ORIGINAL PAPER

A multifacet hierarchical sentiment-topic model with application to multi-brand online review analysis

Qiao Liang[1,2] · Xinwei Deng[3]

Abstract

Multi-brand analysis based on review comments and ratings is a commonly used strategy to compare different brands in marketing. It can help consumers make more informed decisions and help marketers understand their brand's position in the market. In this work, we propose a multifacet hierarchical sentiment-topic model (MH-STM) to detect brand-associated sentiment polarities towards multiple comparative aspects from online customer reviews. The proposed method is built on a unified generative framework that explains review words with a hierarchical brand-associated topic model and the overall polarity score with a regression model on the empirical topic distribution. Moreover, a novel hierarchical Pólya urn (HPU) scheme is proposed to enhance the topic-word association among topic hierarchy, such that the general topics shared by all brands are separated effectively from the unique topics specific to individual brands. The performance of the proposed method is evaluated on both synthetic data and two real-world review corpora. Experimental studies demonstrate that the proposed method can be effective in detecting reasonable topic hierarchy and deriving accurate brand-associated rankings on multi-aspects.

1 Introduction

As online activities and intelligence advance, an increasing amount of online reviews provide a practical way to understand customers' feedback to multiple aspects of online products and services. Typically, a customer tends to express positive or negative opinions on different aspects of a product type or brand. The multifacet information in reviews enables a data-driven comparative analysis of different online entities, including but not limited to different products (He et al. 2023), brands (Alzate et al. 2022), platforms (Zhang et al. 2019), and even cultures (Brand et al. 2022). The primary interest of this study is to conduct a thorough modeling and analysis on the customer reviews for automatic multi-brand comparison, namely, to compare multifacet features and advantages of different brands in marketing. Note that the proposed method can be extended to other review-based multifacet comparison across different entities of interest.

Overall, the existing researches on brand aspect detection from online reviews are mainly built on classical topic models, where all topics are organized in a flat structure that assumes equal information granularities among topics. However, in practice, the aspects embedded in reviews tend to demonstrate different granularities, which naturally form a tree structure. For example, Fig. 1 illustrates online customer reviews on several laptop brands. Each review is tagged with a certain brand and composed of a plain text that describes customer's opinions on multi-aspects of the bought laptop as well as an overall rating that indicates the general sentiment polarity. The multi-aspect information in brand-associated reviews can be captured by a tree structure of product aspect nodes shared by all brands, where different brands present different distributions on the aspect nodes, and each aspect node is linked to multiple brand-specific sentiment polarities that summarize the brand competitive performance on that aspect and influence the overall ratings observed in brand reviews.

✉ Xinwei Deng
  xdeng@vt.edu

1 Joint Laboratory of Data Science and Business Intelligence,
   Southwestern University of Finance and Economics,
   Chengdu, China

2 School of Statistics, Southwestern University of Finance and
   Economics, Chengdu, China

3 Department of Statistics, Virginia Tech, Blacksburg, VA, USA

Fig. 1 Illustration of laptop reviews with respect to various brands. Each review is composed of a plain text and an overall rating. Different brands share multiple comparative aspects represented by a tree structure of product aspect nodes with respective distributions on that, where each aspect node is linked to multiple brand-specific sentiment polarities that influence the overall ratings in brand reviews
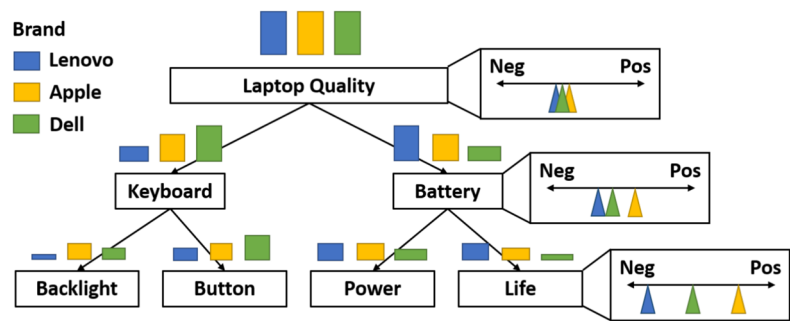
For the purpose of multi-brand comparison, a hierarchical structure of comparative aspects rather than its flat counterpart is preferred, on account of enhanced brand differentiation with deepened levels in the hierarchy. A flat structure assumes the global background and local ratable aspects in reviews are mixed up, such that the local aspects are smoothed by their shared background (Titov and McDonald 2008). In contrast, a hierarchical structure presents a varying degree of smoothness among multi-level aspects, where upper aspects of hierarchy summarize product background and general property shared among all brands, and deeper aspects capture fine-grained characteristics specific to individual brands. Therefore, the brand-associated strengths and weaknesses can be degeneralized from their shared background by digging into the hierarchy, which enables a salient comparison.

We propose a multifacet hierarchical sentiment-topic model (MH-STM) for automatic multi-aspect brand comparison based on the brand-associated review corpora including both review texts and overall ratings, as illustrated in Fig. 1. The proposed MH-STM employs a unified framework that combines the extraction of hierarchical aspects from reviews and acquiring of brand-associated sentiment scores specific to multi-aspects. For deriving the hierarchical comparative aspects among brands, the MH-STM borrows the idea of hierarchical topic models (HTMs; Liu et al. 2016) that are natural extension of flat topic models. In the MH-STM, we adopt the hierarchical topic assumptions of hierarchical latent Dirichlet allocation (hLDA; Blei et al. 2010) for the tree-structured product aspects as in Fig. 1, and brand-associated reviews are distributed respectively over the shared topic tree.

However, a challenge of applying hLDA and its variants to our problem directly is that the detection of topic hierarchy in these models relies only on word co-occurrence rules

without considering the general-to-specific characteristics of words explicitly. Some researches (Kang et al. 2012; Xu et al. 2018) have shown that hLDA could produce unreasonable hierarchy. For example, some semantically general words are assigned with the topics of lower levels if those words do not occur frequently enough. In contrast, some semantically specific but frequently occurring words are wrongly assigned to topics of upper levels. This mismatch brings more uncertainties in differentiating the brand-associated properties from their shared background. To address this challenge, we develop a hierarchical Pólya urn (HPU) scheme that can be properly applied to hierarchical topic models for enhancing the topic level assignments of words based on their general-to-specific semantics, namely, a general word is more bursty around the root topic, and a specific word is more bursty around a leaf topic.

Overall, the major contributions of this work are two-fold. First, we propose a multifacet hierarchical sentiment-topic model that detects brand-associated comparative aspects in hierarchy and customers' aspect-specific polarities towards different brands jointly from online reviews. Compared to other multi-brand models, it is constructed on a hierarchical structure of multiple product aspects for enhanced brand differentiation with deepened levels in hierarchy. Second, we introduce a novel hierarchical Pólya urn (HPU) scheme in the inference of topic hierarchy among multiple brands. By incorporating different burstiness of review words in topic hierarchy according to their general-to-specific characteristics, it improves the separation of general aspects shared by all brands from the unique aspects specific to individual brands, resulting in a more accurate multi-brand comparison. The proposed model produces a ranked list of brands on various customer-concerned aspects that enables a straightforward brand comparison.

The outline of this article is organized as follows. Section 2 provides a literature review in the related field. Section 3 details the proposed MH-STM with inference via a hierarchical Pólya urn scheme. Section 4 elaborates performance of the proposed method through simulation. Section 5 evaluates model performance using two real-world online review corpora. We conclude this work with discussions in Sect. 6.

2 Literature review

2.1 Review-based brand comparison

Review-based comparison of multiple brands can help consumers make more informed purchasing decisions and help marketers understand their brand's position in the market (Colladon 2018; Mitra and Jenamani 2020). Recent researches mainly aim to derive brand-associated polarities from user-generated reviews. Several works considered a two-stage procedure (Zhang et al. 2015; Barry et al. 2018; Sajid et al. 2022), which first detects brand-associated evaluation aspects from reviews by implementing probabilistic topic models (Blei et al. 2003; Vayansky and Kumar 2020), and subsequently tracks customers' orientations to the extracted aspects or topics of different brands via sentiment analysis. In the first stage, the hierarchical topic models (HTMs; Liu et al. 2016) can be applied to derive the hierarchical comparative aspects among brands. Commonly used HTMs include two mainstream structures: Hierarchical latent Dirichlet allocation (hLDA; Blei et al. 2010) and hierarchical Pachinko allocation model (hPAM; Mimno et al. 2007). Specifically, hLDA defines a tree-structure hierarchy where each child topic node has only one parent, and hPAM is built on a directed acyclic multilevel graph where a lower-level topic correlates with multiple upper-level topics. An obvious limitation of the two-stage method is that the topic extraction is unsupervised and separated from the subsequent sentiment analysis, leading to reduced predictive power for customers' brand-associated polarities (Blei and McAuliffe 2007).

Joint sentiment-topic modeling of review texts and ratings provides a powerful solution to the problem of two-stage methods (Liang et al. 2023). There are also several works using a joint approach that derives brand-associated comparative aspects and customer polarities towards the brands simultaneously. For example, the text-based ideal point (TBIP) model (Vafa et al. 2020) is constructed based on an unsupervised Poisson factorisation to infer brand-level polarity scores that influence the word distributions of polarity-bearing topics shared among brands. Moreover, the brand topic model (BTM) by Zhao et al. (2021) extended the TBIP model by incorporating supervision from the document-level sentiment labels, leading to improved performance in brand ranking. And Zhao et al. (2023) further proposed a dynamic

variant of BTM for tracking the latent brand polarity scores over time. However, these joint models often focus on the overall brand polarity comparison. They do not provide brand rankings under multiple aspects directly, which is less informative for the decision making of brand market strategies.

2.2 Pólya urn scheme in topic modeling

The simple Pólya urn scheme (SPU) can be naturally combined with topic modeling for enhancing the topic-word association (Mahmoud 2008). The SPU scheme presents a self-reinforcing property known as "the rich get richer", which helps to capture the word burstiness in a document, i.e., if a word appears once, it is more likely to appear again (Doyle and Elkan 2009). This property enables SPU to show superior performance in the topic modeling of text documents (Madsen et al. 2005). Recently, some enhanced Pólya urn schemes have been proposed for improving the topic-word associations in topic inference, such as generalized Pólya urn (GPU; Li et al. 2016) and weighted Pólya urn (WPU; Wang et al. 2020). But these methods are mainly designed for non-hierarchical topic models.

For adaption to hierarchical topic relations, Xu et al. (2018) combined GPU with hierarchical topic sampling to produce a topic hierarchy with improved coherence and reasonable structure. However, it relies on the prior knowledge extracted from multiple domains' corpora. Liang (2024) developed an enhanced Pólya urn scheme by adapting the sampling weights of words to their general-to-specific characteristics in hierarchy without requiring any external knowledge. However, it still adopts equal burstiness of words across hierarchy following the assumption of SPU. In comparison, our proposed HPU scheme considers different burstiness of words in topic hierarchy relying only on their general-to-specific characteristics, namely, a general word is more bursty around the root topic, and a specific word is more bursty around a leaf topic.

3 A multifacet hierarchical sentiment-topic model

For extracting comparative aspects among different brands as well as brand-associated sentiment scores for multi-aspects, Sect. 3.1 introduces a multifacet hierarchical sentiment-topic model (MH-STM), and Sect. 3.2 combines it with a novel hierarchical Pólya urn (HPU) scheme for enhancing the topic-word association in topic hierarchy. Finally, Sect. 3.3 demonstrates a stochastic EM algorithm for model inference.

3.1 Model formulation

Consider the data consisting of product reviews each indexed by $d \in \{1, 2, \cdots, D_b\}$ belonging to each of competitive brands $b \in \{1, 2, \cdots, B\}$ in a market. Without loss of generality, each review d contains a normalized overall polarity score $y_d \in [0, 1]$ and S_d review sentences each indexed by $s \in \{1, 2, \cdots, S_d\}$. Suppose a review sentence s in review d contains $N_{d,s}$ words denoted as $\{w_{d,s,n}\}_{n=1}^{N_{d,s}}$, thus having the total number of words in review d denoted as $N_d = \sum_{s=1}^{S_d} N_{d,s}$, and each word in the observed sentence is assumed to be from the vocabulary indexed by $v \in \{1, 2, \cdots, V\}$.

The MH-STM employs a unified frame of probabilistic generative process that explains the observed words and overall polarity scores in brand-associated reviews jointly. As illustrated in Fig. 1, the reviews are explained by mixtures of latent topics that are organized in a tree with a pre-defined depth L and unbounded branches. Each topic node k is represented by a V-dimension multinomial word distribution $\boldsymbol{\phi}_k$ drawn from a symmetric Dirichlet prior: $\boldsymbol{\phi}_k \sim \mathrm{Dir}(\boldsymbol{\eta})$, where $\boldsymbol{\eta} = (\eta, \cdots, \eta)^T$. Overall, a topic located deeper in the tree represents a more fine-grained aspect that customers are concerned with.

Given a topic node k in the tree, all direct child nodes of k compose its child set $\Lambda(k)$ that corresponds to each subdivision of the topic, e.g., $\Lambda(Keyboard) = \{Backlight, Button\}$ in Fig. 1. And k's ancestor set $V(k)$ is composed of all other nodes on the way from the node k to the root, e.g., $V(Backlight) = \{Keyboard, LaptopQuality\}$ in Fig. 1. A path is defined as a sequence of distinct topic nodes connected by parent–child edges in the tree: $\boldsymbol{c} = (c_1, c_2, \cdots, c_L)^T$, $c_l \in \Lambda(c_{l-1})$, $\forall l \in \{2, \cdots, L\}$, where c_1 is the root node, and c_L is a leaf node. Each path denotes a coherent semantic theme of reviews with different information granularities captured by each level's node in the path, e.g., $\boldsymbol{c} = (LaptopQuality, Keyboard, Button)^T$ in Fig. 1.

To organize latent topics in an unbounded tree, we employ the nested Chinese restaurant process (nCRP; Blei et al. 2003) that is widely used in hierarchical topic modeling to place a nonparametric prior partitions on possible topic trees. It is assumed that different brands share the same hierarchical structure of topic nodes but with respective proportions over the nodes, which means each brand's reviews are partitioned by their respective nCRPs. The partitioning process of nCRP specific to each brand is achieved by recursively traversing the entire tree from a specified root node. For example, let c_1 be the root topic and selected as the beginning node, once given the previous $i - 1$ document assignments of brand b, the ith subsequent document of brand b traverses each level of tree by nCRP with parameter γ, resulting a path of topics $\boldsymbol{c} = (c_1, \cdots, c_L)^T \sim \mathrm{nCRP}(\gamma, b)$. Specifically, for each

level $l \in \{2, \cdots, L\}$, it draws a direct child topic node under c_{l-1} with the following distribution:

$$
p(c_l = k | c_{l-1})
$$
$$
= \begin{cases} \dfrac{M_{b,k}+1}{\gamma+M_{b,c_{l-1}}+|\Lambda(c_{l-1})|} & \text{if } k \text{ is an existing child node of } c_{l-1}, \\[2ex] \dfrac{\gamma}{\gamma+M_{b,c_{l-1}}+|\Lambda(c_{l-1})|} & \text{if } k \text{ is a new child node,} \end{cases}
$$
$$(1)$$

where $M_{b,k}$ is the number of previous document visits at topic node k in the corpus of brand b, specially having $M_{b,c_1} = i - 1$ in the ith trial. The hyperparameter γ governs the likelihood of creating a new child node in topic hierarchy. Laplace smoothing is additionally incorporated to prevent zero probability values, where $|\Lambda(c_{l-1})|$ represents the number of existing child nodes descending from c_{l-1}. The nCRP allows continuous change and branching of topic groups in a hierarchy, where the total number of topics is decided by data.

Each review is allowed to have multiple paths through the tree as there could be multiple semantic themes in the same review. Specifically, each review sentence is assigned to a single path with L-level topics based on the brand-specific nCRP prior. Once a path $\boldsymbol{c}_{d,s}$ is selected for a review sentence indexed by (d, s), the generative process of words in the sentence reduces to a flat LDA model (Blei et al. 2003). In this case, sentence words $\{w_{d,s,n}\}_{n=1}^{N_{d,s}}$ are respectively assigned to different levels $\{l_{d,s,n}\}_{n=1}^{N_{d,s}}$ along the selected path based on an L-dimension multinomial distribution $\mathrm{Multi}(\boldsymbol{\theta}_{d,s})$. $\boldsymbol{\theta}_{d,s}$ is drawn from a symmetric Dirichlet prior on level distributions: $\boldsymbol{\theta}_{d,s} \sim \mathrm{Dir}(\boldsymbol{\alpha})$, where $\boldsymbol{\alpha} = (\alpha, \cdots, \alpha)^T$. The review sentence path $\boldsymbol{c}_{d,s}$ and each word's level assignment $l_{d,s,n}$ jointly specifies the corresponding topic assignment of word: $k_{d,s,n} \equiv c_{d,s,l_{d,s,n}}$. After that each sentence word is drawn from the word distribution $\boldsymbol{\phi}_{k_{d,s,n}}$ independently conditioned on its topic assignment.

We consider the polarity score y_d accompanying each review d as a response variable in the regression, which is referred to as polarity response hereafter. It relies on brand-specific regression parameters $\boldsymbol{\beta}^b$ corresponding to the extracted topics $\{k\}$, and uses the empirical distribution of review d over topics as regressors: $x_k^d = \frac{1}{N_d} \sum_{s=1}^{S_d} \sum_{n=1}^{N_{d,s}} \mathbb{I}(k_{d,s,n} = k)$. The fitting of polarity response y_d, in turn, helps to inform the recovery of topics by incorporating their association with extremes on these topics. Specifically, the brand-specific topic regression parameters $\boldsymbol{\beta}^b$ indicate customers' sentiment polarities towards various topics or aspects of each brand, which can be used for ranking brand performances in multiple aspects.

A generative procedure summarizing the proposed MH-STM is illustrated in Algorithm 1. Figure 2 shows a plate notation of the generative procedure accordingly. It is

assumed that only review words $\{w_{d,s,n}\}$ and overall polarity responses $\{y_d\}$ are observed, and they are both affected by an infinite number of tree-structured latent topics that are positioned by corresponding assignments of path $c_{d,s} \sim$ nCRP(γ, b) and level $l_{d,s,n} \sim$ Multi$(\theta_{d,s})$ in the tree. The observed review words are generated by multinomial word distributions ϕ_k conditioned on their topic assignments, and review's polarity response variables are explained by the topic regression coefficients β_k^b specific to each brand.

Algorithm 1: Generative procedure of reviews in MH-STM

- For each topic node k in the tree:

  1. Draw a multinomial word probability $\phi_k \sim$ Dir(η) specific to the topic k.
  2. For each brand $b \in \{1, \cdots, B\}$:
     (a) Draw a brand-specific regression parameter $\beta_k^b \sim \mathcal{N}(\mu, \sigma^2)$.

- For each review $d \in \{1, \cdots, D_b\}$ of brand b:

  1. For each sentence $s \in \{1, \cdots, S_d\}$ of review d:
     (a) Draw a path $c_{d,s} \sim$ nCRP(γ, b), where $c_{d,s} = (c_{d,s,1}, \cdots, c_{d,s,L})^T$ includes each level's topic along the path.
     (b) Draw an L-dimensional topic level multinomial distribution $\theta_{d,s} \sim$ Dir(α).
     (c) For each word token $n \in \{1, \cdots, N_{d,s}\}$ in the review-sentence pair (d, s):
        (i) Draw topic level $l_{d,s,n} \sim$ Multi$(\theta_{d,s})$, which specifies the corresponding topic $k_{d,s,n} \equiv c_{d,s,l_{d,s,n}}$ along the selected path.
        (ii) Draw word $w_{d,s,n} \sim$ Multi$(\phi_{k_{d,s,n}})$.
  2. Draw polarity response $y_d \sim \mathcal{N}((x^d)^T \beta^b, \rho^2)$, where $x_k^d = \frac{1}{N_d} \sum_{s=1}^{S_d} \sum_{n=1}^{N_{d,s}} \mathbb{I}(k_{d,s,n} = k)$.

3.2 Enhancing topic-word association via Pólya urn scheme

As illustrated in Fig. 2, the multinomial word distribution ϕ_k conditioned on topic k is usually drawn from a symmetric Dirichlet prior: Dir(η), where $\eta = (\eta, \ldots, \eta)^T$. In practice, however, it is common to deal directly with the "collapsed" distribution by integrating over the topic-word multinomial ϕ_k (Mimno et al. 2011):
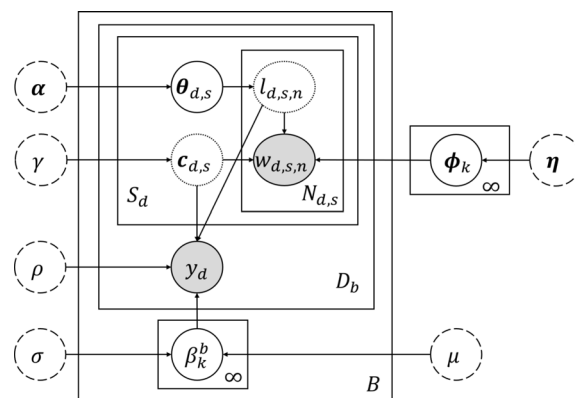

Fig. 2 Graphical representation of MH-STM. The shaded circles represent observable variables, dashed circles represent hyperparameters, dotted circles represent latent variables, and the rest solid circles represent model parameters

$$p(w \mid k, \eta, \mathcal{W}) = \frac{N_{k,w} + \eta}{\sum_{v=1}^{V} (N_{k,v} + \eta)}, \tag{2}$$

which is a function of the hyperparameter η and the number of each word assigned to that topic, i.e., $N_{k,v}$. This collapsed distribution is known as Dirichlet compound multinomial (DCM) or multivariate Pólya distribution (Doyle and Elkan 2009). It breaks the assumption of conditional independence between word tokens given topics, but has been proven to show substantially better performance than traditional multinomials in the modeling of text documents (Madsen et al. 2005). In the following parts, we start with introducing the existing simple Pólya urn scheme that constructs DCM in flat topic models, and then we propose a novel hierarchical Pólya urn scheme that adapts to topic hierarchy.

3.2.1 Simple Pólya urn scheme

The process for generating a sequence of words following the DCM distribution is known as the simple Pólya urn (SPU) scheme (Mahmoud 2008), which employs an analogy of urns with colored balls to interpret the topic-word distribution. Let $u_k^0 = \eta$ be an urn initially containing an equal number of η standard balls in one of V distinct colors, and $\mathbf{U}^0 = [u_1^0, u_2^0, \cdots, u_K^0]^T$ represent K urns in the same initial states. In this context, the urns and different colored balls represent K topics and V distinct words, respectively. Starting at time $i = 0$, each time a colored ball is randomly drawn from an urn, it is returned to the urn along with another ball of the same color. If $\{(k_i, v_i)\}$ is a sequence of balls each in color of $v_i \in \{1, \cdots, V\}$ and drawn from the urn $k_i \in \{1, \cdots, K\}$, then the state of urns can be describe as a recurrent process as:

$$\mathbf{U}^{i+1} = \mathbf{U}^i + \mathbf{E}^i, \tag{3}$$

where $\mathbf{E}^i$ is a $K \times V$ matrix with 1 in dimension of (k_i, v_i) and 0 elsewhere. This process is equivalent to the marginal distribution in Eq. (2) with a Dirichlet prior and a multinomial likelihood, where the proportion of colored balls in an urn reflects the word distribution under a topic. As a sampled color is more likely to be sampled from the same urn again in the next round, the SPU scheme enables to capture the word burstiness in a document, i.e., if a word appears once, it is more likely to appear again (Doyle and Elkan 2009).

3.2.2 Hierarchical Pólya urn scheme

The SPU scheme is designed for flat topic structure, where all word tokens are equally bursty under each of topics while ignoring the hierarchical relationship among them. Liang (2024) has developed an enhanced Pólya urn scheme that adapts the sampling weights of balls to their general-to-specific characteristics in hierarchy. However, it still adopts the assumption of equal burstiness across hierarchy, i.e., it still follows the stochastic process in Eq. (3). Therefore, we introduce a hierarchical Pólya urn (HPU) scheme in this part for incorporating adaptive word burstiness in hierarchical topic modeling.

Formally, given a sequence of balls $\{(k_i, v_i)\}$, we propose the HPU scheme with urns' state process as

$$\mathbf{U}^{i+1} = \mathbf{U}^i + \mathbf{E}^i + \tilde{\mathbf{E}}^i \odot \mathbf{A}^i, \tag{4}$$

where $\tilde{\mathbf{E}}^i$ is a $K \times V$ matrix with 1 in each dimension of $\{(k, v_i) : k \in \mathrm{V}(k_i)\}$ and 0 elsewhere, $\mathbf{A}^i$ is a $K \times V$ addition weight matrix that stores the weights of additional balls put to all ancestor urns in $\mathrm{V}(k_i)$, and $\odot$ performs Hadamard product of two matrices. According to the process specified by Eq. (4), when a colored ball is drawn from a certain urn, the selected ball and a new ball in the same color are put back to the urn, and additionally some weighted balls in the same color will be placed to the urn's all ancestor urns. By applying this HPU scheme in a hierarchical topic model, the consistency between parent and child topics is enhanced such that a child topic's frequent word is also likely to appear in its parent.

The additionally placed balls are weighted by their general-to-specific characteristics along the hierarchy. Specifically, we define the addition weight matrix $\mathbf{A}^i = [a^i_{kv}]_{K \times V}$ as

$$
\begin{aligned}
a^i_{kv} &= \min \left(\frac{\mathrm{Ent}(\Lambda(k) \mid v)}{\mathrm{Ent}(\Lambda(k))}, 1 \right), \\
\mathrm{Ent}(\Lambda(k)) &= - \sum_{k' \in \Lambda(k)} p(k') \log p(k'), \\
\mathrm{Ent}(\Lambda(k) \mid v) &= - \sum_{k' \in \Lambda(k)} p(k' \mid v) \log p(k' \mid v),
\end{aligned}
\tag{5}
$$

where we use the index variables k and v for denoting the corresponding indexed urn (topic) and ball (word) respectively. $\mathrm{Ent}(\Lambda(k) \mid v)$ and $\mathrm{Ent}(\Lambda(k))$ represent Shannon's entropy on the set of k's direct children $\Lambda(k)$ with and without the specified word v respectively, which measures the uncertainties of distinguishing one child topic from another in $\Lambda(k)$.

The addition weights specified above range in $[0, 1]$, and their values rely only on intermediate topic sampling results with no requirement for external knowledge. When the information gain $\mathrm{Ent}(\Lambda(k)) - \mathrm{Ent}(\Lambda(k) \mid v) \leq 0$, it triggers the maximum addition weight $a_{kv} = 1$. It is desired that a word should be more weighted in the parent node k if it is shared by each child node in $\Lambda(k)$. In contrast, it triggers the minimum addition weight $a_{kv} = 0$ when $\mathrm{Ent}(\Lambda(k) \mid v) = 0$, which means a word should be less weighted in the parent node k if it is associated mainly with a specific child topic in $\Lambda(k)$. Under this rule, we achieve a desired concept tree that places semantic notions of generality to the root and the counterpart of specificity to the leaves (Tumpa and Ali 2018).

A sequence of words under topics generated by the HPU process defines a set of topic-specific word distributions as:

$$p(w \mid k, \eta, \mathcal{W}) = \frac{W_{k,w} + \eta}{\sum_{v=1}^{V} \left(W_{k,v} + \eta \right)}, \tag{6}$$

where $W_{k,v}$ denotes the cumulative weights of each word v under the topic k. It can be seen as a weighted version of the classical DCM in Eq. (2).

Figure 3 illustrates some sampling instances of the proposed HPU scheme in comparison with SPU, which are both applied to a two-level hierarchy. Compared to SPU, HPU adds additional words to the parent topics with weights that are consistent with the general-to-specific characteristics of different words. For example, a general word v shared by different child topics of k leads to a lower information gain and a higher addition weight of a_{kv} according to Eq. (5), which makes it more bursty to the parent topic k (e.g., the left case in Fig. 3b). On the contrary, a word specific to a certain child topic $k' \in \Lambda(k)$ brings higher information gain in differentiating the branches under parent topic k, leading to a lower value of a_{kv} that prevents v to be assigned to the parent topic k (e.g., the right case in Fig. 3b).

Overall, SPU assumes equal word burstiness under various topics, while HPU scheme produces word-specific burstiness that adapts to topic hierarchy, where the general words are more likely to repeat within a document around the root topic. Therefore, the root node tends to place probability mass on general words, and the leaf nodes place on specific words, such that the neutral topics at the top that are general to all brands are distinct from the polarised topics at the bottom that are specific to different brands, leading to accurate brand comparison in the shared topic hierarchy.

Fig. 3 Comparison of simple and hierarchical Pólya urn schemes in hierarchical topic modeling. Urns represent topics and balls with different numbers/colors represent different words. Grey balls (No.1–2) represent general words that are shared among child topics, while balls in color of blue (No.3–5), green (No.6–8), and orange (No.9–11) are specialized words for each child topic, respectively

(a) Simple Pólya Urn (SPU)

(b) Hierarchical Pólya Urn (HPU)
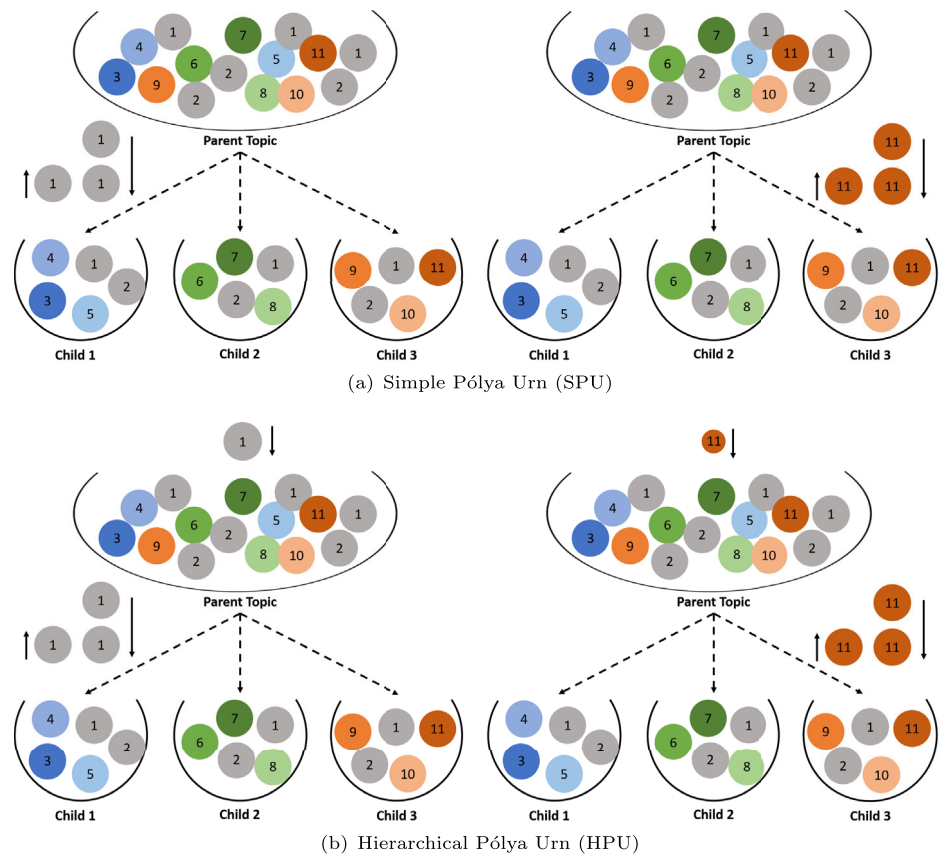
3.3 Model inference

The proposed model is parameterized by $\Theta = (\boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{\beta})$. Overall, we employ a probabilistic frame of stochastic EM that is commonly adopted for the inference of supervised topic models (Boyd-Graber and Resnik 2010; Nguyen et al. 2013). The stochastic EM algorithm alternates between a Gibbs sampling E-step and an optimization M-step. Specifically, in the E-step, we conduct a collapsed Gibbs sampling that employs the proposed HPU scheme to construct a Markov chain over two groups of latent variables, i.e., the path assignment $\boldsymbol{c}$ of each review sentence and the topic level assignment l of each word. During the sampling process, one latent group in $\{\boldsymbol{c}, l\}$ is sampled conditioned on the other by a set of conditional independence assumptions illustrated in Fig. 2, and all other variables such as $\boldsymbol{\phi}, \boldsymbol{\theta}$ are integrated out. In the M-step, we optimize the multi-level topic regression parameters $\boldsymbol{\beta}$ by maximum likelihood estimation (MLE) given samples of $\{\boldsymbol{c}, l\}$ in E-step. Let us provide a detailed explanation of the two steps below.

Sampling $\boldsymbol{c}$ in E-step: Conditioned on the observed words $\boldsymbol{w}_{d,s}$ and its corresponding polarity response y_d of a certain review-sentence pair (d, s) under brand b, as well as fixed topic levels $\boldsymbol{l}_{d,s}$ of the words $\boldsymbol{w}_{d,s}$, the probability of assigning (d, s) to a path $\boldsymbol{c}$ is

$$
\begin{aligned}
p(\boldsymbol{c}_{d,s} = \boldsymbol{c} | &\boldsymbol{c}^{-(d,s)}, \boldsymbol{l}_{d,s}, \boldsymbol{w}_{d,s}, y_d, b, \boldsymbol{\beta}^b) \\
&\propto p(\boldsymbol{w}_{d,s} | \boldsymbol{c}_{d,s} = \boldsymbol{c}, \boldsymbol{l}_{d,s}) \\
&\quad \times p(y_d | \boldsymbol{c}_{d,s} = \boldsymbol{c}, \boldsymbol{l}_{d,s}, \boldsymbol{\beta}^b) \\
&\quad \times p(\boldsymbol{c}_{d,s} = \boldsymbol{c} | \boldsymbol{c}^{-(d,s)}, b),
\end{aligned}
\tag{7}
$$

where $\boldsymbol{c}^{-(d,s)}$ represents the path assignments of all review sentences other than (d, s) in review corpus. From the perspective of Bayes' rule, the first two terms can be seen as the likelihood of observed words and polarity response given a selected path, and the third term is the prior on path $\boldsymbol{c}_{d,s}$ implied by the nCRP under brand b. Specifically, the prior probability of $\boldsymbol{c}_{d,s}$ under brand b is derived by the nCRP in Eq. (1) as

$$p(\boldsymbol{c}_{d,s} = \boldsymbol{c}|\boldsymbol{c}^{-(d,s)}, b) \propto \begin{cases} \prod_{l=2}^{L} \frac{M_{b,c_l}^{-(d,s)}+1}{\gamma + M_{b,c_{l-1}}^{-(d,s)} + |\Lambda(c_{l-1})|} & \text{if } \boldsymbol{c} \text{ is an existing path,} \\ \frac{\gamma}{\gamma + M_{b,c_{l*}}^{-(d,s)} + |\Lambda(c_{l*})|} \prod_{l=2}^{l^*} \frac{M_{b,c_l}^{-(d,s)}+1}{\gamma + M_{b,c_{l-1}}^{-(d,s)} + |\Lambda(c_{l-1})|} & \text{if } \boldsymbol{c} \text{ is a new path that branches} \\ & \text{under an internal node at level } l^*, \end{cases} \tag{8}$$

where $M_{b,c_l}^{-(d,s)}$ is the total number of sentence visits at the topic node c_l in brand b's corpus excluding the review-sentence pair (d, s). Moreover, the word likelihood of $\boldsymbol{w}_{d,s}$ is derived by the collapsed multivariate Pólya distribution that integrates out the topic-word multinomials $\boldsymbol{\phi}$:

$$p(\boldsymbol{w}_{d,s}|\boldsymbol{c}_{d,s} = \boldsymbol{c}, \boldsymbol{l}_{d,s})$$
$$= \prod_{n=1}^{N_{d,s}} p(w_{d,s,n} \mid k = c_{l_{d,s,n}}, \boldsymbol{w}^{-(d,s)}, \boldsymbol{k}^{-(d,s)})$$
$$= \prod_{n=1}^{N_{d,s}} \int_{\boldsymbol{\phi}_k} \phi_{k,w_{d,s,n}} p(\boldsymbol{\phi}_k \mid \boldsymbol{w}^{-(d,s)}, \boldsymbol{k}^{-(d,s)}) \, d\boldsymbol{\phi}_k \tag{9}$$
$$= \prod_{n=1}^{N_{d,s}} \mathbb{E}(\phi_{k,w_{d,s,n}} \mid \boldsymbol{w}^{-(d,s)}, \boldsymbol{k}^{-(d,s)})$$
$$= \prod_{n=1}^{N_{d,s}} \frac{W_{k=c_{l_{d,s,n}},w_{d,s,n}}^{-(d,s)} + \eta}{\sum_{v=1}^{V} \left(W_{k=c_{l_{d,s,n}},v}^{-(d,s)} + \eta \right)},$$

where $W_{k,v}^{-(d,s)}$ represents the cumulative weights of word token v under the specified topic k in the entire corpus excluding review-sentence pair (d, s). It is analogous to the introduced Pólya urn process that chooses a ball in vth color from the urn indexed by k based on the weighted proportions of existing balls in it, and η measures the initial weights of balls at the beginning of sampling. Specifically, the cumulative weights $[W_{k,v}]_{K \times V}$ of different tokens in a set of hierarchically structured topics are updated dynamically following the proposed HPU scheme, which is elaborated in Algorithm 2.

Finally, the conditional density of y_d is derived by a normal linear regression based on brand-associated regression parameters:

$$y_d \mid \boldsymbol{c}_{d,s} = \boldsymbol{c}, \boldsymbol{l}_{d,s}, \boldsymbol{\beta}^b \sim \mathcal{N} \left((\boldsymbol{x}^d)^T \boldsymbol{\beta}^b, \rho^2 \right), \tag{10}$$

where each $x_k^d = \frac{1}{N_d} \sum_{s'=1}^{S_d} \sum_{n'=1}^{N_{d,s'}} \mathbb{I}(k_{d,s',n'} = k)$.

Sampling $\boldsymbol{l}$ in E-step: Alternately, when the path $\boldsymbol{c}_{d,s}$ is selected for review-sentence pair (d, s) under brand b, we assign each word token $w_{d,s,n}$ in (d, s) to a level l based on the following probability:

$$p(l_{d,s,n} = l|\boldsymbol{c}_{d,s}, \boldsymbol{l}_{d,s,-n}, w_{d,s,n}, y_d, b, \boldsymbol{\beta}^b)$$
$$\propto p(w_{d,s,n}|l_{d,s,n} = l, \boldsymbol{c}_{d,s}) \times p(y_d|l_{d,s,n} = l, \boldsymbol{c}_{d,s}, \boldsymbol{\beta}^b)$$
$$\times p(l_{d,s,n} = l|\boldsymbol{l}_{d,s,-n}). \tag{11}$$

The prior of $l_{d,s,n}$ can be derived by integrating out $\boldsymbol{\theta}_{d,s}$ via Dirichlet-multinomial conjugacy:

$$p(l_{d,s,n} = l|\boldsymbol{l}_{d,s,-n})$$
$$= \int_{\boldsymbol{\theta}_{d,s}} p(l_{d,s,n} = l \mid \boldsymbol{\theta}_{d,s}) p(\boldsymbol{\theta}_{d,s} \mid \boldsymbol{l}_{d,s,-n}) \, d\boldsymbol{\theta}_{d,s}$$
$$= \int_{\boldsymbol{\theta}_{d,s}} \theta_{d,s,l} p(\boldsymbol{\theta}_{d,s} \mid \boldsymbol{l}_{d,s,-n}) \, d\boldsymbol{\theta}_{d,s} \tag{12}$$
$$= \mathbb{E}(\theta_{d,s,l} \mid \boldsymbol{l}_{d,s,-n})$$
$$= \frac{N_{d,s,l}^{-(d,s,n)} + \alpha}{\sum_{l'=1}^{L} \left(N_{d,s,l'}^{-(d,s,n)} + \alpha \right)},$$

where $N_{d,s,l}^{-(d,s,n)}$ represents the total number of words assigned to the lth level in (d, s) excluding $w_{d,s,n}$. Similarly, the likelihood of word $w_{d,s,n}$ given its level assignment with all observations excluding (d, s, n) is derived as

$$p(w_{d,s,n} \mid l_{d,s,n} = l, \boldsymbol{c}_{d,s})$$
$$= \frac{W_{k=c_{d,s,l},w_{d,s,n}}^{-(d,s,n)} + \eta}{\sum_{v=1}^{V} \left(W_{k=c_{d,s,l},v}^{-(d,s,n)} + \eta \right)}. \tag{13}$$

And the conditional density of y_d is in the same form with Eq. (10).

Optimizing $\boldsymbol{\beta}$ in M-step: Given empirical regressors $x_k^d = \frac{1}{N_d} \sum_{s=1}^{S_d} \sum_{n=1}^{N_{d,s}} \mathbb{I}(k_{d,s,n} = k)$ from the samples of $(\boldsymbol{c}, \boldsymbol{l})$ in E-step, we update the estimation of multi-level topic regression parameters $\boldsymbol{\beta}$ via MLE on observed polarity responses $\boldsymbol{y}$. That is $\hat{\boldsymbol{\beta}} = \arg\max_{\boldsymbol{\beta}} \mathcal{L}(\boldsymbol{y}, \boldsymbol{\beta})$ with

$$\mathcal{L}(\boldsymbol{y}, \boldsymbol{\beta}) = -\frac{1}{2\rho^2} \sum_{b=1}^{B} \sum_{d=1}^{D_b} \left(y_d - (\boldsymbol{x}^d)^T \boldsymbol{\beta}^b \right)^2. \tag{14}$$

The stopping criterion of algorithm is defined on the difference between log-likelihoods of observed review documents $\mathcal{D} = \{\boldsymbol{y}, \mathcal{W}\}$ in two successive steps: $\mathcal{L}(\mathcal{D}, \Theta^{i+1}) - \mathcal{L}(\mathcal{D}, \Theta^i) < \epsilon$. The overall procedure of model inference

is summarized in Algorithm 2. It requires four hyperparameters: γ, η, α, ρ. We set the variance of polarity response as $\rho^2 = 0.5$ following Nguyen et al. (2013), a symmetric Dirichlet parameter $\eta = 0.1$ following Blei et al. (2003), and update other hyperparameters γ, α of hierarchical topic modeling via Bayesian optimization in practice.

In each iteration of model inference, the computation burden falls mainly on the E-step that is affected by the vocabulary size V and tree depth L. The vocabulary size V mainly influences the update of addition weight matrix $\mathbf{A}$ in E-step. As the computation complexity of $\mathbf{A}$'s update is $\mathcal{O}(KV)$, this step will scale linearly with V. Moreover, the depth L of topic tree mainly influences the path and level samplings in E-step. Specifically, the total computation complexity of E-step with respect to the tree depth L is $\mathcal{O}(\sum_{b=1}^{B} \sum_{d=1}^{D_b} \sum_{s=1}^{S_d} (1 + N_{d,s})L)$, and this step will scale linearly with L.

4 Simulation study

In this section, we conduct different simulation scenarios for evaluating the performance of our proposed model extensively. We first consider a small-sized demonstration in Sect. 4.1 for visualization and qualitative analysis. Then we conduct quantitative comparison in two aspects on larger simulated corpora in Sect. 4.2: one is the accuracy of multi-aspect brand rankings, and the other is the quality of extracted topic hierarchy.

Data was generated to reasonably resemble the observed review corpora under various hierarchical topic structures. Considering the hyperparameter η governs the smoothing degree of each topic over vocabulary, e.g., a larger η tends to generate a more smoothed topic-word distribution, we also demonstrate the simulation results under different values of η. Moreover, we compare the proposed method with the following baseline models:

M1: SLDA (Blei and McAuliffe 2007) is a supervised LDA model that extends classical LDA with a response variable predicted by latent topics. We employed separate regressions on each brand corpus to obtain brand-associated aspect polarities. The fixed number of topics in SLDA is set equal to the number of leaf nodes in the hierarchical model.

M2: SHLDA (Nguyen et al. 2013) is a supervised hierarchical LDA model. We also employed separate regressions on each brand corpus for brand comparison.

Algorithm 2: Stochastic EM algorithm for MH-STM with HPU scheme.

Input: review collection $\{\{\boldsymbol{w}_d, y_d\}_{d=1}^{D_b}\}_{b=1}^{B}$, hyperparameters γ, η, α, ρ.

Output: Topic hierarchy with brand-associated regression parameters of topics.

Initialize root topic of the tree and the addition weight matrix $\mathbf{A}^0 = \mathbf{A}^1 = [0]_{K \times V}$;

for *each iteration* $i \in \{1, \cdots, I\}$ **do**

  /* Sampling paths in E-step */

  for *each review-sentence pair* (d, s) *under brand b* **do**

    Exclude (d, s) associated with its path assignment $\boldsymbol{c}_{d,s}$ from count variables:

    $M_{b,c_{d,s,l}}^{-(d,s)} \leftarrow M_{b,c_{d,s,l}} - 1$ for $\forall l \in \{1, \ldots, L\}$;

    $W_{k_{d,s,n},w_{d,s,n}}^{-(d,s)} \leftarrow W_{k_{d,s,n},w_{d,s,n}} - 1$ for $\forall n \in \{1, \cdots, N_{d,s}\}$;

    $W_{k,w_{d,s,n}}^{-(d,s)} \leftarrow W_{k,w_{d,s,n}} - a_{kw_{d,s,n}}^{i-1}$ for $\forall n \in \{1, \cdots, N_{d,s}\}, \forall k \in \mathrm{V}(k_{d,s,n})$;

    Sample a new path $\boldsymbol{c}_{d,s}$ for $\boldsymbol{w}_{d,s}$ via Equations (7)–(10);

    Update count variables with the new path $\boldsymbol{c}_{d,s}$:

    $M_{b,c_{d,s,l}} \leftarrow M_{b,c_{d,s,l}}^{-(d,s)} + 1$ for $\forall l \in \{1, \ldots, L\}$;

    $W_{k_{d,s,n},w_{d,s,n}} \leftarrow W_{k_{d,s,n},w_{d,s,n}}^{-(d,s)} + 1$ for $\forall n \in \{1, \cdots, N_{d,s}\}$;

    $W_{k,w_{d,s,n}} \leftarrow W_{k,w_{d,s,n}}^{-(d,s)} + a_{kw_{d,s,n}}^{i}$ for $\forall n \in \{1, \cdots, N_{d,s}\}, \forall k \in \mathrm{V}(k_{d,s,n})$;

  end

  /* Sampling levels in E-step */

  for *each review-sentence pair* (d, s) **do**

    for *each word* $w_{d,s,n}, n = 1, \ldots, N_{d,s}$ **do**

      Exclude $w_{d,s,n}$ associated with its level assignment $l_{d,s,n}$ from count variables:

      $N_{d,s,l_{d,s,n}}^{-(d,s,n)} \leftarrow N_{d,s,l_{d,s,n}} - 1$;

      $W_{k_{d,s,n},w_{d,s,n}}^{-(d,s,n)} \leftarrow W_{k_{d,s,n},w_{d,s,n}} - 1$;

      $W_{k,w_{d,s,n}}^{-(d,s,n)} \leftarrow W_{k,w_{d,s,n}} - a_{kw_{d,s,n}}^{i}$ for $\forall k \in \mathrm{V}(k_{d,s,n})$;

      Sample a new level $l_{d,s,n}$ for $w_{d,s,n}$ via Equations (11)–(13);

      Update count variables with the new level $l_{d,s,n}$:

      $N_{d,s,l_{d,s,n}} \leftarrow N_{d,s,l_{d,s,n}}^{-(d,s,n)} + 1$;

      $W_{k_{d,s,n},w_{d,s,n}} \leftarrow W_{k_{d,s,n},w_{d,s,n}}^{-(d,s,n)} + 1$;

      $W_{k,w_{d,s,n}} \leftarrow W_{k,w_{d,s,n}}^{-(d,s,n)} + a_{kw_{d,s,n}}^{i}$ for $\forall k \in \mathrm{V}(k_{d,s,n})$;

    end

  end

  Update addition matrix $\mathbf{A}^{i+1}$ via Equation (5);

  /* M-step */

  Update $\boldsymbol{\beta}$ by MLE on Equation (14);

end

4.1 Qualitative comparison with topic tree visualization

We first illustrate the effectiveness of the proposed MH-STM model on a synthetic corpus with a small vocabulary. We generate a synthetic corpus that consists of 10 brands each having 200 documents from a 9-term vocabulary represented by a 3×3 grid. The observed words are generated by a three-level topic tree with 1, 3, 9 topics at the first level, second level, and third level, respectively. The root topic has a uniform distribution over the entire vocabulary grid. Topics at the second level are uniformly distributed over the terms in each row of the grid. Topics at the third level have full

probability concentrated at a single term from the row of its parent.

The results of topic hierarchy inferred from the synthetic data above are visualized in Fig. 4. The proposed MH-STM model successfully recovers the original hierarchical structure, and the inferred topics by MH-STM are almost identical to the original topics. By contrast, in the tree derived by SHLDA, the first and second level topics are almost identical to the original topics, while the third level topics are disturbed by some noise words. In addition, the topic tree inferred by SHLDA exhibits higher disorders in hierarchical affinity. For example in Fig. 4b, the topic 1-1-3 is more close to the second-level topic 1-2 than its direct parent topic 1-1. This observation is consistent with the quantitative compar-
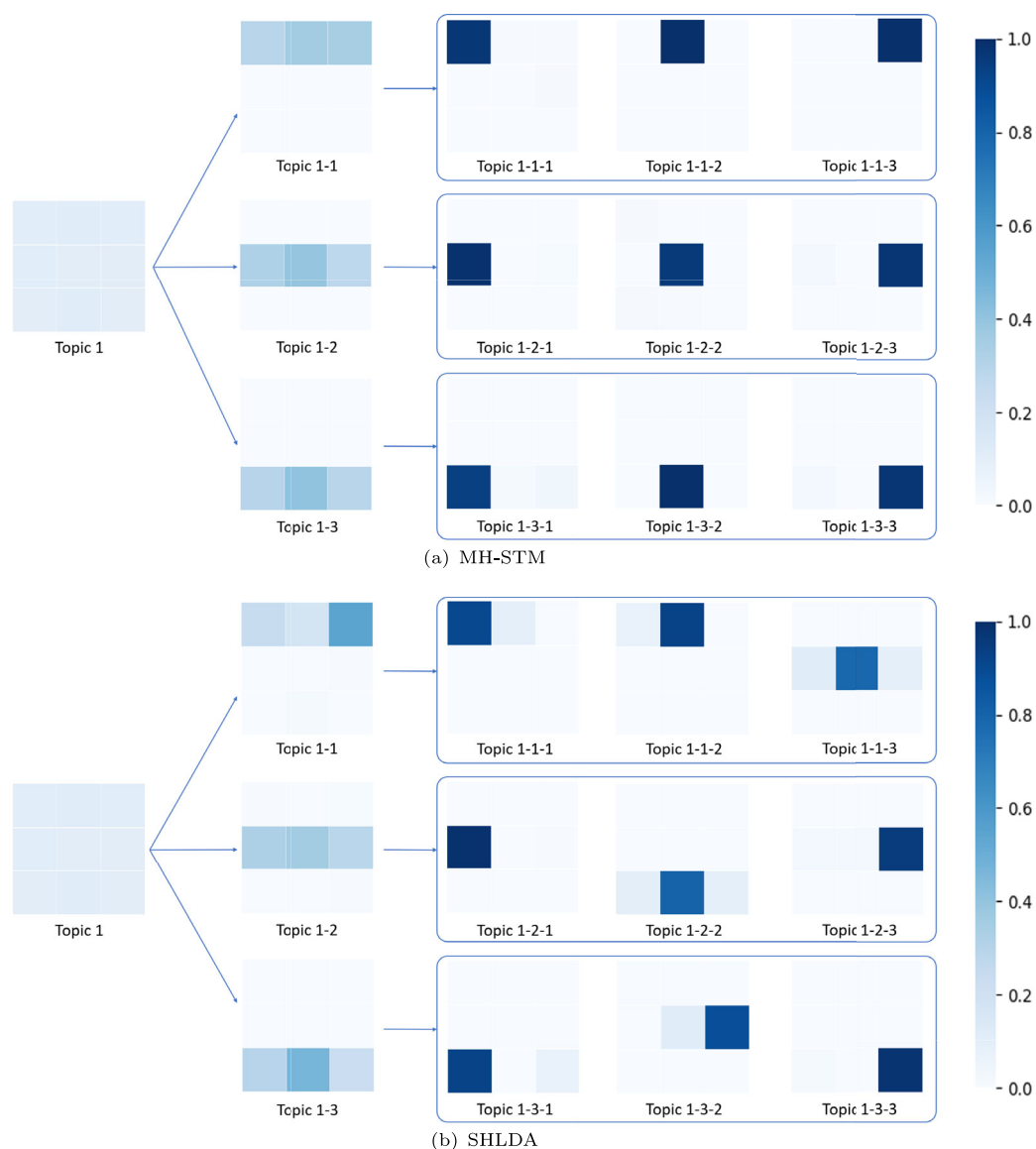

Fig. 4 Topic trees inferred from synthetic data. Each cell corresponds to a single word, and is shaded with intensity proportional to the probability of each word in the topic

ison results of hierarchical affinity scores in the following Tables 2 and 4, that MH-STM is much preferred for reasonable parent–child relations in extracted hierarchy.

4.2 Quantitative analysis

To have a comprehensive examination of the proposed model, we evaluate the proposed method on a set of synthetic corpora that are generated under different scenarios of topic structures and distribution parameters. Each experimental corpus is generated from a three-level topic hierarchy with a vocabulary of 100 terms. We consider different probability distributions specific to totally $B = 10$ brands on the topic hierarchy, and generate $D_b = 200$ synthetic reviews respectively for each brand relying on the generative process in Sect. 3.1, with each review including $S_d = 5$ sentences in average length of $N_{d,s} = 10$ words.

In each simulation scenario, we consider the brand topic regression parameter $\beta_k^b \sim \mathcal{N}(\mu, \sigma_{l(k)}^2)$, where $\mu = 0$ and $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \sigma_3)^T = (1, 2, 3)^T$, depending on the topic level $l(k) \in \{1, 2, 3\}$ in the three-level hierarchy. This formulation induces higher variance of brand polarities towards leaf topics, which is consistent with the real-world observation in Fig. 5. We assume a symmetric Dirichlet with hyperparameter $\boldsymbol{\alpha} = (1.0, 1.0, 1.0)^T$ for the 3-dimensional topic proportion vector $\boldsymbol{\theta} \sim \text{Dir}(\boldsymbol{\alpha})$ of each sentence. We also set $\rho = 1$ as the standard deviation for generating polarity responses in all of simulation cases.

To assess the effectiveness of the proposed model in a variety of experimental scenarios, we designed hierarchical structures following the simulation settings similar to the settings in other hierarchical topic models (Blei et al. 2003, 2010). For example, the scales of their ground-truth topic hierarchies are generally 3-level tree structures, ranging 3–5 branches from the root. Similarly, we designed several 3-level topic hierarchies denoted as $n(t_1, \ldots, t_n)$: 3(3,2,4), 4(6,5,2,4), 5(3,5,5,2,1). Here n is the number of branches from the root and t_j is the number of children under the jth branch. Given the fixed tree structure with finite paths of topics, the path assignments of each brand's reviews are decided by brand-specific distributions over paths that are randomly generated. Moreover, we consider different smoothing parameters $\eta \in \{0.01, 0.1, 1.0\}$ for drawing the topic components.

4.2.1 Multi-aspect brand ranking evaluations

We report the multi-aspect brand ranking results based on several evaluation metrics that are commonly used in ranking tasks: (1) Two correlation coefficients respectively for Spearman's correlations and Kendall's Tau. Both metrics measure how well the predicted brand aspect rankings correlate with

their ground-truth aspect ratings under each leaf topic. (2) Average precision (AP@K) that evaluates the model's top-K recommendation performance in ranking different brands for a specified aspect. We treat each leaf topic as a query, and the top K of brands ranked by ground-truth aspect ratings as the relevant items. The AP@K measures how many of the top-K brands recommended by a model are included in the relevant items and how highly they are ranked. A higher AP@K indicates better top-K recommendation performance of a model. Specifically, we report the average results of AP@5 corresponding to top-5 recommendation for each leaf topic.

For each model, we repeat the inference algorithm 20 times with various initializations driven by different random seeds and take the average results. Table 1 summarizes the average results of different metrics as well as their standard deviations for each model, where the best results are in bold. Overall, our proposed MH-STM model achieves the best performance in multi-aspect brand comparison. Although the performance decays with larger values of η, MH-STM still dominates other alternatives. The performance decay under large η is expected, as large values of η produce highly smoothed word distributions across topics, leading to difficulty in separating one topic from another based on their word co-occurrence rules.

4.2.2 Recovery of topic structure

We quantitatively evaluate the performance of different methods in topic recovery by four measures, including: (1) Topic accuracy that quantifies the proportion of correct topic assignments in synthetic documents. Considering the issue of label switching between the learned topics and the ground-truth ones, we find the optimal alignment between the learned and the ground-truth topics by Kuhn-Munkres algorithm (Lovász and Plummer 2009). (2) Held-out likelihood that evaluates how well the trained model explains the held-out data. It is defined as the average log-likelihood per word of a held-out corpus given the trained model, which is estimated by the importance sampling methods in Wallach et al. (2009). (3) The coherence scores of top-5 words under each topic. A higher coherence score indicates a superior topic that better captures the word co-occurrence rules in corpus. We adopt the topic coherence metric by Mimno et al. (2011):

$$Coherence\left(k; V^{(k)}\right) = \sum_{i=2}^{5} \sum_{j=1}^{i-1} \log \frac{DF\left(v_i^{(k)}, v_j^{(k)}\right) + 1}{DF\left(v_j^{(k)}\right)},$$

where $V^{(k)} = \left(v_1^{(k)}, \cdots, v_5^{(k)}\right)$ represents the top 5 words assigned with topic k, $DF\left(v_j^{(k)}\right)$ is the document frequency of word $v_j^{(k)}$, and $DF\left(v_i^{(k)}, v_j^{(k)}\right)$ is the co-document fre-

Table 1 Comparison results of multi-aspect brand rankings on simulated corpora

Hierarchy	η	Metrics	SLDA	SHLDA	MH-STM
3(3,2,4)	0.01	Spearman's	0.602 (0.013)	0.560 (0.021)	**0.680 (0.016)**
		Kendall's tau	0.511 (0.011)	0.492 (0.018)	**0.580 (0.014)**
		AP@5	0.760 (0.009)	0.781 (0.012)	**0.842 (0.008)**
	0.1	Spearman's	0.584 (0.015)	0.626 (0.022)	**0.732 (0.016)**
		Kendall's tau	0.477 (0.013)	0.540 (0.018)	**0.628 (0.014)**
		AP@5	0.745 (0.013)	0.828 (0.011)	**0.887 (0.011)**
	1	Spearman's	0.234 (0.022)	0.246 (0.024)	**0.406 (0.015)**
		Kendall's tau	0.180 (0.018)	0.190 (0.018)	**0.309 (0.012)**
		AP@5	0.517 (0.016)	0.602 (0.019)	**0.698 (0.011)**
4(6,5,2,4)	0.01	Spearman's	0.573 (0.012)	0.575 (0.014)	**0.640 (0.015)**
		Kendall's tau	0.444 (0.010)	0.461 (0.012)	**0.517 (0.013)**
		AP@5	0.783 (0.008)	0.766 (0.010)	**0.802 (0.012)**
	0.1	Spearman's	0.479 (0.012)	0.563 (0.015)	**0.667 (0.015)**
		Kendall's tau	0.361 (0.010)	0.457 (0.012)	**0.542 (0.012)**
		AP@5	0.676 (0.009)	0.777 (0.012)	**0.847 (0.008)**
	1	Spearman's	0.104 (0.016)	0.146 (0.017)	**0.214 (0.015)**
		Kendall's tau	0.090 (0.011)	0.106 (0.014)	**0.172 (0.009)**
		AP@5	0.524 (0.011)	0.553 (0.008)	**0.612 (0.008)**
5(3,5,5,2,1)	0.01	Spearman's	0.529 (0.011)	0.505 (0.018)	**0.591 (0.011)**
		Kendall's tau	0.419 (0.010)	0.402 (0.014)	**0.467 (0.011)**
		AP@5	0.769 (0.008)	0.783 (0.010)	**0.825 (0.010)**
	0.1	Spearman's	0.497 (0.011)	0.530 (0.009)	**0.642 (0.011)**
		Kendall's tau	0.378 (0.010)	0.422 (0.008)	**0.528 (0.009)**
		AP@5	0.730 (0.008)	0.776 (0.009)	**0.837 (0.007)**
	1	Spearman's	0.087 (0.017)	0.196 (0.020)	**0.235 (0.013)**
		Kendall's tau	0.063 (0.013)	0.147 (0.016)	**0.176 (0.010)**
		AP@5	0.538 (0.016)	0.619 (0.010)	**0.642 (0.006)**

quency of both words. (4) The hierarchical affinity score that measures the goodness of a topic hierarchy (Kim et al. 2013). It is defined as the ratio of the average cosine similarity between the second-level topics and their direct children topics to the average cosine similarity between the second-level topics and their non-children topics at the third level. A lower hierarchical affinity score, especially lower than 1, indicates higher disorders in tree's parent–child affinity that a parent topic node tends to violate its direct children in semantic consistency.

The average values and standard deviations of different performance metrics for alternative models are presented in Table 2, where the best results are in bold. Overall, our proposed MH-STM model achieves the highest topic accuracy and coherence, indicating its superiority in the unsupervised discovery of latent topics. In comparison with SHLDA, MH-STM produces uniformly higher hierarchical affinity scores, which demonstrates that the discovered topic trees by MH-STM show higher consistency among parent–child relations, and this result is also in accord with the qualitative comparison on visualized hierarchies in Sect. 4.1. Likewise, the affinity performance degrades under a large value of smoothing parameter η, as large values of η produce highly smoothed word distributions across topics, leading to difficulty in separating one topic from another and higher disorders in parent–child relations. In terms of held-out likelihood, the hierarchical models (i.e., SHLDA and MH-STM) produce generally better results than the flat model of SLDA. However, the proposed MH-STM is overtaken by SHLDA in likelihood comparison. This loss is induced by the enhanced

Table 2 Comparison results of topic quality on simulated corpora

Hierarchy	η	Metrics	SLDA	SHLDA	MH-STM
3(3,2,4)	0.01	Topic accuracy	0.566 (0.006)	0.546 (0.010)	**0.583 (0.010)**
		Held-out likelihood	−2.713 (0.002)	**−2.654 (0.003)**	−2.663 (0.004)
		Coherence	−1.456 (0.021)	−1.608 (0.072)	**−1.265 (0.020)**
		Hierarchical Affinity	-	2.748 (0.486)	**3.185 (0.298)**
	0.1	Topic accuracy	0.438 (0.004)	0.475 (0.009)	**0.513 (0.008)**
		Held-out likelihood	−3.924 (0.003)	**−3.881 (0.002)**	−3.902 (0.003)
		Coherence	−1.772 (0.009)	−1.737 (0.010)	**−1.700 (0.017)**
		Hierarchical Affinity	-	1.654 (0.155)	**1.994 (0.081)**
	1	Topic accuracy	0.146 (0.001)	0.192 (0.003)	**0.225 (0.004)**
		Held-out likelihood	−4.520 (0.001)	**−4.489 (0.001)**	−4.503 (0.001)
		Coherence	−2.165 (0.009)	−2.170 (0.006)	**−2.144 (0.006)**
		Hierarchical Affinity	-	1.015 (0.023)	**1.183 (0.010)**
4(6,5,2,4)	0.01	Topic accuracy	0.507 (0.007)	0.472 (0.006)	**0.546 (0.007)**
		Held-out likelihood	−2.952 (0.003)	**−2.869 (0.003)**	−2.899 (0.003)
		Coherence	−2.147 (0.05)	−1.872 (0.069)	**−1.614 (0.028)**
		Hierarchical Affinity	-	1.847 (0.192)	**3.032 (0.177)**
	0.1	Topic accuracy	0.393 (0.005)	0.396 (0.005)	**0.427 (0.005)**
		Held-out likelihood	−4.045 (0.003)	**−3.985 (0.002)**	−4.013 (0.002)
		Coherence	−1.994 (0.008)	−1.984 (0.009)	**−1.868 (0.010)**
		Hierarchical Affinity	-	1.249 (0.059)	**1.926 (0.062)**
	1	Topic accuracy	0.082 (0.001)	0.117 (0.001)	**0.127 (0.001)**
		Held-out likelihood	−4.544 (0.000)	**−4.490 (0.001)**	−4.502 (0.001)
		Coherence	−2.275 (0.005)	−2.286 (0.004)	**−2.241 (0.004)**
		Hierarchical Affinity	-	0.997 (0.006)	**1.233 (0.034)**
5(3,5,5,2,1)	0.01	Topic accuracy	0.495 (0.007)	0.493 (0.010)	**0.548 (0.008)**
		Held-out likelihood	−2.932 (0.002)	**−2.815 (0.004)**	−2.864 (0.004)
		Coherence	−2.063 (0.048)	−1.832 (0.057)	**−1.634 (0.041)**
		Hierarchical Affinity	-	2.400 (0.192)	**3.778 (0.210)**
	0.1	Topic accuracy	0.407 (0.006)	0.411 (0.006)	**0.459 (0.006)**
		Held-out likelihood	−4.009 (0.002)	**−3.933 (0.002)**	−3.968 (0.002)
		Coherence	−1.957 (0.009)	−1.950 (0.009)	**−1.849 (0.008)**
		Hierarchical Affinity	-	1.425 (0.073)	**2.488 (0.091)**
	1	Topic accuracy	0.087 (0.001)	0.113 (0.001)	**0.132 (0.001)**
		Held-out likelihood	−4.544 (0.000)	**−4.487 (0.001)**	−4.501 (0.001)
		Coherence	−2.264 (0.006)	−2.260 (0.003)	**−2.239 (0.004)**
		Hierarchical Affinity	-	0.988 (0.007)	**1.251 (0.021)**

sampling process of HPU on the proposed model, during which the observed words are sampled based on not only their full probabilities but also their general-to-specific char-acteristics, such that the final results deviate slightly from a maximum likelihood. However, a slight loss of likelihood is worth for more interpretable topics and higher predictive

accuracy in brand ranking. Moreover, the loss of held-out likelihood by MH-STM is consistent with the observations in previous studies on topic comparison (Chang et al. 2009; Mimno et al. 2011) that higher likelihood of held-out documents could violate the judgementof topic coherence in an interpretable latent space.

5 Real-data case studies

In this section, we evaluate the proposed method under two representative review corpora in the real case studies: laptop review corpus and beer review corpus.

Laptop review corpus (Data 1): This corpus (Ni et al. 2019) consists of customer reviews with corresponding overall ratings on Amazon that belong to four laptop brands: Dell, HP, Lenovo, and Apple. After preprocessing, it contains 100,931 reviews with a vocabulary of 1,642 unique terms.

Beer review corpus (Data 2): The beer review corpus (McAuley et al. 2012) consists of beer reviews on RateBeer that belong to 26 brewers each regarded as a brand. Specifically, each review is accompanied by both an overall rating and four aspect-specific ratings that correspond to four performance aspects of beer: appearance, aroma, palate, and taste. Given the four aspects and their aspect-specific ratings as the ground truth, the multi-aspect brand ranking results can be validated. After preprocessing, the corpus contains 89,309 reviews with a vocabulary of 1,318 terms.

For both corpora, we implemented a series of pre-processing procedures. The data pre-processing on review texts includes sentence and word tokenizations, removing terms such as punctuations, digits, stop words, infrequent words, and stemming each word to its root for a dense vocabulary. As the overall ratings and various aspect ratings originally range in different scales, we map them uniformly to [0, 1] by performing a min-max normalization.

All models are implemented on individual sentences of reviews for leveraging the sentence-level information. Following the practice in Nguyen et al. (2013), we set regression parameters of the root node to zero for both hierarchical models, i.e., SHLDA and MH-STM. It is reasonable to assume that root node would not change the response variable since it is associated with every document.

5.1 Multi-aspect brand rankings

As the laptop review corpus does not provide ground truth aspect-specific ratings, we examine the multi-aspect brand ranking results only on the beer review corpus (Data 2). Specifically, all models were trained on the entire beer corpus (Data 2) with only overall ratings, and the aspect-specific ratings were held out for validation. To align the topics extracted freely from corpus to the four specified aspects, we adopt

Table 3 Multi-aspect brand ranking evaluations on beer review corpus

Beer aspect	Metrics	SLDA	SHLDA	MH-STM
Appearance	Spearman's	0.246	0.458	**0.586**
	Kendall's tau	0.182	0.323	**0.434**
	AP@10	0.438	0.642	**0.674**
	AP@5	0.357	0.457	**0.547**
Aroma	Spearman's	0.636	0.815	**0.875**
	Kendall's tau	0.471	0.637	**0.723**
	AP@10	0.602	0.897	**0.918**
	AP@5	0.543	**0.910**	**0.910**
Palate	Spearman's	0.385	0.855	**0.936**
	Kendall's tau	0.268	0.673	**0.785**
	AP@10	0.534	0.969	**0.990**
	AP@5	0.384	0.843	**0.960**
Tastea	Spearman's	0.480	0.904	**0.957**
	Kendall's tau	0.372	0.766	**0.858**
	AP@10	0.588	0.931	**0.966**
	AP@5	0.457	0.960	**0.955**

four hold-out corpora each composed of review sentences annotated with corresponding aspect labels. Each annotated corpus is matched with the model's extracted topics and represented as a topic distribution specific to that aspect. After that, the brand rankings in an aspect are predicted by their mean polarity responses with respect to the aspect's distribution over topics.

We report the multi-aspect brand ranking results of different models in Table 3, where the best results are in bold. Similarly, the model performance in multi-aspect brand ranking is quantitatively evaluated by the following metrics: (1) The average results of Spearman's correlations and Kendall's Tau for ranking different brands in each aspect. (2) Mean average precision (AP@5 and AP@10) that evaluates the model's top-5 and top-10 recommendation performances in each aspect. It can be seen from Table 3 that MH-STM achieves the best performance in aspect-specific brand rankings. Overall, hierarchical models (i.e., SHLDA and MH-STM) outperform the flat model (i.e., SLDA) in producing accurate brand rankings for a specified aspect. Compared to SHLDA, MH-STM can provide an improved brand differentiation by integrating brand preference on topic distributions and an enhanced HPU scheme.

5.2 Quality of discovered topics

We also compare the performance of different models in terms of topic findings that are measured by several quantita-

Table 4 Topic quality evaluations on two corpora

Dataset	Metrics	SLDA	SHLDA	MH-STM
Beer	Coherence ↑	−3.140	−3.885	**−3.134**
	Held-out likelihood ↑	−5.640	**−5.241**	−5.246
	Hierarchical Affinity ↑	–	2.247	**4.804**
Laptop	Coherence ↑	−3.306	−3.184	**−3.068**
	Held-out likelihood ↑	−6.164	**−5.974**	−6.021
	Hierarchical Affinity ↑	–	1.450	**4.639**

tive metrics commonly used in topic models: (1) Coherence score for top 10 words of each topic. (2) Held-out likelihood. We perform a 5-fold cross validation on the experimental corpora and report their average results. (3) Hierarchical affinity for measuring the goodness of a topic hierarchy.

Table 4 summarizes the topic quality metrics from different models on the two corpora. It is seen that MH-STM achieves the best coherence on both corpora, where the best results are in bold, and it exceeds the SHLDA model in discovered hierarchical affinity. The held-out likelihoods of MH-STM outperform the flat model of SLDA but slightly

Fig. 5 Portions of 3-level topic hierarchy discovered from laptop reviews by SHLDA and MH-STM. Each topic is represented by its top 10 words with corresponding word probabilities under the topic. The numbers on the right side of each topic are brand-specific regression parameters learned with that topic
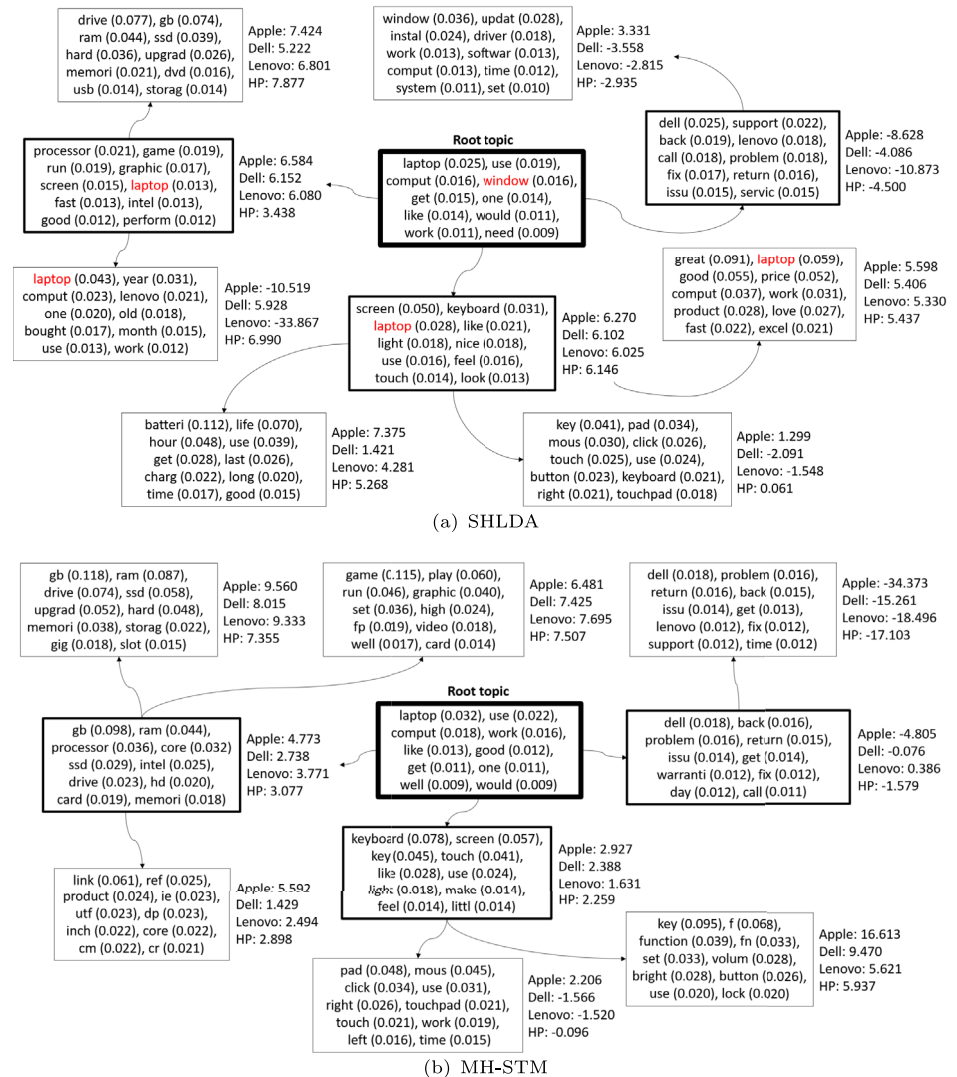
fall behind the SHLDA on both corpora, which is consistent with the simulation observations in Table 2.

5.3 Qualitative analysis on topic hierarchy

It is also interesting to visualize the topic hierarchies obtained from the proposed method. Figure 5 illustrates portions of 3-level topic hierarchies discovered from the laptop corpus respectively by SHLDA and MH-STM, where the top words with their probabilities and the brand regression parameters specific to each topic are presented. Both hierarchies present a general-to-specific tendency that root topics are general, and topics close to the leaves are more specific. Moreover, the regression parameters close to leaf topics learned by both models present overall higher variations among different brands, indicating that customers' polarities towards different brands are more differentiated in lower-level topics. One can obtain aspect-specific brand rankings by referring to the brand regression parameters of that aspect Fig. 5.

One can see that MH-STM provides a better hierarchical topic-word association in comparison with SHLDA. For example, the general word "laptop" is mixed in each level of the hierarchy by SHLDA, while it is concentrated on the root topic by MH-STM. On the other hand, the aspect-specific words are more assigned with lower-level topics by MH-STM. For instance, the top word "window" under the root topic by SHLDA is degraded in the root topic by MH-STM. Overall, this qualitative comparison result demonstrates the effectiveness of the proposed HPU scheme, in which words are assigned with different weights along the hierarchy, forcing topic-word associations to coincide with their general-to-specific characteristics.

5.4 Runtime comparison

The runtime of different methods on the two real-world datasets are investigated in Fig. 6. All methods were implemented in Python 3.10 and conducted on a 64-bit Windows 11 machine with Intel Core Ultra 7 processor and 32GB memory. It shows the runtime of MH-STM ranks between the other two methods. Specifically, compared to the other hierarchical method SHLDA, the proposed model requires an additional update of weight matrix in each iteration, leading to relatively higher computational cost than the SHLDA model.

6 Conclusion

In this work, proposed a multifacet hierarchical sentiment-topic model (MH-STM) for multi-aspect brand comparison automatically from online reviews. The proposed model is capable of extracting both hierarchically structured topics
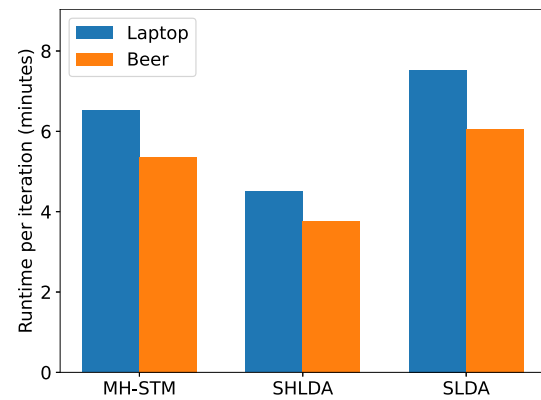

Fig. 6 Runtime of different methods on two datasets

shared among various brands and topic-specific sentiment polarities across brand competitors via a unified generative framework. Moreover, the hierarchical topic modeling is combined with a novel hierarchical Pólya urn scheme that enhances the topic-word association by incorporating different burstiness of words among the topic hierarchy based on their general-to-specific characteristics.

Through two real case studies from Amazon's laptop review corpus and RateBeer's beer review corpus, it is found that the proposed MH-STM have the merits in two major aspects. First, the extracted topics show higher coherence and more reasonable hierarchical relations. Second, the proposed method produces more accurate brand rankings on a specified aspect. Note that the proposed method can be easily adapted to a broad range of online review corpora in different application areas, serving as an effective tool of online brand positioning.

There are several directions for the future work. It is worthwhile to consider a scalable extension of the MH-STM model. The current model applies a fully collapsed Gibbs sampler and requires an additional update of weight matrix in each iteration. It will be interesting to incorporate some efficient partially collapsed Gibbs sampling algorithms for hierarchical topic models (Chen et al. 2018) to develop scalable algorithms. In the current simulation, we conduct model evaluation on moderate-scale topic hierarchies to achieve a balance between the efficiency and effectiveness of validation. With a scalable extension, model evaluation can be conducted efficiently on a larger-scaled topic hierarchy with extended breadth and depth, which can better approximate the real topic hierarchies in practical applications. Another potential extension is to handle unbalanced topic hierarchies. The current model could overlook some scarce topics when there are extremely unbalanced topic distributions. One remedy for improvement is to incorporate prior knowledge such as topic seeds (Jagarlamudi et al. 2012), or uncover topics from word co-occurrence networks where the distribution over topics is less skewed (Zuo et al. 2016).

Data and code availability statements Data and code to perform the experiments in this work are publicly available at GitHub: https://github.com/LiangQiao94/MH-STM.

Declarations

Conflict of interest The authors have no conflict of interest to declare that are relevant to the content of this article.

References

Alzate, M., Arce-Urriza, M., Cebollada, J.: Mining the text of online consumer reviews to analyze brand image and brand positioning. J. Retail. Consum. Serv. **67**, 102989 (2022)

Blei, D.M., Griffiths, T.L., Jordan, M.I.: The nested chinese restaurant process and bayesian nonparametric inference of topic hierarchies. J. ACM **57**(2), 1–30 (2010)

Boyd-Graber, J., Resnik, P.: Holistic sentiment analysis across languages: Multilingual supervised latent dirichlet allocation. In: Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing, pp. 45–55. Association for Computational Linguistics, USA (2010)

Blei, D.M., Jordan, M.I., Griffiths, T.L., Tenenbaum, J.B.: Hierarchical topic models and the nested chinese restaurant process. In: Proceedings of the 16th International Conference on Neural Information Processing Systems, pp. 17–24. MIT Press, Cambridge, MA, USA (2003)

Brand, B.M., Kopplin, C.S., Rausch, T.M.: Cultural differences in processing online customer reviews: holistic versus analytic thinkers. Electron. Mark. **32**(3), 1039–1060 (2022)

Blei, D.M., McAuliffe, J.D.: Supervised topic models. In: Proceedings of the 20th International Conference on Neural Information Processing Systems, pp. 121–128. Curran Associates Inc., Red Hook, NY, USA (2007)

Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent dirichlet allocation. J. Mach. Learn. Res. **3**(Jan), 993–1022 (2003)

Barry, A.E., Valdez, D., Padon, A.A., Russell, A.M.: Alcohol advertising on twitter-a topic model. Am. J. Health Educ. **49**(4), 256–263 (2018)

Chang, J., Boyd-Graber, J., Gerrish, S., Wang, C., Blei, D.M.: Reading tea leaves: How humans interpret topic models. In: Proceedings of the 22nd International Conference on Neural Information Processing Systems, pp. 288–296. Curran Associates Inc., Red Hook, NY, USA (2009)

Colladon, A.F.: The semantic brand score. J. Bus. Res. **88**, 150–160 (2018)

Chen, J., Zhu, J., Lu, J., Liu, S.: Scalable training of hierarchical topic models. Proc. VLDB Endow. **11**(7), 826–839 (2018)

Doyle, G., Elkan, C.: Accounting for burstiness in topic models. In: Proceedings of the 26th Annual International Conference on Machine Learning, pp. 281–288 (2009)

He, Z., Zheng, L., He, S.: A novel approach for product competitive analysis based on online reviews. Electron. Commer. Res. **23**(4), 2259–2290 (2023)

Jagarlamudi, J., Daumé III, H., Udupa, R.: Incorporating lexical priors into topic models. In: Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics, pp. 204–213 (2012)

Kang, J.-H., Ma, J., Liu, Y.: Transfer topic modeling with ease and scalability. In: Proceedings of the 2012 SIAM International Conference on Data Mining, pp. 564–575 (2012). SIAM

Kim, S., Zhang, J., Chen, Z., Oh, A., Liu, S.: A hierarchical aspect-sentiment model for online reviews. In: Proceedings of the Twenty-Seventh AAAI Conference on Artificial Intelligence, pp. 526–533 (2013)

Liang, Q.: Tree-based data filtering for online user-generated reviews. IISE Trans. **56**(8), 824–840 (2024). https://doi.org/10.1080/24725854.2023.2228861

Lovász, L., Plummer, M.D.: Matching Theory. American Mathematical Soc, Providence, RI (2009)

Liang, Q., Ranganathan, S., Wang, K., Deng, X.: JST-RR model: Joint modeling of ratings and reviews in sentiment-topic prediction. Technometrics **65**(1), 57–69 (2023). https://doi.org/10.1080/00401706.2022.2063187

Liu, L., Tang, L., He, L., Zhou, W., Yao, S.: An overview of hierarchical topic modeling. In: 2016 8th International Conference on Intelligent Human-Machine Systems and Cybernetics, pp. 391–394 (2016). IEEE

Li, C., Wang, H., Zhang, Z., Sun, A., Ma, Z.: Topic modeling for short texts with auxiliary word embeddings. In: Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 165–174 (2016)

Mahmoud, H.: Pólya Urn Models. Chapman and Hall/CRC, Boca Raton, FL (2008)

Mitra, S., Jenamani, M.: Obim: a computational model to estimate brand image from online consumer review. J. Bus. Res. **114**, 213–226 (2020)

Madsen, R.E., Kauchak, D., Elkan, C.: Modeling word burstiness using the dirichlet distribution. In: Proceedings of the 22nd International Conference on Machine Learning, pp. 545–552 (2005)

McAuley, J., Leskovec, J., Jurafsky, D.: Learning attitudes and attributes from multi-aspect reviews. In: 2012 IEEE 12th International Conference on Data Mining, pp. 1020–1025 (2012). IEEE

Mimno, D., Li, W., McCallum, A.: Mixtures of hierarchical topics with pachinko allocation. In: Proceedings of the 24th International Conference on Machine Learning, pp. 633–640 (2007)

Mimno, D., Wallach, H., Talley, E., Leenders, M., McCallum, A.: Optimizing semantic coherence in topic models. In: Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing, pp. 262–272 (2011)

Nguyen, V.-A., Boyd-Graber, J., Resnik, P.: Lexical and hierarchical topic regression. In: Proceedings of the 26th International Conference on Neural Information Processing Systems, pp. 1106–1114. Curran Associates Inc., Red Hook, NY, USA (2013)

Ni, J., Li, J., McAuley, J.: Justifying recommendations using distantly-labeled reviews and fine-grained aspects. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), pp. 188–197 (2019)

Sajid, S., Volkova, N., Wilson, J.A., Opoku-Asante, E.: Using text mining and crowdsourcing platforms to build employer brand in the us banking industry. Glob. Bus. Organ. Excell. **41**(4), 6–27 (2022)

Tumpa, S.N., Ali, M.M.: Document concept hierarchy generation by extracting semantic tree using knowledge graph. In: 2018 IEEE International WIE Conference on Electrical and Computer Engineering (WIECON-ECE), pp. 83–86 (2018). IEEE

Titov, I., McDonald, R.: Modeling online reviews with multi-grain topic models. In: Proceedings of the 17th International Conference on World Wide Web, pp. 111–120 (2008)

Vayansky, I., Kumar, S.A.: A review of topic modeling methods. Inform. Syst. **94**, 101582 (2020)

Vafa, K., Naidu, S., Blei, D.: Text-based ideal points. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pp. 5345–5357 (2020)

Wallach, H.M., Murray, I., Salakhutdinov, R., Mimno, D.: Evaluation methods for topic models. In: Proceedings of the 26th Annual International Conference on Machine Learning, pp. 1105–1112 (2009)

Wang, R., Zhou, D., He, Y.: Optimising topic coherence with weighted polya urn scheme. Neurocomputing **385**, 329–339 (2020)

Xu, Y., Yin, J., Huang, J., Yin, Y.: Hierarchical topic modeling with automatic knowledge mining. Expert Syst. Appl. **103**, 106–117 (2018)

Zhao, R., Gui, L., Pergola, G., He, Y.: Adversarial learning of poisson factorisation model for gauging brand sentiment in user reviews. In: Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume, pp. 2341–2351 (2021)

Zhao, R., Gui, L., Yan, H., He, Y.: Tracking brand-associated polarity-bearing topics in user reviews. Trans. Assoc. Comput. Linguistics **11**, 404–418 (2023)

Zhang, H., Kim, G., Xing, E.P.: Dynamic topic modeling for monitoring market competition from online text and image data. In: Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1425–1434 (2015)

Zhang, C., Tong, T., Bu, Y.: Examining differences among book reviews from various online platforms. Online Inform. Rev. **43**(7), 1169–1187 (2019)

Zuo, Y., Zhao, J., Xu, K.: Word network topic model: a simple but general solution for short and imbalanced texts. Knowl. Inf. Syst. **48**, 379–398 (2016)