影响。以往文本研究强调文本应该易于理解，

**收稿日期：**2019-11-07；**修订日期：**2020-07-17

**基金项目：**国家自然科学基金资助项目(71872106, 71572099)；上海财经大学研究生创新基金(CXJJ-2019-325, CXJJ-2018-401)

**通讯作者简介：**刘德文(1991-)，男(汉族)，上海财经大学商学院博士研究生，研究方向：文本处理、电影营销，E-mail: dewen0731@126.com

即可读性要高。本文从电影营销实践出发,提炼并提出了独立于可读性的另外一个特征:吸引力。随着营销文本研究的逐渐兴起,本文也为该领域的理论研究提供了有益补充。第二,在测量文本可读性和吸引力方面,本文采用了机器学习的方法,对上述两个指标进行度量,机器学习克服了以往文本度量的噪音问题,为后续的定量分析奠定基础。第三,作为电影文本简介的开拓性研究,本文的研究结果对于电影营销理论研究具有重要意义。电影简介撰写应同时考虑提升文本的可读性和吸引力,以促进短期和长期的票房增长。第四,2018年中国票房首次超越美国,成为全球第一大票仓。但与中国迅速崛起的电影实践相比,针对中国电影市场的研究较为匮乏。中美电影市场诸多的不同点使得研究结论难以直接复制。从电影营销宣发的角度出发,本文的研究也为中国电影市场实践提供借鉴。

## 2 文献回顾及假设提出

### 2.1 启发式-系统式模型

个体对外界信息(如,文本)的学习和处理通常遵循两条并行路径:利用情绪和先验经验的启发式信息处理和以认知和深层推理为基础的系统式信息处理。当前者被激活时,个体提取信息的表层或者边缘特征,以情感和经验为标准来对信息进行判断和处理;而当后者被激活时,个体提取信息的深层或者中心特征,以认知和推演对信息进行决策<sup>[8]</sup>。大量学者采用该模型对文字信息的说服效果进行研究<sup>[9]</sup>。例如,Guo等<sup>[10]</sup>认为在线评论的情感内容启发了阅读者的启发式处理,并通过情绪感染的过程影响了消费者购买的可能性,而评论可信性和可诊断性则扮演着系统式线索对消费者决策施加影响。本文认为,营销情境下的电影简介应具有以下两方面的功能:在认知层面,它要提供给消费者清晰、易懂、且较为完备的电影情节、特征和其他相关信息,启动消费者的系统式信息处理路径,使得消费者可通过阅读文本就能对电影商品有较为完整的认识;在情绪层面上,文本内容要能够使得消费检索到相关信息,调动消费者的好奇心,引发消费者的共鸣,激活消费者的启发式信息处理路径,最终使得消费行为得以发生。

长期以来,学者们就文本中何种线索是

启发式及系统式存在着争议<sup>[9]</sup>。Chaiken<sup>[11]</sup>指出,信息线索自身不存在启发式信息线索和系统式信息线索之分,个体在决策时对信息线索的努力成本决定了到底启动何种信息处理模式。例如,有学者用信息来源作为启发式的代替变量<sup>[9]</sup>,但对于一个生僻的信息来源,个体甚至需要用更多的认知成本去思考处理这条线索,这反而启动了个体的系统式处理。上述的观点也得到了Sherman<sup>[12]</sup>的支持,即基于双系统理论进行的研究不能将内容和信息处理过程混淆。到底一条信息线索引发的是启发式信息处理还是系统式信息处理是个体的“心理黑箱”。启发式信息处理之所以能够减少认知努力,其中重要一条原因是:在消费者进行决策的过程中,某些信息线索可能会被消费者更容易地检索到<sup>[13]</sup>。从而,这些信息“主宰”了消费者的决策系统,简化了消费者赋予不同信息线索的规则。因此,本文使用一个能够整合克服以往研究中混淆线索和处理路径的缺陷,从而更为精准地判断消费者信息决策过程的研究方法势在必行。

### 2.2 文本的可读性

可读性意指读者是否能够轻松理解文本,对应了认知层面上消费者的系统式信息处理路径<sup>[14]</sup>。文本的可读性则是个体理解信息并改变态度的前提。大量研究针对可读性对于个体态度和行为的改变进行了有益探索。在公司治理文献中,有研究指出文本撰写者为了达到某些目的会对文本内容进行操纵,从而对文本的可读性产生影响,继而最终影响阅读文本者的决策。例如,业绩较差的公司为了隐瞒自己的不良经营状况,会刻意在年报中使用大量的专业术语、逻辑词和生僻字,以此增加文本的复杂性,达到获取额外薪酬,提高公司估值的目的<sup>[14]</sup>。营销学的相关研究也指出,可读性利于信息的广泛传播和接受,从而对信息接受者起到劝说作用<sup>[15]</sup>。Burtch等<sup>[2]</sup>的研究指出,高可读性的项目描述会吸引更多的投资者;类似研究结论也在中文情境中得到了复制<sup>[16]</sup>,借款文本的可读性有利于减少信息噪音,从而向借款者传递积极的信号。

在衡量可读性上,通常的做法是利用文



本的特定属性内容来对其进行表征。例如，文档的大小可以直接用来表征可读性；文本中特定的词，如逆接成分密度、专业术语密度和次常用词密度等；公式，如 Fog 指数，Flesh 易读指数。本文认为，消费情境下的文本更具异质性和分散性<sup>[17]</sup>。文本的可读性不仅取决于组成文本的每个单词，也与词和词之间的相对属性有所关联。已有学者开始考虑用更综合的可读性衡量方式。例如，Moon 等<sup>[5]</sup>采用对评论文本进行聚类并计算评论之间的相似性距离对可读性进行计算。文本的诸多特征使得单一地选取某几个量化指标对可读性进行评价的方法解释力较弱，随着自然语言技术处理方法的兴起，利用机器学习中支持向量机的方法将文本中的词集按照规则抽象为具有  $n$  维特征的向量，继而通过优化求解得到一个分割超平面来预测文本可读性算法。这种方法的适用性更优，且对语言的依赖性较小，并已经在研究中得到了初步运用。

既往研究指出，可读性较高的文本会加快个体的信息处理速度，并进一步降低个体的风险感知<sup>[18]</sup>。对于文本而言，其主要功能就是有效地提供信息，因而，准确、便于理解和无逻辑冲突的文本才能达到改变消费者态度的目的。本文认为，电影简介的文本可读性代表了简介文本信息是否清晰、明了，易于消费者理解电影梗概。之所以可读性会对电影票房产生作用主要基于以下两方面原因：第一，文本的可读性在一定程度上反应了电影制作方或发行方的规模和实力，因为只有大公司才能出具更加规范的文本，小公司的简介文本通常排版不规范，甚至还有语法错误和错别字。继而大公司的电影会使得消费者感知更低水平的风险。Tausczik 和 Pennebaker<sup>[19]</sup>针对借贷市场的研究就曾指出，借款人撰写的文本信息可读性与借款人的社会地位、教育水平等息息相关，因而可读性高的文本的作者更有可能完成借款。第二，由于可读性较高，消费者处理文本信息的程度大大降低，从而方便消费者记忆并将此信息传播给更多的人。Agnihotri 和 Bhattacharya<sup>[20]</sup>的研究就指出，评论的可读性能够显著影响消费者的点赞

数。基于此，本文提出如下假设：

H1：控制其他因素之后，可读性越高的电影简介，电影票房越高。

## 2.3 文本的吸引力

吸引力意指文本信息是否能够让唤起消费者的情绪共鸣和好奇心，启动了消费者的启发式信息处理路径。在传统的文本研究中，学者们认为公司在撰写文本时需要尽量使用平述、直白的字词以达到高效沟通的目的<sup>[14,16]</sup>。不同于上述文本，商品简介则是企业向消费者呈现的文本信息，除了提供商品信息之外，企业更希冀消费者能够通过阅读文本做出购买决策，故而可读性并不是吸引消费者的首要因素<sup>[5]</sup>。例如，以往的文本研究从可读性出发，指出文本需要尽量少用或者不用诸如“但是”等转折连词<sup>[2,16]</sup>，然而，这些连词却引发了阅读者的遐想<sup>[21]</sup>，驱使其想要一探究竟的兴趣，继而消费者在阅读此种商品描述后就更有可能去进行消费。同样，问号、省略号、语气词等也被视为削弱文本可读性的指标，但是这些词却在一定程度上增加了文本对读者的吸引。事实上，已有学者开注意到这一问题。Herzenstein 等<sup>[22]</sup>发现文本叙事中的身份宣示词汇，如可信的、成功的等会对贷款成功率和还款表现有差异化影响。其原因在于这些信息从情感上“打动”了贷款人，从而使得贷款人被借款人的故事所吸引。Xiang 等<sup>[23]</sup>的研究发现，企业在网站发布的众筹信息一方面要准确、清晰地描述企业所生产的产品（诉诸信息），同时也要生动地描述其创业理想和愿景（诉诸情感），这两个因素对于后续的众筹支持率有重要影响。本文认为，在电影简介文本中也存在着类似的效应。即除了用简洁的字句交代电影主旨内容后，也要能够在叙述上打动消费者，唤起消费者的购买欲。这一论述也得到了电影营销实践的佐证，目前越来越多的电影公司在试图创作高概念电影（high concept movie），以简单的情节叙述辅之以商业化、奇观化的电影概念以获取商业成功。根据电影学者贾斯汀·怀亚特的定义，高概念意指“一个显著的、容易的、可简化的叙述，同时也是提供了高程度可销售

性的叙述”，这种电影制作模式同样反应在电影的简介文本的撰写风格中。因此，本文认为，Brand 等<sup>[24]</sup>提出的文本吸引力在电影简介的文本情境下依然存在，即电影简介存在着独立于可读性的另一个特性：吸引力。文本吸引力的概念也得到了品牌生成内容研究的佐证。不同于消费者生成内容，品牌生成内容在于吸引消费者持续的互动；因此，有趣性和生动性较高的文本更有可能吸引消费者。吸引力高的文本可以使得信息受众在阅读过程中产生高水平的满足感，从而影响信息受众的互动行为<sup>[25]</sup>。在企业和消费者互动的社交媒体中，这种效应还会驱使消费者形成更加积极的品牌偏好和品牌态度。在衡量吸引力上。基于语义学的视角，Huang 和 Lo<sup>[25]</sup>指出文本的内容吸引力应包括文本的良好编排、内容优化以及信息融合，并通过模糊层次算法来提取文本中的关键信息以计算该指标；Brand 等<sup>[24]</sup>则采用了实验的方法让被试在阅读完文本后进行打分。也有研究改编了品牌吸引力的量表，让受访者对文本进行打分以测量其吸引力<sup>[26]</sup>。

作为一种典型的体验品，消费者的观影决策发生在瞬时<sup>[6]</sup>。研究指出，品牌吸引对消费者的购买意愿、支付溢价、甚至进一步的品牌认同、品牌忠诚等的积极影响。这是因为品牌吸引能满足消费者的自我定义动机，从而使消费者产生对品牌的积极态度和行为支持<sup>[27]</sup>。之所以简介文本的吸引力亦能对电影票房产生积极影响，主要基于下述理由：第一，简介文本的吸引力代表着文本中的某些线索能够在消费者记忆中被检索到，这可能是某个特定词性的词语或某个电影元素的表达，这使得消费者采取认知捷径来完成决策，从而观影决策就容易发生。Zhao 等<sup>[28]</sup>的研究就指出，为了唤起消费者的熟悉感，制作方在取名时往往借鉴以往成功电影的经验来获取票房成功。第二，文本的吸引力本身也能够激起消费者的情绪唤起，从而驱动消费者完成自己的消费目标。消费者观看电影是为了收获积极的情绪体验，故基于吸引力的文本往往能唤起消费的情感诉求。这种效应在传统的商品情景中已得到证实。如，杨学成等<sup>[29]</sup>的研究指出，内容生动有趣

的新浪微博更容易引发消费者的积极互动。此外，吸引力诱发的消费者积极情绪亦会导致其更积极的自发口碑传播。基于此，本文提出如下假设：

H2：控制其他因素之后，吸引力越高的电影简介，电影票房越高。

### 3 研究设计

#### 3.1 研究样本和数据来源

本文选取了 2018 年中国大陆地区上映的电影作为研究样本来研究商品描述对销量的影响。本文选择电影行业主要基于以下考虑：第一，中国电影市场迅速发展，已经于 2018 年取代美国成为世界第一大票仓。而线上购票占比超过 90%，因此对于网络购票终端的设计就尤为重要；第二，电影是典型的体验品，新产品特征明显，消费之前难以获得商品信息<sup>[6]</sup>，因此来自商家的推介成为影响消费者决策的重要因素；第三，不同于其他种类产品在实证层面上很难获得全国范围内的销售数据，而电影票房等公开信息为文本的研究奠定了良好的数据基础。具体，电影简介的文本信息来自猫眼电影。电影的票房信息来自于 CBO，电影排片数据来自于糯米网。此外，电影评分信息及其类型、导演、主演等信息来自于猫眼电影。在样本构建上，本文剔除了未在全国广泛上映的电影、检索不到相关信息的电影、上映时间少于 3 天的电影。通过以上步骤，本文共得到 264 份数据。

#### 3.2 主要变量衡量

##### 3.2.1 可读性和吸引力

为了研究文本可读性和吸引力对票房的影响，本文必须构建适合中文环境下的对可读性和吸引力的衡量方式。对于可读性而言，国外学者们提出了大量的对可读性进行测度的方式和字典。但是上述分析方法是基于英文语境的，中英文语言之间的差异较大，因此上述衡量方式难以复制到非英文书写的文本环境中。有学者提出了以现有汉语字典的常见和非常见字和词的出现频率作为文本可读性的一个维度<sup>[16]</sup>。然而，这样的做法仅仅关注于文本中单个的字与词，却忽略了字与字，词与词之间关系及相对位置对结果变量的影响。此外，仅仅用单个或几个

指标对可读性进行衡量可能会造成测量误差。对于吸引力而言,目前学术界就这一概念的表现还未达成共识,仅有 Brand 等<sup>[24]</sup>、Huang 和 Lo<sup>[25]</sup>几位学者提出了初步的量化方法。吸引力反映了消费者在阅读完文本后是否会对商品产生兴趣<sup>[24,25]</sup>。在电影营销实践中,大多数电影在撰写梗概文本时使用大量的转折词、助词、反问句式、疑问句式、省略号等来营造出一种“神秘感”,并希望简介中有所保留地呈现信息。然而,按照传统的可读性度量方法,上述这些词语大大削弱了文本可读性。

近年来,自然语言处理(NLP)结合机器学习模型(ML)逐渐进入文本研究中。为了更好地对文本的可读性和吸引力进行衡量,本文首先利用中科院心理研究所开发文心(TextMind)实现对电影简介的文本分词,它针对简体中文的语言特点,参照LIWC2007和正体中文C-LIWC词库实现了对文本进行语言分析的解决方案。在将文本导入系统之前,本文对电影简介文本进行了处理。文心根据所输入文本,自动析出103个的文本相对特征,包括各种词性词的数量、各种时代词的数量、各种标点符号数量等。随后,我们邀请若干名商科博士生对部分电影简介进行人工阅读,并对文本的可读性和吸引力进行打分,在打分之前我们也向其陈述了概念的定义。邀请商科博士生的原因在于他们对本文的两个概念有清晰的认识,而且他们平时鲜有时间去看电影(使得标签与特征尽可能直接关联)。根据上述人工打分的标签,使用机器学习的方法计算出评委团的打分模型,继而利用该模型给出所有电影简介的可读性和吸引力水平。由于在衡量文本可读性和吸引力时综合考虑了百个特征,本文的衡量方法也更加全面。而机器学习模型可以克服自然语言障碍,使得可读性和吸引力的衡量推广变得可能。

实现方式上,本文选择SVM模型进行文本预测和处理。相比较神经网络等算法而言,SVM模型本身具有较好有抗过拟合的能力,从而过滤掉数据中的噪声。通过设置不同核函数参数得到具有鲁棒性的分类器,在处理线性和非线性数据集中都具有良好表

现。Sun 等<sup>[30]</sup>使用SVM处理了存在训练样本不平衡分布的文本分类问题。其结果发现SVM获得最优决策超平面的概率最大。SVM模型因其具有强大的预测性和实用性,已被学者们广泛接受和运用。例如,Kübler<sup>[31]</sup>在提取消费者生成内容情感的研究中,就指出SVM模型可以较好地度量消费者在消费体验品时候的情绪和心态,且对后续消费者实际的购买倾向起到预测作用。

考虑到本文研究中涉及的两个主要指标(可读性和吸引力)本身带有较强的主观成分,即受到评判者较强主观情绪的影响较大。在该类学习任务下,如果使用追求较高精度的神经网络等学习器会导致模型对噪音也赋予了较高的权重,从而影响后续的影响因素剥离等任务。如果使用SVM分类器则可在损失较低精度的条件下,学习到指标的主要预测机制,对少部分带有较强主观情绪的投票结果进行适当过滤,从而获得一个具有鲁棒性的模型,有利于后续分析任务的推进。

具体而言,本文将第*i*部电影的简介文本抽象成高维向量 $x_i$ 。 $x_i = (x_{i1}, x_{i2}, \dots, x_{ij})'$ 表示该文本的*j*个属性特征,将该文本的可读性和吸引力定义为标签 $y_i$ ,因此样本电影简介就是一个特征空间上的训练数据集:

$$T = \{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}$$

其中 $x \in X = \mathfrak{R}^n, y \in Y, i = 1, 2, \dots, N$ ,在简化模型下,将标签空间 $Y$ 设定为 $\{-1, +1\}$ ,且令 $j = 2$ ,一个模拟的分类情形如图1所示,最终得到分割超平面的方程 $H$ ,对于一个新电影样本 $x_u$ ,根据该样本的属性 $(x_{u1}, x_{u2}, \dots, x_{u3})'$ 可以通过方程 $H$ 预测出该简介的 $y_u$ ,即预测该部电影简介的可读性和吸引力水平。

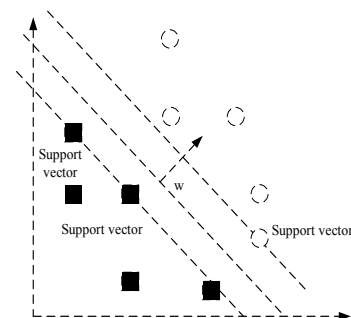


图1 SVM分类



### 3.2.2 其他变量

为了探究商品描述对销量的影响。本文选取了每部电影的首周票房和总票房作为因变量，以判别文本对于电影票房的短期和

长期影响。此外，参照以往电影的研究<sup>[6]</sup>，本文纳入了以下控制变量：是否续集、制作国别、排片、评分、档期、类型、是否明星参与。具体的变量说明见表 1。

表 1 变量说明及描述性统计

变量名称	变量测量	均值	标准差	p25	p50	p75
Readability	采用机器学习的方法衡量，简言之由提取文本特征、人工标记和训练机器学习并生成可读性或吸引力水平几个步骤生成	4.77	1.12	4.33	4.80	5.36
Attractiveness	电影的首周票房（万元，人民币）	5.28	1.48	4.33	5.33	6.32
First week BO	电影的总票房（万元，人民币）	8153.56	18370.46	243.60	1655.20	6324.85
Total BO	电影的首周排片数量（万块）	21570.71	52210.93	395.60	2736.20	13300
First week Screens	电影的总排片数量（万块）	1.70	2.48	0.17	0.81	2.15
Total Screens	电影是否为续集电影（是=1，否=0）	3.95	6.38	0.29	1.35	4.48
Sequel	电影的制作国家是否为中国大陆（是=1，否=0）	0.12	0.32	0	0	0
Nation	电影的评分（10 分制，猫眼评分）	0.45	0.50	0	0	1
Rating	电影上映是否热门档期（是=1，否=0），热门档期为贺岁档、五一档、暑期档、国庆档	7.85	1.27	7.4	8.2	8.7
Schedule	电影的主创团队是否有明星（是=1，否=0），明星的衡量为其主创团队成员是否在 2017 年度新浪的明星势力榜单上	0.25	0.43	0	0	0
Star	电影的类型（惊悚、爱情、动作、剧情、喜剧、动画）	0.33	0.47	0	0	1
Genre		-	-	-	-	-

## 4 数据分析与假设检验

### 4.1 描述性统计

首先，本文对涉及到的所有变量进行描述性统计（表 1）。对于可读性和吸引力而言，数值越高表示其文本在该项上更好。可读性和吸引力的范围在 1-9，均值和中位数均超过了 4.5，说明电影简介文本的可读性和吸引力都较优。从均值、标准差、分位数的数值关系可以大致推测可读性和吸引力分布较为均匀。此外，配对样本 t 检验显示（ $t(263) = 6.444$ ），可读性和吸引力的差异显著，这为后续检验提供了基础。

### 4.2 实证检验

在上一节描述性分析部分，本文发现所研究的因变量（票房）方差均值差异较大，

因此本文将因变量取对数后并使用稳健标准误进行回归来验证假设，回归结果见表 2。模型 1 和 5 是仅放入控制变量的结果，从中可以发现，排片、评分、档期、是否有明星对电影首周票房影响显著。然而，电影制作国家和是否续集对首周票房影响不显著。这有可能反应了中国电影愈发成熟，平均来讲，好莱坞大片在中国并不会获取超额收益。模型 2/6 在模型 1/5 的基础上加入了可读性，模型 3/7 加入了吸引力，模型 4/8 同时纳入两个变量。实证结果表明，简介文本的可读性和吸引力对首周票房和总票房都有积极影响。从系数上看，可读性对总票房的影响弱于其对首周票房的影响，而吸引力对首周票房和总票房的影响差别不大。因此，本文的两个假设基本得到了验证。

表 2 简介文本可读性和吸引力对票房的影响

变量	模型 1	模型 2	模型 3	模型 4	模型 5	模型 6	模型 7	模型 8
	因变量: First week Box Office				因变量: Total Box Office			
Readability	-	0.385*** (0.07)	-	0.252*** (0.07)	-	0.333*** (0.06)	-	0.197** (0.07)
Attractiveness	-	-	0.351*** (0.05)	0.238*** (0.06)	-	-	0.332*** (0.05)	0.246*** (0.06)
First week Screens	0.413*** (0.06)	0.413*** (0.06)	0.350*** (0.06)	0.370*** (0.06)	-	-	-	-
Total Screens	-	-	-	-	0.170*** (0.02)	0.168*** (0.02)	0.149*** (0.01)	0.154*** (0.02)
Sequel	0.244 (0.23)	0.284 (0.22)	0.189 (0.23)	0.233 (0.22)	0.456* (0.23)	0.495* (0.22)	0.370 (0.23)	0.415 (0.22)
Nation	0.242 (0.17)	0.216 (0.16)	0.101 (0.16)	0.129 (0.16)	0.295 (0.17)	0.271 (0.16)	0.128 (0.16)	0.156 (0.16)
Rating	0.359*** (0.92)	0.297** (0.09)	0.318*** (0.08)	0.291** (0.09)	0.322*** (0.08)	0.272** (0.08)	0.294*** (0.08)	0.272** (0.08)
Schedule	0.288 (0.16)	0.358* (0.15)	0.166 (0.15)	0.250* (0.15)	0.313* (0.16)	0.376* (0.16)	0.218 (0.15)	0.280* (0.15)
Star	1.375*** (0.16)	1.214*** (0.15)	1.126*** (0.15)	1.101*** (0.15)	1.474*** (0.15)	1.345*** (0.15)	1.236*** (0.15)	1.221*** (0.15)
Genre	控制	控制	控制	控制	控制	控制	控制	控制
_cons	2.705*** (0.75)	1.486* (0.68)	1.614* (0.81)	1.164 (0.66)	3.467*** (0.75)	2.385** (0.66)	2.334** (0.73)	1.984** (0.65)
N	264	264	264	264	264	264	264	264
R <sup>2</sup>	0.739	0.775	0.777	0.789	0.770	0.794	0.801	0.808
Adj.R <sup>2</sup>	0.73	0.76	0.77	0.78	0.76	0.78	0.79	0.80

注：括号内为稳健标准误，\*\*\*、\*\*、\*分别表示  $p < 0.001$ 、 $p < 0.01$ 、 $p < 0.05$ ，下同。

### 4.3 进一步的分析

#### 4.3.1 明星的调节作用

在电影产业中，一部电影中明星的出现会给制作方带来大量收益。之所以明星能起到这么大的作用，其中的一条原因是明星在一定程度上产生了大量的曝光。在自媒体时代，明星的一举一动（如，发电影相关或非相关微博）都会引起消费者的注意，从而使消费者仅仅因为明星的存在就做出观影决策<sup>[32]</sup>。因而，为了更进一步研究明星对电影票房的作用，本文猜想若电影中存在明星。消费者对于电影描述的内容关注就会减少，

从而使得文本可读性和吸引力对票房的影响削弱。为了验证上述猜想，本文将文本可读性和吸引力中心化后构建了分别以首周票房和总票房为因变量的模型 9-模型 14。从表 3 的结果来看，在模型 11 和模型 14 中，交互项的系数都显著为负。这表明，当电影没有明星参与时，简介文本的可读性和吸引力对电影票房影响力的权重减少了。此外，对比首周票房和总票房，明星的存在对于文本可读性影响力的削弱效果在首周较大，对于文本吸引力影响力的削弱效果在整个上映周期更大。



表 3 明星的调节效应

变量	模型 9	模型 10	模型 11	模型 12	模型 13	模型 14
	因变量: First week Box Office			因变量: Total Box Office		
Readability	0.518*** (0.07)	-	0.329*** (0.08)	0.459*** (0.07)	-	0.261** (0.08)
Attractiveness	-	0.449*** (0.06)	0.288*** (0.07)	-	0.427*** (0.05)	0.299*** (0.07)
Star	1.336*** (0.15)	1.256*** (0.15)	1.268*** (0.14)	1.443*** (0.15)	1.357*** (0.15)	1.369*** (0.15)
Readability ×Star	-0.449** (0.13)	-	-0.292* (0.14)	-0.422** (0.12)	-	-0.254 (0.13)
Attractiveness ×Star	-	-0.360*** (0.09)	-0.215* (0.10)	-	-0.340*** (0.10)	-0.220** (0.11)
Control Variables	控制	控制	控制	控制	控制	控制
_cons	3.335*** (0.73)	3.618*** (0.88)	3.690*** (0.70)	4.013*** (0.72)	4.282*** (0.79)	4.343*** (0.68)
N	264	264	264	264	264	264
R <sup>2</sup>	0.786	0.786	0.799	0.803	0.809	0.816
Adj.R <sup>2</sup>	0.77	0.77	0.79	0.79	0.80	0.80

#### 4.3.2 可读性和吸引性的交互分析

Chaiken 和 Maheswaran<sup>[8]</sup>提出某些情况下,系统式信息处理可能会稀释启发式信息处理的作用,当个体拥有更高的动机去处理信息时,个体更加依赖于系统式信息处理而非启发式信息处理,两种信息处理存在着衰减效应。故,本文推测,若简介吸引力启动了消费者的启发式处理,在认知上,文本的可读性可能会被消费者忽略;反之亦然。继而,为印证该假设,本文分析了文本可读性和文本吸引力对首周票房及总票房的交互影响作用。模型 15 和模型 16 以全样本进

行分析,结果显示可读性和吸引性的交互作用对首周票房以及总票房的作用不显著。进一步,本文以电影评分的中位数为切割点,将总样本分为了高/低评分电影两个样本。模型 17 和模型 18 的因变量为首周票房;模型 19 和模型 20 的因变量为总票房。从表 4 中可看出,只有在高电影评分组中,可读性和吸引性的交互作用对首周票房( $\beta=-0.143$ ,  $p<0.05$ )/总票房( $\beta=-0.122$ ,  $p<0.05$ )的影响显著。这说明,在高质量电影的样本中,文本的吸引力所产生的效果会被文本可读性削弱。

表 4 可读性和吸引性的交互分析

变量	模型 16	模型 17	模型 18	模型 19	模型 20
Readability	0.193*** (0.07)	0.161 (0.10)	0.390* (0.11)	0.145 (0.10)	0.292** (0.09)
Attractiveness	0.232*** (0.06)	0.192 (0.10)	0.275*** (0.07)	0.151 (0.09)	0.315*** (0.07)
Readability×Attractiveness	-0.031 (0.03)	-0.024 (0.05)	-0.143* (0.06)	-0.026 (0.04)	-0.122* (0.05)
Control Variables	控制	控制	控制	控制	控制
_cons	4.242*** (0.66)	2.030* (1.01)	-0.414 (2.68)	3.892*** (0.92)	-1.216 (2.94)
N	264	139	125	139	125
R <sup>2</sup>	0.808	0.791	0.781	0.810	0.812
Adj.R <sup>2</sup>	0.80	0.77	0.75	0.79	0.79

#### 4.4 稳健性检验

为确保结果的稳健,本文利用以下四种方法分别以首周票房和总票房作为因变量进行稳健性检验:第一,将观影人次作为因变量(模型 21/22);第二,利用 Gauss 核函数(RBF)

对可读性和吸引性重新建模(模型 23/24);第三,将出品公司数作为制作成本的代理变量进行控制,重新运行线性回归(模型 25/26);第四,以截尾负二项回归重新运行回归模型(模型 27/28)。从表 5 中可知,主要变量的结果与上文保持一致。因此,本文的研究具有良好的稳健性。

表 5 稳健性检验

变量	模型 21	模型 22	模型 23	模型 24	模型 25	模型 26	模型 27	模型 28
Readability	0.286*** (0.08)	0.229** (0.07)	0.177** (0.06)	0.135* (0.06)	0.247** (0.07)	0.195** (0.07)	0.178** (0.06)	0.169* (0.07)
Attractiveness	0.221** (0.07)	0.232*** (0.06)	0.204*** (0.06)	0.214*** (0.05)	0.231*** (0.06)	0.243*** (0.06)	0.196** (0.06)	0.232*** (0.06)
Control Variables	控制	控制	控制	控制	控制	控制	控制	控制
_cons	-2.420** (0.73)	-1.583* (0.68)	1.483* (0.66)	2.268*** (0.63)	1.046 (0.64)	1.921** (0.63)	3.38*** (0.70)	2.49*** (0.54)
N	264	264	264	264	264	264	264	264
R <sup>2</sup>	0.766	0.796	0.781	0.803	0.791	0.808	0.097	0.095
Adj.R <sup>2</sup>	0.75	0.78	0.77	0.79	0.78	0.80	-	-

注:奇数模型因变量为首周票房,偶数模型因变量为总票房;模型 27/28 的 R<sup>2</sup> 为 Pseudo R<sup>2</sup>

## 5 结语

基于启发式-系统式模型,本文以 2018 年上映的电影为样本,采用机器学习的方式度量了 2018 年上映电影的简介文本的可读性和吸引性水平,并检验其对票房的影响。

主要结论如下:第一,电影简介的可读性和吸引性越高,电影的首周票房和总票房越高,而且吸引性对票房的影响作用更大。第二,电影主创团队中存在明星将会削弱上可读性和吸引性对首周票房和总票房的影响。而且在高评分电影中,可读性和吸引性的交互效应对票房存在弱的负向作用。

本文的主要贡献在于：第一，通过文献梳理，并结合营销情景的实践总结出商品简介文本的一个新属性：吸引力。这无疑是对既往研究的扩展。同时有助于学者后续针对营销情景下的文本分析提供理论依据。第二，现有文献针对文本某个属性（如，可读性）的衡量大多是以文本的几个特征来进行表征。该方法割裂了文本之间各属性间的关系，也忽视了文本的属性依赖于阅读者的评判。本文借助机器学习的方法可以得到阅读群体的评判模型，在一定程度上避免了某些阅读者因个人原因对某些电影做出不是基于文本特征的判断，最大限度地降低了评判标准的主观性，为文本特征的测量提供了新思路。第三，以往的电影研究中，票房的驱动因素大多集中于电影上映后的消费者端，如口碑、影评等。随着中国电影产业的崛起，如何在电影的制作和发行阶段为未来票房提供保证是需要回答的问题。继而，本文的研究丰富了中国电影的理论研究。

本文对中国电影市场的营销实践有以下两点启示：第一，电影发行公司在撰写简介文本时需要同时将可读性和吸引力考虑。更进一步，电影发行方可以开发自己的机器学习模型，对自己在网络购票平台上简介文本的可读性和吸引力进行综合评判，以达到吸引消费者，提高电影票房的目的；第二，本文亦发现，电影中存在明星将会削弱文本属性对票房的影响。然而，需要注意的是，这并不代表明星的存在就使得文本对票房的影响不再显著，而是依然显著。因此，电影制片方也可考虑剧情和明星的动态关系。此外，考虑到在高评分电影组可读性和吸引力的交互项负向影响票房，这需要公司。

本文的局限在于：第一，本文仅以电影作为研究对象，探究商品的简介文本描述的可读性和吸引力对销量的影响。事实上，其他实物商品也提供了商品描述信息，未来若能获取这些商品的销量数据，可以在上述对象上进行研究。第二，本文利用机器学习的方法提炼出文本的可读性和吸引力水平，从建模结果来看，上述两个指标对首周票房和总票房有着差异化影响，但相应的影响路径尚未详细研究，有待未来研究进一步挖掘。

第三，本文的研究基于历史数据展开，未来学者可考虑采用不同水平的可读性和吸引力简介文本来进行田野实验，以进一步验证本文结果的有效性。

#### 参考文献：

- [1] Ghose A, Ipeirotis P G, Li B. Designing ranking systems for hotels on travel search engines by mining user-generated and crowd-sourced content[J]. *Marketing Science*, 2012, 31(3):493-520.
- [2] Burtch, G, Ghose A, Wattal S. An empirical examination of the antecedents and consequences of contribution patterns in crowd-funded markets[J]. *Information Systems Research*, 2013, 24(3), 499-519.
- [3] 王伟, 陈伟, 祝效国, 等. 众筹融资成功率与语言风格的说服力——基于 Kick-starter 的实证研究[J]. *管理世界*, 2016(5):81-98.
- [4] Yin D, Bond S D, Zhang H. Anxious or angry? effects of discrete emotions on the perceived helpfulness of online reviews[J]. *MIS Quarterly*, 2014, 38(2): 539-560.
- [5] Moon S, Park Y, Kim Y S. The impact of text product reviews on sales[J]. *European Journal of Marketing*, 2014, 48(11/12):2176-2197.
- [6] 郝晓玲, 陈晓梦. 体验型产品消费行为的羊群效应及机理研究——基于电影行业消费行为的实证解释[J]. *中国管理科学*, 2019, 27(11):176-188.
- [7] 罗伯特·麦基. 故事：材质·结构·风格和银幕剧作的原理[M]. 天津人民出版社, 2014.
- [8] Chaiken S, Maheswaran D. Heuristic processing can bias systematic processing: effects of source credibility, argument ambiguity, and task importance on attitude judgment[J]. *Journal of Personality & Social Psychology*, 1994,



- 66(3):460-473.
- [9] Qahri-Saremi H, Montazemi A R. Factors affecting the adoption of an electronic word of mouth message: a meta-analysis[J]. *Journal of Management Information Systems*, 2019, 36(3): 969-1001.
- [10] Guo J, Wang X, Wu Y. Positive emotion bias: Role of emotional content from online customer reviews in purchase decisions[J]. *Journal of Retailing and Consumer Services*, 2020, 52(1): 20-31.
- [11] Chaiken S. Heuristic versus systematic information processing and the use of source versus message cues in persuasion [J]. *Journal of Personality and Social Psychology*, 1980, 39(5):752-766.
- [12] Sherman J W. On building a better process model: it's not only how many, but which ones and by which means?[J]. *Psychological Inquiry*, 2006, 17(3): 173-184.
- [13] Shah A K, Oppenheimer D M. Heuristics made easy: An effort-reduction framework[J]. *Psychological Bulletin*, 2008, 134(2):207-222.
- [14] 王克敏, 王华杰, 李栋栋, 等. 年报文本信息复杂性与管理者自利——来自中国上市公司的证据[J]. *管理世界*, 2018 (12) :120-132.
- [15] Sawyer A G, Laran J, Xu J. The readability of marketing journals: are award-winning articles better written[J]. *Journal of Marketing*, 2008, 72(1): 108-117.
- [16] 陈霄, 叶德珠, 邓洁. 借款描述的可读性能够提高网络借款成功率吗?[J]. *中国工业经济*, 2018 (3) :174-192.
- [17] Smith E R, Decoster J. Dual-process models in social and cognitive psychology: conceptual integration and links to underlying memory systems[J]. *Personality & Social Psychology Review*, 2000, 4(2):108-131.
- [18] Alter A L, Oppenheimer D M. Suppressing secrecy through meta-cognitive ease: cognitive fluency encourages self-disclosure[J]. *Psychological Science*, 2010, 20(11):1414-1420.
- [19] Tausczik Y R, Pennebaker J W. The psychological meaning of words: LIWC and computerized text analysis methods[J]. *Journal of Language and Social Psychology*, 2010, 29(1):24-54.
- [20] Agnihotri A, Bhattacharya S. Online review helpfulness: role of qualitative factors[J]. *Psychology & Marketing*, 2016, 33(11):1006-1017.
- [21] 刘佳平. 转折连词“但是”与“不过”的比较[J]. *语文学刊*, 2008(9):47-50.
- [22] Herzenstein M, Sonenshein S, Dholakia U M. Tell me a good story and I may lend you money: the role of narratives in Peer-to-Peer lending decisions[J]. *Journal of Marketing Research*, 2011, 48(SI): S138-S149.
- [23] Xiang D, Zhang L, Tao Q, et al.,. Informational or emotional appeals in crowd-funding message strategy: an empirical investigation of backers' support decisions[J]. *Journal of the Academy of Marketing Science*, 2019, 47(6): 1046-1063.
- [24] Brand R J, Bonatsos A, D'Orazio R, et al. What is beautiful is good, even online: Correlations between photo attractiveness and text attractiveness in men's online dating profiles[J]. *Computers in Human Behavior*, 2012, 28(1):166-170.
- [25] Huang Y, Lo Y. What makes blogging attractive to bloggers: a case of college-level constituency users[J]. *Journal of Computer Assisted Learning*, 2012, 28(3): 208-221.
- [26] Hudson S, Huang L, Roth M S, et al. The influence of social media interactions on consumer-brand relationships: a three-country study of brand perceptions and

- marketing behaviors[J]. International Journal of Research in Marketing, 2016, 33(1):27-41.
- [27] He H, Li Y, Harris L. Social identity perspective on brand loyalty[J]. Journal of Business Research, 2012, 65(5): 648-657.
- [28] Zhao E Y , Ishihara M, Lounsbury M. Overcoming the illegitimacy discount: cultural entrepreneurship in the US feature film industry[J]. Organization Studies, 2013, 34(12):1747-1776.
- [29] 杨学成, 兰冰, 孙飞. 品牌微博如何吸引粉丝互动——基于 CMC 理论的实证研究 [J]. 管理评论, 2015, 27(1): 158-168.
- [30] Sun A, Lim E P, Liu Y. On strategies for imbalanced text classification using SVM: a comparative study[J]. Decision Support Systems, 2009, 48(1): 191-201.
- [31] Kübler R V, Colicev A, Pauwels K H. Social media's impact on the consumer mindset: when to use which sentiment extraction tool?[J]. Journal of Interactive Marketing, 2020, 50(5):136-155.
- [32] Kupfer A K , Holte N P, Kübler R V, et al. The role of the partner brand's social media power in brand alliances[J]. Journal of Marketing, 2018, 82(5):25-44.

## The impact of readability and attractiveness on product sales——Text analysis based on movie introduction

**Gao Wei-he<sup>1</sup>, Liu De-wen<sup>1</sup>, Min Liang-yu<sup>2</sup>**

(1. College of Business, Shanghai University of Finance and Economics, Shanghai 200433, China; 2. School of Information Management & Engineering, Shanghai University of Finance and Economics, Shanghai 200433, China)

**Abstract:** Product introduction plays an important role in consumer purchase decisions. It is of practical significance to understand how to design and craft the text introduction so as to study consumer purchase decision making behavior of such products. However, previous marketing text analysis researches has mainly focused on consumer-side variables (e.g., online reviews), and didn't pay enough attention on producer-side variables (e.g., product introduction). Also, the outcome variables in previous text analysis studies included only non-purchase variables such as shares, comments and likes. Due to the difficulty of matching product sales at an aggregate level, few studies have used actual secondary data to quantify the impact of text-based outcomes.

In order to overcome the mentioned research gap, based on relevant marketing practices and heuristic-systemic model, two characteristics of movie introduction (readability and attractiveness) were proposed. Then, used a machine learning SVM model, the readability and attractiveness level of the introduction of 264 films released in mainland China in 2018 were computed and its impact on the box office in short term and long term were tested .

It is found that: movie introduction's readability and attractiveness had a different impact on first week's box office and total box office. Readability had a greater impact on first week's box office, and attractiveness had a greater impact on total box office. Beside, the presence of stars

undermined the impact of readability and attractiveness on the box office, and readability and attractiveness have significant negative interactions effect on first week's box office and total box office in the high-rating movie group. The research conclusion has great practical significance for the development of the Chinese movie industry and reasonable guidance for movie managers in promoting movie sales.

**Key Words:** movie introduction; readability; attractiveness; machine learning

