Adversarial Learning of Poisson Factorisation Model for Gauging Brand Sentiment in User Reviews

Runcong Zhao, Lin Gui, Gabriele Pergola, Yulan He
Department of Computer Science, University of Warwick, UK
{runcong.zhao,lin.gui,gabriele.pergola,yulan.he}@warwick.ac.uk

Abstract

In this paper, we propose the Brand-Topic Model (BTM) which aims to detect brand-associated polarity-bearing topics from product reviews. Different from existing models for sentiment-topic extraction which assume topics are grouped under discrete sentiment categories such as '*positive*', '*negative*' and '*neural*', BTM is able to automatically infer real-valued brand-associated sentiment scores and generate fine-grained sentiment-topics in which we can observe continuous changes of words under a certain topic (e.g., '*shaver*' or '*cream*') while its associated sentiment gradually varies from negative to positive. BTM is built on the Poisson factorisation model with the incorporation of adversarial learning. It has been evaluated on a dataset constructed from Amazon reviews. Experimental results show that BTM outperforms a number of competitive baselines in brand ranking, achieving a better balance of topic coherence and uniqueness, and extracting better-separated polarity-bearing topics.

1 Introduction

Market intelligence aims to gather data from a company's external environment, such as customer surveys, news outlets and social media sites, in order to understand customer feedback to their products and services and to their competitors, for a better decision making of their marketing strategies. Since consumer purchase decisions are heavily influenced by online reviews, it is important to automatically analyse customer reviews for online brand monitoring. Existing sentiment analysis models either classify reviews into discrete polarity categories such as '*positive*', '*negative*' or '*neural*', or perform more fine-grained sentiment analysis, in which aspect-level sentiment label is predicted, though still in the discrete polarity category space. We argue that it is desirable to be able to detect subtle topic changes under continuous sentiment scores. This allows us to identify, for example, whether customers with slightly negative views share similar concerns with those holding strong negative opinions; and what positive aspects are praised by customers the most. In addition, deriving brand-associated sentiment scores in a continuous space makes it easier to generate a ranked list of brands, allowing for easy comparison.

Existing studies on brand topic detection were largely built on the Latent Dirichlet Allocation (LDA) model (Blei et al., 2003) which assumes that latent topics are shared among competing brands for a certain market. They however are not able to separate positive topics from negative ones. Approaches to polarity-bearing topic detection can only identify topics under discrete polarity categories such as '*positive*' and '*negative*'. We instead assume that each brand is associated with a latent real-valued sentiment score falling into the range of $[-1, 1]$ in which -1 denotes negative, 0 being neutral and 1 positive, and propose a Brand-Topic Model built on the Poisson Factorisation model with adversarial learning. Example outputs generated from BTM are shown in Figure 1 in which we can observe a transition of topics with varying topic polarity scores together with their associated brands.

More concretely, in BTM, a document-word count matrix is factorised into a product of two positive matrices, a document-topic matrix and a topic-word matrix. A word count in a document is assumed drawn from a Poisson distribution with its rate parameter defined as a product of a document-specific topic intensity and its word probability under the corresponding topic, summing over all topics. We further assume that each document is associated with a brand-associated sentiment score and a latent topic-word offset value. The occurrence count of a word is then jointly determined

Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics, pages 2341–2351
April 19 - 23, 2021.

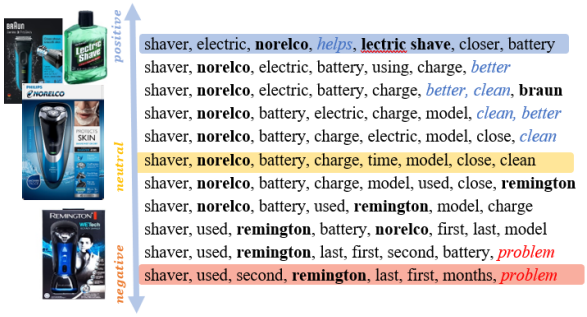

Figure 1: Example topic results generated from proposed Brand-Topic Model. We observe a transition of topics with varying topic polarity scores. Besides the change of sentiment-related words (e.g., '*problem*' in negative topics and '*better*' in positive topics), we could also see a change of their associated brands. Users are more positive about BRAUN, negative about REMINGTON, and have mixed opinions on NORELCO.

by both the brand-associated sentiment score and the topic-word offset value. The intuition behind is that if a word tends to occur in documents with positive polarities, but the brand-associated sentiment score is negative, then the topic-word offset value will have an opposite sign, forcing the occurrence count of such a word to be reduced. Furthermore, for each document, we can sample its word counts from their corresponding Poisson distributions and form a document representation which is subsequently fed into a sentiment classifier to predict its sentiment label. If we reverse the sign of the latent brand-associated sentiment score and sample the word counts again, then the sentiment classifier fed with the resulting document representation should generate an opposite sentiment label.

Our proposed BTM is partly inspired by the recently developed Text-Based Ideal Point (TBIP) model (Vafa et al., 2020) in which the topic-specific word choices are influenced by the ideal points of authors in political debates. However, TBIP is fully unsupervised and when used in customer reviews, it generates topics with mixed polarities. On the contrary, BTM makes use of the document-level sentiment labels and is able to produce better separated polarity-bearing topics. As will be shown in the experiments section, BTM outperforms TBIP on brand ranking, achieving a better balance of topic coherence and topic uniqueness measures.

The contributions of the model are three-fold:

- We propose a novel model built on Poisson Factorisation with adversarial learning for brand topic analysis which can disentangle the sentiment factor from the semantic latent representations to achieve a flexible and controllable topic generation;
- We approximate word count sampling from Poisson distributions by the Gumbel-Softmax-based word sampling technique, and construct document representations based on the sampled word counts, which can be fed into a sentiment classifier, allowing for end-to-end learning of the model;
- The model, trained with the supervision of review ratings, is able to automatically infer the brand polarity scores from review text only.

The rest of the paper is organised as follows. Section 2 presents the related work. Section 3 describes our proposed Brand-Topic Model. Section 4 and 5 discusses the experimental setup and evaluation results, respectively. Finally, Section 5 concludes the paper and outlines the future research directions.

2 Related Work

Our work is related to the following research:

Poisson Factorisation Models Poisson factorisation is a class of non-negative matrix factorisation in which a matrix is decomposed into a product of matrices. It has been used in many personalise application such as personalised budgets recommendation (Guo et al., 2017), ranking (Kuo et al., 2018), or content-based social recommendation (Su et al., 2019; de Souza da Silva et al., 2017).

Poisson factorisation can also be used for topic modelling where a document-word count matrix is factorised into a product of two positive matrices, a document-topic matrix and a topic-word matrix (Gan et al., 2015; Jiang et al., 2017). In such a setup, a word count in a document is assumed drawn from a Poisson distribution with its rate parameter defined as a product of a document-specific topic intensity and its word probability under the corresponding topic, summing over all topics.

Polarity-bearing Topics Models Early approaches to polarity-bearing topics extraction were built on LDA in which a word is assumed to be generated from a corpus-wide sentiment-topic-word distributions (Lin and He, 2009). In order to be able to separate topics bearing different polarities, word prior polarity knowledge needs to be

incorporated into model learning. In recent years, the neural network based topic models have been proposed for many NLP tasks, such as information retrieval (Xie et al., 2015), aspect extraction (He, 2017) and sentiment classification (He et al., 2018). Most of them are built upon Variational Autoencode (VAE) (Kingma and Welling, 2014) which constructs a neural network to approximate the topic-word distribution in probabilistic topic models (Srivastava and Sutton, 2017; Sønderby et al., 2016; Bouchacourt et al., 2018). Intuitively, training the VAE-based supervised neural topic models with class labels (Chaidaroon and Fang, 2017; Huang et al., 2018; Gui et al., 2020) can introduce sentiment information into topic modelling, which may generate better features for sentiment classification.

Market/Brand Topic Analysis The classic LDA can also be used to analyse market segmentation and brand reputation in various fields such as finance and medicine (Barry et al., 2018; Doyle and Elkan, 2009). For market analysis, the model proposed by Iwata et al. (2009) used topic tracking to analyse customers' purchase probabilities and trends without storing historical data for inference at the current time step. Topic analysis can also be combined with additional market information for recommendations. For example, based on user profiles and item topics, Gao et al. (2017) dynamically modelled users' interested items for recommendation. Zhang et al. (2015) focused on brand topic tracking. They built a dynamic topic model to analyse texts and images posted on Twitter and track competitions in the luxury market among given brands, in which topic words were used to identify recent hot topics in the market (e.g. *Rolex watch*) and brands over topics were used to identify the market share of each brand.

Adversarial Learning Several studies have explored the application of adversarial learning mechanics to text processing for style transferring (John et al., 2019), disentangling representations (John et al., 2019) and topic modelling (Masada and Takasu, 2018). In particular, Wang et al. (2019) has proposed an Adversarial-neural Topic Model (ATM) based on the Generative Adversarial Network (GAN) (Goodfellow et al., 2014), that employees an adversarial approach to train a generator network producing word distributions indistinguishable from topic distributions in the training set. (Wang et al., 2020) further extended the ATM model with a Bidirectional Adversarial Topic (BAT) model, using a bidirectional adversarial training to incorporate a Dirichlet distribution as prior and exploit the information encoded in word embeddings. Similarly, (Hu et al., 2020) builds on the aforementioned adversarial approach adding cycle-consistent constraints.

Although the previous methods make use of adversarial mechanisms to approximate the posterior distribution of topics, to the best of our knowledge, none of them has so far used adversarial learning to lead the generation of topics based on their sentiment polarity and they do not provide any mechanism for smooth transitions between topics, as introduced in the presented Brand-Topic Model.

3 Brand-Topic Model (BTM)

We propose a probabilistic model for monitoring the assessment of various brands in the beauty market from Amazon reviews. We extend the Text-Based Ideal Point (TBIP) model with adversarial learning and Gumbel-Softmax to construct document features for sentiment classification. The overall architecture of our proposed BTM is shown in Figure 2. In what follows, we will first give a brief introduction of TBIP, followed by the presentation of our proposed BTM.

3.1 Background: Text-Based Ideal Point (TBIP) model

TBIP (Vafa et al., 2020) is a probabilistic model which aims to quantify political positions (i.e. ideal points) from politicians' speeches and tweets via Poisson factorisation. In its generative processes, political text is generated from the interactions of several latent variables: the per-document topic intensity θ_{dk} for K topics and D documents, the V-vectors representing the topics β_{kv} with vocabulary size $|V|$, the author's ideal point s expressed with a real-valued scalar x_s and the ideological topic expressed by a real-valued V-vector η_k. In particular, the ideological topic η_k aligns the neutral topic (e.g. *gun*, *abortion*, etc.) according to the author's ideal point (e.g. *liberal*, *neutral*, *conservative*), thus modifying the prominent words in the original topic (e.g. '*gun violence*', or '*constitutional rights*'). The observed variables are the author a_d for a document d, and the word count for a term v in d encoded as c_{dv} .

The TBIP model places a Gamma prior on β

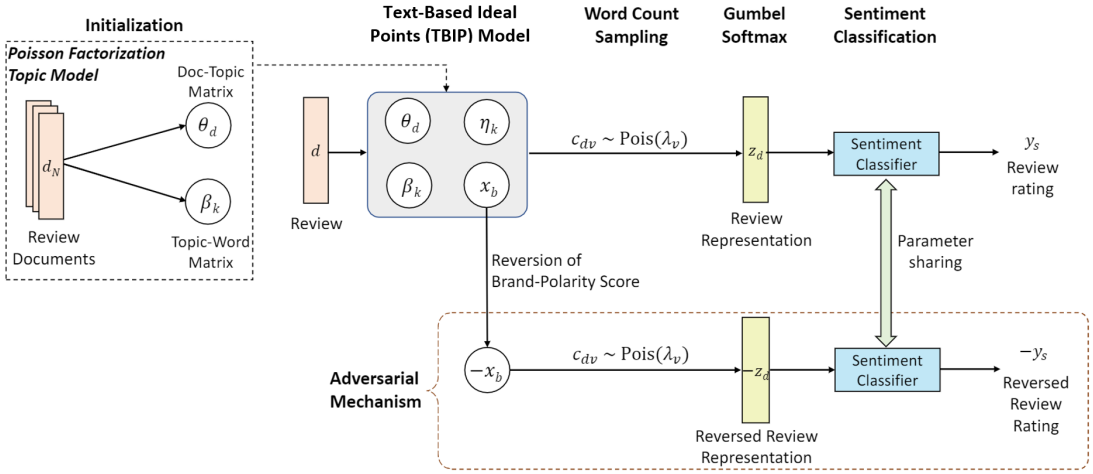

Figure 2: The overall architecture of the Brand-Topic Model.

and $\boldsymbol{\theta}$, which is the assumption inherited from the Poisson factorisation, with m, n being hyper-parameters.

$$\theta_{dk} \sim \text{Gamma}(m, n) \quad \beta_{kv} \sim \text{Gamma}(m, n)$$

It places instead a normal prior over the ideological topic $\boldsymbol{\eta}$ and ideal point $\boldsymbol{x}$:

$$\eta_{kv} \sim \mathcal{N}(0, 1) \quad x_s \sim \mathcal{N}(0, 1)$$

The word count for a term v in d, c_{dv}, can be modelled with Poisson distribution:

$$c_{dv} \sim \text{Pois}(\sum_k \theta_{dk}\beta_{kv} \exp\{x_{a_d}\eta_{kv}\}) \quad (1)$$

3.2 Brand-Topic Model (BTM)

Inspired by the TBIP model, we introduce the Brand-Topic Model by reinterpreting the ideal point x_s as brand-polarity score x_b expressing an *ideal feeling* derived from reviews related to a brand, and the ideological topics η_{kv} as *opinionated topics*, i.e. polarised topics about brand qualities.

Thus, a term count c_{dv} for a product's reviews derives from the hidden variable interactions as $c_{dv} \sim Pois(\lambda_{dv})$ where:

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

with the priors over $\boldsymbol{\beta}$, $\boldsymbol{\theta}$, $\boldsymbol{\eta}$ and $\boldsymbol{x}$ initialised according to the TBIP model.

The intuition is that if a word tends to frequently occur in reviews with positive polarities, but the brand-polarity score for the current brand is negative, then the occurrence count of such a word would be reduced since x_{b_d} and η_{kv} have opposite signs.

Distant Supervision and Adversarial Learning Product reviews might contain opinions about products and more general users' experiences (e.g. delivery service), which are not strictly related to the product itself and could mislead the inference of a reliable brand-polarity score. Therefore, to generate topics which are mainly characterised by product opinions, we provide an additional distant supervision signal via their review ratings. To this aim, we use a sentiment classifier, a simple linear layer, over the generated document representations to infer topics that are discriminative of the review's rating.

In addition, to deal with the imbalanced distribution in the reviews, we design an adversarial mechanism linking the brand-polarity score to the topics as shown in Figure 3. We contrastively sample adversarial training instances by reversing the original brand-polarity score ($x_b \in [-1, 1]$) and generating associated representations. This representation will be fed into the shared sentiment classifier with the original representation to maximise their distance in the latent feature space.

Gumbel-Softmax for Word Sampling As discussed earlier, in order to construct document features for sentiment classification, we need to sample word counts from the Poisson distribution. However, directly sampling word counts from the Poisson distribution is not differentiable. In order to enable back-propagation of gradients, we apply Gumbel-Softmax (Jang et al., 2017; Joo et al., 2020), which is a gradient estimator with the reparameterization trick.

For a word v in document d, its occurrence count,

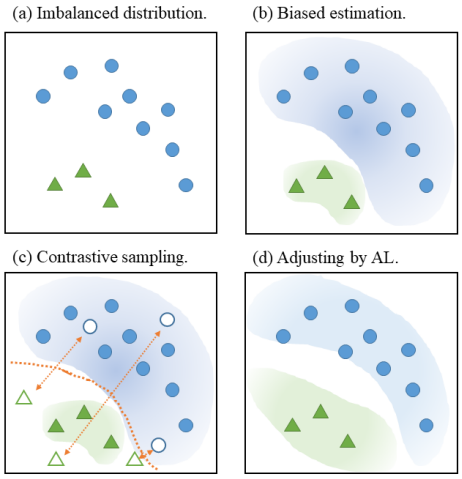

Figure 3: Process of Adversarial Learning (AL): (a) The imbalanced distribution of different sentiment categories; (b) The biased estimation of distribution from training samples; (c) Contrastive sample generation (white triangles) by reversing the sampling results from biased estimation (white dots); (d) Adjusting the biased estimation of (b) by the contrastive samples.

$c_{dv} \sim \text{Pois}(\lambda_{dv})$, is a non-negative random variable with the Poisson rate λ_{dv}. We can approximate it by sampling from the truncated Poisson distribution, $c_{dv_n} \sim \text{TruncatedPois}(\lambda_{dv}, n)$, where

$$\pi_k = Pr(c_{dv} = k) = \frac{\lambda_{dv}^k e^{-\lambda_{dv}}}{k!}$$
$$\pi_{n-1} = 1 - \sum_k \pi_k \quad \text{for} \quad k \in \{0, 1, ..., n-2\}.$$

We can then draw samples z_{dv} from the categorical distribution with class probabilities $\pi = (\pi_0, \pi_1, \cdots, \pi_{n-1})$ using:

$$u_i \sim \text{Uniform}(0,1) \quad g_i = -\log(-\log(u_i))$$
$$w_i = \text{softmax}\big((g_i + \log \pi_i)/\tau\big) \quad z_{dv} = \sum_i w_i c_i$$

where τ is a constant referred to as the temperature, c is the outcome vector. By using the average of weighted word account, the process is now differentiable and we use the sampled word counts to form the document representation and feed it as an input to the sentiment classifier.

Objective Function Our final objective function consists of three parts, including the Poisson factorisation model, the sentiment classification loss, and the reversed sentiment classification loss (for adversarial learning). For the Poisson factorisation modelling part, mean-field variational inference is used to approximate posterior distribution (Jordan et al., 1999; Wainwright and Jordan, 2008; Blei et al., 2017).

$$q_\phi(\theta, \beta, \eta, x) = \prod_{d,k.b} q(\theta_d) q(\beta_k) q(\eta_k) q(x_b) \quad (3)$$

For optimisation, to minimise the approximation of $q_\phi(\theta, \beta, \eta, x)$ and the posterior, equivalently we maximise the evidence lower bound (ELBO):

$$ELBO = \mathbb{E}_{q_\phi}[log\ p(\theta, \beta, \eta, x)] + log\ p(y|\theta, \beta, \eta, x) - log\ q_\phi(\theta, \beta, \eta, x)] \quad (4)$$

The Poisson factorization model is pre-trained by applying the algorithm in Gan et al. (2015), which is then used to initialise the varational parameters of θ_d and β_k. Our final objective function is:

$$Loss = ELBO + \lambda(L_s + L_a) \quad (5)$$

where L_s and L_a are the cross entropy loss of sentiment classification for sampled documents and reversed sampled documents, respectively, and λ is the weight to balance the two parts of loss, which is set to be 100 in our experiments.

4 Experimental Setup

Datasets We construct our dataset by retrieving reviews in the Beauty category from the Amazon review corpus[1] (He and McAuley, 2016). Each review is accompanied with the rating score (between 1 and 5), reviewer name and the product meta-data such as product ID, description, brand and image. We use the product meta-data to relate a product with its associated brand. By only selecting brands with relatively more and balanced reviews, our final dataset contains a total of 78,322 reviews from 45 brands. Reviews with the rating score of 1 and 2 are grouped as negative reviews; those with the score of 3 are neutral reviews; and the remaining are positive reviews. The statistics of our dataset is shown in Table 1[2]. We can observe that our data is highly imbalanced, with the positive reviews far more than negative and neutral reviews.

Baselines We compare the performance of our model with the following baselines:

[1] http://jmcauley.ucsd.edu/data/amazon/

[2] The detailed rating score distributions of brands and their average rating are shown in Table A1 in the Appendix.

Dataset	Amazon-Beauty Reviews
Documents per classes	
Neg / Neu / Pos	9,545 / 5,578 / 63,199
Brands	45
Total #Documents	78,322
Avg. Document Length	9.7
Vocabulary size	$\sim$ 5000

Table 1: Dataset statistics of reviews within the Amazon dataset under the *Beauty* category.

- Joint Sentiment-Topic (JST) model (Lin and He, 2009), built on LDA, can extract polarity-bearing topics from text provided that it is supplied with the word prior sentiment knowledge. In our experiments, the MPQA subjectivity lexicon[3] is used to derive the word prior sentiment information.

- SCHOLAR (Card et al., 2018), a neural topic model built on VAE. It allows the incorporation of meta-information such as document class labels into the model for training, essentially turning it into a supervised topic model.

- Text-Based Ideal Point (TBIP) model, an unsupervised Poisson factorisation model which can infer latent brand sentiment scores.

Parameter setting Since documents are represented as the bag-of-words which result in the loss of word ordering or structural linguistics information, frequent bigrams and trigrams such as '*without doubt*', '*stopped working*', are also used as features for document representation construction. Tokens, i.e., n-grams ($n = \{1, 2, 3\}$), occurred less than twice are filtered. In our experiments, we set aside 10% reviews (7,826 reviews) as the test set and the remaining (70,436 reviews) as the training set. For hyperparameters, we set the batch size to 1,024, the maximum training steps to 50,000, the topic number to 30, the temperature in the Gumbel-Softmax equation in Section 3.2 to 1. Since our dataset is highly imbalanced, we balance data in each mini-batch by oversampling. For a fair comparison, we report two sets of results from the baseline models, one trained from the original data, the other trained from the balanced training data by oversampling negative reviews. The latter results in an increased training set consisting of 113,730 reviews.

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

5 Experimental Results

In this section, we will present the experimental results in comparison with the baseline models in brand ranking, topic coherence and uniqueness measures, and also present the qualitative evaluation of the topic extraction results. We will further discuss the limitations of our model and outline future directions.

5.1 Comparison with Existing Models

Model	Spearman's		Kendall's tau	
	corr	p-val	corr	p-val
JST	0.241	0.111	0.180	0.082
JST*	0.395	0.007	0.281	0.007
SCHOLAR	-0.140	0.358	-0.103	0.318
SCHOLAR*	0.050	0.743	0.046	0.653
TBIP	0.361	0.016	0.264	0.012
BTM	**0.486**	0.001	**0.352**	0.001

Table 2: Brand ranking results generated by various models based on the test set. We report the correlation coefficients *corr* and its associated two-sided p-values for both Spearman's correlations and Kendall's tau. * indicates models trained on balanced training data.

Brand Ranking We report in Table 2 the brand ranking results generated by various models on the test set. The two commonly used evaluation metrics for ranking tasks, Spearman's correlations and Kendall's Tau, are used here. They penalise inversions equally across the ranked list. Both TBIP and BTM can infer each brand's associated polarity score automatically which can be used for ranking. For both JST and SCHOLAR, we derive the polarity score of a brand by aggregating the sentiment probabilities of its associated review documents and then normalising over the total number of brand-related reviews. It can be observed from Table 2 that JST outperforms both SCHOLAR and TBIP. Balancing the distributions of sentiment classes improves the performance of JST and SCHOLAR. Overall, BTM gives the best results, showing the effectiveness of adversarial learning.

Topic Coherence and Uniqueness Here we choose the top 10 words for each topics to calculate the context-vector-based topic coherence scores (Röder et al., 2015). In the topics generated by TBIP and BTM, we can vary the topic polarity scores to generate positive, negative and neutral

subtopics as shown in Table 4. We would like to achieve high topic coherence, but at the same time maintain a good level of topic uniqueness across the sentiment subtopics since they express different polarities. Therefore, we additionally consider the topic uniqueness (Nan et al., 2019) to measure word redundancy among sentiment subtopics, $TU = \frac{1}{LK}\sum_{l=1}^{K}\sum_{l=1}^{L}\frac{1}{cnt(l,k)}$, where $cnt(l,k)$ denotes the number of times word l appear across *positive*, *neutral* and *negative* topics under the same topic number k. We can see from Table 3 that both TBIP and BTM achieve higher coherence scores compared to JST and SCHOLAR. TBIP slightly outperforms BTM on topic coherence, but has a lower topic uniqueness score. As will be shown in Table 4, topics extracted by TBIP contain words significantly overlapped with each other among sentiment subtopics. SCHOLAR gives the highest topic uniqueness score. However, it cannot separate topics with different polarities. Overall, our proposed BTM achieves the best balance between topic coherence and topic uniqueness.

Model	Topic Coherence	Topic Uniqueness
JST	0.1423	0.7699
JST*	0.1317	0.7217
SCHOLAR	0.1287	**0.9640**
SCHOLAR*	0.1196	0.9256
TBIP	**0.1525**	0.8647
BTM	0.1407	0.9033

Table 3: Topic coherence/uniqueness measures of results generated by various models.

5.2 Example Topics Extracted from Amazon Reviews

We illustrate some representative topics generated by TBIP and BTM in Table 4. It is worth noting that we can generate a smooth transition of topics by varying the topic polarity score gradually as shown in Figure 1. Due to space limit, we only show topics when the topic polarity score takes the value of -1 (*negative*), 0 (*neutral*) and 1 (*positive*). It can be observed that TBIP fails to separate subtopics bearing different sentiments. For example, all the subtopics under 'Duration' express a positive polarity. On the contrary, BTM shows a better-separated sentiment subtopics. For 'Duration', we see positive words such as '*comfortable*' under the positive subtopic, and words such as '*stopped working*' clearly expressing negative sentiment under the negative subtopic. Moreover, top words under different sentiment subtopics largely overlapped with each other for TBIP. But we observe a more varied vocabulary in the sentiment subtopics for BTM.

TBIP was originally proposed to deal with political speeches in which speakers holding different ideal points tend to use different words to express their stance on the same topic. This is however not the case in Amazon reviews where the same word could appear in both positive and negative reviews. For example, '*cheap*' for lower-priced products could convey a positive polarity to express value for money, but it could also bear a negative polarity implying a poor quality. As such, it is difficult for TBIP to separate words under different polarity-bearing topics. On the contrary, with the incorporation of adversarial learning, our proposed BTM is able to extract different set of words co-occurred with '*cheap*' under topics with different polarities, thus accurately capturing the contextual polarity of the word '*cheap*'. For example, '*cheap*' appears in both positive and negative subtopics for 'Brush' in Table 4. But we can find other co-occurred words such as '*pretty*' and '*soft*' under the positive subtopic, and '*plastic*' and '*flimsy*' under the negative subtopic, which help to infer the contextual polarity of '*cheap*'.

TBIP also appears to have a difficulty in dealing with highly imbalanced data. In our constructed dataset, positive reviews significantly outnumber both negative and neutral ones. In many sentiment subtopics extracted by TBIP, all of them convey a positive polarity. One example is the 'Duration' topic under TBIP, where words such as '*great*', '*great price*' appear in all positive, negative and neutral topics. With the incorporation of supervised signals such as the document-level sentiment labels, our proposed BTM is able to derive better separated polarised topics.

As an example shown in Figure 1, if we vary the polarity score of a topic from -1 to 1, we observe a smooth transition of its associated topic words, gradually moving from negative to positive. Under the topic (*shaver*) shown in this figure, four brand names appeared: REMINGTON, NORELCO, BRAUN and LECTRIC SHAVE. The first three brands can be found in our dataset. REMINGTON appears in the negative side and it indeed has the lowest review score among these 3 brands; NORELCO appears most and it is indeed a popular

Topic Label	Sentiment Topics	Top Words
		BTM
Brush	Positive	brushes, cheap, came, pay, *pretty*, brush, *okay*, case, glue, *soft*
	Neutral	cheap, feel, set, buy, *cheaply made*, feels, made, worth, spend, bucks
	Negative	plastic, made, cheap, parts, feels, *flimsy*, money, *break*, metal, bucks
Oral Care	Positive	teeth, taste, mouth, strips, crest, mouthwash, tongue, using, *white*, rinse
	Neutral	teeth, *pain*, mouth, strips, using, taste, used, crest, mouthwash, *white*
	Negative	*pain*, *issues*, causing, teeth, caused, removing, wore, *burn*, little, cause
Duration	Positive	stay, pillow, *comfortable*, string, tub, mirror, stick, back, months
	Neutral	months, year, *lasted*, *stopped working*, *sorry*, n, worked, working, u, last
	Negative	months, year, last, *lasted*, battery, warranty, *stopped working*, *died*, *less*
		TBIP
Brush	Positive	love, *favorite*, products, *definitely recommend*, forever, carry, brushes
	Neutral	love, brushes, *cute*, *favorite*, *definitely recommend*, soft, *cheap*
	Negative	love, brushes, cute, *soft*, *cheap*, set, case, quality price, buy, bag
Oral Care	Positive	teeth, strips, crest, mouth, mouthwash, taste, *white*, *whitening*, sensitivity
	Neutral	teeth, strips, mouth, crest, taste, work, *pain*, using, *white*, mouthwash
	Negative	teeth, strips, mouth, crest, taste, work, *pain*, using, *white*, mouthwash
Duration	Positive	great, *love shampoo*, *great price*, *great product*, *lasts long time*
	Neutral	*great*, *great price*, *lasts long time*, *great product*, price, *works expected*
	Negative	quality, *great*, *fast shipping*, *great price*, *low price*, price quality, hoped

Table 4: Example topics generated by BTM and TBIP on Amazon reviews. The topic labels are assigned by manual inspection. Positive words are highlighted with the blue colour, while negative words are marked with the red colour. BTM generates better-separated sentiment topics compared to TBIP.

brand with mixed reviews; and BRAUN gets the highest score in these 3 brands, which is also consistent with the observations in our data. Another interesting finding is the brand LECTRIC SHAVE, which is not one of the brands we have in the dataset. But we could predict from the results that it is a product with relatively good reviews.

5.3 Limitations and Future work

Our model requires the use of a vanilla Poisson factorisation model to initialise the topic distributions before applying the adversarial learning mechanism of BTM to perform a further split of topics based on varying polarities. Essentially topics generated by a vanilla Poisson factorisation model can be considered as parent topics, while polarity-bearing subtopics generated by BTM can be considered as child topics. Ideally, we would like the parent topics to be either neutral or carrying a mixed sentiment which would facilitate the learning of polarised sub-topics better. In cases when parent topics carry either strongly positive or strongly negative sentiment signals, BTM would fail to produce polarity-varying subtopics. One possible way is to employ earlier filtering of topics with strong polarities. For example, topic labeling (Bhatia et al., 2016) could be employed to obtain a rough estimate of initial topic polarities; these labels would be in turn used for filtering out topics carrying strong sentiment polarities.

Although the adversarial mechanism tends to be robust with respect to class imbalance, the disproportion of available reviews with different polarities could hinder the model performance. One promising approach suitable for the BTM adversarial mechanism would consist in decoupling the representation learning and the classification, as suggested in Kang et al. (2020), preserving the original data distribution used by the model to estimate the brand score.

6 Conclusion

In this paper, we presented the Brand-Topic Model, a probabilistic model which is able to generate polarity-bearing topics of commercial brands. Compared to other topic models, BMT infers real-valued brand-associated sentiment scores and extracts fine-grained sentiment-topics which vary

smoothly in a continuous range of polarity scores. It builds on the Poisson factorisation model, combining it with an adversarial learning mechanism to induce better-separated polarity-bearing topics. Experimental evaluation on Amazon reviews against several baselines shows an overall improvement of topic quality in terms of coherence, uniqueness and separation of polarised topics.

Acknowledgements

This work is funded by the EPSRC (grant no. EP/T017112/1, EP/V048597/1). YH is supported by a Turing AI Fellowship funded by the UK Research and Innovation (UKRI) (grant no. EP/V020579/1).

References

Adam E. Barry, Danny Valdez, Alisa A.Padon, and Alex M. Russel. 2018. Alcohol advertising on twitter—a topic model. *American Journal of Health Education*, pages 256–263.

Shraey Bhatia, Jey Han Lau, and Timothy Baldwin. 2016. Automatic labelling of topics with neural embeddings. In *Proceedings of the 26th International Conference on Computational Linguistics*, pages 953–963.

David M Blei, Alp Kucukelbir, and Jon D McAuliffe. 2017. Variational inference: A review for statisticians. *Journal of the American Statistical Association*, 112(518):859–877.

David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent dirichlet allocation. *Journal of Machine Learning Research*, 3(2003):993–1022.

Diane Bouchacourt, Ryota Tomioka, and Sebastian Nowozin. 2018. Multi-level variational autoencoder: Learning disentangled representations from grouped observations. In *The 32nd AAAI Conference on Artificial Intelligence*, pages 2095–2102.

Dallas Card, Chenhao Tan, and Noah A. Smith. 2018. Neural models for documents with metadata. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, pages 2031–2040.

Suthee Chaidaroon and Yi Fang. 2017. Variational deep semantic hashing for text documents. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 75–84.

G. Doyle and C. Elkan. 2009. Financial topic models. In *In NIPS Workshop for Applications for Topic Models: Text and Beyond*.

Zhe Gan, Changyou Chen, Ricardo Henao, David E. Carlson, and Lawrence Carin. 2015. Scalable deep poisson factor analysis for topic modeling. In *Proceedings of the 32nd International Conference on Machine Learning*, volume 37, pages 1823–1832.

Li Gao, Jia Wu, Chuan Zhou, and Yue Hu. 2017. Collaborative dynamic sparse topic regression with user profile evolution for item recommendation. In *Proceedings of the 31st AAAI Conference on Artificial Intelligence*, pages 1316–1322.

Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative adversarial nets. In *Advances in Neural Information Processing Systems 27*, pages 2672–2680.

Lin Gui, Leng Jia, Jiyun Zhou, Ruifeng Xu, and Yulan He. 2020. Multi-task mutual learning for joint sentiment classification and topic detection. In *IEEE Transactions on Knowledge and Data Engineering*, pages 1–1.

Yunhui Guo, Congfu Xu, Hanzhang Song, and Xin Wang. 2017. Understanding users' budgets for recommendation with hierarchical poisson factorization. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence*, pages 1781–1787.

Ruidan He. 2017. An unsupervised neural attention model for aspect extraction. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, pages 388–397.

Ruidan He, Wee Sun Lee, Hwee Tou Ng, and Daniel Dahlmeier. 2018. Effective attention modeling for aspect-level sentiment classification. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 1121–1131.

Ruining He and Julian McAuley. 2016. Ups and downs: Modeling the visual evolution of fashion trends with one-class collaborative filtering. In *Proceedings of the 25th International Conference on World Wide Web*.

Xuemeng Hu, Rui Wang, Deyu Zhou, and Yuxuan Xiong. 2020. Neural topic modeling with cycle-consistent adversarial training. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*.

Minghui Huang, Yanghui Rao, Yuwei Liu, Haoran Xie, and Fu Lee Wang. 2018. Siamese network-based supervised topic modeling. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 4652–4662.

Tomoharu Iwata, Shinji Watanabe, Takeshi Yamada, and Naonori Ueda. 2009. Topic tracking model for analyzing consumer purchase behavior. In *Proceedings of the 21st International Joint Conference on Artificial Intelligence*, pages 1427–1432.

Eric Jang, Shixiang Gu, and Ben Poole. 2017. Categorical reparameterization with gumbel-softmax. In *International Conference on Learning Representations*.

Haixin Jiang, Rui Zhou, Limeng Zhang, Hua Wang, and Yanchun Zhang. 2017. A topic model based on poisson decomposition. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, pages 1489–1498.

Vineet John, Lili Mou, Hareesh Bahuleyan, and Olga Vechtomova. 2019. Disentangled representation learning for non-parallel text style transfer. In *Proceedings of the 57th Conference of the Association for Computational Linguistics*, pages 424–434.

Weonyoung Joo, Dongjun Kim, Seungjae Shin, and Il-Chul Moon. 2020. Generalized gumbel-softmax gradient estimator for various discrete random variables. *Computing Research Repository*, arXiv:2003.01847v2.

Michael I Jordan, Zoubin Ghahramani, Tommi S Jaakkola, and Lawrence K Saul. 1999. An introduction to variational methods for graphical models. *Machine Learning*, 37(2):183–233.

Bingyi Kang, Saining Xie, Marcus Rohrbach, Zhicheng Yan, Albert Gordo, Jiashi Feng, and Yannis Kalantidis. 2020. Decoupling representation and classifier for long-tailed recognition. In *the 8th International Conference on Learning Representations*.

Diederik P. Kingma and Max Welling. 2014. Auto-encoding variational bayes. In *2nd International Conference on Learning Representations*.

Li-Yen Kuo, Chung-Kuang Chou, and Ming-Syan Chen. 2018. Personalized ranking on poisson factorization. In *Proceedings of the 2018 SIAM International Conference on Data Mining*, pages 720–728.

Chenghua Lin and Yulan He. 2009. Joint sentiment/topic model for sentiment analysis. In *Proceedings of the 18th ACM Conference on Information and Knowledge Management*, pages 375–384.

Tomonari Masada and Atsuhiro Takasu. 2018. Adversarial learning for topic models. In *Proceedings of the 14th International Conference on Advanced Data Mining and Applications*, volume 11323 of *Lecture Notes in Computer Science*, pages 292–302.

Feng Nan, Ran Ding, Ramesh Nallapati, , and Bing Xiang. 2019. Topic modeling with wasserstein autoencoders. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, page 6345–6381.

Michael Röder, Andreas Both, and Alexander Hinneburg. 2015. Exploring the space of topic coherence measures. In *The 8th ACM International Conference on Web Search and Data Mining*, pages 399–408.

Eliezer de Souza da Silva, Helge Langseth, and Heri Ramampiaro. 2017. Content-based social recommendation with poisson matrix factorization. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, volume 10534 of *Lecture Notes in Computer Science*, pages 530–546.

Casper Kaae Sønderby, Tapani Raiko, Lars Maaløe, Søren Kaae Sønderby, and Ole Winther. 2016. Ladder variational autoencoders. In *The Annual Conference on Neural Information Processing Systems*, pages 3738–3746.

Akash Srivastava and Charles Sutton. 2017. Autoencoding variational inference for topic models. In *International Conference on Learning Representations*.

Yijun Su, Xiang Li, Wei Tang, Daren Zha, Ji Xiang, and Neng Gao. 2019. Personalized point-of-interest recommendation on ranking with poisson factorization. In *International Joint Conference on Neural Networks*, pages 1–8.

Keyon Vafa, Suresh Naidu, and David M. Blei. 2020. Text-based ideal points. In *Proceedings of the 2020 Conference of the Association for Computational Linguistics*, pages 5345–5357.

Martin J Wainwright and Michael I Jordan. 2008. Graphical models, exponential families, and variational inference. *Foundations and Trends in Machine Learning*, 1(1-2):1–305.

Rui Wang, Xuemeng Hu, Deyu Zhou, Yulan He, Yuxuan Xiong, Chenchen Ye, and Haiyang Xu. 2020. Neural topic modeling with bidirectional adversarial training. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 340–350.

Rui Wang, Deyu Zhou, and Yulan He. 2019. ATM: Adversarial-neural topic model. *Information Processing & Management*, 56(6):102098.

Pengtao Xie, Yuntian Deng, and Eric P. Xing. 2015. Diversifying restricted boltzmann machine for document modeling. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1315–1324.

Hao Zhang, Gunhee Kim, and Eric P. Xing. 2015. 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*, pages 1425–1434.

A Appendix

Brand	Average Rating	Number of Reviews	Distribution of Ratings				
			1	2	3	4	5
General	3.478	1103	236	89	144	180	454
VAGA	3.492	1057	209	116	133	144	455
Remington	3.609	1211	193	111	149	282	476
Hittime	3.611	815	143	62	110	154	346
Crest	3.637	1744	352	96	159	363	774
ArtNaturals	3.714	767	138	54	65	143	368
Urban Spa	3.802	1279	118	105	211	323	522
GiGi	3.811	1047	151	79	110	184	523
Helen Of Troy	3.865	3386	463	20	325	472	1836
Super Sunnies	3.929	1205	166	64	126	193	666
e.l.f	3.966	1218	117	85	148	241	627
AXE PW	4.002	834	85	71	55	169	454
Fiery Youth	4.005	2177	208	146	257	381	1185
Philips Norelco	4.034	12427	1067	818	1155	2975	6412
Panasonic	4.048	2473	276	158	179	419	1441
SilcSkin	4.051	710	69	49	58	135	399
Rimmel	4.122	911	67	58	99	160	527
Avalon Organics	4.147	1066	111	52	82	145	676
L'Oreal Paris	4.238	973	88	40	72	136	651
OZ Naturals	4.245	973	79	43	74	142	635
Andalou Naturals	4.302	1033	58	57	83	152	683
Avalon	4.304	1344	132	62	57	108	985
TIGI	4.319	712	53	32	42	93	492
Neutrogena	4.331	1200	91	55	66	142	846
Dr. Woods	4.345	911	60	42	74	83	652
Gillette	4.361	2576	115	94	174	555	1638
Jubujub	4.367	1328	53	42	132	238	863
Williams	4.380	1887	85	65	144	347	1246
Braun	4.382	2636	163	85	147	429	1812
Italia-Deluxe	4.385	1964	96	73	134	336	1325
Booty Magic	4.488	728	28	7	48	144	501
Greenvida	4.520	1102	55	33	51	108	855
Catrice	4.527	990	49	35	34	99	773
NARS	4.535	1719	60	36	107	237	1279
Astra	4.556	4578	155	121	220	608	3474
Heritage Products	4.577	837	25	18	52	96	646
Poppy Austin	4.603	1079	36	31	38	115	859
Aquaphor	4.633	2882	100	58	106	272	2346
KENT	4.636	752	23	8	42	74	605
Perfecto	4.801	4862	44	36	81	523	4178
Citre Shine	4.815	713	17	5	3	43	645
Bath & Body Works	4.819	2525	60	27	20	95	2323
Bonne Bell	4.840	1010	22	9	6	35	938
Yardley	4.923	788	3	4	3	31	747
Fruits & Passion	4.932	776	3	2	3	29	739
Overall	4.259	78322	5922	3623	5578	12322	50877

Table A1: Brand Statistics. The table shows the average rating score, the total number of associated reviews, and the distribution of the number of reviews for ratings ranging between 1 star to 5 stars, for each of the 45 brands.