# Tr追踪用户评论中品牌关联的带极性主题 s

**Runcong Zhao**[1,2], **Lin Gui**[1],**严汉奇**[2],**Yulan He**[1,2,3]伦敦国王学院， [2]华威大学， [3]
艾伦·图灵研究所 runcong.zhao@warwick.ac.uk, yulan.he@kcl.ac.uk

arXiv:2301.07183v1 [cs.IR] 3 Jan 2023

## 摘要

商业组织监测在线客户评论，对于衡量客户满意度和更好地管理自身声誉至关重要。本文提出一种新颖的动态品牌-主题模型（dBTM），能够从按时间顺序划分的时间区间内整理的产品评论中，自动检测并追踪品牌相关情感得分和带极性的主题。dBTM通过高斯状态空间模型对潜在品牌极性得分和主题-词分布随时间的演变进行建模。此外，该模型还融入了一种元学习策略，用于控制各时间区间内主题-词分布的更新，从而确保主题平滑转变并提升品牌得分预测效果。我们在一个由MakeupAlley评论构建的数据集和一个酒店评论数据集上对其进行了评估。实验结果表明，dBTM在品牌排名方面优于多种有竞争力的基线方法，在主题连贯性与唯一性之间实现了良好平衡，并能在不同时间区间提取出区分度高的带极性主题[1]。

## 1 引言

随着社交媒体平台日益普及，客户倾向于在网上分享自己对产品的亲身体验。追踪网上的客户评论有助于企业衡量客户满意度，并更好地管理自身声誉。可以通过动态主题模型（Blei 和 Lafferty，2006；Wang 等人，2008年；Dieng等人，2019年）监测评论中与品牌相关的主题变化。例如，动态联合情感主题模型（dJST）（He 等人，2014）通过假设不同时间片之间的情感-主题-词分布存在依赖关系，能够提取随时间演变的带极性的主题。然而，这类方法需要纳入词语先验极性信息，并假设主题与离散极性类别相关联。此外，它们无法直接推断品牌极性分数。

近期提出的品牌-主题模型（BTM）（Zhao 等人，2021）能够从评论中自动推断实值品牌相关情感得分，并通过逐渐将相关情感得分从负面调整到正面，生成一组情感主题。这使用户能够检测例如强烈正面的主题或略微负面的主题。然而，BTM 假设所有文档都在模型学习之前可用，因此无法追踪主题演化和品牌极性随时间的变化。

本文提出一种受元学习启发的新型框架，元学习广泛用于分布适应任务（Suo 等人，2020）。在对按时间排序且划分为多个时间片的文档进行训练时，我们假设，在每个时间片中提取带极性的主题并推断品牌极性分数，可以视为一个新的子任务；模型学习的目标是在新的时间片中，学会适应与不同品牌极性分数相关联的主题-词分布。我们使用 BTM 作为基础模型，并将学习到的参数存储在记忆库中。在每个时间片，我们根据模型生成的品牌排名结果，在验证集上评估模型性能。评估结果用于早停，并通过元学习动态初始化下一个时间片的模型参数。所得模型称为动态品牌主题建模（dBTM）。

dBTM 的最终结果如图 1所示，它能够同时追踪主题演化并推断潜在品牌极性

[1]数据和代码可在此处获取 https://github.com/BLPXSPG/dBTM。

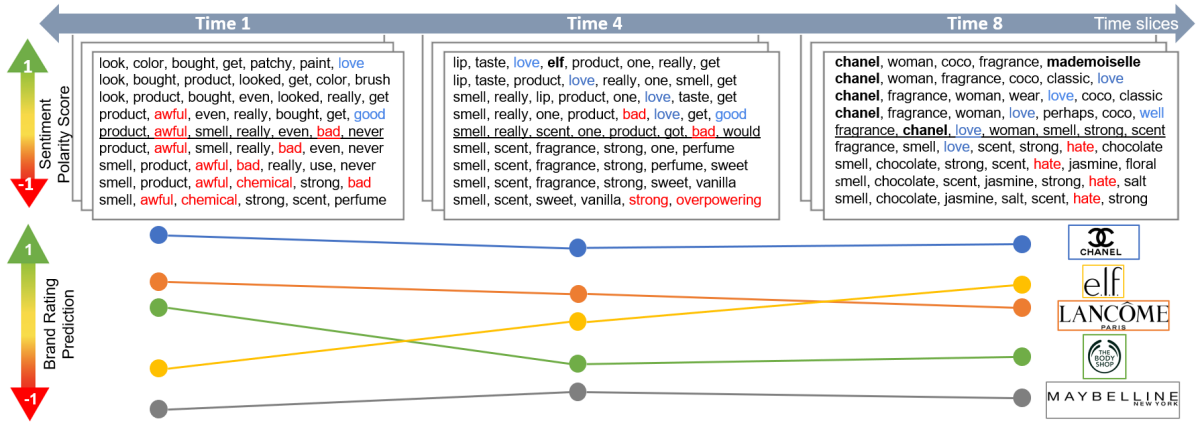



图1：我们提出的模型对品牌关联的带极性主题进行追踪。横轴展示了从时间片1、4和8中提取的一个示例主题的高频单词。在每个时间片中，我们可以看到一组主题，其关联的情感分数从 -1（负面）逐渐变化到 1（正面 ，如纵轴所示。为便于查看，正面单词以蓝色突出显示，负面单词则以红色突出显示。我们可以看到，在时间片1中，负面主题主要围绕对香水化学气味的抱怨，而正面主题则是对产品外观的赞美。从时间片1到时间片8，我们还可以看到负面主题所涉及的方面不断演变，从抱怨香水气味过于浓烈，转变为抱怨甜香过于浓郁。在图的下半部分，我们展示了推断出的三个品牌的极性分数。例如，Channel 的排名通常高于兰蔻，而兰蔻的得分又高于美体小铺。

得分随时间的变化。此外，通过逐步改变品牌极性分数，它还能够在每个时间片生成细粒度的带极性的主题。从本质上说，我们可以从两个维度观察主题转变：沿着离散的时间维度，或沿着连续的品牌极性分数维度。

我们在一个由 MakeupAlley 构建的评论数据集上评估了 dBTM[2]，，该数据集包含超过 61.1 万条评论，时间跨度超过 9 年；此外，我们还在从 HotelRec 抽取的酒店评论数据集上进行了评估（An-tognini and Faltings，2020），其中包含 7 年间最受欢迎的 25 家酒店的评论。我们将其性能与多种具有竞争力的基线方法进行了比较，发现它能够生成更好的品牌排名结果，更准确地预测品牌分数时间序列，并生成极性分明、主题连贯性和多样性更均衡的主题。更值得注意的是，我们在一个更困难的设置中评估了 dBTM：监督标签信息（即评论评分）仅在第一个时间片提供，此后 dBTM 在不使用评论评分的情况下以无监督方式训练。在这种设置下，与在监督设置下训练的基线方法相比，dBTM 仍能更准确地生成跨时间片的品牌排名结果。这一特性令人称道，因为 dBTM 最初在一小组标注数据上训练后，便能以无监督方式利用流式数据自适应调整其参数。

我们的贡献主要有三点：

- 我们提出了一种名为 dBTM 的新模型，该模型基于高斯状态空间模型，并结合元学习，用于动态品牌主题和极性分数跟踪；

- 我们提出了一种新颖的元学习策略，在每个时间片动态初始化模型参数，以便更好地捕捉评分变化，进而生成总体质量更高的主题；

- 实验结果表明，在初始时间片使用评论评分监督训练的 dBTM，能够以无监督的方式利用流式数据自适应调整参数，并且相较于有监督基线，仍能取得更好的品牌排名结果。

## 2 相关工作

我们的工作与以下研究相关：

### 2.1 动态主题模型

主题模型，例如潜在狄利克雷分配（Blei 等人，2003）模型，是其中最

[2]https://www.makeupalley.com/

成功的文档集合统计分析方法之一。动态主题模型旨在分析大型文档集合中主题随时间的演变。基于潜在狄利克雷分配的早期方法包括动态主题模型（DTM）（Blei 和 Lafferty，2006），该模型使用卡尔曼滤波器对主题随时间的变化进行建模；以及连续时间动态主题模型（Wangetal.，2008年），该模型以连续推广形式取代了 DTM 的离散状态空间模型。近年来，研究者将 DTM 与词嵌入相结合，以便在文档流中生成更多样且更连贯的主题（Dieng等人，2019年）。

除了常用的潜在狄利克雷分配外，泊松因子分解也可用于主题建模，它将文档-词计数矩阵分解为文档-主题矩阵与主题-词矩阵的乘积。通过卡尔曼滤波器（Charlin 等人，2015）、神经网络（Gong 和 Huang，2017），或将文档-词计数上的泊松分布扩展为随时间变化的非齐次泊松过程（Hosseini 等人，2018年），即可捕捉时间步之间的依赖关系，从而将其扩展到分析序列计数向量，例如包含单个词计数矩阵且每个时间间隔对应一列的文档语料库。

虽然上述模型通常用于无监督设置，但联合情感-主题（JST）模型（Lin 和 He，2009；Lin 等人，2012）将极性词先验纳入模型学习，从而能够提取归属于不同情感类别的主题。此后，JST 被扩展为动态版本，称为 dJST；该模型通过假设当前时刻的情感-主题词分布由一个狄利克雷分布生成，而该分布的参数由前一时间区间的情感-主题词分布确定，来追踪主题和情感随时间的变化（He 等人，2014）。

### 2.2 市场/品牌主题分析

潜在狄利克雷分配及其变体已应用于营销研究。例如，通过分析消费者购买行为来检测用户兴趣（Gao 等人，2017；Sun 等人，2021），通过挖掘推特数据，追踪给定品牌中奢侈品市场的竞争对手（Zhang 等人，2015），以及识别新兴应用中的问题（Yang 等人，2021）。能够提取全局信息的矩阵分解也用于产品推荐（Zhou 等人，2020）和评论摘要（Cui 和 Hu，2021）。通过结合基于采样的方法进行近似，并利用情感先验知识（例如情感词典），可以对主题与极性之间的交互进行建模（Lin 和 He，2009；Lin 等人，2012）。然而，这类先验知识具有很强的领域特异性。已知极性的种子词，或根据形态学信息生成的种子词（Brody 和 Elhadad，2010），是获取主题极性的另一种常用方法。但这些方法侧重于分析现有主题的极性。近期，有研究提出了基于泊松因子分解构建的品牌-主题模型（Zhao 等人，2021），该模型能够推断品牌极性分数，并生成细粒度的带极性的主题。BTM 的详细说明见第 3 节。

### 2.3 元学习

元学习，也称为学会学习，大体上可分为基于度量的学习和基于优化的学习。基于度量的学习旨在学习训练实例之间的距离函数，从而在学习到的嵌入空间中，将测试实例与训练实例进行比较并据此分类（Sung 等人，2018）。基于优化的学习通常会将有标签样本划分为训练集和验证集。其基本思想是在训练集上微调参数以获得更新后的参数，然后在验证集上进行评估，得到误差，并将其转换为损失值，以优化原始参数（Finn 等人，2017；Jamal 和 Qi，2019）。元学习已应用于许多任务，包括文本分类（Geng 等人，2020）、主题建模（Song 等人，2020）、知识表示（Zheng 等人，2021）、推荐系统（Neupane 等人，2021；Dong 等人，2020；Lu 等人，2020）和事件检测（Deng 等人，2020）。尤其是，基于元学习的方法在分布适应方面取得了显著成果（Suo 等人，2020；Yu 等人，2021）。我们在此提出一种元学习策略，用于学习如何在每个时间片中自动初始化模型参数。

## 3 初步介绍：品牌主题模型

品牌-主题模型（BTM）（Zhao 等人，2021 年），如图2中间部分所示，使用配有文档级情感类别标签（例如“正面”、“负面”和“中性”）的评论文档进行训练。它能够自动推断实值的品牌相关极性分数，并生成细粒度的情感主题；在这些主题中，随着相关情感逐渐变化，可以观察到单词的连续变化。该模型部分受到基于文本的理想点模型（TBIP）（Vafa 等人，2020）的启发，后者旨在通过泊松因子分解对文本生成过程建模。具体而言，对于输入的词袋数据，词项 $v$ 在文档 $d$ 中的计数表示为词项计数 $c_{dv}$，并假设其服从泊松分布 $c_{dv} \sim$ 泊松($\lambda_{dv}$)，其中速率参数 $\lambda_{dv}$ 可分解为：

$$\lambda_{dv} = \sum_k \theta_{dk}\beta_{kv} \tag{1}$$

其中，$\theta_{dk}$ 表示文档级主题强度，$\beta_{kv}$ 表示主题-词分布。我们有 $\boldsymbol{\theta} \in \mathbb{R}_+^{D\times K}$、$\boldsymbol{\beta} \in \mathbb{R}_+^{K\times V}$，其中 $D$ 是语料库中的文档总数，$K$ 是主题数，$V$是词汇量。随后，将品牌极性分数$x_{b_d}$和主题-词偏移量 $\eta_{kv}$添加到模型中：

$$\lambda_{dv} = \sum_k \theta_{dk} \exp(\log \beta_{kv} + x_{b_d}\eta_{kv}) \tag{2}$$

$x_{b_d}$ 是品牌 $b$ 的文档 $d$ 对应的品牌极性分数，我们有 $\boldsymbol{\eta} \in \mathbb{R}_+^{K\times V}$、$\boldsymbol{x} \in \mathbb{R}$。为便于展示，模型将输出中的品牌极性分配归一化为 [-1,1]。

上述公式的直觉在于，捕捉品牌极性分数的潜变量$x_{b_d}$ 可以是正值，也可以是负值。如果某个词倾向于频繁出现在极性为正面的评论中，但当前品牌的极性得分为负面，那么通过使$x_{b_d}$和 $\eta_{kv}$ 符号相反，该词的出现次数就会减少。

对 $\boldsymbol{\theta}$ 和 $\boldsymbol{\beta}$ 赋予伽马先验，其中 $a$，$b$，$c$，$d$ 为超参数；而对品牌极性分数 $\boldsymbol{x}$ 和主题-词计数偏移量 $\boldsymbol{\eta}$ 赋予正态先验。

$$\theta_{dk} \sim \text{Gamma}(a,b), \quad \beta_{kv} \sim \text{Gamma}(c,d),$$
$$x_{b_d} \sim \mathcal{N}(0,1), \quad \eta_{kv} \sim \mathcal{N}(0,\boldsymbol{I}).$$

BTM 使用 Gumbel-Softmax（Jang 等人，2017）构造用于情感分类的文档特征。这是因为直接从泊松分布中采样词频不可微。Gumbel-Softmax 是一种结合重参数化技巧的梯度估计器，可用于实现梯度的反向传播。更多详情请参见（Zhao 等人，2021）。

## 4 动态品牌主题模型（dBTM）

为了追踪顾客评论中与品牌相关的主题动态，我们将文档划分为多个时间片，每个时间片的时间段可以任意设定，例如一周、一个月或一年。在每个时间片中，我们都有一系列 $M$ 文档 $\{d_1, \cdots, d_M\}$，这些文档按照发布时间戳排序。时间片 $d$ 中的文档 $t$ 以词袋（BoW）表示形式输入。我们扩展 BTM 以处理流式文档，假设当前时间片中的文档会受到过去文档的影响。由此得到的模型称为动态品牌-主题模型（dBTM），其架构如图2所示。

### 4.1 初始化

在原始 BTM 模型中，需要推断的潜变量包括文档-主题分布 $\boldsymbol{\theta}$、主题-词分布 $\boldsymbol{\beta}$、品牌相关极性得分 $\boldsymbol{x}$以及与极性相关的主题-词偏移量 $\boldsymbol{\eta}$。在时间片 0 中，我们将该时间片内的所有文档表示为文档-词计数矩阵。随后，我们采用坐标上升变分推断进行泊松因子分解（Gopalan 等人，2015），以推导出 $\boldsymbol{\theta}$和 $\boldsymbol{\beta}$（见式（1））。主题-词计数偏移量$\boldsymbol{\eta}$ 和品牌极性分数 $\boldsymbol{x}$ 从标准正态分布中采样得到。

### 4.2 状态空间模型

在时间片 $t$，我们可以通过高斯状态空间模型来建模潜在品牌相关极性得分 $\boldsymbol{x}^t$ 和极性相关主题-词偏移量 $\boldsymbol{\eta}^t$ 随时间的演变：

$$\boldsymbol{x}^t|\boldsymbol{x}^{t-1} \sim \mathcal{N}(\boldsymbol{x}^{t-1}, \sigma_x^2\boldsymbol{I}) \tag{3}$$
$$\boldsymbol{\eta}^t|\boldsymbol{\eta}^{t-1} \sim \mathcal{N}(\boldsymbol{\eta}^{t-1}, \sigma_\eta^2\boldsymbol{I}) \tag{4}$$

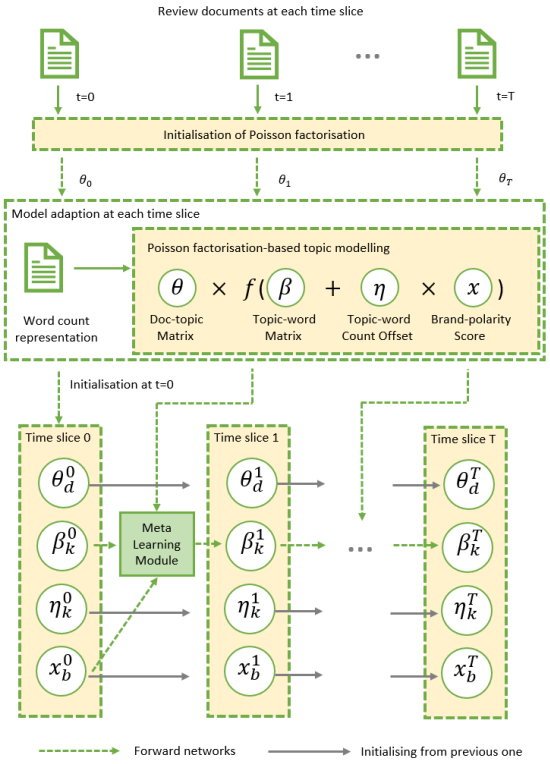



图2：动态品牌-主题模型（dBTM）的整体架构。该模型扩展了上方方框中所示的品牌-主题模型（BTM），以处理流式文档。具体而言，在时间片$t$，文档-主题分布 $\boldsymbol{\theta}^t$ 由标准泊松因子分解模型初始化；潜在品牌相关极性得分 $\boldsymbol{x}^t$ 和极性相关主题-词偏移量 $\boldsymbol{\eta}^t$ 的演变则由两个独立的高斯状态空间模型建模。主题-词分布 $\boldsymbol{\beta}^t$ 的先验设置依据前两个时间片中模型在品牌排名结果上的性能趋势。灰色线表示由高斯状态空间模型关联的参数，绿色线表示前向计算。

对于主题-词分布 $\boldsymbol{\beta}$，我们采用类似的高斯状态空间模型，但使用对数正态分布：

$$\boldsymbol{\beta}^t|\boldsymbol{\beta}^{t-1} \sim \mathcal{LN}(\boldsymbol{\beta}^{t-1}, \sigma_\beta^2 \boldsymbol{I}) \tag{5}$$

虽然主题-词分布可以从前一个时间片继承，但由于每个时间片的文档集合各不相同，文档-主题分布 $\boldsymbol{\theta}^t$ 需要在每个时间片开始时重新初始化。我们建议在每个时间片进行模型适配之前，先运行简单的泊松因子分解，以推导出 $\boldsymbol{\theta}^t_{(p)}$ 的初始值：这里，前一个时间片的主题-词分布 $\boldsymbol{\beta}^{t-1}{}_{(p)}$ 成为当前时间片中主题-词分布的先验 $\boldsymbol{\beta}^t_{(p)}$，如公式 (5) 所定义。我们使用下标 $(p)$ 表示这些参数是在每个时间片开始时的泊松因子分解初始化阶段推导得出的。

本质上，在每个时间片 $t$，我们将 BTM 模型的文档-主题分布 $\boldsymbol{\theta}^t$ 初始化为 $\boldsymbol{\theta}^t_{(p)}$，该分布通过对 $t$中的文档-词计数矩阵进行泊松因子分解得到。对于主题-词分布，在 BTM 中，我们可以将 $\boldsymbol{\beta}^t$ 设为继承自 $\boldsymbol{\beta}^{t-1}$，如式（5）所定义；此外，我们还有 $\boldsymbol{\beta}^t_{(p)}$，它是通过直接对当前时间片中的文档-词计数矩阵进行泊松因子分解得到的。接下来，我们将介绍如何通过元学习初始化 $\boldsymbol{\beta}^t$ 的值。

### 4.3 元学习

我们注意到，尽管每个时间间隔的参数通过高斯状态空间模型与前一个时间间隔的参数相联系，但各时间间隔生成的结果并不稳定。受元学习启发，我们将每个时间间隔的潜在品牌得分预测和情感主题提取视为一个新的子任务，并提出一种学习策略，根据前一个时间间隔验证集上的品牌评分预测性能，动态初始化每个时间间隔的模型参数。具体而言，我们从每个时间间隔的训练数据中留出 10% 作为验证集，并使用斯皮尔曼等级相关系数，将模型推断出的品牌排名结果与黄金标准进行比较。默认情况下，当前时间间隔的主题-词分布 $\boldsymbol{\beta}^t$ 的先验设为前一个时间间隔中学习到的 $\boldsymbol{\beta}^{t-1}$。但是，如果前一个时间间隔的品牌排名结果较差，那么 $\boldsymbol{\beta}^t$ 将被初始化为 $\boldsymbol{\beta}^{t-1}$ 与当前时间间隔中泊松分解初始化阶段得到的主题-词分布 $\boldsymbol{\beta}^t_{(p)}$ 的加权插值。其理由是，如果模型在前一个时间间隔表现不佳，那么它学到的主题-词分布对当前时间间隔参数的影响就应该较小。具体来说，我们首先在 $t-1$ 的验证集上评估模型在时间片 $t-1$ 返回的品牌排名结果：

$$\rho^{t-1} = \text{SpearmanRank}(\hat{\boldsymbol{r}}^{t-1}, \boldsymbol{r}^{t-1}) \tag{6}$$

**算法 1：** dBTM 的训练过程

**输入：** 主题数 $K$、品牌数 $B$、时间片 $t \in \{0, 1, 2, \cdots, T\}$，文档-词计数矩阵流 $\boldsymbol{C} = \{\boldsymbol{c}^0, \cdots, \boldsymbol{c}^T\}$

**输出：** 文档-主题强度 $\{\boldsymbol{\theta}^t\}_{t=0}^T$、主题-词矩阵 $\{\boldsymbol{\beta}^t\}_{t=0}^T$、品牌分数 $\{x_1^t, \cdots, x_B^t\}_{t=0}^T$

初始化：

通过泊松因子分解 $\boldsymbol{c}^0 \sim \text{Poisson}(\boldsymbol{\theta}^0\boldsymbol{\beta}^0)$，初始化 $\boldsymbol{\theta}^0$，$\boldsymbol{\beta}^0$　　$s\boldsymbol{x}^0 \sim \mathcal{N}(0, \boldsymbol{I})$，　　$\boldsymbol{\eta}^0 \sim \mathcal{N}(0, \Sigma_\eta^2 \boldsymbol{I})$，　　$\gamma^0 = 0$

通过最小化式中定义的损失来更新参数。 12

根据推断得到的品牌极性分数，在验证集上得出品牌排名 $\hat{}\boldsymbol{r}^0$ $\{x^0$ $_b\}_{b=1}^B$

计算斯皮尔曼等级相关系数 $\rho^0 = \text{SpearmanRank}(\hat{}\boldsymbol{r}^0, \boldsymbol{r}^0)$

设置权重 $\gamma^1 = \max(0.05, \Phi_{z_{\rho^0}}(0))$

训练：

**对** $t = 1$ **到** $T$ **执行**

　预训练：$\boldsymbol{c}^t \sim$ 泊松分布$(\boldsymbol{\theta}_{(p)}^t \boldsymbol{\beta}_{(p)}^t)$

　每个周期初始化：

　$\boldsymbol{x}^t \sim \mathcal{N}(\boldsymbol{x}^{t-1}, \sigma_x^2 \boldsymbol{I})$，$\boldsymbol{\eta}^t \sim \mathcal{N}(\boldsymbol{\eta}^{t-1}, \sigma_\eta^2 \boldsymbol{I})$，　$\boldsymbol{\theta}^t = \boldsymbol{\theta}_{(p)}^t$，　$\boldsymbol{\beta}^t = (1-\gamma^t)\boldsymbol{\beta}^{t-1} + \gamma^t \boldsymbol{\beta}_{(p)}^t$

　**对于** $i = 0$ **到**最大迭代次数**执行**

　　通过最小化式中定义的损失来更新参数 12

　　**如果**检查点**那么**

　　根据推断得到的品牌极性分数 $\{x^t$，得出验证集上的品牌排名 $\hat{}\boldsymbol{r}^t$ $_b\}_{b=1}^B$

　　计算斯皮尔曼等级相关系数 $\rho^t = \text{SpearmanRank}(\hat{}\boldsymbol{r}^t, \boldsymbol{r}^t)$

　　如果$(\Phi_{z_{\rho^t}}(\rho^{t-1}) > 0.95)$**然后**

　　中断# 根据公式，零假设被上四分位数拒绝。 7

　　**end**

　　设置权重 $\gamma^{t+1} = \max(0.05, \Phi_{z_{\rho^t}}(\rho^{t-1}))$

　**end**

**end**

**结束**

其中，$\hat{}\boldsymbol{r}^{t-1}$ 表示根据模型在时间片 $t-1$ 预测的潜在品牌极性得分 $\hat{}\boldsymbol{x}^{t-1}$ 得出的品牌排名结果，$\boldsymbol{r}^{t-1}$ 是金标准品牌排名，$\rho^{t-1}$ 是斯皮尔曼等级相关系数。为判断品牌排名结果是否变差，我们将其与前一个时间间隔中的品牌排名评估结果 $\rho^{t-2}$ 进行比较。具体而言，我们首先对 $\rho^{t-1}$ 进行费舍尔 z 变换 $z_{\rho^{t-1}}$，并假设其服从高斯分布：

$$z_{\rho^{t-1}} \sim \mathcal{N}\left(\ln(\frac{1+\rho^{t-1}}{1-\rho^{t-1}})^{0.5}, \frac{1}{B-3}\right) \tag{7}$$

其中，$B$ 表示品牌总数。然后，我们计算上述正态分布的累积分布函数（CDF），记为 $\Phi_{z_{\rho^{t-1}}}$，并计算$\Phi_{z_{\rho^{t-1}}}(\rho^{t-2})$，其本质上返回 $\Pr(z_{\rho^{t-1}} \leq \rho^{t-2})$。$\Phi_{z_{\rho^{t-1}}}(\rho^{t-2})$的值越小，表示模型在 $t-1$ 生成的品牌排名结果优于前一个时间片 $t-2$中的结果。这相当于执行假设检验，将排名评估结果 $\rho^{t-1}$ 与 $\rho^{t-2}$ 进行比较，以检验模型在 $t-1$ 的表现是否优于在 $t-2$的表现。假设检验的结果可用于设置权重 $\gamma^t$，以确定如何初始化主题在 $t$处的词分布，$\boldsymbol{\beta}^t$:

$$\gamma^t = \max\left(0.05, \Phi_{z_{\rho^{t-1}}}(\rho^{t-2})\right) \tag{8}$$

$$\boldsymbol{\beta}^t = (1-\gamma^t)\boldsymbol{\beta}^{t-1} + \gamma^t \boldsymbol{\beta}_{(p)}^t \tag{9}$$

上述公式表明，如果在 $t-1$ 训练的模型生成的品牌排名结果显著优于前一个时间片中的结果（$p$ 值为 $> 0.05$），那么根据估计的 $\Pr(z_{\rho^{t-1}} > \rho^{t-2}) = 1 - \gamma^t$ 概率，我们就更有信心主要基于 $\boldsymbol{\beta}^{t-1}$ 初始化 $\boldsymbol{\beta}^t$。我们必须重新初始化 $\boldsymbol{\beta}^t$，主要依据当前区间内泊松因子分解初始化阶段得到的主题-词分布 $\boldsymbol{\beta}_{(p)}^t$。

### 4.4 参数推断

我们通过最大化证据下界（ELBO），使用平均场变分分布来近似潜变量 $\boldsymbol{\theta}$，$\boldsymbol{\beta}$，$\boldsymbol{\eta}$，$\boldsymbol{x}$ 在给定观测到的文档词频数据 $\boldsymbol{c}$ 时的后验分布：

$$\begin{aligned}\mathcal{L}_{\text{ELBO}} = \mathbb{E}_{q_\phi}[\log\ p(\boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{x})] + \\ \log\ p(\boldsymbol{c}|\boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{x}) - \log\ q_\phi(\boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{x})]\end{aligned} \tag{10}$$

其中

$$q_\phi(\boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{x}) = \prod_{d,k,b} q(\theta_d) q(\beta_k) q(\eta_k) q(x_b) \tag{11}$$

此外，对于每篇文档 $d$，我们使用 Gumbel-Softmax 从上述学习到的参数中采样词频，构建其表示 $z_d$，并将其输入情感分类器，以预测类别分布 $\hat{y}_d$。我们还通过反转文档 $d$ 所关联品牌的推断极性得分的符号来进行对抗学习，并生成对抗表示 $\tilde{z}_d$。随后也将其输入同一个情感分类器，以生成另一个预测类别分布 $\tilde{y}_d$。我们通过最小化预测类别分布与实际类别分布之间的 Wasserstein距离来训练模型。最终损失函数由证据下界和 Wasserstein距离损失组合而成：

$$\mathcal{L} = -\mathcal{L}_{\text{ELBO}} + \frac{1}{M}\sum_{d=1}^{M}\big(\mathcal{L}_{\text{WD}}(\hat{y}_d, y_d) + \mathcal{L}_{\text{WD}}(\tilde{y}_d, \bar{y}_d)\big) \quad (12)$$

其中，$\mathcal{L}_{\text{WD}}(\cdot)$表示Wasserstein距离，$y_d$ 是金标准类别分布，$\bar{y}_d$是根据反转文档评分得出的类别分布。通过反转文档评分，我们实质上平衡了文档评分分布：对于每个正面文档，我们也会创建一个合成负面文档，反之亦然。

## 5 实验设置

**数据集** Yelp 和亚马逊产品等热门数据集（Ni 等人，2019年）以及多领域情感数据集（Blitzer 等人，2007）都是通过从亚马逊或 Yelp 随机抽取评论构建的，没有考虑这些评论在不同品牌和不同时间段上的分布。因此，我们从美妆产品评论网站 MakeupAlley 抓取排名前 25 个品牌的评论，构建了自己的数据集。每条评论都附有评分、产品类型、品牌和发布时间。按照 BTM 中的标签设置，我们将评分为 1 和 2 的评论归为负面类，评分为 3 的评论归为中性类，其余评分为 4 和 5 的评论归为正面类。整个数据集包含 611,128 条评论，覆盖 9 年（2005年至2013年）。我们将每年视为一个时间片，并将评论划分为 9 个时间片。平均评论长度为 123 个单词。除 MakeupAlley-Beauty 外，我们还在 HotelRec 上进行了实验（Antognini 和 Faltings，2020），通过选取 7 年间（2012年至2018年）排名前 25 家酒店的评论构建数据集。我们的数据集统计信息如表1所示。可以看出，该数据集类别分布不平衡：MakeupAlley-Beauty 中正面评论数量是负面评论的三倍多，而 HotelRec 中则接近十倍。

| **数据集** | **MakeupAlley-Beauty评论** |
|---|---|
| 每个类别的文档数 | |
| 负面 / 中性 / 正面 | 114,837 / 88,710 / 407,581 |
| 品牌数 | 25 |
| 文档总数 | 611,128 |
| 时间片数 | 9 |
| 平均评论长度（单词数） | 123 |
| 每个切片的平均文档数 | $\sim 68k$ |
| 词汇量 | $\sim 4500$ |
| **数据集** | **HotelRec 评论** |
| 每个类别的文档数 | |
| 负面 / 中性 / 正面 | 14,600 / 20,629 / 150,265 |
| 酒店数 | 25 |
| 文档总数 | 185,496 |
| 时间片数 | 7 |
| 平均评论长度（单词数） | 204 |
| 每个切片的平均文档数 | $\sim 26k$ |
| 词汇量 | $\sim 7000$ |

表 1：评论数据集统计。

**用于比较的模型**我们使用以下模型进行实验：

- 动态联合情感主题模型（dJST）（He 等人，2014），该模型基于潜在狄利克雷分配（LDA），并融入词语先验情感知识，可从文本中检测和追踪带极性的主题。在我们的实验中，MPQA 主观性词典[3] 用于推导词语先验情感信息。

- 基于文本的理想点模型（TBIP）（Vafa 等人，2020），这是一种无监督的泊松因子分解模型，可推断潜在品牌情感得分。

- 品牌主题模型（BTM）（Zhao 等人，2021），这是一种在 TBIP 基础上扩展的有监督泊松因子分解模型，并融入了文档级情感标签。

- dBTM，我们提出的动态品牌主题模型，该模型使用

[3] https://mpqa.cs.pitt.edu/lexicons/

| 时间片 | dJST | | TBIP | | BTM | | O-dBTM | | dBTM | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Corr | p 值 | Corr | p 值 | Corr | p 值 | Corr | p 值 | Corr | p 值 |
| *MakeupAlley-Beauty* | | | | | | | | | | |
| 1 | -0.249 | 0.230 | -0.567 | 0.003 | **0.552** | 0.004 | 0.454 | 0.023 | 0.402 | 0.046 |
| 2 | -0.437 | 0.029 | **0.527** | 0.007 | 0.488 | 0.013 | 0.459 | 0.021 | 0.438 | 0.029 |
| 3 | -0.327 | 0.111 | -0.543 | 0.005 | -0.384 | 0.058 | 0.504 | 0.010 | **0.523** | 0.007 |
| 4 | -0.127 | 0.545 | -0.431 | 0.032 | -0.428 | 0.033 | 0.448 | 0.025 | **0.453** | 0.023 |
| 5 | 0.112 | 0.596 | -0.347 | 0.089 | 0.402 | 0.047 | **0.438** | 0.028 | 0.394 | 0.051 |
| 6 | -0.118 | 0.573 | -0.392 | 0.053 | 0.432 | 0.031 | 0.402 | 0.047 | **0.433** | 0.031 |
| 7 | -0.203 | 0.330 | 0.400 | 0.048 | **0.417** | 0.038 | 0.400 | 0.048 | 0.402 | 0.047 |
| 8 | -0.552 | 0.004 | 0.348 | 0.089 | 0.363 | 0.074 | 0.359 | 0.078 | **0.364** | 0.074 |
| *HotelRec* | | | | | | | | | | |
| 1 | 0.097 | 0.645 | 0.121 | 0.565 | -0.508 | 0.009 | **0.356** | 0.081 | 0.285 | 0.168 |
| 2 | -0.242 | 0.244 | **0.443** | 0.027 | -0.337 | 0.100 | 0.196 | 0.347 | 0.382 | 0.059 |
| 3 | -0.112 | 0.596 | -0.392 | 0.053 | 0.318 | 0.121 | **0.419** | 0.037 | 0.355 | 0.082 |
| 4 | -0.362 | 0.076 | 0.276 | 0.181 | 0.301 | 0.144 | **0.349** | 0.087 | 0.315 | 0.126 |
| 5 | -0.045 | 0.829 | 0.292 | 0.156 | 0.225 | 0.279 | 0.323 | 0.115 | **0.364** | 0.074 |
| 6 | 0.222 | 0.285 | 0.298 | 0.148 | 0.306 | 0.137 | 0.294 | 0.154 | **0.312** | 0.130 |

表 2：在时间片 $t$ 上训练并在时间片 $t+1$上测试的各模型生成的品牌排名结果。我们报告相关系数 相关性及其对应的双侧 $p$值。

每个时间片的文档级情感标签进行训练。

- O-dBTM 是我们模型的一个变体，仅使用第一个时间片（记为0时间片）的监督式评论级情感标签进行训练。在随后的时间片中，它在无监督设置下进行训练。在这种情况下，除0时间片外，我们不再有金标准品牌排名。我们不直接计算 Spearman 秩相关系数，而是测量相邻时间片中品牌排名结果的差异，并用其设置式（8）中的权重 $\gamma^t$ 。

**参数设置** 除了单元组之外，我们还将高频二元组和三元组作为特征加入文档表示中[4] 。在实验中，我们使用当前时间片的数据训练模型，并使用下一时间片的全部数据测试模型性能。训练期间，我们在每个时间片中留出 10% 的数据作为验证集。超参数设置如下：批量大小为 256，最大训练步数为 50,000，主题数为 50[5]。值得注意的是，由于 TBIP 等静态模型并未显式建模主题动态而且 BTM 在不同时间片中提取的主题并未直接关联。

## 6 实验结果

本节将展示实验结果，并从品牌评分、主题连贯性/独特性度量以及生成的主题的定性评估等方面与基线模型进行比较。为确保公平比较，基线方法基于此前所有时间片进行训练，并在当前时间片上进行预测。

### 6.1 品牌评分

TBIP、BTM 和 dBTM 可以自动推断与各品牌相关的极性得分。对于 dJST，我们通过汇总与品牌相关的评论文档的标签分布，并按品牌相关评论总数进行归一化，得出品牌评分。品牌 $b$ 在时间片 $t$ 的文档级评分平均值用作品牌评分的真实值 $x_b^t$。我们从两个方面评估品牌评分：

**品牌排名结果** 我们在表 2中报告了通过斯皮尔曼相关系数衡量的品牌排名结果，该系数反映预测的品牌评分与真实值之间的相关性，同时还报告了相应的双侧 $p$斯皮尔曼相关系数的值

主题模型变体（如 dJST、TBIP 和 BTM）得出的品牌排名结果与真实排名结果呈正相关或负相关。我们可以看到，BTM 的相关性

[4]使用 NLTK 过滤掉了常见但信息量较低的 $n$元词组，例如“实际上买了”。[5]主题数根据第0个时间片的验证集经验性地设定。

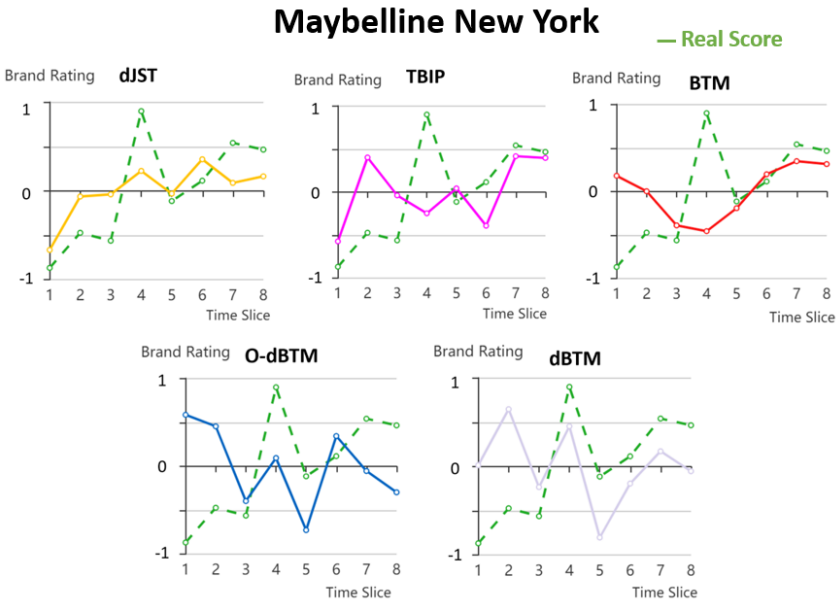



图 3：“Maybeline New York”的评分时间序列。评分分数在 $[-1, 1]$范围内进行了归一化，正值表示正面情感，负值表示负面情感。每个子图中的虚线表示实际评分分数。

在不同时间片之间交替呈现正相关和负相关。借助高斯状态空间模型，我们提出的模型 dBTM 及其变体 O-dBTM 得出了更稳定的结果。在 MakeupAlley-Beauty 上，dBTM 在 8 个时间片中的 4 个取得了最佳结果。有趣的是，在仅第一个时间片提供监督信息的情况下，O-dBTM 在 8 个时间片中的 3 个优于 BTM 等静态模型，展现了我们提出的架构在追踪品牌评分动态方面的有效性。在 HotelRec 上也可得出类似结论：O-dBTM 在 6 个时间片中的 5 个表现优于 BTM。在第 2 个时间片中，除 TBIP 外，O-dBTM 和 dBTM 均优于其他基线方法。

总而言之，在 dBTM 中，品牌评分被视为一个潜变量（即 $x_{b_d}$ 见公式 2），并直接从数据中推断。相反，dJST 等模型需要经过后处理，通过汇总文档级情感标签来计算品牌评分，因此不如 dBTM。这表明，与传统的动态主题模型相比，我们提出的 dBTM 在品牌排名方面具有优势。

**品牌评分时间序列** 品牌评分时间序列旨在比较各模型追踪品牌评分趋势的能力。为便于比较，我们对各模型生成的评分进行归一化，使图表仅反映评分随时间的波动。图 3 展示了各模型在不同时间片上，根据 MakeupAlley-Beauty 测试集生成的品牌“Maybeline New York”评分。可以观察到，TBIP 和 BTM 生成的品牌评分与实际评分的相关性不高。dJST 的评分趋势更为吻合，但其预测未能捕捉一些短期变化，例如时间片 7 的品牌评分峰值。相比之下，dBTM 正确预测了品牌评分的总体趋势。弱监督模型 O-dBTM 能够跟随总体趋势，但遗漏了一些短期变化，例如时间片 1 到 2 以及时间片 6 到 7 的上升趋势。

### 6.2 主题评估结果

我们使用每个主题的前10个词来计算基于上下文向量的主题连贯性得分（Röder等人，2015），以及主题唯一性（Nan等人，2019年），后者衡量主题之间的词语重叠比例。我们希望在主题连贯性和多样性之间取得平衡。因此，我们将主题连贯性和主题多样性结合起来，作为衡量主题总体质量的指标（Dieng等人，2020）。由于在我们的实验中，主题连贯性的结果为负值，即绝对值越小越好，因此我们将总体质量定义为

| 时间片 | dJST | | | TBIP | | | BTM | | | O-dBTM | | | dBTM | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | coh | uni | 质量 | coh | uni | 质量 | coh | uni | 质量 | coh | uni | 质量 | coh | uni | 质量 |
| *MakeupAlley-Beauty* | | | | | | | | | | | | | | | |
| 1 | -3.087 | 0.564 | 0.183 | -3.653 | 0.861 | 0.236 | -3.836 | 0.862 | 0.225 | -3.486 | 0.820 | **0.235** | -3.685 | 0.833 | 0.226 |
| 2 | -3.008 | 0.513 | 0.170 | -4.043 | 0.850 | 0.210 | -3.867 | 0.864 | 0.223 | -3.360 | 0.807 | **0.240** | -3.642 | 0.829 | 0.228 |
| 3 | -3.286 | 0.552 | 0.168 | -3.949 | 0.843 | 0.214 | -3.716 | 0.851 | 0.229 | -3.369 | 0.787 | **0.234** | -3.611 | 0.823 | 0.228 |
| 4 | -3.004 | 0.515 | 0.172 | -3.629 | 0.808 | 0.223 | -3.837 | 0.846 | 0.220 | -3.457 | 0.771 | 0.223 | -3.549 | 0.799 | **0.225** |
| 5 | -3.112 | 0.560 | 0.180 | -4.168 | 0.838 | 0.201 | -4.023 | 0.839 | 0.208 | -3.412 | 0.793 | **0.232** | -3.523 | 0.818 | **0.232** |
| 6 | -3.139 | 0.542 | 0.173 | -4.100 | 0.841 | 0.205 | -3.976 | 0.846 | 0.213 | -3.433 | 0.761 | 0.222 | -3.577 | 0.814 | **0.228** |
| 7 | -3.269 | 0.521 | 0.159 | -4.049 | 0.854 | 0.211 | -3.675 | 0.845 | 0.230 | -3.330 | 0.772 | **0.232** | -3.667 | 0.825 | 0.225 |
| 8 | -3.060 | 0.560 | 0.183 | -3.942 | 0.843 | 0.214 | -3.715 | 0.837 | 0.225 | -3.589 | 0.789 | 0.220 | -3.546 | 0.818 | **0.231** |
| 平均值 | -3.120 | 0.541 | 0.173 | -3.942 | 0.842 | 0.214 | -3.831 | 0.849 | 0.222 | -3.430 | 0.788 | **0.230** | -3.600 | 0.820 | 0.228 |
| *HotelRec* | | | | | | | | | | | | | | | |
| 1 | -3.749 | 0.615 | 0.164 | -4.024 | 0.767 | 0.191 | -3.935 | 0.851 | 0.216 | -4.051 | 0.812 | 0.201 | -3.716 | 0.818 | **0.220** |
| 2 | -4.020 | 0.633 | 0.158 | -3.577 | 0.753 | 0.211 | -3.960 | 0.813 | 0.205 | -3.851 | 0.803 | 0.209 | -3.696 | 0.809 | **0.219** |
| 3 | -3.667 | 0.593 | 0.162 | -3.905 | 0.817 | 0.209 | -4.078 | 0.844 | 0.207 | -3.861 | 0.819 | 0.212 | -3.854 | 0.820 | **0.213** |
| 4 | -4.008 | 0.644 | 0.161 | -3.747 | 0.808 | 0.216 | -3.946 | 0.859 | 0.218 | -3.637 | 0.814 | **0.224** | -3.681 | 0.794 | 0.216 |
| 5 | -3.751 | 0.691 | 0.184 | -4.057 | 0.800 | 0.197 | -3.953 | 0.823 | 0.208 | -3.705 | 0.804 | 0.217 | -3.547 | 0.817 | **0.230** |
| 6 | -3.916 | 0.697 | 0.178 | -3.770 | 0.810 | 0.215 | -4.061 | 0.855 | 0.210 | -3.510 | 0.800 | **0.228** | -3.705 | 0.821 | 0.222 |
| 平均值 | -3.852 | 0.645 | 0.168 | -3.847 | 0.793 | 0.206 | -3.989 | 0.841 | 0.211 | -3.769 | 0.809 | 0.215 | -3.700 | 0.813 | **0.220** |

表3：不同模型生成结果的主题连贯性（一致性）和唯一性（唯一性）指标。我们还将这两个分数结合起来，得出提取主题的总体质量。

主题的 $q = \frac{topic\ uniqueness}{|topic\ coherence|}$。表3 展示了主题评估结果。总体而言，主题连贯性与主题多样性之间存在权衡。平均来说，dJST的连贯性最高，但独特性分数最低；而TBIP的独特性较高，但连贯性数值最低。O-dBTM和dBTM均在连贯性与独特性之间取得了良好平衡，且总体质量优于其他模型。

### 6.3 不同时段的主题示例

我们展示了 dBTM 在不同时间片生成的一些代表性主题。为了便于检查，我们从语料库中为每个主题检索一个代表性句子。对于一个句子，我们通过对其组成单词的 GloVe嵌入取平均值，得到该句子的表示。对于一个主题，我们也对其相关的高频单词的 GloVe 嵌入取平均值，但以主题词概率作为权重。然后选取余弦相似度最高的句子。

生成的主题示例，涉及 眼部产品’ 和 ‘护肤’，来自 MakeupAlley——Beauty，见图 4。我们可以观察到，对于主题 ‘眼部产品’，关于 ‘眼部清洁剂’ 的负面评论中的高频词，从皮肤反应（例如 ‘刺痛’、‘灼痛’）逐渐转向清洁能力（例如 ‘去除’、‘残留物’）。我们还可以看到，正面主题逐渐变化：在时间片 1 中赞扬产品改善 ‘黑眼圈’ 的功效，到时间片 4 中转而关注眼影的品质，并在时间片 8 中关注眼部妆前乳。此外，我们观察到时间片 4 的正面主题中出现了品牌名称 *M.A.C.* ，这与其真实评分一致。对于主题 ‘护肤’，可以观察到，负面主题逐渐从对洁面产品的抱怨转向对防晒霜质地厚重的抱怨，而正面主题则始终更多地围绕对M.A.C粉底液遮盖力的赞扬。结果表明，dBTM 能够生成区分明确的带极性的主题，还能追踪主题随时间的变化。

图 5 展示了 HotelRec 中与“房间状况”和“食物”相关的生成主题示例。我们可以看到，对于“房间状况”这一主题，高频单词逐渐从描述清洁程度（例如，正面评论中的“干净”，以及负面评论中的“脏”）转向描述房间类型和大小（例如，正面评论中的“行政房”和“别墅”，以及负面评论中对房间“小”的担忧）。对于“食物”这一主题，人们关注的食物随时间从饮品（例如，“咖啡”、“茶”）转变为餐食（例如，“鸡蛋”、“吐司”）。负面评论主要关注食物质量问题（例如，“冷”），而正面评论则普遍赞扬食物和服务（例如，“喜欢”、“不错”）。

### 6.4 消融研究

我们通过消融研究来探究元学习组件（即公式 8 和 9 ）的作用，结果如表 4 所示。总体而言，我们可以观察到，移除元

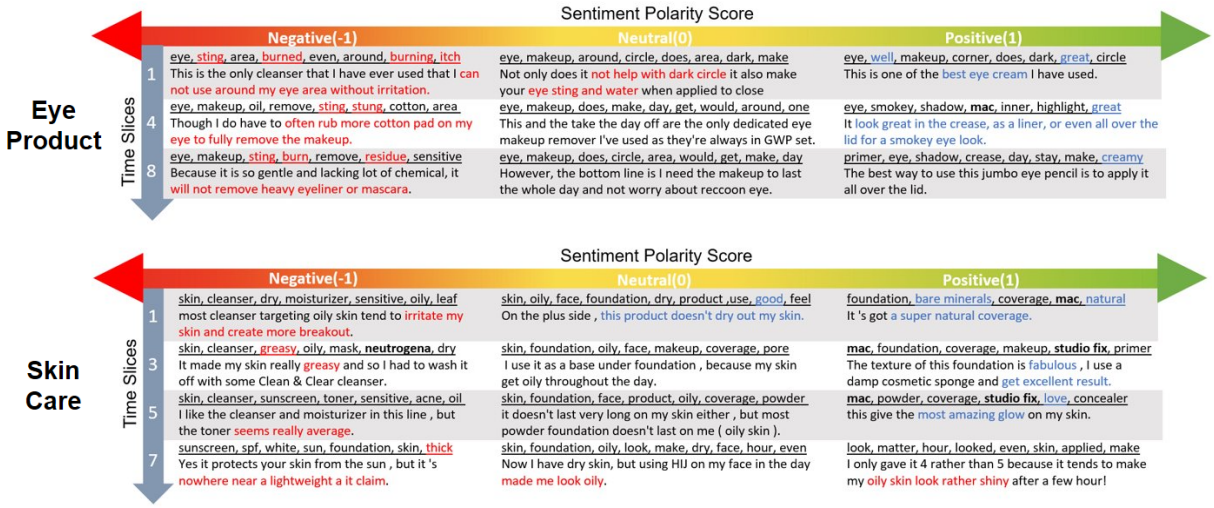



图 4：以列表形式展示的生成主题示例，列表中列出了 MakeupAlley 数据集中不同时间片内排名靠前的相关单词（带下划线）。为便于查看，我们还展示了每个主题下最具代表性的句子。每个时间片内的负面、中性和正面主题，是通过将品牌极性分数分别设为 -1、0 和 1 生成的。正面单词/短语以蓝色突出显示，负面单词/短语以红色突出显示，品牌名称则以粗体显示。

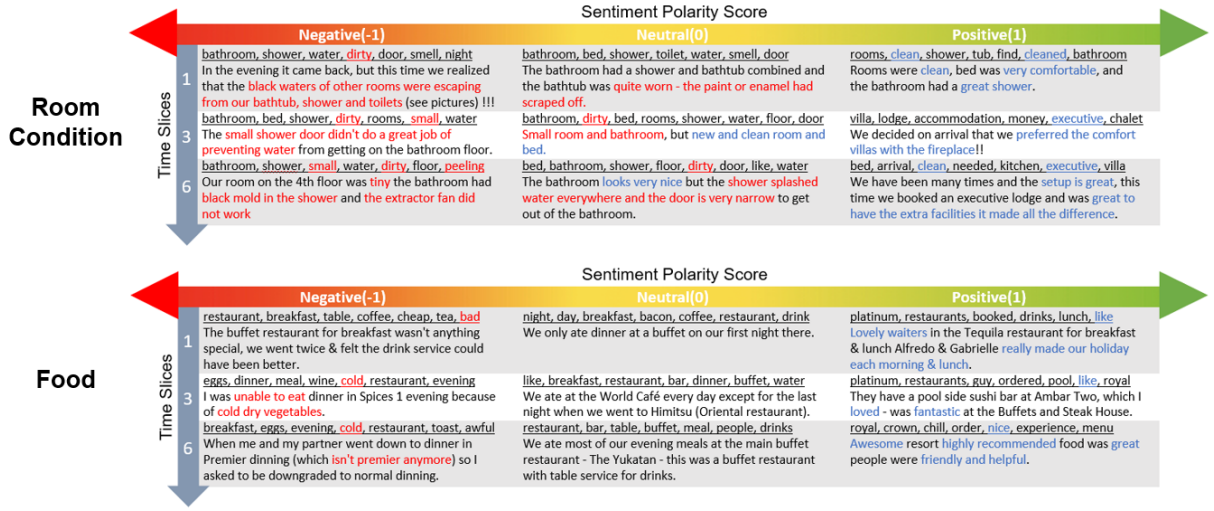



图 5：示　例：以列表形式展示的生成主题，列表中列出了不同时间片内排名靠前的相关单词（带下划线）　es 来自 H　otelRec 数据集。为便于查看，还展示了每个主题的代表性句子。

学习会导致 MakeupAlley-Beauty 数据集在所有时间片上的品牌排名相关性显著降低。在主题质量方面，我们观察到，未使用元学习时，连贯性分数有所降低，但独特性分数略有提高，因此大多数时间片的主题质量分数整体下降。

对于 HotelRec，我们可以看到，移除元学习也会导致品牌排名结果下降，但其影响小于 MakeupAlley-Beauty。就主题质量而言，我们观察到一致性有所提高，但唯一性变差，因此在没有元学习的情况下，主题质量略有下降，且在大多数时间片中都是如此。一个主要原因是，与化妆品品牌会随着时间推移推出新产品、进而导致评论中讨论的主题发生变化不同，酒店评论的主题-词分布在不同时间片之间变化不大。因此，是否使用元学习对结果的影响较小。

### 6.5 训练时间复杂度

所有实验均在一块配备 11GB 记忆库的 GeForce 1080 GPU 上运行。各模型在不同时间片上的训练时间如图所示——

| 时间片 | dBTM | | | | dBTM（无元学习） | | | |
|---|---|---|---|---|---|---|---|---|
| | cor | coh | uni | 质量 | cor | coh | uni | 质量 |
| *MakeupAlley-Beauty* | | | | | | | | |
| 1 | 0.402 | -3.685 | 0.833 | **0.226** | **0.435** | -3.972 | 0.861 | 0.217 |
| 2 | **0.438** | -3.642 | 0.829 | **0.228** | 0.189 | -3.704 | 0.840 | 0.227 |
| 3 | **0.523** | -3.611 | 0.823 | **0.228** | -0.162 | -3.873 | 0.828 | 0.214 |
| 4 | **0.453** | -3.549 | 0.799 | 0.225 | -0.042 | -3.745 | 0.849 | **0.227** |
| 5 | **0.394** | -3.523 | 0.818 | **0.232** | 0.086 | -3.990 | 0.832 | 0.209 |
| 6 | **0.433** | -3.577 | 0.814 | **0.228** | -0.029 | -3.958 | 0.856 | 0.216 |
| 7 | **0.402** | -3.667 | 0.825 | 0.225 | -0.042 | -3.587 | 0.842 | **0.235** |
| 8 | **0.364** | -3.546 | 0.818 | **0.231** | -0.125 | -3.920 | 0.847 | 0.216 |
| *HotelRec* | | | | | | | | |
| 1 | **0.285** | -3.716 | 0.818 | 0.220 | 0.222 | -3.559 | 0.791 | **0.222** |
| 2 | **0.382** | -3.696 | 0.809 | **0.219** | 0.210 | -3.796 | 0.801 | 0.211 |
| 3 | **0.355** | -3.854 | 0.820 | **0.213** | 0.285 | -3.763 | 0.790 | 0.210 |
| 4 | 0.315 | -3.681 | 0.794 | 0.216 | **0.408** | -3.597 | 0.817 | **0.227** |
| 5 | **0.364** | -3.547 | 0.817 | **0.230** | 0.362 | -3.657 | 0.793 | 0.217 |
| 6 | **0.312** | -3.705 | 0.821 | **0.222** | 0.262 | -3.694 | 0.809 | 0.219 |

表4：dBTM 在启用和未启用元学习组件时的结果。

6。可以观察到，随着时间片数量增加，dJST 和 BTM 的训练时间迅速增长。TBIP 和 dBTM 的训练时间则明显更短。TBIP 只是在每个时间片上独立执行泊松因子分解，无法跟踪随时间变化的主题/情感变化。相比之下，我们提出的 dBTM 和 O-dBTM 能够监测主题/情感的演变，而且训练时间甚至比 TBIP 更短。主要原因之一是，借助我们提出的元学习方法，dBTM 和 O-dBTM 能自动调整迭代次数，因此训练效率更高。

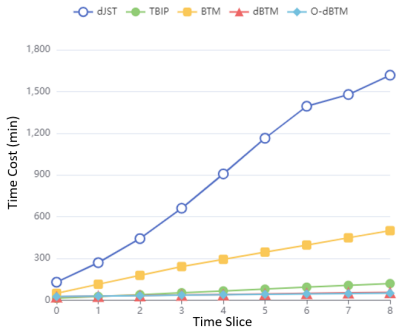



图6：模型在各时间片上的训练时间。

## 7 结论

我们提出了 dBTM，它能够自动检测和追踪与品牌相关的主题及情感分数。基于 MakeupAlley 和 HotelRec 评论开展的实验评估表明，dBTM 在品牌排名和动态主题提取方面优于以往模型。dBTM 的变体 O-dBTM 仅使用第一个时间片中的文档级情感标签进行训练，在品牌排名方面优于基线模型，并在主题质量评估中取得了最佳总体结果。这表明，所提出的架构能够有效地对品牌分数和主题随时间区间变化的情况进行建模。

我们的模型目前仅考虑评论评分，但现实应用还可能涉及其他因素（例如用户偏好）。一种可行的解决方案是探索同时对用户偏好进行建模，以提取个性化的品牌极性主题。

## 致谢

本研究部分获得英国工程与自然科学研究委员会的资助（项目编号：EP/T017112/1、EP/V048597/1、EP/X019063/1）。YH 获得由英国研究与创新署资助的图灵人工智能奖学金（项目编号：EP/V020579/1）。

## 参考文献

Diego Antognini 和 Boi Faltings。2020。Hotel-rec：一个新颖的大规模酒店推荐数据集。载于第*12*届语言资源与评估会议论文集，第4917–4923页。

David M Blei 和 John D Lafferty。2006。动态主题模型。载于第*23*届国际机器学习会议论文集，第113–120页。

David M. Blei、Andrew Y. Ng 和 Michael I. Jordan。2003。潜在狄利克雷分配。《机器学习研究杂志》，3:993–1022。

John Blitzer、Mark Dredze 和 Fernando Pereira。2007。人物传记、宝莱坞、音响和搅拌机：用于情感分类的领域自适应。载于第*45*届计算语言学协会年会论文集，第440–447页。

Samuel Brody 和 Noémie Elhadad。2010。用于在线评论的无监督方面情感模型。在 计算语言学协会北美分会*2010*年年会论文集，第804–812页。

Laurent Charlin、Rajesh Ranganath、James McIn- erney 和 David M Blei。2015。动态泊松分解。在 第*9*届*ACM*推荐系统会议论文集，第155–162页。

Peng Cui 和 Le Hu。2021。主题引导的抽象式多文档摘要。载于计算语言学协会论文集：*EMNLP 2021*，第 1463–1472 页。

Shumin Deng、Ningyu Zhang、Jiaojian Kang、Yichi Zhang、Wei Zhang 和 Huajun Chen。2020。采用动态记忆库原型网络进行小样本事件检测的元学习。载于第十三届*ACM*国际网络搜索与数据挖掘会议论文集，第151–159页。Adji B. Dieng、Francisco J. R. Ruiz 和 David M. Blei。2020。嵌入空间中的主题建模。载于计算语言学协会会刊论文集，第439–453页。Adji B Dieng、Francisco JR Ruiz 和 David M Blei。2019。动态嵌入主题模型。*arXiv*预印本 *arXiv:1907.05545*。Manqing Dong、Feng Yuan、Lina Yao、Xiwei Xu 和 Liming Zhu。2020。MAMO：面向冷启动推荐的记忆增强元优化。载于第*26*届*ACM SIGKDD*知识发现与数据挖掘会议论文集，第688–697页。Chelsea Finn、Pieter Abbeel 和 Sergey Levine。2017。用于深度网络快速适应的模型无关元学习。载于第*34*届国际机器学习会议论文集，第1126–1135页。Li Gao、Jia Wu、Chuan Zhou 和 Yue Hu。2017。结合用户画像演化的协同动态稀疏主题回归，用于物品推荐。载于第*31*届*AAAI*人工智能会议论文集，第1316–1322页。Ruiying Geng、Binhua Li、Yongbin Li、Jian Sun 和 Xiaodan Zhu。2020。用于小样本文本分类的动态记忆诱导网络。载于计算语言学协会第*58*届年会论文集，第1087–1094页。Chengyue Gong 和 Win-bin Huang。2017。深度动态泊松分解模型。载于神经信息处理系统国际会议第*31*届会议论文集，第1665–1673页。

Prem Gopalan、Jake Hofman 和 David Blei。2015。基于分层泊松因子分解的可扩展推荐。载于论文集 $f$ 第 *31* 届人工智能不确定性，第 326–335 页。

Yulan He、Chenghua Lin、Wei Gao 和 Kam-Fai Wong。2014。动态联合情感-主题模型。*ACM* 智能系统与技术汇刊，5:1–21。Seyed Abbas Hosseini、Ali Khodadadi、Keivan Alizadeh、Ali Arabzadeh、Mehrdad Farajtabar、Hongyuan Zha 和 Hamid R Rabiee。2018年。用于时间推荐的循环泊松分解。*IEEE*知识与数据工程汇刊，32(1):121–134。Muhammad Abdullah Jamal 和 Guo-Jun Qi。2019年。用于小样本学习的任务无关元学习。载于*IEEE*计算机视觉与模式识别会议论文集，第11719–11727页。Eric Jang、Shixiang Gu 和 Ben Poole。2017。使用 Gumbel-Softmax 的类别重参数化。载于国际学习表征会议。Chenghua Lin 和 Yulan He。2009。用于情感分析的联合情感/主题模型。载于*ACM*信息与知识管理会议论文集，第375–384页。Chenghua Lin、Yulan He、Richard Everson 和 Stefan Ruger。2012。基于文本的弱监督联合情感-主题检测。*IEEE*知识与数据工程汇刊，24(6):1134–1145。Yuanfu Lu、Yuan Fang 和 Chuan Shi。2020。异构信息网络上的元学习，用于冷启动推荐。载于第*26*届*ACM SIGKDD*知识发现与数据挖掘会议论文集，第1563–1573页。

Feng Nan、Ran Ding、Ramesh Nallapati、和 Bing Xiang。2019年。使用 Wasserstein 自编码器进行主题建模。载于第*57*届计算语言学协会年会论文集，第6345–6381页。

Krishna Prasad Neupane、Ervine Zheng 和 Qi Yu。2021年。Metaedl：面向不确定性感知冷启动推荐的元证据学习。载于第*21*届*IEEE*国际数据挖掘会议论文集，第1258–1263页。

Jianmo Ni、Jiacheng Li 和 Julian McAuley。2019年。利用远程标注评论和细粒度方面为推荐提供依据。载于自然语言处理实证方法会议论文集，第188–197页。

Michael Röder、Andreas Both 和 Alexander Hinneburg。2015。探索主题连贯性度量的空间。载于第*8*届*ACM*国际网络搜索与数据挖掘会议论文集，第399–408页。

Yuanfeng Song、Yongxin Tong、Siqi Bao、Di Jiang、Hua Wu 和 Raymond Chi-Wing Wong。2020。Topicocean：一种结合元学习、不断扩展的主题模型。载于第*20*届*IEEE*国际数据挖掘会议论文集，第1262–1267页。

Wu-Jiu Sun、Xiao Fan Liu 和 Fei Shen。2021年。学习动态用户交互，以预测在线论坛评论。在第*21*届*IEEE*国际数据挖掘会议论文集，第1342–1347页。

Flood Sung、Yongxin Yang、Li Zhang、Tao Xiang、Philip H. S. Torr 和 Timothy M. Hospedales。2018年。学习比较：用于小样本学习的关系网络。在*IEEE*计算机视觉与模式识别会议论文集，第1199–1208页。

Qiuling Suo、Jingyuan Chou、Weida Zhong 和 Aidong Zhang。2020。Tadanet：用于图增强元学习的任务自适应网络。在第*26*届*ACM SIGKDD*知识发现与数据挖掘会议论文集，第1789–1799页。

Keyon Vafa、Suresh Naidu 和 David M. Blei。2020。基于文本的理想点。载于会议论文集 计算语言学协会会议，第 5345–5357 页。

Chong Wang、DavidBlei 和 David Heckerman。2008年。连续时间动态主题模型。在第*24*届人工智能不确定性会议论文集，第579–586页。

Tianyi Yang、Cuiyun Gao、Jingya Zang、David Lo 和 Michael R. Lyu。2021。TOUR：用于辅助应用发布的用户评论动态主题与情感分析。载于*2021*年网络会议论文集，第708–712页。

Runsheng Yu、Yu Gong、Xu He、Yu Zhu、Qing- wen Liu、Wenwu Ou 和 Bo An。2021。用于冷启动用户偏好预测的个性化自适应元学习。载于第*35*届人工智能促进协会会议论文集，第10772–10780页。

Hao Zhang、Gunhee Kim 和 Eric P. Xing。2015。利用在线文本和图像数据进行动态主题建模，以监测市场竞争。载于第*21*届国际知识发现与数据挖掘会议论文集，第1425–1434页。

Runcong Zhao、Lin Gui、Gabriele Pergola 和 Yulan He。2021。用于衡量用户评论中品牌情感的泊松分解模型对抗学习。载于第*16*届欧洲计算语言学协会分会会议论文集，第2341–2351页。

Wenbo Zheng、Lan Yan、Chao Gou 和 Fei- Yue Wang。2021 年。知识即力量：面向艺术领域视觉推理的层次化知识嵌入元学习。在《第 *27* 届 *ACM SIGKDD* 知识发现与数据挖掘会议论文集》中，第 2360–2368 页。

Kun Zhou、Yuanhang Zhou、Wayne Xin Zhao、Xi- aoke Wang 和 Ji-Rong Wen：2020 年。迈向主题引导的对话式推荐系统。在《第 *28* 届国际计算语言学会议论文集》中，第 4128–4139 页。