

Targeted Topic Modeling for Focused Analysis

Shuai Wang[†], Zhiyuan Chen[†], Geli Fei[†], Bing Liu[†] and Sherry Emery[‡]

[†]Department of Computer Science

[‡]Institute of Health Research and Policy

University of Illinois at Chicago

Chicago, IL, USA

shuaiwanghk@gmail.com, czyuanacm@gmail.com,

gfei2@uic.edu, liub@cs.uic.edu, slemery@uic.edu

ABSTRACT

One of the overarching tasks of document analysis is to find what topics people talk about. One of the main techniques for this purpose is topic modeling. So far many models have been proposed. However, the existing models typically perform full analysis on the whole data to find all topics. This is certainly useful, but in practice we found that the user almost always also wants to perform more detailed analyses on some specific aspects, which we refer to as *targets* (or *targeted aspects*). Current full-analysis models are not suitable for such analyses as their generated topics are often too coarse and may not even be on target. For example, given a set of tweets about e-cigarette, one may want to find out what topics under discussion are specifically related to *children*. Likewise, given a collection of online reviews about a camera, a consumer or camera manufacturer may be interested in finding out all topics about the camera's *screen*, the targeted aspect. As we will see in our experiments, current full topic models are ineffective for such targeted analyses. This paper studies this problem and proposes a novel *targeted topic model* (TTM) to enable focused analyses on any specific aspect of interest. Our experimental results demonstrate the effectiveness of the TTM.

Keywords

Targeted Topic Modeling; Targeted Aspect; Targeted Analysis; Focused Analysis; Targeted Modeling; Topic Modeling

1. INTRODUCTION

One of the important text mining tasks is to discover the topics discussed in a collection of text documents (or a corpus). Topic modeling is one of the main techniques used for this purpose. So far numerous topic models have been proposed in the literature, which may mine topics only or jointly mine topics and other types of useful information, for example, sentiment information [17, 5, 32].

However, existing models typically perform full analysis

on the entire corpus to discover all topics. This is certainly useful, but it is inevitably coarse. In practice we found that the user almost always also wants to perform deeper and more focused analysis on some specific aspects of the data, which we refer to as *targets*, or *targeted aspects* in this paper. For example, given a set of tweets about *e-cigarette*, the user (or researcher) wants to gain insight into topics that have been discussed about *children*. Here *children* is the targeted aspect. If a topic model can find topics such as *regulations* and *fears* about children smoking e-cigarette that are specifically related to this target, it will be very useful. Formally, the proposed targeted analysis problem is defined as follows (note that we use targeted analysis and focused analysis in this paper interchangeably).

Problem Definition: Given a corpus C of documents of a broad area/domain, discover related topics T of a user-interested aspect (called targeted aspect) represented with a set of keywords S provided by the user.

To solve this problem, a natural approach to start with is to use a regular full-analysis topic model such as LDA [4]. Following the previous example, we can see that e-cigarette is a broad area (with corpus C provided) and *children* is the targeted aspect. Note that the target *children* is represented by some keywords S (e.g., $S = \{\text{"children," "kids"}\}$). After applying LDA on the corpus C to produce a set of topics Q , we find those topics in Q that contain some keywords from S in their top ranked words and study them in order to find the target-related topics T ($T \subseteq Q$). However, this approach is often unsatisfactory due to a few reasons (or *issues*).

1. The user does not know all the keywords that can represent a targeted aspect. In the above example, if the user specifies "children" as the only keyword and miss out other related keywords such as "young" and "minors", he may lose some important topics.

2. It may not find any topics for the user-interested aspect. Because there may be many other more prevalent or dominating topics in the data, the model may not find the related topics of some less frequent aspects. For example, when our targeted aspect is *weight* represented by the keyword "weight" in camera reviews, the word "weight" has a relatively low occurrence frequency as people often mention this concept implicitly in the sentences like "the lens is so heavy" or "its battery is light". A full-analysis topic model such as LDA may not find any topic about *weight* at all.

3. Even if the keywords are frequent, one still may not find good coherent topics due to two reasons: (1) *Topic suppression*: Since the targeted aspect is only one of many aspects discussed in a broad area and a full-analysis topic

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

KDD '16, August 13-17, 2016, San Francisco, CA, USA

© 2016 ACM. ISBN 978-1-4503-4232-2/16/08...\$15.00

DOI: <http://dx.doi.org/10.1145/2939672.2939743>

model generates all topics for all aspects, the related topics of the targeted aspect may be suppressed. Many general words may be ranked at the top. For example, a topic about *children* (the targeted aspect) with topical words like “children”, “kids”, “young” is not informative. (2) *Word intrusion*: Words from other non-targeted aspects’ topics may be *intruder* words appearing in the related topics of the targeted aspect, which makes the detection and understanding of the target-related topics difficult.

The cause of the above problems is related to some properties of topic modeling. First, useful information may not be easily detected under the condition of data sparsity or small data size. That is because classic topic models are unsupervised and governed by the phenomena called *higher order co-occurrence* [13]. As a result, some informative but infrequent words may be ranked low or even can not be correctly grouped. Second, the existing models are not targeted towards any user interest. As identified in [6], the objective functions of topic models may not correlate well with human judgments and needs. For example, a given broad corpus often cover a large number of topics, the topics for the user-interested aspect (i.e., target) may be mixed up with other non-target related topics and become incoherent. Although increasing the number of topics may help, the problem remains and other problems like information fragmentation may show up (the problems will be further discussed in Section 4.4). One may use knowledge-based models for targeted modeling [23, 32], but these models only try to put words related to the user specified keywords in the same topic. They do not distill the topics related to target aspect represented by keywords as we do. They still suffer from the aforementioned issues.

Another intuitive approach to solving the proposed problem is to select a subset of documents from the corpus C that contains at least one keyword $s \in S$ (denoted by C') and apply LDA to the resulting documents in C' . Clearly, this approach has the problem as illustrated in *Issue 1*. It also has *Issue 2* but manifested differently. For example, when the keyword set S is {“weight”} the number of document C' might be so small that a topic model is unable to produce many good topics. Moreover, since it discards many potentially relevant documents, it diminishes the quality of topics and also loses many potentially related topics.

Recently, topic modeling with sparsity has been proposed. Sparse topic models such as those in [8, 33, 18] can (a) identify focused topics of a document, or (b) extract focused words of a topic. However, they are still full-analysis models and not for targeted modeling, and thus still suffer from the aforementioned issues. More discussions will be given in Section 2. Inspired by them, we employ the sparsity idea in designing our new model for targeted/focused analysis.

To address the proposed problems, we designed a new model called the *targeted topic model* (TTM). It is used for focused analysis as it can directly generate related topics of a given targeted aspect. The novelty of TTM is that it models using the entire corpus C while targeted at the user-specified aspect. This enables TTM to discover more related topics and also improve the topic quality because it can better exploit the information from other relevant documents in C that do not contain the given keywords in S .

In summary, this paper makes the following contributions:

1. It proposes the new problem of targeted topic modeling to discover only topics that are related to a user-specified

aspect. To the best of our knowledge, no existing topic model can perform this task. Such targeted/focused analysis is important because not everyone is interested in everything in a corpus. When one is interested in a particular aspect, he/she often wants to perform deeper and focused analysis.

2. It proposes a new probabilistic topic model called *Targeted Topic Model* (TTM) that is able to perform the proposed focused analysis, which is also the first such model.

3. Our experimental results using five real-life datasets and a set of aspects show the effectiveness of the proposed model. It outperforms state-of-art baseline models markedly.

2. RELATED WORK

To our knowledge, there is no existing topic model that is able to perform the proposed targeted analysis as we do. Our work is, however, clearly related to the classic topic models such as PLSA [14] and LDA [4] and their variants. These models have been used to discover hidden thematic structures in a collection of documents or corpus. There are numerous existing models (e.g., [25, 24, 34, 23, 8, 11, 27]). They either identify topics only or jointly identify both topics and other types of information. For example, while both LDA and PLSA only identify topics, [30, 22] jointly model both topics and ratings in reviews. [24, 10] model labeled data with the class information. [19, 15] conduct time-series analysis of topics. However, as we indicated in the introduction section, all these models and their variants are full analysis models. They aim to find all topics in the corpus, and none of them is able to perform targeted analysis based on only a specific aspect that is of interest to users. Existing research also proposed several knowledge-based topic models, which can incorporate prior domain knowledge in topic modeling [1, 23, 9, 32] to generate better results. But they are also full-analysis models, and do not help discover related topics of the user interested aspect.

Our work is also related to sparse topic models. The modeling of sparsity can represent skewed distributions, e.g., a topic usually focuses on a narrow range of words instead of a wide range of them in the vocabulary [18]. This type of models is inspired by the influential power of the asymmetric Dirichlet prior [28]. Existing sparse topic models can be categorized into two types: (1) those that discover the salient (focused) topics of a specific document [8, 33], and (2) those that discover representative (focused) terms of a particular topic [29]. Researchers also tried to achieve both at the same time in a hybrid manner [18, 2]. However, it is important to note that focused topics or terms in sparse models are entirely different from our targeted analysis because they aim to achieve very high probabilities for a small number of topics in a document or for a small set of words/terms in a topic. They are still full-analysis models and cannot focus their modeling only on a user-specified aspect.

Although not directly related, there are other works of non-probabilistic models with sparse coding [35, 20, 7, 31]. However, they function quite differently in the goal or the methodology compared to probabilistic generative models. For example, they do not need to tackle the knotty issue of decoupling sparsity and smoothing [29] for sparsity realization in probabilistic models. Furthermore, they have been shown with poorer performances as the number of topics increases and are inferior compared to the probabilistic sparse topic models reported in [18]. Most importantly, like other models, they do not do targeted analysis like we do.

topic t is chosen from θ^r . The total number of topics is $|T|$ and these topics are the topics related to the target. After that a word is emitted from the selected topic by φ_t^r . When $r=0$ the generative process is similar but TTM does not further generate multiple topics of the non-targeted aspects because they are not related to the user-interested aspect. Thus a word w_i is emitted directly from φ^{ir} . In other words, the words in the irrelevant documents are drawn from only one (irrelevant) topic.

3.2 Targeted Modeling with Sparsity

This sub-section presents how TTM achieves the targeted modeling by exploiting the idea of sparsity. Traditionally, sparsity in topic modeling indicates that a topic usually focuses on a narrow range of words rather than a wide range of them in the vocabulary, or a document focuses on a very small number of topics [29, 18]. Similarly, in TTM we use a similar concept called *aspect sparsity*, which consists of two parts: *document-aspect sparsity* and *targeted-aspect sparsity*, which are detailed below.

Since a sentence is regarded as a document, TTM already assumes that each document focuses on only one aspect. The document-aspect sparsity is then naturally achieved. Additionally, the idea of targeted modeling with sparsity is based on two importance observations.

First, the targeted aspect may only be a small part or a minority among all aspects in a given corpus. For example, *Children* is only one aspect in the e-cigarette domain, which has a large number of other aspects such as *Elderly*, *Vaporizer* and *Health*. This observation gives raise to the targeted-aspect sparsity.

Second, since each document is coupled to one aspect (using document-aspect sparsity), if we can better represent the targeted aspect, we should be able to better extract its relevant documents. If we can extract more relevant documents for the targeted aspect, we are likely to discover other more important words to better represent the target. As a result, better and more topics are more likely to be generated because of better representative words and more relevant documents about the target are identified. Thus, we believe jointly modeling of relevance status r and the targeted aspect can benefit the topic discovery.

As introduced in section 2, sparsity has been used to represent the skewed data distribution in topic modeling. It provides a possible way to represent the property of minority (the first observation above) and to model the targeted aspect with document relevance in a unified framework (the second observation above). We therefore follow this direction and propose the idea of *biased sparsity* to help achieve the targeted modeling and the aspect sparsity.

3.2.1 Biased Sparsity

Note that a target (aspect) is essentially represented by words. So the problem is how to automatically discover representative words of the targeted aspect in a joint modeling manner. As discussed above, a targeted aspect is sparse among all aspects. From it we can further posit that (in most cases) a targeted aspect is also sparse compared to the combination of all non-targeted aspects. When combining this statement with the scenario that *a targeted aspect is represented by words*, we can conclude that the number of important words for distinguishing the targeted aspect from non-targeted aspects is in a narrow range. That is,

the representative words for the targeted aspect are sparse. These words are denoted by $V_{r=1}$. However, the representative words from the combination of all other non-targeted aspects are probably not sparse because that combination needs to contain almost all possible words for describing all other aspects. These words are denoted by $V_{r=0}$. Now we introduce the proposed *biased sparsity*, which is used for realizing the above idea.

The proposed *biased sparsity* approach is to make the word sparsity biased much more towards the targeted aspect (i.e., $|V_{r=1}| \ll |V|$) and keep the non-targeted aspects almost non-sparse (i.e., $|V_{r=0}| \approx |V|$). In other words, the word distribution of the targeted aspect (denoted by φ^r) only focuses on a small number of representative words. But the word distribution of non-targeted aspects (denoted by φ^{ir} , where ir means a unified irrelevant topic) contains almost all possible words. With this setting, only those words that simultaneously satisfy both the following two conditions can be selected as representative words for the targeted aspect: (a) they are semantically correlated with the known words of the target (achieved by the power of topic modeling) and (b) they can distinguish the targeted aspect from non-targeted aspects (constrained by *biased sparsity*).

3.2.2 Targeted Modeling

This sub-section demonstrates the targeted modeling in TTM, with the incorporation of biased sparsity. Here we follow the previous examples in section 3.1 to illustrate. Recall that $S=\{\text{“children”}\}$ and the document $d1$ containing the keyword ($x_{d1}=1$) is known to be relevant ($r_{d1}=1$). Although the word “children” is not in document $d2$, the word “gateway” (also in $d1$) serves as a bridge to connect the words in $d2$ via topic modeling. If the “gateway” has been identified as a discriminative word for target *children*, it makes $d2$ more probable to be relevant ($r_{d2}=1$) as $d2$ also contains the word “gateway”. In this case, even though $d2$ has no keyword “children” it still has a high probability to be relevant. In contrast, although “smoking” appears in both documents $d1$ and $d3$, if “smoking” is not identified as a discriminative word for aspect *children*, $d3$ will be mostly treated as irrelevant ($r_{d3}=0$).

Biased sparsity is the key to identify discriminative words for the targeted aspect, which is done by imposing the sparse constraint ($|V_{r=1}| \ll |V|$ and $|V_{r=0}| \approx |V|$) discussed above. Specifically, when the target is *children*, “gateway” will be a word included in φ^r and “smoking” will not be. An explanation is that “smoking” is also widely used by other non-targeted aspects like *Elderly* and *Health*, while “gateway” is used in *Children* more often. Thus, “gateway” is more discriminative for *children* but “smoking” is too general.

Another crucial factor is the relevance status r . When more discriminative words (like “gateway”) are included in φ^r , they in return increase the overall probability of the documents (like $d2$) that contain those words (like “gateway”) to be relevant ($r=1$). When those documents are identified as relevant, new representative words (like “kids” in $d2$) for the target aspect are also found and added to φ^r . Note that the growing of words in φ^r will not be an endless process because of the sparsity constraint. After that these new words (like “kids”) in φ^r will help detect other relevant documents. This process shows the strength of joint modeling of the target aspect (using biased sparsity) and the relevance status r in TTM. It also explains why TTM can finally shape bet-

ter topics. Because it is able to exploit the information in other relevant documents in C even without knowing the additional keywords such as “kids”, “young” and “minors”, as they are automatically found and involved in the modeling process.

In a nutshell, the relevance variable r and the biased-sparsity related variable φ^r and φ^{ir} function together to ensure the property of aspect sparsity. Specifically, the encoding of φ^r and φ^{ir} to achieve *biased sparsity* is based on the implementation of spike-and-slab prior.

3.2.3 Spike-and-Slab Prior

The spike-and-slab prior is incorporated in the probabilistic topic model to realize a switcher-like “on” and “off” selector. This prior is first introduced in [21] and is recently reported as an effective way to reduce the model uncertainty [16]. It can also be used to decouple the sparsity and smoothing problem in Dirichlet distribution [3]. It has been used in topic modeling for word sparsity [29, 18]. In brief, “spike” controls the selection of a word, while “slab” smooths the word selected by “spike”.

The spike-and-slab prior and other related components are together formulated as random variables in our model for the implementation of the biased sparsity. The related variables are ω , p , q , β^r , δ , ϵ and φ^r . For every word $v \in V$, $\beta_v^r \in \{0, 1\}$ is the specific word selector. When $\beta_v^r=1$ a word is selected and $\beta_v^r=0$ a word is not selected. Note that in TTM, for a keyword $s \in S$ the β_s^r is set to 1. It is intuitive as those keywords are given by the user for specifying the targeted aspect. Different from β^r , β^{ir} is a predefined prior. It is symmetric for all other words except the known keywords $s \in S$, because the known keywords given by the user is very unlikely to be irrelevant to the target.

Unlike non-probabilistic models, one knotty issue for probabilistic models is the decoupling of sparsity and smoothing [29]. Motivated by [18], we incorporate a weak smoothing prior ϵ in addition to the regular smoothing prior δ .

Double Sparsity: In addition to the *target sparsity*, the words for a specific topic related to the targeted aspect is also sparse. That is reasonable because TTM aim at discovering more fine-grained topics and the words for describing such a topic are naturally sparse. Likewise, the $\varphi_{t,v}^r$ where $t \in T$ and $v \in V$ is also encoded with a spike-and-slab prior in TTM. Together with the target sparsity, the sparsity of topical words enables the model to generate more coherent topics of the targeted aspect.

3.3 Inference

We use Gibbs Sampling [12] for model inference. The conditional distributions are shown in Equations 1, 2, 3, 4 and 5 (see Table 1 for the meanings of notations). The * symbol means the summation of all instances of the corresponding variable, for example, $C_{r,t,*}^{RTW(-i)}$ is the number of all words under relevance status r and topic t except the word in position i .

First, we sample the relevance status r for every document m , where $r \in R$ and $m \in M$.

$$P(r_m = c | x_m = d, \Theta), \Theta = \{\mathbf{r}_m^{-i}, \mathbf{w}, \boldsymbol{\pi}, \boldsymbol{\beta}^r, \boldsymbol{\beta}^{ir}, \boldsymbol{\gamma}, \delta, \epsilon\}$$

$$\propto \begin{cases} g(c, \Theta) & d = 0, c = 0 \text{ or } 1 \\ \lambda \times g(c, \Theta) & d = 1, c = 1 \\ (1 - \lambda) \times g(c, \Theta) & d = 1, c = 0 \end{cases} \quad (1)$$

$$g(c, \Theta) = \begin{cases} \frac{C_c^{R(-m)} + \gamma}{C_*^{R(-m)} + |R|\gamma} \times \frac{\prod_v \Gamma(\beta_v^r C_{c,v}^{RW(-m)} + \beta_v^r \delta + \epsilon + f_{c,m,v})}{\Gamma(\sum_v (\beta_v^r C_{c,v}^{RW(-m)} + f_{c,m,v}) + |\beta_*^r| \delta + |V|\epsilon)} & c = 1 \\ \frac{C_c^{R(-m)} + \gamma}{C_*^{R(-m)} + |R|\gamma} \times \frac{\prod_v \Gamma(C_{c,v}^{RW(-m)} + \beta_v^{ir} + f_{c,m,v})}{\Gamma(\sum_v C_{c,v}^{RW(-m)} + f_{c,m,v} + \beta_*^{ir})} & c = 0 \end{cases} \quad (2)$$

Second, we sample the term selector β_v^r , where $v \in V$. $|\beta_*^r|$ is the sum of the values of all such term selectors.

$$P(\beta_v^r = b | \boldsymbol{\beta}^{r(-v)}, \mathbf{r}, \mathbf{w}, \delta, \epsilon, p, q) \propto \begin{cases} \Gamma(C_{r=1,*}^{RW} + \delta + \epsilon) \times \Gamma(|\beta_*^{r(-v)}| \delta + |V|\epsilon + C_{r=1,*}^{RW(-v)}) \\ \times \Gamma(|\beta_*^{r(-v)}| \delta + \delta + |V|\epsilon) \times (p + |\beta_*^{r(-v)}|) & b = 1 \\ \Gamma(\delta + \epsilon) \times \Gamma(|\beta_*^{r(-v)}| \delta + \delta + |V|\epsilon + C_{r=1,*}^{RW(-v)}) \\ \times \Gamma(|\beta_*^{r(-v)}| \delta + |V|\epsilon) \times (q + |V| - |\beta_*^{r(-v)}| - 1) & b = 0 \end{cases} \quad (3)$$

Third, we sample a topic of the word in position i . We do it for all words in the corpus. $|\beta_{t,*}^r|$ is the sum of the values of all such term selectors in topic t .

$$P(z_i = t | \mathbf{z}^{-i}, \mathbf{r}, \mathbf{w}, \alpha, \boldsymbol{\beta}^r, \boldsymbol{\beta}^{ir}, \delta, \epsilon) \propto \begin{cases} \frac{C_{r_i,m,t}^{RMT(-i)} + \alpha}{C_{r_i,m,*}^{RMT(-i)} + |T|\alpha} \times \frac{(\beta_{t,v}^r C_{r_i,t,*}^{RTW(-i)} + \beta_{t,v}^r \delta + \epsilon)}{C_{r_i,t,*}^{RTW(-i)} + |\beta_{t,*}^r| \delta + |V|\epsilon} & r_i = 1 \\ \frac{C_{r_i,m,t}^{RMT(-i)} + \alpha}{C_{r_i,m,*}^{RMT(-i)} + |T|\alpha} \times \frac{(C_{r_i,t,*}^{RTW(-i)} + \beta_{t,v}^{ir})}{C_{r_i,t,*}^{RTW(-i)} + |\beta_{t,*}^{ir}|} & r_i = 0 \end{cases} \quad (4)$$

Last, we sample the term selector $\beta_{t,v}$, where $t \in T$ and $v \in V$.

$$P(\beta_{t,v}^r = s | \boldsymbol{\beta}_t^{r(-v)}, \mathbf{z}, \mathbf{r}, \mathbf{w}, \delta, \epsilon, p, q) \propto \begin{cases} \Gamma(C_{r=1,t,*}^{RTW} + \delta + \epsilon) \times \Gamma(|\beta_{t,*}^{r(-v)}| \delta + |V|\epsilon + C_{r=1,t,*}^{RTW(-v)}) \\ \times \Gamma(|\beta_{t,*}^{r(-v)}| \delta + \delta + |V|\epsilon) \times (p + |\beta_{t,*}^{r(-v)}|) & s = 1 \\ \Gamma(\delta + \epsilon) \times \Gamma(|\beta_{t,*}^{r(-v)}| \delta + \delta + |V|\epsilon + C_{r=1,t,*}^{RTW(-v)}) \\ \times \Gamma(|\beta_{t,*}^{r(-v)}| \delta + |V|\epsilon) \times (q + |V| - |\beta_{t,*}^{r(-v)}| - 1) & s = 0 \end{cases} \quad (5)$$

4. EXPERIMENT

4.1 Experimental Setup

Data and targeted aspects: Five real-world data sets in different domains are used in our experiments, namely, E-Cigarette, Cigar, Camera, Cell-Phone and Computer. The first two data sets are tweets collected from Twitter in October 2014. Specifically, E-Cigarette and Cigar are two types of tobacco-related products, which are the research areas of the last author from health science. The last three datasets are product reviews of three popular electronic products. We crawled the reviews from Amazon.com. More detailed information about the data sets is presented in Table 2.

Three targeted aspects are picked from each domain for targeted analysis. The aspects cover a wide range of diverse areas: Some of them are typical or frequent aspects in its domain like *Children* in E-Cigarette and *Lens* in Camera. Some are small or infrequent aspects like *Weight*, *Warranty* and *Horse*¹. Note that one infrequent topic is specially chosen for each domain, listed as the last one in Table 2.

Dataset (Area)	Source	Size	Targeted Aspects
E-Cigarette	Twitter	50k	Children, Vape, Health
Cigar	Twitter	50k	Event, Box, Horse ¹
Camera	Amazon	10k	Lens, Screen, Weight
Cell-Phone	Amazon	10k	Sound, Case, Battery
Computer	Amazon	10k	Monitor, Software, Warranty

Table 2: Five datasets, targeted aspects, and initial documents (tweets or review sentences)

Parameter Setting: For the hyper-parameter setting, we place: $q = p = 1$ for a uniform Beta; $\gamma = 1$, $\alpha = 1$, $\beta^{tr} = \delta = 0.001$ and $\epsilon = 1.0 \times 10^{-7}$. The models for comparison are also with the same setting. λ in Equation 1 is set to 1 for TTM.

4.2 Baseline Models for Comparison

To evaluate our proposed targeted topic model TTM, we compare it with the following baseline models:

LDA: LDA is a well-known topic model [4]. It finds all topics in a corpus. To identify topics that are relevant to the targeted aspect, we manually inspect all resulting topics from LDA and find the subset of relevant topics. Note that the targeted aspect may be split into multiple topics by LDA. This is a labor-intensive and tedious process if there is a large number of topics from LDA.

LDA*: We still use LDA for topic generation, but instead of manual inspection to find related topics, we use keywords in S to search for relevant topics. We refer to this as the *search strategy*, which eliminates the tedious manual process. In this approach the number of topics T can be set large because only the retrieved topics will be analyzed. We search only the top 20 words of each topic.

DS-LDA: This is a state-of-art probabilistic sparse topic model [18] that models both the sparsity of topic mixtures (mining salient topics of a document) and topical words (mining representative words of a topic). We follow the implementation in [18] and refer to this as dual-sparse topic model, DS-LDA, in this paper. Like that for LDA, the relevant topics for DS-LDA are found via manual inspection.

DS-LDA*: Like LDA*, we use keywords S and adopt the *search strategy* to find possible relevant topics from all topics generated by DS-LDA.

SS-LDA: We also use a single-sparse topic model named SS-LDA in this paper for comparison, to see whether the direct injection of the word sparsity can help our task, because it enables the topical word to be more focused. The realization is similar to DS-LDA but it only addresses the sparsity of topical words. Likewise, it requires manual inspection.

SS-LDA*: Like DS-LDA*, a *search strategy* is adopted to find possible relevant topics from all topics.

LDA-PD: This model runs LDA only on the documents in each dataset that contains one or more keywords from S . For instance, to find topics about targeted aspect *Children*, we use the keyword “children” to search for tweets in the *E-Cigarette* tweet corpus. After that we run LDA on the resulting tweets to find topics. We name this approach LDA-PD (where PD means Partial Data) for short.

¹A special topic: there was an well-known race horse called “Cigar” who died in October 2014. It is covered in the *Cigar* dataset as people talked about it in social media using its name. It is an evidently infrequent aspect in the Cigar data.

4.3 Quantitative Evaluation

Since our goal is to discover topics for a given targeted aspect but the correct number of topics and the number of terms/words under each topic are unknown, a natural evaluation metric is to give the precision results at different rank position n , called *Precision@n* (or $P@n$ for short).

For the evaluation of the first two datasets, two experts from our health science collaborator’s team who are specialized in the tobacco-related products and social media were invited to judge the results. Two human labelers who are familiar with Amazon product reviews labeled the results from the other three datasets. Cohen’s Kappa scores of the two genres are 0.792 and 0.822 respectively.

Evaluation Measure: We use a normalized form of precision (defined in Equation 6) that can evaluate both the correctness of the topical words and the number of detected topics in a unified manner as both are important. Specifically, after the human judgment of all models results, the number of Maximum Unique Topics (MUT) can be obtained. For example, when the target is X , *Model-1* finds one unique topic A (which is correct), *Model-2* finds two unique topics A, B and *Model-3* finds three unique topics A, B, C then MUT is 3. In another case, when *Model-1* finds two unique topics A, B, *Model-2* finds two unique topics B, C and *Model-3* finds topics B, C then the MUT is also 3.

$$P_{(i)}@n = \frac{\sum_{st}^{ST} \#C_{(i)st}(correct@n)}{\sum_{mt}^{MUT} \#C_{mt}(words@n)} \quad (6)$$

In Equation 6, $P_{(i)}@n$ indicates the *precision@n* for model (i), given the targeted aspect. $\#C_{(i)st}(correct@n)$ is the number of correct words found in the topic st , given that there are ST topics found by model i . $\#C_{mt}(words@n)$ is the maximum number of correct words from all models.

This evaluation measure is fair and reasonable because a model may only find one correct topic with high topical word precision but miss some correct topics. Note that if there are more than one identical topics generated by a model, their average score is used. If there are multiple topics mixed in a single topic generated by a model, we use the best topic based on the number of relevant words in the top 20 words.

Two Comparison Settings: Two different experiment settings are used for comparison due to different properties of the candidate models. The reasons and differences will be elaborated in the following sub-sections.

4.3.1 Precision in Setting One

Here we compare LDA, DS-LDA, SS-LDA and LDA-PD with TTM. In this setting the *search strategy* is not used. Instead, the annotators were asked to go through all top (20) words in the generated topics. For LDA-PD, we use the target keyword itself (e.g., “children”) to extract documents. For TTM, we also use the target keyword for targeted modeling. The topic number T in LDA-PD and TTM is set to 5 or 10 because we have no prior information about the number of target-related topics but we know that it intuitively depends on the prevalence of the targeted aspect (e.g., in the Camera domain the targeted aspect *Screen* is likely to be frequent and have more topics than the infrequent aspect *Weight*). In general, T is set to 5 for *infrequent* targets and to 10 for other more *frequent* targets. Although T could be 5 or 10 according to the targeted aspect, the same value of T is used for the comparison of results for both LDA-PD and TTM. Likewise, for LDA, DS-LDA and SS-LDA, T is

Data Set	Aspects	LDA (T-15 30)			DS-LDA (T-15 30)			SS-LDA (T-15 30)			LDA-PD (T-5 10)			TTM (T-5 10)		
		P@5	P@10	P@20	P@5	P@10	P@20	P@5	P@10	P@20	P@5	P@10	P@20	P@5	P@10	P@20
E-Cig	children	0.15	0.14	0.13	0.15	0.10	0.07	0.15	0.11	0.09	0.48	0.35	0.27	0.73	0.59	0.41
	vape	0.23	0.30	0.24	0.16	0.10	0.09	0.36	0.32	0.22	0.59	0.42	0.22	0.76	0.57	0.30
	health	0.48	0.39	0.30	0.32	0.19	0.18	0.38	0.25	0.17	0.38	0.28	0.20	0.55	0.42	0.30
Cigar	box	0.46	0.35	0.22	0.15	0.13	0.08	0.20	0.23	0.15	0.53	0.51	0.34	0.65	0.54	0.41
	event	0.30	0.28	0.31	0.24	0.14	0.09	0.36	0.18	0.15	0.30	0.39	0.33	0.64	0.58	0.45
	horse	0.27	0.20	0.26	0.15	0.10	0.06	0.55	0.43	0.24	0.73	0.46	0.34	0.88	0.61	0.37
Camera	screen	0.51	0.35	0.35	0.12	0.15	0.13	0.30	0.25	0.15	0.44	0.31	0.27	0.74	0.61	0.40
	lens	0.25	0.16	0.13	0.21	0.17	0.12	0.31	0.27	0.19	0.36	0.34	0.30	0.53	0.36	0.34
	weight	0.20	0.20	0.15	0.04	0.02	0.04	0.16	0.10	0.08	0.56	0.44	0.25	0.72	0.60	0.44
Cell-Phone	sound	0.58	0.40	0.30	0.27	0.17	0.14	0.50	0.35	0.21	0.52	0.37	0.29	0.75	0.60	0.40
	case	0.56	0.59	0.53	0.14	0.18	0.15	0.23	0.16	0.14	0.61	0.44	0.38	0.66	0.56	0.40
	battery	0.41	0.30	0.20	0.30	0.20	0.14	0.16	0.12	0.12	0.48	0.38	0.25	0.60	0.42	0.42
Computer	monitor	0.72	0.60	0.42	0.11	0.11	0.06	0.57	0.39	0.26	0.66	0.54	0.43	0.70	0.56	0.45
	software	0.48	0.39	0.35	0.24	0.21	0.14	0.36	0.30	0.33	0.36	0.34	0.30	0.60	0.39	0.35
	warranty	0.31	0.23	0.19	0.03	0.03	0.04	0.23	0.13	0.08	0.47	0.38	0.31	0.63	0.53	0.43
average score		0.39	0.32	0.27	0.18	0.13	0.10	0.32	0.24	0.17	0.50	0.40	0.30	0.68	0.53	0.39
improvement by TTM		+0.28	+0.21	+0.12	+0.50	+0.40	+0.29	+0.35	+0.29	+0.22	+0.18	+0.13	+0.09	n/a	n/a	n/a

Table 3: Precisions of setting one. The last two rows are (a) the average scores of all targeted aspects of all topics and (b) the improvement achieved by TTM over other models respectively

set to 15 or 30 (and choose the one produces higher precision results). The numbers are larger because they do not directly generate topics for the targeted aspect like LDA-PD and TTM. They also produce topics for other non-targeted aspects.

The precision results at the rank position of 5, 10 and 20 are reported in Table 3 and we observe the following:

1. TTM significantly outperforms other models. The average scores and the improvements by TTM are presented in the last two rows in Table 3. Among them, LDA-PD obtains the second best scores. Compared to LDA-PD, TTM has two key advantages, which are also the main reasons for the higher scores. (a) TTM can identify and use other relevant documents in modeling the topics of the targeted aspect with the power of biased sparsity, while LDA-PD can only generate topics from the subset of documents selected by user-specified keywords. (b) TTM can find more or better topics for the targeted aspect than LDA-PD because LDA-PD discards many relevant documents, resulting in information loss. Concrete examples will be further given in sub-section 4.4.
2. LDA-PD is the second best because it rules out the irrelevant documents that can interfere with good topic identification. But then it also loses relevant information as we discussed above. Thus, it results in inferior performance to TTM. Moreover, LDA-PD works poorly when it comes to some *infrequent* targets because relevant documents about these targets are not easy to extract as they may not contain the keywords.
3. LDA is neither as good as TTM nor LDA-PD. But it is better than both DS-LDA and SS-LDA. The DS-LDA that models both the sparsity of topic mixture and topical words has the worse performance. We will analyze the reason in the next sub-section.

One might argue that we can keep increasing T to a larger number to make further improvement. That is possible, but it becomes rather messy and impractical (labor intensive and time consuming) for the user to manually inspect all generated topics. An alternative is to apply the *search strategy*

with increased T , which leads to our following evaluation of setting two.

4.3.2 Precision in Setting Two

This subsection compares LDA*, DS-LDA* and SS-LDA* against TTM. Different from setting one, here the annotators utilize the *search strategy* to identify relevant topics from all models except TTM. That is, the targeted aspect keyword is used to search in the top 20 topical words in each topic to find possibly relevant topics to the target. Only those resulting topics are evaluated. Since search is used, the number of topics T can be large. For LDA* we set T to 15, 30 and 50. For DS-LDA* and SS-LDA* we follow the same setting but their results are not good as LDA*. Due to the limited space we show the $T=50$ for DS-LDA* and SS-LDA* only.

The precision results at the rank position of 5, 10 and 20 are reported in Table 4 and we observe the following:

1. TTM again outperforms the other models by a large margin. We can see that the results from the search strategy is also poor. TTM generates better topics without human invention.
2. The increase of T for LDA* helps improve the performance but the enhancement is in a decreasing trend. Although the precision scores improve with increased T , the improvement from LDA*(T-30) to LDA*(T-50), which are 5.0%, 2.5%, 0.3% for $P@5$, $P@10$, $P@15$, drops from LDA*(T-15) to LDA*(T-30), which are 7.1%, 5.5%, 3.7% respectively. One reason is that with the increment of T , although more possible relevant topics may be generated, the *search strategy* can only find a subset of the relevant topics because the complete set of keywords for the targeted aspect is unknown. For instance, for aspect *Children* there might be new topics formed about children with a larger T but they do not contain the keyword “children”. Instead, the new topics may contain other related keywords such as “kids”, “young”, and “minors”.
3. Neither SS-LDA nor SS-LDA* can produce better results than LDA or LDA*. The reason is that when

Data Set	Aspects	LDA*(T-15)			LDA*(T-30)			LDA*(T-50)			DS-LDA*(T-50)			SS-LDA*(T-50)		
		P@5	P@10	P@20	P@5	P@10	P@20	P@5	P@10	P@20	P@5	P@10	P@20	P@5	P@10	P@20
E-Cig	children	0.08	0.05	0.04	0.13	0.13	0.11	0.20	0.13	0.08	0.05	0.30	0.04	0.16	0.12	0.08
	vape	0.12	0.14	0.13	0.17	0.14	0.17	0.28	0.30	0.23	0.08	0.06	0.06	0.36	0.32	0.19
	health	0.13	0.13	0.12	0.17	0.12	0.06	0.32	0.23	0.16	0.14	0.11	0.11	0.13	0.15	0.8
Cigar	box	0.30	0.25	0.20	0.20	0.24	0.20	0.43	0.26	0.16	0.08	0.06	0.08	0.33	0.29	0.18
	event	0.16	0.08	0.12	0.24	0.20	0.18	0.19	0.18	0.14	0.12	0.08	0.05	0.8	0.60	0.70
	horse	0.05	0.10	0.11	0.25	0.18	0.24	0.50	0.42	0.31	0.31	0.05	0.04	0.03	0.5	0.28
Camera	screen	0.08	0.06	0.08	0.08	0.06	0.06	0.08	0.05	0.06	0.23	0.20	0.14	0.15	0.11	0.06
	lens	0.13	0.10	0.08	0.18	0.16	0.12	0.12	0.11	0.10	0.17	0.13	0.11	0.27	0.16	0.11
	weight	0.08	0.08	0.08	0.16	0.14	0.10	0.12	0.10	0.07	0.04	0.04	0.03	0.13	0.1	0.13
Cell-Phone	sound	0.20	0.15	0.13	0.47	0.37	0.27	0.63	0.45	0.28	0.17	0.12	0.10	0.3	0.2	0.15
	case	0.38	0.35	0.31	0.55	0.51	0.36	0.39	0.37	0.30	0.13	0.11	0.21	0.2	0.2	0.18
	battery	0.16	0.12	0.08	0.32	0.22	0.16	0.32	0.20	0.16	0.16	0.08	0.06	0.12	0.12	0.11
Computer	monitor	0.59	0.56	0.40	0.61	0.51	0.38	0.61	0.46	0.38	0.20	0.10	0.09	0.31	0.24	0.18
	software	0.32	0.24	0.19	0.12	0.11	0.09	0.18	0.15	0.11	0.08	0.04	0.04	0.16	0.12	0.1
	warranty	0.10	0.08	0.05	0.30	0.23	0.18	0.33	0.30	0.20	0.07	0.03	0.03	0.23	0.18	0.11
average score		0.19	0.17	0.14	0.26	0.22	0.18	0.31	0.25	0.18	0.12	0.10	0.08	0.28	0.21	0.22
improvement by TTM		+0.49	+0.37	+0.26	+0.41	+0.31	+0.22	+0.36	+0.28	+0.22	+0.56	+0.43	+0.32	+0.40	+0.32	+0.18

Table 4: Precisions of setting two. The last two rows are (a) the average scores of all targeted aspects of all topics and (b) the improvement achieved by TTM over other models respectively.

the word sparsity can find focused words to represent a topic, the number of related topics (to the target) found by SS-LDA is fewer than that by LDA. In other words, SS-LDA might form better topics but they are less related to the targeted aspect. DS-LDA has the worse results because it may find more focused topics for individual documents but not for the targeted aspect of user interest.

4.4 Qualitative Evaluation

This section presents the qualitative evaluation. We show several resulting topics in Tables 5 and 6 to give a flavor of each system. The domain and the targeted aspect are shown at top of the tables above the model names and topic names (given by us). Incorrect words of a topic are italicized and marked in red. Notice that the targeted aspect keyword itself is also shown but it is excluded from the computation of precision in the previous section because it is already known.

4.4.1 Example One: E-Cigarette and Children

E-Cigarette (e-cig) is a key area studied by our collaborators from the health science and *Children* is one aspect that they are highly interested in and thus want to know its topics discussed on Twitter. Table 5 shows the topical words discovered by different models. The models are attached with the setting in the previous section so the topic number T is not explicitly given in the table. We explain the results below.

(1) **Topic fears:** It means the fears or concerns about children using e-cig. By comparison, we see that TTM generates a clearer and more reasonable topic. While the other methods generate incoherent topics with many wrong words, the words in TTM are more informative and interpretable, because the words like “nicotine”, “gateway”, and “sale” can better indicate the reasons of the fears as well as what people are actually concerned about for the children. That is, people worry about the “nicotine” in e-cig that is bad for kids; they are afraid that e-cig becomes a “gateway” for their kids to smoking, and the increasing “sale” to the young.

(2) **Topic regulation:** It is about the regulation of e-cig

for children, mostly about the policy of purchase. Similar to topic *fears*, TTM finds more informative words like “laws” and “safe”. Particularly, the word “young” is also included in the topic *regulation* of TTM, which is meaningful. It can infer that other documents that contain the unknown/unprovided keyword “young” are probably identified by TTM as relevant to the targeted aspect children (i.e., $r=1$) and thus involved in the topic generation of the target. Because one short sentence is unlikely to mention the keyword “children” and unknown keyword “young” at the same time (they are semantically similar). The documents containing the word “young” is unlikely to also contain “children” and vice versa. This is an advantage that LDA-PD cannot achieve. Those documents (containing “young”) identified by TTM as relevant help generate a more coherent topic for the targeted aspect.

Here we also analyze the topics formed by LDA and LDA* (T-50) to demonstrate some aforementioned problems. (a) Although the topic *regulation* in LDA looks good with many good words grouped, the topic is actually not very informative as the words “children, kids, minors” all ranked high but they do not tell anything about a topic related to e-cig and children. (b) Since these words are grouped together, it may lead to missing of some good topics of the targeted aspect. (c) When comes to LDA*, although T is set to 50 this method actually finds only 2 topics that contain “children” by using the *search strategy*. (d) Also in the *regulation* topics, one may notice that the word “minors” which is in LDA (where its topic number is 15 or 30) is not included in LDA*(T-50) any more. However, since we did not previously know all those keywords (e.g., “minors”, “young”), the topic is not found with $T = 50$, i.e., a related topic is unfortunately lost.

4.4.2 Example Two: Camera, Screen and Weight

The previous sub-section presents the example from tweets, which mainly reveals the trend of a discussed topic in Twitter. Now let us take a look at online reviews in the Camera domain. The data is from Amazon.com and the analysis is also related to opinion mining and sentiment analysis.

Domain: E-Cigarette											
Targeted aspect: Children											
TTM		LDA-PD		LDA* (T-50)		LDA		DS-LDA		SS-LDA	
fears	regulation	fears	regulation	fear	regulation	fears	regulation	fears	regulation	fears	regulation
children	states	children	buy	fears	buy	fears	buy	children	<i>vaping</i>	<i>lacks</i>	regulation
fears	buy	fears	states	<i>research</i>	legally	<i>city</i>	kids	<i>evidence</i>	regulation	fears	<i>trending</i>
nicotine	live	smoked	children	concerns	children	<i>ban</i>	legally	nicotine	<i>vaporizer</i>	<i>report</i>	<i>two</i>
nightmare	legally	sale	live	<i>cloud</i>	kids	<i>council</i>	children	<i>pour</i>	<i>view</i>	<i>stories</i>	<i>kanavape</i>
sales	<i>atlanta</i>	<i>real</i>	legally	<i>parents</i>	purchase	<i>media</i>	minors	<i>starter</i>	<i>prisoner</i>	<i>watch</i>	washington
data	young	<i>report</i>	<i>online</i>	<i>fulled</i>	state	<i>social</i>	<i>columbia</i>	<i>council</i>	<i>cut</i>	prison	safer
<i>age</i>	smoked	easy	<i>press</i>	<i>batteries</i>	live	<i>public</i>	<i>district</i>	<i>fire</i>	<i>part</i>	<i>slapped</i>	<i>drinking</i>
gateway	<i>online</i>	<i>figures</i>	<i>washington</i>	experimenting	<i>nicotine</i>	<i>study</i>	<i>states</i>	<i>television</i>	washington	<i>world</i>	<i>thing</i>
reason	laws	<i>top</i>	<i>smokeless</i>	<i>progress</i>	<i>nightmare</i>	addition	purchase	virus	state	evaporation	<i>cells</i>
fight	safe	<i>make</i>	<i>post</i>	smoked	<i>press</i>	<i>bans</i>	live	<i>business</i>	<i>social</i>	<i>set</i>	data

Table 5: Topics of aspect *children* under E-Cig. Errors are italicized and marked in red.

Domain: Camera											
Targeted aspect: Screen						Targeted aspect: Weight					
TTM			LDA-PD		LDA	TTM			LDA-PD		DS-LDA*
picture	menu	imaging	picture	menu	picture	lens	battery	carrying	lens	battery	-
picture	easy	image	picture	screen	picture	lens	battery	easy	weight	weight	<i>picture</i>
color	control	shutter	photo	<i>touch</i>	photo	weight	weight	carry	lens	battery	compact
shot	manual	speed	<i>computer</i>	control	<i>easy</i>	<i>canon</i>	light	<i>picture</i>	<i>zoom</i>	light	<i>stabilization</i>
sharp	photo	stabilization	video	menu	<i>video</i>	size	<i>life</i>	compact	<i>extra</i>	<i>image</i>	weight
feature	function	shooting	black	button	feature	heavy	travel	case	size	<i>easy</i>	<i>place</i>
clear	menu	<i>lens</i>	<i>show</i>	easy	clear	smaller	<i>rechargeable</i>	small	<i>add</i>	<i>shot</i>	<i>absolute</i>
resolution	condition	picture	full	<i>small</i>	pic	long	lightweight	<i>extra</i>	light	<i>life</i>	<i>ease</i>
grainy	macro	frame	<i>turn</i>	<i>comcorder</i>	<i>fun</i>	<i>digital</i>	<i>video</i>	big	<i>carry</i>	<i>control</i>	<i>instruction</i>
<i>action</i>	<i>learn</i>	<i>feature</i>	<i>record</i>	<i>change</i>	<i>wonderful</i>	<i>zoom</i>	<i>image</i>	light	heavy	<i>focus</i>	<i>average</i>
<i>video</i>	<i>situation</i>	produce	<i>print</i>	<i>lcd</i>	<i>beautiful</i>	<i>trap</i>	<i>control</i>	bulky	<i>battery</i>	<i>flash</i>	<i>release</i>

Table 6: Topics of two aspects *screen* and *weight* under Camera. Errors are italicized and marked in red.

Since a full/comprehensive comparison has been done in the E-Cig domain, this subsection shows two different aspects (a popular aspect and an infrequent aspect) instead of repeating the full comparison with all models. Thus we only select some good models for comparison. Since LDA-PD achieves the second best score most of the time, it is included in our comparison. In addition, we pick one additional model that can also find the same topic. We analyze the two aspects *screen* and *weight* in the camera domain.

(1) **Aspect Screen:** When the target is screen, we first pick up the topic of *picture* for analysis. It discusses the features of the picture displayed on the *Screen*. From Table 6 we observe that TTM produces a good result. LDA also has a topic about *picture* but it has an issue (though it might look good at the first glance). The problem is that it groups the words “picture”, “photo”, “pic” together as they are synonyms but the topic becomes vague because these words may belong to different fine-grained topics. In addition, while LDA finds some more general features like “wonderful”, “beautiful” and “fun” (these words can be regarded as *general* opinion words) because they actually can modify many different aspects) but TTM finds more specific (or coherent) features like “sharp” and “clear”. This is an important property in opinion mining because people always want to know the specific reasons for opinions. Further illustration about this property in opinion mining can be found in our previous work in [32].

The quality of commonly found topics like *menu* is also improved. Additionally, it is worth noting that TTM iden-

tifies a new unique topic called *imaging*, which is not found by any other model, but is in fact related to aspect *screen*.

(2) **Infrequent Aspect Weight:** Only LDA-PD and TTM find meaningful related topics for the aspect *Weight*. Here *lens* and *battery* are two detected topics for demonstration. Because people often complain the heavy weight of these two components. However, they usually mention it in an implicit manner like “the lens is so heavy” or “the heavy weight battery is annoying”. On the contrary, the topic generated by DS-LDA*(T-50) is not so clear. Moreover, TTM solely detects other interesting topics like “carrying”, which is also closely related to the aspect *Weight*.

5. CONCLUSION

In this paper, we studied the novel problem of targeted modeling. Instead of finding all topics from a corpus like existing models based on full modeling, the proposed model focuses on finding topics of a targeted aspect to help the user perform deeper or finer-grained analysis. This is motivated by real-life applications that researchers are often not interested in everything in a corpus but only some aspects of it in order to answer their research questions. Existing full models are not the most effective methods for such focused analysis because their results are often too coarse and they may not find topics that the user is really interested in and/or miss many details. Experimental results showed that this is indeed the case and the proposed new model outperforms the state-of-the-art existing models markedly.

6. ACKNOWLEDGEMENT

Shuai Wang and Sherry Emery’s research was supported by the National Science Foundation (NSF) under award number NSF1524750 and the National Cancer Institute of the National Institutes of Health (NIH) and FDA Center for Tobacco Products (CTP) under Award Number P50CA179546. Bing Liu’s research was supported in part by a grant from National Science Foundation (NSF) under grant no. IIS-1407927, a NCI grant under grant no. R01CA192240, and a gift from Bosch. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NSF, NIH, FDA, or Bosch.

7. REFERENCES

- [1] D. Andrzejewski, X. Zhu, and M. Craven. Incorporating domain knowledge into topic modeling via dirichlet forest priors. In *ICML*, pages 25–32. ACM, 2009.
- [2] C. Archambeau, B. Lakshminarayanan, and G. Bouchard. Latent ibp compound dirichlet allocation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 37(2):321–333, 2015.
- [3] Y. Bengio, A. C. Courville, and J. S. Bergstra. Unsupervised models of images by spike-and-slab rbms. In *ICML*, pages 1145–1152, 2011.
- [4] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. *the Journal of machine Learning research*, 3:993–1022, 2003.
- [5] S. Brody and N. Elhadad. An unsupervised aspect-sentiment model for online reviews. In *NAACL*, pages 804–812. Association for Computational Linguistics, 2010.
- [6] J. Chang, S. Gerrish, C. Wang, J. L. Boyd-Graber, and D. M. Blei. Reading tea leaves: How humans interpret topic models. In *NIPS*, pages 288–296, 2009.
- [7] X. Chen, Y. Qi, B. Bai, Q. Lin, and J. G. Carbonell. Sparse latent semantic analysis. In *SDM*, pages 474–485. SIAM, 2011.
- [8] X. Chen, M. Zhou, and L. Carin. The contextual focused topic model. In *KDD*, pages 96–104. ACM, 2012.
- [9] Z. Chen and B. Liu. Mining topics in documents: standing on the shoulders of big data. In *KDD*, pages 1116–1125. ACM, 2014.
- [10] J. Eisenstein, A. Ahmed, and E. P. Xing. Sparse additive generative models of text. 2011.
- [11] D. Griffiths and M. Tenenbaum. Hierarchical topic models and the nested chinese restaurant process. *NIPS*, 16:17, 2004.
- [12] T. L. Griffiths and M. Steyvers. Finding scientific topics. *PNAS*, 101(suppl 1):5228–5235, 2004.
- [13] G. Heinrich. A generic approach to topic models. In *ECML-PKDD*, pages 517–532. Springer, 2009.
- [14] T. Hofmann. Probabilistic latent semantic indexing. In *SIGIR*, pages 50–57. ACM, 1999.
- [15] L. Hong, D. Yin, J. Guo, and B. D. Davison. Tracking trends: incorporating term volume into temporal topic models. In *KDD*, pages 484–492. ACM, 2011.
- [16] H. Ishwaran and J. S. Rao. Spike and slab variable selection: frequentist and bayesian strategies. *Annals of Statistics*, pages 730–773, 2005.
- [17] Y. Jo and A. H. Oh. Aspect and sentiment unification model for online review analysis. In *WSDM*, pages 815–824. ACM, 2011.
- [18] T. Lin, W. Tian, Q. Mei, and H. Cheng. The dual-sparse topic model: mining focused topics and focused terms in short text. In *WWW*, pages 539–550. ACM, 2014.
- [19] Q. Mei and C. Zhai. Discovering evolutionary theme patterns from text: an exploration of temporal text mining. In *KDD*, pages 198–207. ACM, 2005.
- [20] K. Min, Z. Zhang, J. Wright, and Y. Ma. Decomposing background topics from keywords by principal component pursuit. In *CIKM*, pages 269–278. ACM, 2010.
- [21] T. J. Mitchell and J. J. Beauchamp. Bayesian variable selection in linear regression. *Journal of the American Statistical Association*, 83(404):1023–1032, 1988.
- [22] S. Moghaddam and M. Ester. The flda model for aspect-based opinion mining: Addressing the cold start problem. In *WWW*, 2013.
- [23] A. Mukherjee and B. Liu. Aspect extraction through semi-supervised modeling. In *ACL*, pages 339–348. Association for Computational Linguistics, 2012.
- [24] D. Ramage, D. Hall, R. Nallapati, and C. D. Manning. Labeled lda: A supervised topic model for credit attribution in multi-labeled corpora. In *EMNLP*, pages 248–256. Association for Computational Linguistics, 2009.
- [25] M. Rosen-Zvi, T. Griffiths, M. Steyvers, and P. Smyth. The author-topic model for authors and documents. In *UAI*, pages 487–494. AUAI Press, 2004.
- [26] I. Titov and R. McDonald. Modeling online reviews with multi-grain topic models. In *WWW*, pages 111–120. ACM, 2008.
- [27] H. M. Wallach. Topic modeling: beyond bag-of-words. In *ICML*, pages 977–984. ACM, 2006.
- [28] H. M. Wallach, D. M. Mimno, and A. McCallum. Rethinking lda: Why priors matter. In *NIPS*, pages 1973–1981, 2009.
- [29] C. Wang and D. M. Blei. Decoupling sparsity and smoothness in the discrete hierarchical dirichlet process. In *NIPS*, pages 1982–1989, 2009.
- [30] H. Wang, Y. Lu, and C. Zhai. Latent aspect rating analysis on review text data: a rating regression approach. In *KDD*, pages 783–792. ACM, 2010.
- [31] Q. Wang, J. Xu, H. Li, and N. Craswell. Regularized latent semantic indexing. In *SIGIR*, pages 685–694. ACM, 2011.
- [32] S. Wang, Z. Chen, and B. Liu. Mining aspect-specific opinion using a holistic lifelong topic model. In *WWW*, pages 167–176. WWW, 2016.
- [33] S. Williamson, C. Wang, K. A. Heller, and D. M. Blei. The ibp compound dirichlet process and its application to focused topic modeling. In *ICML*, pages 1151–1158, 2010.
- [34] W. X. Zhao, J. Jiang, H. Yan, and X. Li. Jointly modeling aspects and opinions with a maxent-lda hybrid. In *EMNLP*, pages 56–65. Association for Computational Linguistics, 2010.
- [35] J. Zhu and E. P. Xing. Sparse topical coding. *arXiv preprint arXiv:1202.3778*, 2012.