

doi:10.19920/j.cnki.jmsc.2025.09.004

# 融合文本和图像的个性化需求预测方法<sup>①</sup>

## ——基于有限偏好视角

姜元春<sup>1,2</sup>, 李 怡<sup>1,2</sup>, 钱 洋<sup>1,3\*</sup>, 柴一栋<sup>1,3</sup>, 刘业政<sup>1,3</sup>

(1. 合肥工业大学管理学院, 合肥 230009; 2. 网络空间行为与管理安徽省哲学社会科学重点实验室, 合肥 230009; 3. 过程优化与智能决策教育部重点实验室, 合肥 230009)

**摘要:** 随着电子商务应用由广度到深度的转变, 个性化成为电子商务服务模式创新的重要方向。为了准确预测消费者的个性化需求, 本研究融合产品描述文本和展示图像信息, 提出一种有限偏好约束的个性化需求预测方法。基于有限注意力假设, 该方法对消费者有限偏好进行建模, 将图像特征和文本特征相融合, 构建有限偏好约束下的融合文本和图像的稀疏图文链接主题模型。该模型通过群体兴趣建模、个体偏好建模和购买决策建模等过程, 对消费者的个性化需求进行预测。基于亚马逊公开数据集的实验表明, 所提模型可以有效预测消费者的群体兴趣和个性化偏好, 文本和图像信息的融合提高了个性化需求预测的可解释性。

**关键词:** 个性化; 需求预测; 有限偏好; 稀疏图文链接主题模型

中图分类号: C93 文献标识码: A 文章编号: 1007-9807(2025)09-0052-13

## 0 引 言

随着消费者需求的差异化趋势越来越明显, 如何面向网络消费者个性化需求创新服务模式, 已经成为电子商务理论与实践创新的重要方向。准确预测消费者的个性化需求是个性化服务有效实施的前提和关键所在。电子商务环境下, 由于无法进行现场体验, 描述文本和展示图像成为消费者了解产品的主要方式。这些文本和图像在影响消费者购买决策的同时, 也蕴含着丰富的个性化偏好信息。如何融合产品描述文本和展示图像构建个性化需求预测方法是电子商务领域研究的重要问题。

为了更好地说明融合图像和文本信息预测个性化偏好的必要性, 图1给出了两位消费者各自购买的4件商品及其相关信息。如图1所示, 如果仅利用产品的ID信息进行个性化偏好预测, 由于

两位消费者购买记录中无相同产品, 将会得到其偏好完全不同的结论。如果考虑产品的展示图像, 可以看出两位消费者在服装款式上具有相似的偏好, 即都喜欢纯色且无袖的连衣裙。如果再考虑产品的标题文本, 可以发现其偏好仍有所差异, 即消费者1喜欢真丝材质, 而消费者2偏好棉麻材质。图1实例表明, 融合文本和图像信息有助于细粒度地刻画消费者偏好, 提高个性化需求预测的准确性和可解释性。

文本和图像信息在消费者行为研究中已经得到了广泛应用。在基于图像数据的消费者行为研究方面, 研究者将图像转换成亮度、饱和度等低维度向量, 或通过图像内容识别等方法将图像转换为文本描述, 在此基础上, 研究图像信息对于品牌形象的影响<sup>[1]</sup>, 考虑缩略图特征的视频内容营销<sup>[2]</sup>以及人脸特征等图像信息对广告效果的影响<sup>[3]</sup>等问题。另一研究思路是利用表征学习方法

① 收稿日期: 2021-04-09; 修订日期: 2023-02-07。

基金项目: 国家自然科学基金资助项目(72171071; 72342011; 72101072; 72271084); 中央高校基本科研业务费专项资金资助项目(JZ2024HGTG0316); 安徽省重大科技攻关专项资助项目(2023zd020006)。

通讯作者: 钱 洋(1995—), 男, 安徽蚌埠人, 博士, 副教授。Email: soberqian@hfut.edu.cn

将图像转换成高维度向量,研究基于图像表征学习的高影响力图片识别<sup>[4]</sup>、融合视觉信息的推荐系统冷启动缓解<sup>[5]</sup>、基于图像信息识别替代品和互补品实现服装搭配推荐<sup>[6]</sup>等问题.在基于文本数据的消费者行为研究方面,相关研究聚焦于探讨电影名称对于观众观影意愿的影响<sup>[7]</sup>,基于主题模型研究网民关注点的变化<sup>[8]</sup>,利用社交媒体发布内容探究在线用户评论行为产生机理<sup>[9]</sup>,结

合用户评论利用混合成员模型解释消费者偏好的异质性<sup>[10]</sup>,研究用户生成内容对视频平台的定价决策和投资策略的影响<sup>[11]</sup>等问题.融合文本、图像等多模态数据进行需求预测和产品推荐是近年来个性化推荐领域的主要趋势.研究者综合利用产品描述、展示图像、在线评论等数据对消费者和产品进行表征<sup>[12, 13]</sup>,进而构建基于矩阵分解、深度学习等方法的个性化推荐模型<sup>[14, 15]</sup>.

消费者1				消费者2			
D9002711	5531	D9002926A	73086	3035#D233A	0820d183	17345	8733
①	②	③	④	①	②	③	④
1. 夏装新款 真丝连衣裙圆领无袖桑蚕丝修身裙子女装 2. 夏装新款 无袖真丝连衣裙欧美品牌女装无袖连衣裙 3. 名媛气质夏季新款无袖系带宽松显瘦纯色A字真丝连衣裙女 4. 夏季大码女装打底背心A字裙无袖宽松中长款真丝连衣裙女				1. 夏季新款大码女装连衣裙韩版宽松显瘦棉麻无袖连衣裙女 2. 新款夏季无袖棉麻连衣裙女森系文艺范双层打底背心裙 3. 夏季新款外贸女装小清新文艺系带收腰纯色棉麻无袖连衣裙背心裙 4. 欧洲站夏季新款欧装无袖棉麻连衣裙女收腰打底裙裙子			

图 1 消费者购买记录举例

Fig. 1 Examples of consumer purchase records

综上所述,文本和图像数据在消费者行为分析和个性化产品推荐中得到了广泛应用.基于矩阵分解、深度学习等方法的个性化推荐研究成果可以为本研究提供有效的理论借鉴,但是,如何融合多模态信息构建可解释性强的需求预测方法仍是目前理论探索的难点.心理学和营销科学研究表明,消费者注意力通常是有限的.消费者的购买决策不会考虑市场中的所有产品,而是考虑极少部分感兴趣的产品<sup>[16]</sup>.受时间成本和认知能力的限制,消费者通常只关注产品少量的突出属性,在此基础上做出购买决策<sup>[17, 18]</sup>,即消费者在产品选择和属性关注上具有有限偏好特点.现有研究通常将消费者兴趣域分布于全部产品和全部属性,如何基于消费者有限偏好构建更加有效的需求预测与产品推荐模型,缺乏深入探讨.

融合文本和图像数据,本研究提出一种有限偏好约束的个性化需求预测方法.与现有方法假设消费者对所有产品均会存在潜在兴趣不同,本

研究基于有限偏好假设,利用 Spike-and-Slab 先验<sup>②</sup>对消费者有限偏好进行建模<sup>[19, 20]</sup>,将图像特征和文本特征相融合,构建有限偏好约束下的稀疏图文链接主题模型,对消费者个性化需求进行预测.本研究基于亚马逊公开数据集<sup>[6]</sup>进行实验以验证模型的有效性,该实验数据集包含 4 913 个消费者针对 11 414 件产品的购买记录.研究结果表明,所提模型可以有效预测消费者的群体兴趣和个性化偏好,文本和图像信息的综合利用有助于提高个性化需求预测的可解释性.

## 1 面向个性化需求预测的稀疏图文链接主题模型

面向个性化需求预测的稀疏图文链接主题模型(Sparse Text-Image Linked Topic Model, STILT)如图 2 所示.假设消费者集合为  $U = \{1, 2, \dots,$

② Spike-and-Slab 先验是一种用于贝叶斯框架下的稀疏信号处理模型,通过区分显著特征(Spike)和噪声(Slab)实现参数的稀疏化.

$|U|$ }, 其中,  $|U|$  为消费者数量. 每个消费者购买的产品有其文本描述和图像展示, 文本描述由一系列单词构成. 假设消费者  $u$  购买的所有产品对应  $N_u$  个单词及  $M_u$  张产品图像, 产品词汇集合表示为  $W_u \in \{W_{un}\}_{n=1}^{N_u}$ , 产品图像特征集合表示为  $V_u \in \{V_{um}\}_{m=1}^{M_u}$ .  $W_{un}$  为消费者  $u$  购买记录中的第  $n$  个词汇,  $V_{um}$  表示消费者  $u$  购买记录中的第  $m$  张图片的图像特征. 本研究用  $E = \{1, 2, \dots, e, \dots, |E|\}$  表示购买记录中所有不重复的词汇,  $L = \{1, 2, \dots, l, \dots, |L|\}$  表示购买记录中所有不重复的图像.

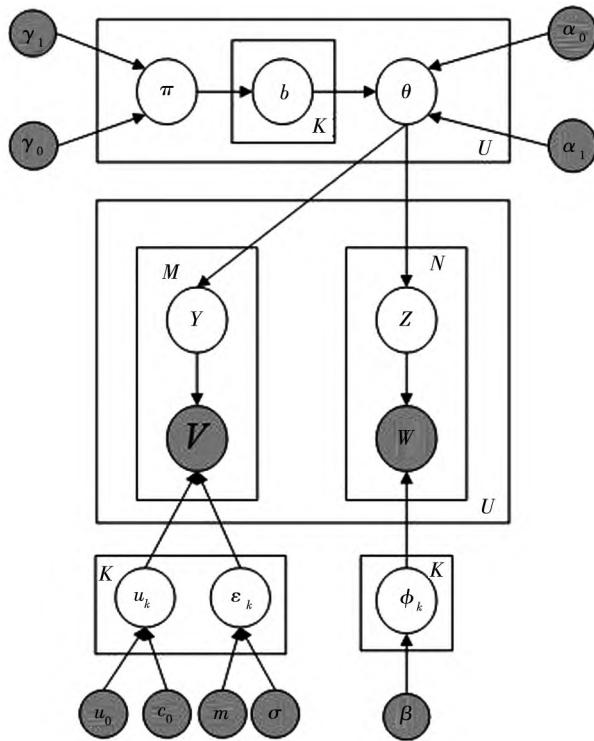


图 2 STILT 图模型

Fig. 2 The graphical model of STILT

STILT 模型分别利用 vMF 分布和狄利克雷分

布建模消费者的图像特征选择偏好  $Y$  和文本内容选择偏好  $Z$ , 基于 Spike-and-Slab 先验构造二进制选择变量  $b$  以约束消费者的偏好分布  $\theta$ , 探索偏好分布的稀疏结构, 在此基础上对消费者的个性化需求进行建模和预测. 下面给出 STILT 建模及模型参数的详细推断过程.

## 1.1 相关定义

### 定义 1 群体兴趣

群体兴趣是指市场上由某种关联关系聚合而成的产品集合,例如,具有相似特征或相似购买模式的产品集合.反映在产品文本和图像上,群体兴趣  $k \in \{1, 2, \dots, K\}$  由一类图像特征  $(u_k, \varepsilon_k)$  和文本主题  $\phi_k$  共同组成.其中,图像特征由 vMF 分布<sup>[20]</sup>定义,文本内容由多项式分布定义.例如,真丝无袖连衣裙、棉麻军绿迷彩服分别代表市场上的两类群体兴趣.

### 定义 2 个体偏好分布

个体偏好分布度量的是每个消费者对不同群体兴趣的喜好程度. 例如, 某消费者偏爱真丝无袖连衣裙和修身长款毛衣, 而另一消费者则偏爱棉麻军绿迷彩服等. 消费者  $u$  的偏好分布  $\theta_u$  由多项式分布组成, 在有限注意力的假设下, 本研究认为  $\theta_u$  只在少数兴趣上有分布. 其中,  $\theta_{uk}$  表示消费者  $u$  对于兴趣  $k$  偏好的程度.

### 定义 3 偏好选择器

偏好选择器  $b$  定义为二进制变量, 用于确定消费者是否偏好特定兴趣. 当  $b_{uk} = 1$  时, 表示消费者  $u$  对兴趣  $k$  有偏好, 否则, 无偏好. 在  $b_{uk}$  选择结果的基础上, 可以分别采样影响消费者图像特征选择的偏好  $Y_{um}$  和文本内容选择的偏好  $Z_{un}$ , 进而生成消费者购买记录所对应的图像特征  $V_{um}$  与词汇  $W_{un}$ .

本研究模型所用的其余变量解释如下：

表1 变量解释

Table 1. Variable illustration

变量	含义	变量	含义
$\theta_u$	消费者 $u$ 的偏好分布	$(u_0, c_0)$	先验 vMF 分布的超参数
$\phi_k$	兴趣 $k$ 的兴趣-词分布	$(m, \sigma)$	先验对数正态分布的超参数
$\beta$	兴趣-词汇分布的狄利克雷先验超参数	$(u_k, \varepsilon_k)$	兴趣 $k$ 的 vMF 分布参数
$\pi$	$b = 1$ 的期望比例	$(\alpha_0, \alpha_1)$	$\theta_u$ 狄利克雷分布的超参数
b	偏好选择器	$(\gamma_0, \gamma_1)$	贝塔分布的超参数

## 1.2 STILT 模型生成过程

由图2可以看出,STILT模型本质上是面向个性化需求预测的生成模型,其逻辑是市场上存

在着多样的群体兴趣,个体消费者  $u$  仅对部分兴趣产生偏好  $\theta_u$ , 消费者偏好决定了其喜欢的产品特征(如纯色无袖款式和真丝棉麻材质等),而消

费者偏好和产品特征共同决定了消费者的购买决策. STILT 模型将消费者购买记录及其对应的文本和图片作为输入, 模型输出包括反映市场兴趣的兴趣-图像特征分布、兴趣-词汇分布, 以及消费

者个体的消费者-偏好分布. 算法 1 给出了图 2 所示模型的具体生成过程, 主要由消费者群体兴趣建模、消费者个体偏好建模、消费者购买决策建模三部分组成.

算法 1: STILT 生成过程

1) 针对每个群体兴趣 $k \in \{1, 2, \dots, K\}$	3) 针对每个消费者 $u \in \{1, 2, \dots, U\}$
① 抽取文本兴趣: $\boldsymbol{\phi}_k \sim Dir(\beta)$	① 针对每个词汇
② 抽取图像兴趣	A. 抽取生成该词汇的兴趣 $Z_{un} \sim Mult(\boldsymbol{\theta}_u)$
$\mu_k \sim vMF(u_0, c_0)$ $\varepsilon_k \sim logNormal(m, \sigma)$	B. 抽取该词汇: $W_{un} \sim Mult(\boldsymbol{\phi}_k)$
2) 针对每个消费者 $u \in \{1, 2, \dots, U\}$	② 对每个产品的展示图像
① 抽取 $\pi_u \sim Beta(\gamma_0, \gamma_1)$	A. 抽取生成该图像的兴趣 $Y_{um} \sim Mult(\boldsymbol{\theta}_u)$
② 针对每个兴趣	B. 抽取该图像 $V_{um} \sim vMF(\boldsymbol{\mu}_k, \varepsilon_k)$
A. 抽取偏好选择器: $b_{uk} \sim Bernoulli(\pi_u)$	
B. 抽取每个消费者的偏好分布	
$\boldsymbol{\theta}_u \sim Dir(\alpha_0 b_u + \alpha_1 \mathbf{1})$	

### 1.2.1 过程一 群体兴趣模型

STILT 模型中, 群体兴趣由产品的文本内容和图像特征表征组成, 由兴趣-词汇分布和兴趣-图像特征分布进行测度, 本研究利用如下方法对兴趣  $k \in \{1, 2, \dots, K\}$  进行建模.

假设兴趣-词汇分布  $\boldsymbol{\phi}_k$  服从参数为  $\beta$  的狄利克雷先验,  $\boldsymbol{\phi}_k$  为一个  $E$  维向量, 表达兴趣  $k$  在产品描述词汇上的概率.

$$\boldsymbol{\phi}_k \sim Dir(\beta) \quad (1)$$

兴趣-图像特征分布  $(\mu_k, \varepsilon_k)$  服从参数为  $(u_0, c_0)$  的 vMF 先验和参数为  $(m, \sigma)$  的对数正态先验,  $(\mu_k, \varepsilon_k)$  表达了兴趣  $k$  在每个图像特征上的概率.

$$\mu_k \sim vMF(u_0, c_0), \varepsilon_k \sim logNormal(m, \sigma) \quad (2)$$

其中  $\beta$  为狄利克雷分布的超参数,  $(u_0, c_0)$  为 vMF 分布的超参数,  $(m, \sigma)$  为对数正态分布的超参数. vMF 分布广泛应用于向量建模, 具有推理效率高、适用于图像分析等特点<sup>[21, 22]</sup>.

由图 2 可以看出, 兴趣-词汇分布  $\boldsymbol{\phi}_k$  和兴趣-图像特征分布  $(\mu_k, \varepsilon_k)$  共同决定了消费者购买产品的文本描述和图像内容, 即决定了消费者的购买决策. 通过式(1)和式(2), 本研究可以综合不同消费者的购买记录, 得到当前市场上的群体兴趣及不同兴趣对应的具体内容. 对于消费者个体而言, 受有限注意力的影响, 一个消费者往往只会偏爱少数兴趣. 下面对消费者个体的有限偏好分布进行建模分析.

### 1.2.2 过程二 个体偏好模型

为了预测消费者偏好, 现有研究通常假设消费者偏好分布于市场中的所有产品, 使得所构建的偏好分布具有极大稀疏性, 不利于提升需求预测的精度. 为此, 本研究使用 Spike-and-Slab 先验方法对消费者的有限偏好进行建模. Spike-and-Slab 先验是贝叶斯特征选择的重要方法, 该方法基于 Spike-and-Slab 的混合先验分布, 能够从概率角度解决变量选择和稀疏恢复等难题. 在消费者个体的有限偏好建模中, 设置面向群体兴趣的二进制选择变量, 并借助 Spike-and-Slab 混合先验分布计算不同兴趣被选中的概率, 对消费者个体的有限兴趣进行识别, 探索消费者个体偏好分布的潜在稀疏结构, 从而克服将消费者偏好分布于全部产品对需求预测的负面影响.

假设  $b_{uk}$  是决定消费者  $u$  是否对兴趣  $k$  有偏好的选择器.

$$b_{uk} \sim Bernoulli(\pi_u), b_{uk} \in \{0, 1\} \quad (3)$$

其中  $\pi_u$  为超参数,  $\pi_u \sim Beta(\gamma_0, \gamma_1)$ ,  $\gamma_0, \gamma_1$  均为 Beta 分布的超参数.

对消费者  $u$  来说, 当  $b_{uk} = 1$  时, 说明消费者关注兴趣  $k$ , 否则, 不关注. 基于偏好选择器  $b_{uk}$ , 本研究定义消费者  $u$  感兴趣的偏好集合为  $A_u = \{k: b_{uk} = 1, k \in \{1, 2, \dots, K\}\}$ . 消费者感兴趣的偏好分布定义为<sup>③</sup>

$$\boldsymbol{\theta}_u \sim Dir(\alpha_0 b_u + \alpha_1 \mathbf{1}) \quad (4)$$

③ 由于版面限制, 详细推理过程可联系作者备索.

其中  $\mathbf{b}_u = \{b_{uk}\}_{k=1}^K$ , 参数  $\alpha_0$  是平滑先验, 用于控制被选择器选中的兴趣; 参数  $\alpha_1 \ll \alpha_0$ , 数值接近于 0, 是弱平滑先验, 用于控制没有被选择器选中的兴趣. 平滑先验  $\alpha_0$  和弱平滑先验  $\alpha_1$ , 保证了消费者对有限兴趣的关注.

### 1.2.3 过程三 购买决策模型

消费者购买决策是一个复杂的过程, 通过消费者个体的偏好分布建模, 本研究得到消费者偏好的有限兴趣, 结合消费者的购买记录, 下面对消费者购买决策过程进行建模.

对消费者  $u \in \{1, 2, \dots, |U|\}$ , 首先从参数为  $\boldsymbol{\theta}_u$  的多项式分布中选择消费者对文本内容的偏好  $Z_{un}$  和图像特征的偏好  $Y_{un}$ .

$$Z_{un} \sim Mult(\boldsymbol{\theta}_u), Y_{un} \sim Mult(\boldsymbol{\theta}_u) \quad (5)$$

其次, 分别从参数为  $\boldsymbol{\varphi}_k$  的多项式分布, 参数为  $(u_k, \varepsilon_k)$  的 vMF 分布中选择符合偏好  $Z_{un}$  和  $Y_{un}$  的词汇以及图像特征.

$$W_{un} \sim Mult(\boldsymbol{\varphi}_{Z_{un}}), V_{un} \sim vMF(\mu_{Y_{un}}, \varepsilon_{Y_{un}}) \quad (6)$$

由于本研究对反映产品特征的文本和图像进行联合建模, 只有当产品的文本内容和图像内容同时符合消费者的偏好内容时, 才会预测消费者对该产品感兴趣.

## 1.3 模型推导

为了得到消费者群体兴趣和消费者个体在群体兴趣上的偏好分布, 本研究模型需要推断的参数包括  $b_{uk}, Z_{un}, Y_{un}, \boldsymbol{\theta}_{uk}, \boldsymbol{\varphi}_{ke}$  和  $\varepsilon_k$ . 本研究采用吉布斯采样的方法对模型进行近似后验推理<sup>④</sup>, 首先在确定已知变量的情况下, 对潜在变量  $b_{uk}, Z_{un}$  和  $Y_{un}$  进行交替采样, 在这些经验分布收敛后, 再基于相关分布估计参数  $\boldsymbol{\theta}_{uk}, \boldsymbol{\varphi}_{ke}$  和  $\boldsymbol{\varphi}_{kl}$ . 具体参数估计如下.

### 1.3.1 推断偏好选择器

为了采样偏好选择器, 本研究通过式(7)计算  $b_{uk}$  和  $\pi_u$  的联合概率分布.

$$p(\pi_u, \mathbf{b}_u | *) \propto \prod_k p(b_{uk} | \pi_u) p(\pi_u | \gamma_0, \gamma_1) \times \frac{I[B_u \in A_u] \Gamma(|A_u| \alpha_0 + K \alpha_1)}{\Gamma(n_u + |A_u| \alpha_0 + K \alpha_1)} \quad (7)$$

其中  $I[*]$  为指标函数,  $A_u$  定义为所有  $b_{uk} = 1$  的兴趣集合,  $|A_u|$  为  $A_u$  中的兴趣数量,  $B_u = \{k : n_u^k > 0,$

$k \in \{1, 2, \dots, K\}\}$  为用户  $u$  偏爱的兴趣集合,  $n_u^k$  表示消费者  $u$  偏爱的兴趣  $k$  中的图像和词汇的数量之和.

结合上述联合概率分布, 对参数  $\pi_u$  进行积分, 得到  $b_{uk}$  的迭代采样公式如下.

$$p(b_{u,k} = 0 | *) \propto (N_{u,0}^{UB} + \gamma_0) \frac{I[B_u \in A_u] \Gamma(|A_u| \alpha_0 + K \alpha_1)}{\Gamma(N_u + |A_u| \alpha_0 + K \alpha_1)} \quad (8)$$

$$p(b_{u,k} = 1 | *) \propto (N_{u,1}^{UB} + \gamma_1) \frac{I[B_u \in A_u] \Gamma(|A_u| \alpha_0 + K \alpha_1)}{\Gamma(N_u + |A_u| \alpha_0 + K \alpha_1)} \quad (9)$$

其中  $N_{u,b}^{UB}$  表示偏好选择器  $b$  作用于消费者  $u$  的次数.

### 1.3.2 推断图像特征与词汇对应的兴趣

针对产品的图像特征, 本研究将 vMF 分布作为共轭先验分布, 通过式(10)、式(11)采样图像特征对应的兴趣. 采样公式如下.

$$P(Y_{un} = k \mid V_{un} = l, \mathbf{V}_{\neg un}, \mathbf{Y}_{\neg un}, \mathbf{Z}, \alpha_0, \alpha_1, \varepsilon_k, u_0, c_0, m, \sigma) \propto C_D \left( \left\| \varepsilon_k \sum_{j \neq i, Y_j=k} x_j + c_0 u_0 \right\| \right) \times C_D(\varepsilon_k) \times \frac{\left\| \varepsilon_k \sum_{j: Y_j=k} x_j + c_0 u_0 \right\|}{\left\| \varepsilon_k \sum_{j: Y_j \neq k} x_j + c_0 u_0 \right\|} \times (n_{u, \neg(un)}^k + b_{uk} \alpha_0 + \alpha_1) \quad (10)$$

$$P(\varepsilon_k \mid \varepsilon_{-k}, \dots) \propto \frac{C_D(c_0) C_D(\varepsilon_k)^{n_k}}{C_D \left( \left\| \varepsilon_k \sum_{j: Y_j=k} x_j + c_0 u_0 \right\| \right)} \times \log Normal(\varepsilon_k \mid m, \sigma) \quad (11)$$

其中  $\mathbf{Y}_{\neg un}$  表示除图像特征  $V_{un}$  对应兴趣以外的其他图像对应的兴趣,  $V_{\neg un}$  表示除了图像特征  $V_{un}$  以外的所有图像特征,  $n_k$  表示兴趣  $k$  中出现的所有图像特征的数量.

为了对产品描述文本的词汇对应的兴趣进行推断, 本研究通过式(12)采样词汇兴趣.

$$P(Z_{un} = k \mid W_{un} = e, \mathbf{W}_{\neg un}, \mathbf{Z}_{\neg un}, \mathbf{Y}, \alpha_0, \alpha_1, \beta) \propto$$

$$(n_{u, \neg(un)}^k + b_{uk} \alpha_1 + \alpha_1) \times \frac{n_{ke} + \beta}{n_k + E\beta} \quad (12)$$

其中  $\mathbf{W}_{\neg un}$  表示除词汇  $W_{un}$  外的所有词汇,  $\mathbf{Z}_{\neg un}$  表示除词汇  $W_{un}$  对应的兴趣以外的所有兴趣,  $n_u^k$  表示用户  $u$  的购买记录中对应兴趣  $k$  的图像特征或词汇出现的次数,  $n_{ke}$  表示分配到兴趣  $k$  中的词汇

<sup>④</sup> 由于版面限制, 兴趣数量确定过程可联系作者.

$e$  的次数,  $n_k$  表示分配到兴趣  $k$  中的所有词汇数.

### 1.3.4 群体兴趣和个体偏好参数估计

基于式(7)~式(12), 得到消费者群体的文本兴趣  $\phi_k$  以及消费者个体的偏好分布  $\theta_u$ ; 采用经验估计的方法, 计算消费者群体的图像特征兴趣  $\phi_{kl}$ , 具体计算公式如下.

$$\theta_{uk} = \frac{n_{uk} + b_{uk} \alpha_0 + \alpha_1}{N_u + |A_m| \alpha_0 + k \alpha_1} \quad (13)$$

$$\varphi_{ke} = \frac{n_{ke} + \beta}{n_k + E\beta} \quad (14)$$

$$\varphi_{kl} = \frac{n_{kl} + \beta}{n_k + L\beta} \quad (15)$$

## 2 实验分析

为了验证所提模型的有效性, 本研究利用 2018 年亚马逊公开数据集中消费者对服装品类产品的购买记录进行实验. 本研究选择被购买次数大于 10 的产品以及购买次数大于 5 的消费者, 得到 4 913 个消费者针对 11 414 件产品的购买记录. 实验将每个消费者所购买产品的标题文本进行汇总、分词和去停用词等操作, 得到反映消费者偏好的文本数据. 将每个产品在搜索页面对应的展示图像作为产品的图像数据, 利用深度学习方

法 ImageNet 提取产品图片的 2 048 维预训练向量, 汇总每个消费者所购买产品的展示图像特征向量, 得到反映消费者偏好的图像数据. 其中, 文本数据由 11 601 个不重复的词组成, 图像数据由 11 414 个不重复的图像特征组成. 在主题模型中, 确定最优兴趣数量  $K$  是发现群体兴趣的前提. 本研究实验利用困惑度方法得到最优兴趣数量  $K$  为 60.

### 2.1 群体兴趣实验

基于所提模型及其推断结果, 本研究可以得到市场上消费者群体的兴趣. 图 3(a)描绘了群体兴趣及其组成产品的关系网络. 其中, 绿色节点表示群体兴趣, 白色节点表示产品. 从图 3(a)可以看出, 本研究方法得到的群体兴趣往往包含多个产品, 部分产品被划分到不同的兴趣. 与本研究式(10)~式(12)相一致, 在产品图像和词汇采样过程中, 消费者偏好起重要作用. 当存在多个产品的图像和词汇符合消费者的相同偏好时, 该产品图像和产品词汇组成了市场中的某一群体兴趣. 图 3(b)给出了每个兴趣所包含的产品数量情况. 可以看出, 不同兴趣所包含的产品数量存在较大差异. 例如, 基于本研究实验数据得到的兴趣中, 有 8 个兴趣包含 70 款以上的产品, 而有 6 个兴趣包含的产品不到 30 款. 图 3(c)给出了不同

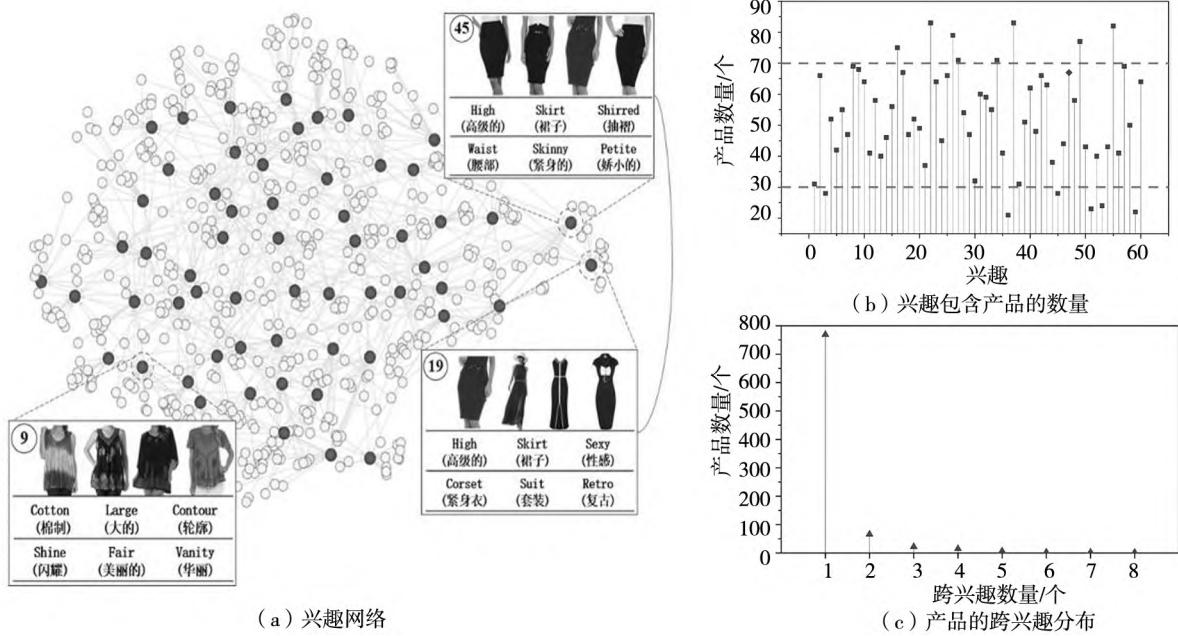


图 3 消费者兴趣分布

Fig. 3 Distribution of consumer interest

产品的跨兴趣分布统计情况,可以看出,由于产品特征的多样性,部分产品存在于多个兴趣之中,即能够满足消费者的多方面需要. 同时,本研究假设消费者购买决策相互独立,不同消费者选择相同产品可能取决于不同的兴趣偏好. 因此,同一产品的图像特征和词汇可能出现在不同的兴趣,实验结果符合本研究对于消费者购买决策的建模过程.

为了更好的理解市场兴趣的具体内容,图 4 举例本研究得到的 4 个兴趣及其包含产品的图像和文本信息. 如图 4 所示,兴趣 18 的展示图像体现了修身、下摆等兴趣特征,描述文本体现

了羊毛、针织等兴趣特征. 兴趣 19 图像体现的是时尚、束腰等兴趣特征,文本体现了经典、复古、套装等兴趣特征. 兴趣 25 的图像内容包含了迷彩、休闲等特点,文本内容包含了军工品牌、军事、斜纹布等信息. 兴趣 47 的图像体现了运动、简洁等特点,文本体现了冠军、棉布、低价等信息. 由图 3 和图 4 可以看出,借助本研究构建的融合图像特征和词汇的兴趣推断方法,本研究可以对市场兴趣进行有效识别. 融合图像和文本信息,本研究方法提高了所得市场兴趣的可解释性,有助于企业更加准确和细粒度的理解市场需求.



图 4 偏好内容展示

Fig. 4 Demonstration of interest

基于得到的市场兴趣,本研究探究不同兴趣的流行度差异,实验统计了购买不同兴趣产品的消费者数量. 如图 5 所示,购买不同兴趣中相关产

品的消费者数量存在显著差异. 其中兴趣 17、兴趣 19、兴趣 36、兴趣 49、兴趣 52 等的产品在市场上拥有更多的消费者. 通过对图 3(b)和图 5 可

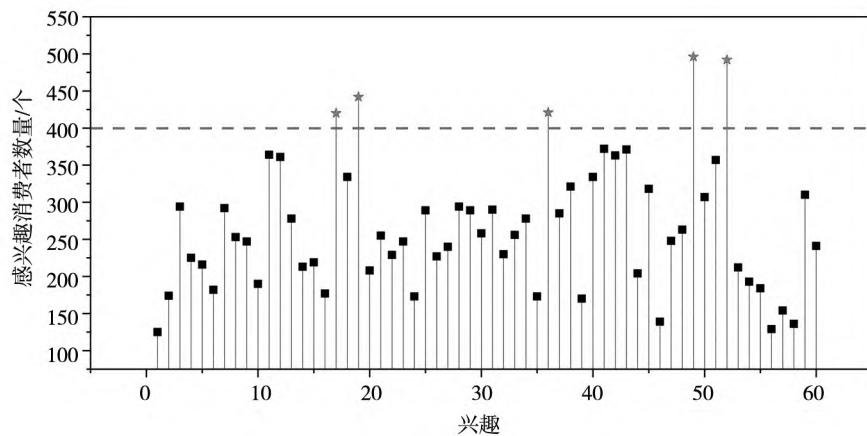


图 5 兴趣受欢迎程度图

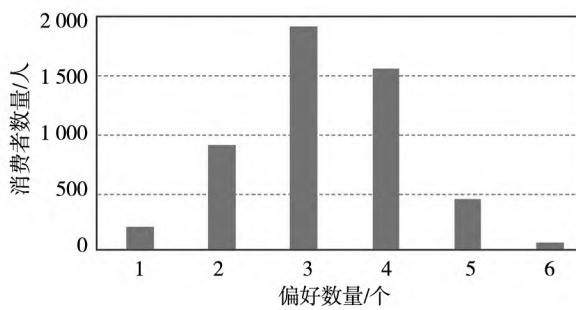
Fig. 5 Graph of interest popularity

以发现,兴趣中的产品数量与市场流行度并不完全一致,包含产品多的兴趣不一定具有更高的流行度.例如,兴趣 36 所反映的女士无袖连衣裙市场,虽然包含的产品数量不多(见图 3(b)),但是每件产品均具有较好的销量,使得该兴趣具有较好的市场流行度.本文研究结果可以为企业产品定位和市场选择决策提供有效建议.

## 2.2 个体偏好实验

基于有限注意力假设,本研究认为消费者的偏好是有限的,并基于 Spike-and-Slab 先验对消费者的有限偏好进行建模.为了对消费者的有限偏好进行分析,参考 Wang 和 Blei<sup>[23]</sup>的工作,本研究设定偏好选择器  $b$  为 0 的比重为偏好稀疏度.对于消费者  $u$ ,其对应的稀疏度计算如下.

$$Sparsity_{ratio(u)} = 1 - \frac{\sum_{k=1}^K b_{uk}}{K} \quad (16)$$



(a) 消费者偏好数量柱状图

基于式(16),实验计算消费者偏好稀疏度的分布情况如图 6(a)所示.由图 6(a)可以看出,与市场上大量的商品和广泛的兴趣相比,消费者偏好的稀疏度较高,平均稀疏度达到 0.98.在本研究得到的 60 个市场兴趣中,大部分消费者偏好的兴趣数量为 2 个~4 个,平均每个用户的偏好数量为 3.24,符合本研究个体偏好模型中关于消费者有限偏好的假设.为了更好地理解消费者的有限偏好,图 6(b)给出了 2 个消费者所购买产品的情况.其中,消费者 1 购买了 18 件产品,所购买产品覆盖袜子、男装、手表、皮带等多个类别,其偏好稀疏度为 0.57.消费者 2 购买了 8 件产品,且产品类别主要为特色鲜明的女装,稀疏度为 0.97.由图 6(b)可以发现,结合式(3)和式(7)所示的有限偏好建模与推断方法,本研究所得偏好分布和稀疏度能够对消费者购买决策进行有效刻画.



(b) 部分消费者购买产品内容展示

图 6 消费者偏好稀疏度

Fig. 6 Sparsity of consumer preference

本研究不仅可以得到消费者个体的偏好,基于式(4)和式(13),本研究还可以得到消费者在不同偏好上的分布.由图 7 可以看出,消费者在不同兴趣上的分布各不相同.虽然消费者 1 对蓝色和卡其色的舒适型套衫感兴趣,但是对卡其色套衫的偏好远高于对蓝色套衫的兴趣.消费者 2 的偏好相对广泛,有弹力、柔软、原创、舒适的男式卫衣和袜子等产品均符合其偏好,而且该消费者对 6 类偏好的喜好程度相对均衡.而消费者 3 是典型的家庭购物的决策者或执行者,不仅偏好柔软舒适的男装也同时关注经典印花的束腰女装.图 6 和图 7 实验结果表明,本研究模型可以对消费者个体的偏好分布进行有效地建模分析,研究

结果有助于深入理解和预测消费者的个性化需求,为企业个性化推荐和营销决策提供理论支撑.

## 2.3 模型准确性对比实验

为了验证所提模型的有效性,本研究对所提 STILT 模型与 LDA、Link-LDA、KGE-LDA 及 SLTM 等模型进行了 6 组对比实验.

1) LDA<sup>[24]</sup>是使用最广泛的主题模型,本研究将文本信息、图像信息以及联合信息分别作为模型输入进行对比实验,文本部分处理与本模型相同,图像部分用视觉词袋方法进行处理.使用文本信息的模型称为 LDA-word 模型,使用图像信息的模型称为 LDA-image 模型,使用图文融合信息的模型称为 LDA-combine 模型.

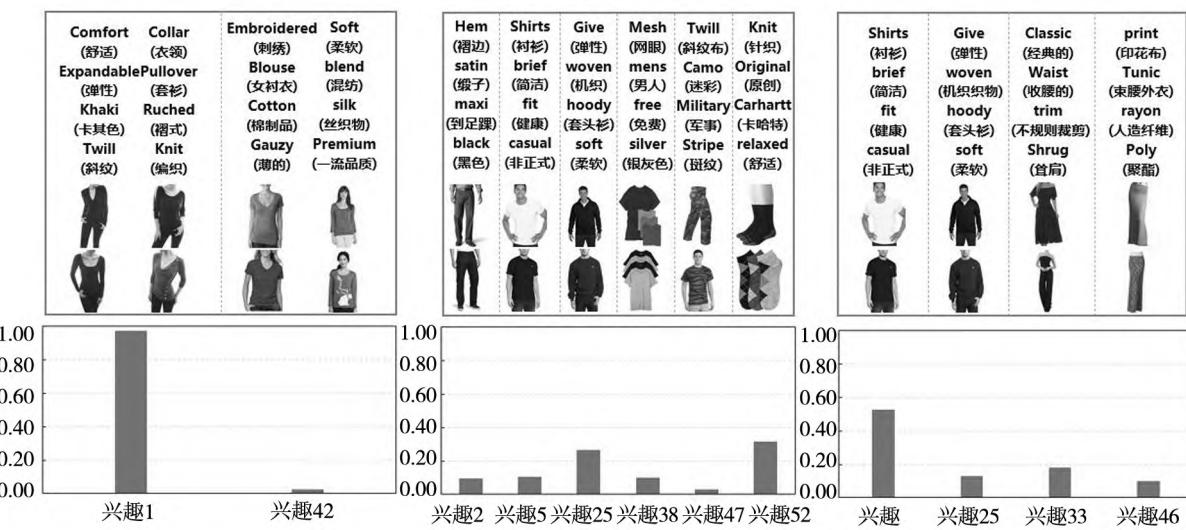


图 7 消费者偏好展示图

Fig. 7 Example of consumer preference

2) Link-LDA<sup>[25]</sup>是经典的多模态主题模型,能够联合不同模态的信息学习主题分布,不同模态信息采样均使用狄利克雷-多项式共轭分布. 与 STILT 相比,该模型不考虑消费者偏好的稀疏性,文本部分处理与本模型相同,图像部分用视觉词袋方法进行处理.

3) KGE-LDA<sup>[26]</sup>是融合知识图谱的多模态主题模型,该模型能够联合不同模态信息学习主题分布. 与 Link-LDA 不同的是,该模型利用 vMF 共轭分布将图像信息表征为向量形式. 该模型不考虑消费者偏好的稀疏性,文本与图像部分的处理均与本模型相同.

4) SLTM<sup>[27]</sup>是 Link-LDA 的变形,该模型能够联合不同模态信息学习主题分布,并考虑了消费者偏好的稀疏性. 与 STILT 相比,该模型不同模态信息采样均使用狄利克雷-多项式共轭分布,文本部分处理与本模型相同,图像部分用视觉词袋方法进行处理.

本研究使用困惑度评分度量不同模型学习消费者偏好的有效性. 困惑度是评估文本主题建模(本研究为消费者兴趣建模)有效性的常用指标,该指标量化了偏好内容分布的良好程度. 为了测试鲁棒性,本研究在不同的兴趣数量( $K$ )下将 STILT 模型与基准模型进行对比实验.

为了验证文本和图像信息融合的有效性,本

研究首先计算 STILT 模型的困惑度,并与 LDA-word 和 LDA-image 的困惑度进行对比(如图 8 (a) 和图 8 (b)). 由图 8 (a) 和图 8 (b) 可以看出, STILT 模型的困惑度优于基于单模态信息的 LDA-word 和 LDA-image 模型. 本研究方法将消费者购买记录与产品展示图像和描述文本相结合,能够有效提高消费者个性化需求预测的效果. 为了验证有限偏好假设的有效性,本研究分别利用 STILT、Link-LDA、KGE-LDA、LDA 模型,融合图像和文本信息对消费者个性化需求进行建模分析,图 8 (c) 给出了不同模型的困惑度对比情况. 由图 8 (c) 可以看出, STILT 的效果显著优于 Link-LDA、KGE-LDA 和 LDA 模型. 为了验证利用 vMF 分布对图像建模的有效性,本研究利用 STILT、SLTM 模型,融合图像和文本信息对消费者有限偏好的个性化需求进行建模,图 8 (d) 给出了不同模型的困惑度对比情况. 由图 8 (d) 可以看出, STILT 的效果显著优于 SLTM 模型,由此可见利用 vMF 分布建模图像信息比视觉词袋的方法更有效. 图 8 所示实验表明,本研究模型考虑消费者有限偏好,将产品图像信息表征为向量以保留更多的图像偏好信息,将图像和文本相融合识别消费者的细粒度偏好,提高消费者需求预测的可解释性,是预测消费者个性化需求的有效思路.

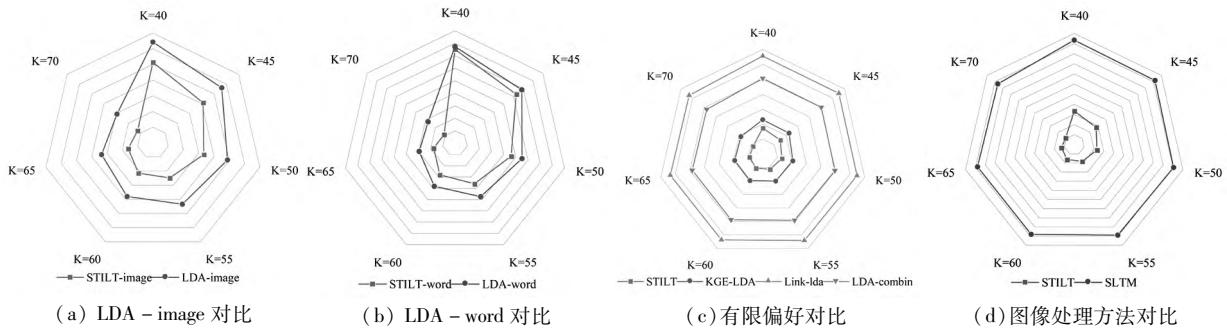


图 8 模型效果对比实验

Fig. 8 Perplexity score of different models

## 2.4 模型推荐效果对比实验

为了验证所提模型在推荐方面的性能,本研究利用所提模型与基准模型得到的消费者偏好内容及兴趣分布为消费者推荐产品,对比所提 STILT 模型与以上基准模型在个性化推

荐方面的表现. 本研究采用了准确率、召回率两种评估指标,指标值越高,表示推荐效果越好. 为了保证结果的稳健性,本研究分别获取长度为 1、5、10、20 的推荐列表. 图 9 展示了实验结果<sup>⑤</sup>.

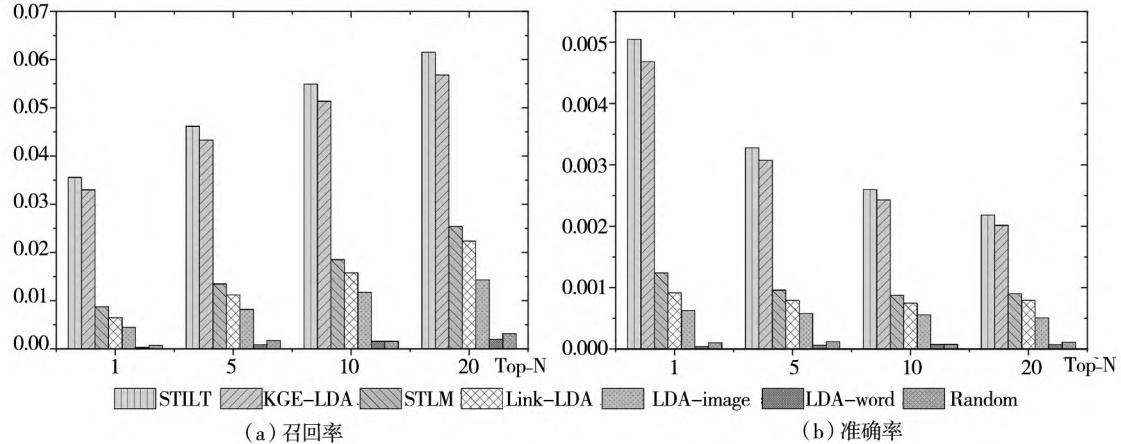


图 9 模型推荐效果对比

Fig. 9 Comparison of model recommendation effect

结果表明,在准确率和召回率这两个推荐指标上,本研究提出的 STILT 模型相比于其他方法取得了更好的性能,证明基于 STILT 模型结果可以获得更优的个性化推荐效果. 如图 9 所示,基于 LDA-image 的推荐效果优于基于 LDA-word 的推荐,表明图像信息能够更好地捕捉消费者的个性化偏好. 融合图像、文本数据的多模态 Link-LDA 模型优于只利用文本数据的 LDA-word 模型与只利用图像数据的 LDA-image 模型,证明了融合图文信息在推荐中的贡献. 值得关注的是,STILT 模

型与 KGE-LDA 模型的推荐性能显著优于其他模型,证明了 vMF 分布建模图像数据的优越性. 同时,有限偏好的重要性可以通过两组模型对比体现,分别是 SLTM 模型对比 Link-LDA 模型与 STILT 模型对比 KGE-LDA 模型. 这两组对比模型中数据处理方法相同,但前者均考虑了有限偏好,后者均未考虑有限偏好. 前者的推荐效果均优于后者,证明了考虑消费者有限偏好在需求获取和个性化推荐中的重要作用. 通过上述实验结果可以发现,STILT 模型效果的提升源于图文信息的

<sup>⑤</sup> 由于版面限制,推荐列表获取过程可联系作者备索.

融合、vMF 分布对于图像数据的建模以及对消费者有限偏好的关注。

### 3 结束语

随着电子商务的广泛应用,个性化已经成为电子商务模式创新的重要趋势。为了准确预测消费者需求,本研究提出了融合产品描述文本和展示图像的个性化需求预测方法,构建了有限偏好约束下的融合文本和图像信息的稀疏图文链接主题模型。基于亚马逊数据集的实验表明,所提模型具有良好的需求预测效果,文本和图像信息的利用提高了个性化需求预测的可解释性。研究结果不仅可以对消费者的群体兴趣进行有效分析,有助于企业更好地掌握市场需求状况;而且可以识

别消费者的个体偏好,为企业产品推荐和个性化营销决策提供依据。

融合多模态信息的个性化需求预测与推荐是管理领域持续探索的热点。本研究模型重点考虑了文本和图像的互补信息,如服装图像中“纯色且无袖”等形状信息和描述文本中的“亚麻”等图像无法反应的材质信息。文本和图像信息在互补的同时,有可能存在信息冗余。冗余信息一方面说明消费者在相关方面具有更强的兴趣,对预测消费者偏好具有正向作用。另一方面,过多的冗余会过分强化消费者的特定兴趣,对预测具有负向作用。后续研究将深入分析产品描述文本和展示图像的语义关联关系,克服图文冗余信息的影响,构建考虑图文语义关联的个性化需求预测模型。

### 参 考 文 献:

- [1] Liu L, Dzyabura D, Mizik N. Visual listening in: Extracting brand image portrayed on social media[J]. *Marketing Science*, 2020, 39(4): 669 – 686.
- [2] Koh B, Cui F. An exploration of the relation between the visual attributes of thumbnails and the view-through of videos: The case of branded video content[J]. *Decision Support Systems*, 2022, 160: 113820.
- [3] Xiao L, Ding M. Just the faces: Exploring the effects of facial features in print advertising[J]. *Marketing Science*, 2014, 33(3): 338 – 352.
- [4] Gomez L, Patel Y, Rusiñol M, et al. Self-supervised learning of visual features through embedding images into text topic spaces[C]. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Hawaii: USA, 2017: 4230 – 4239.
- [5] He R, McAuley J. VBPR: Visual bayesian personalized ranking from implicit feedback[C]. *Proceedings of the AAAI Conference on Artificial Intelligence*, Arizona: USA, 2016: 144 – 150.
- [6] McAuley J, Targett C, Shi Q, et al. Image-based recommendations on styles and substitutes[C]. *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, Oregon: USA, 2015: 43 – 52.
- [7] 王海忠, 欧阳建颖, 陈宣臻. 续集电影的片名策略及其市场效应研究[J]. 管理科学学报, 2019, 22(11): 19 – 32.  
Wang Haizhong, Ouyang Jianying, Chen Xuanzhen. Movie sequel's naming strategies and its effect on market performance [J]. *Journal of Management Sciences in China*, 2019, 22(11): 19 – 32. (in Chinese)
- [8] 曹丽娜, 唐锡晋. 基于主题模型的 BBS 话题演化趋势分析[J]. 管理科学学报, 2014, 11(17): 109 – 121.  
Cao Lina, Tang Xijin. Trends of BBS topics based on dynamic topic model[J]. *Journal of Management Sciences in China*, 2014, 11(17): 109 – 121. (in Chinese)
- [9] 金晓玲, 周中允, 尹梦杰, 等. 在线用户点赞与评论行为的产生机理差异研究——以医疗健康类企业微信公众号为例[J]. 管理科学学报, 2021, 24(4): 54 – 68.

- Jin Xiaoling, Zhou Zhongyun, Yin Mengjie, et al. Understanding antecedent differences across online users' like and comment behaviors: The case of healthcare enterprise WeChat public platform [J]. *Journal of Management Sciences in China*, 2021, 24(4): 54–68. (in Chinese)
- [10] Dotson M R, Büschken J, Allenby G M. Explaining preference heterogeneity with mixed membership modeling [J]. *Marketing Science*, 2020, 39(2): 407–426.
- [11] 尹鹏, 丁栋虹, 豆国威. 视频平台用户生成内容投资和定价决策 [J]. *管理科学学报*, 2020, 23(10): 116–126. Yin Peng, Ding Donghong, Dou Guowei. User generated content investment and pricing decisions in online video markets [J]. *Journal of Management Sciences in China*, 2020, 23(10): 116–126. (in Chinese)
- [12] Khorramfar R, Özaltın O Y, Kempf K G, et al. Managing product transitions: A bilevel programming approach [J]. *INFORMS Journal on Computing*, 2022, 34(5): 2828–2844.
- [13] Cheng Z, Chang X, Zhu L, et al. MMALFM: Explainable recommendation by leveraging reviews and images [J]. *ACM Transactions on Information Systems (TOIS)*, 2019, 37(2): 1–28.
- [14] Paramveer S D, Sinan A. Modeling dynamic user interests: A neural matrix factorization approach [J]. *Marketing Science*, 2021, 40(6): 1059–1080.
- [15] Guan Y, Wei Q, Chen G. Deep learning based personalized recommendation with multi-view information integration [J]. *Decision Support Systems*, 2019, 118: 58–69.
- [16] Lleras J S, Masatlioglu Y, Nakajima D, et al. When more is less: Limited consideration [J]. *Journal of Economic Theory*, 2017, 170: 70–85.
- [17] Li Q, Maggitti P G, Smith K G, et al. Top management attention to innovation: The role of search selection and intensity in new product introductions [J]. *Academy of Management Journal*, 2013, 56(3): 893–916.
- [18] Leung A C M, Agarwal A, Konana P, et al. Network analysis of search dynamics: The case of stock habitats [J]. *Management Science*, 2017, 63(8): 2667–2687.
- [19] Lin T, Tian W, Mei Q, et al. The dual-sparse topic model: Mining focused topics and focused terms in short text [C]. *Proceedings of the 23rd International Conference on World Wide Web*, Seoul, 2014: 539–550.
- [20] Wang C, Blei D. Decoupling sparsity and smoothness in the discrete hierarchical dirichlet process [J]. *Advances in Neural Information Processing Systems*, 2009, 22: 1982–1989.
- [21] Gopal S, Yang Y. Von mises-fisher clustering models [C]. *Proceedings of the International Conference on Machine Learning*, Beijing, 2014: 154–162.
- [22] Batmanghelich K, Saeedi A, Narasimhan K, et al. Nonparametric spherical topic modeling with word embeddings [C]. *Proceedings of the Conference Association for Computational Linguistics Meeting*, Berlin, 2016: 537–542.
- [23] Wang C, Blei D. Decoupling sparsity and smoothness in the discrete hierarchical dirichlet process [C]. *Proceedings of the 22rd International Conference on Neural Information Processing Systems*, British Columbia, 2009: 1982–1989.
- [24] Blei D M, Ng A Y, Jordan M I. Latent dirichlet allocation [J]. *Journal of Machine Learning Research*, 2003, 3: 993–1022.
- [25] Blei D M, Jordan M I. Modeling annotated data [C]. *Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Informaion Retrieval*, Toronto, 2003: 127–134.
- [26] Yao L, Zhang Y, Wei B, et al. Incorporating knowledge graph embeddings into topic modeling [C]. *Proceedings of the AAAI Conference on Artificial Intelligence*, California, 2017: 3119–3126.
- [27] Qian Y, Liu Y, Jiang Y, et al. Detecting topic-level influencers in large-scale scientific networks [J]. *World Wide Web*, 2020, 23(2): 831–851.

## Personalized demand prediction based on text and image information: From the perspective of limited preferences

JIANG Yuan-chun<sup>1, 2</sup>, LI Yi<sup>1, 2</sup>, QIAN Yang<sup>1, 3\*</sup>, CHAI Yi-dong<sup>1, 3</sup>, LIU Ye-zheng<sup>1, 3</sup>

1. School of Management, Hefei University of Technology, Hefei 230009, China;
2. Key Laboratory of Philosophy and Social Sciences for Cyberspace Behaviour and Management, Hefei 230009, China;
3. Key Laboratory of Process Optimization and Intelligent Decision-making, Ministry of Education, Hefei 230009, China

**Abstract:** As e-commerce applications evolve from breadth to depth, personalization has become an important direction for e-commerce service innovation. In order to accurately predict consumers' personalized demand, this paper proposes a personalized demand prediction method with a limited preference constraint by integrating information of product description texts and display images. Inspired by the hypothesis of limited attention, this paper models the limited personalized preference of consumers, and combines the image and text features to construct a Sparse Text-Image Linked Topic Model with a limited preference constraint. The model predicts the individual demands of consumers through group interest modeling, individual preference modeling, and purchase decision modeling. Experiments on the Amazon public data set show that the proposed model can effectively predict group interests and personalized preferences of consumers. The integration of text and image information improves the interpretability of personalized demand prediction.

**Key words:** personalization; demand prediction; limited preference; Sparse Text-Image Linked Topic Model