Joint Author Sentiment Topic Model

Subhabrata Mukherjee*   Gaurab Basu†   Sachindra Joshi‡

Abstract

Traditional works in sentiment analysis and aspect rating prediction do not take *author preferences* and *writing style* into account during rating prediction of reviews. In this work, we introduce Joint Author Sentiment Topic Model (JAST), a generative process of writing a review by an author. Authors have different topic preferences, 'emotional' attachment to topics, writing style based on the distribution of *semantic* (topic) and *syntactic* (background) words and their tendency to *switch topics*. JAST uses Latent Dirichlet Allocation to learn the distribution of author-specific topic preferences and emotional attachment to topics. It uses a Hidden Markov Model to capture short range syntactic and long range semantic dependencies in reviews to capture *coherence* in author writing style. JAST jointly discovers the topics in a review, author preferences for the topics, topic ratings as well as the overall review rating from the point of view of an author. To the best of our knowledge, this is the first work in Natural Language Processing to bring all these dimensions together to have an *author-specific* generative model of a review.

1 Introduction

Sentiment analysis attempts to find customer preferences, likes and dislikes, and potential market segments from reviews, blogs, micro-blogs *etc.* A review may have multiple facets or topics, with a different opinion about each facet. Consider the following movie review: "*This film is based on a true-life story. It sounds like a great plot and the director makes a decent attempt in narrating a powerful story. However, the film does not quite make the mark due to sloppy acting.*"   ... (1)
This movie review is positive with respect to the topics 'direction' and 'story', but negative with respect to 'acting'. The overall rating for this review will differ for different authors depending on their topic preferences. If a reviewer watches a movie for a good story and narration, then his rating for the movie will be different than that if he watches it only for the acting skills of the

protagonists. Although sentiment analysis attempts to mine out customer preferences from data, it has largely overlooked the influence of author preferences during rating or polarity prediction of reviews.

Aspect rating prediction has received a great deal of attention in recent times. However, most of the works try to fit a global model over the entire corpus, *independent of the author of the review*. Traditional generative models not only overlook author topic preferences, but also ignore the author's writing style that is essential for maintaining coherence in reviews by detecting topic switch and semantic-syntactic class transition. For instance, in formal writing, female writing exhibits greater usage of 'involved' features, whereas male writing exhibits greater usage of 'informational' features [3]. Similarly, some authors are very verbose, whereas others make abrupt topic switches. It is essential to detect topic switch to maintain *coherence* in reviews for better association of facets to topics. We refer the association between facets to topics as *semantic dependencies*. In the above review the first two sentences refer to the same topic 'story' with facets like 'plot' and 'narration'. The author makes a topic switch in the next sentence using the discourse particle 'however'. We refer to the connection between facets and background words as *syntactic dependencies* which are required to make the review coherent and grammatically correct.

In this work, we introduce the Joint Author Sentiment Topic Model (JAST), a generative process of writing a review by an author. We use Latent Dirichlet Allocation (LDA) to learn the latent topics, topic-ratings which reflect the 'emotional' attachment of the author to the topics and author-specific topic preferences. LDA models the review as a bag of topics, ignoring the dependencies and coherence in the review writing process. In order to capture the syntactic and semantic dependencies in the review, Hidden Markov Model (HMM) is used to discover the syntactic classes, semantic topics and the author-specific semantic-syntactic class transition. The other purpose served by HMM is to incorporate coherence in the review writing process, where the reviewer dwells in a particular topic for sometime before moving on to some other topic. All the above observations are incorporated in an HMM-LDA based model that generates a review tailored to an author.
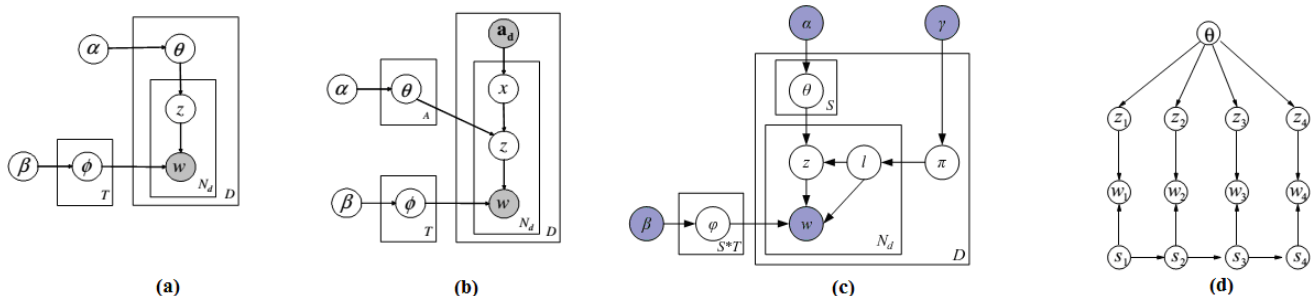
Figure 1: (a) LDA model (b) Author-Topic Model (c) Joint Sentiment Topic Model (JST) (d) Topic-Syntax Model

2 Related Work

Aspect rating prediction has received vigorous interest in recent times. Latent Aspect Rating Analysis Model (LARAM) [23, 24] jointly identifies latent aspects, aspect ratings, and weights placed on the aspects in a review. However, the model ignores author identity and writing style, and learns parameters *per review basis* in contrast to our model which learns the latent parameters *per author basis*.

A shallow dependency parser is used to learn product aspects and aspect-specific opinions in [26] by jointly considering the aspect frequency and the consumers opinions about each aspect. A rated aspect summary of short comments is done in [11]. Similar to LARAM, the statistics are aggregated at the comment level in this work and not at the author-level.

A topic model is used in [22] to assign words to a set of induced topics. The model is extended through a set of maximum entropy classifiers, one per each rated aspect, that are used to predict aspect specific ratings. The authors in [19] jointly learn ranking models for individual aspects by modeling dependencies between assigned ranks by analyzing meta-relations between opinions, such as agreement and contrast.

A joint sentiment topic model (JST) is described in [9] which detects sentiment and topic simultaneously from text. In JST (Figure 1.c), each document has a sentiment label distribution. Topics are associated to sentiment labels, and words are associated to both topics and sentiment labels. In contrast to [22] and some other similar works [23, 24, 26, 19, 11], which require some kind of a supervised setting like aspect ratings or overall review rating [12], JST is fully unsupervised. The CFACTS model [7] extends the JST model to capture facet coherences in a review using Hidden Markov Model. However, all these models do not incorporate any authorship information to incorporate *author preference* for the facets or *author style* information for maintaining coherence in reviews.

All these generative models have their root in La-

tent Dirichlet Allocation Model [1] (Figure 1.a). LDA assumes a document to have a probability distribution over a mixture of topics and topics to have a probability distribution over words. In the Topic-Syntax Model [4] (Figure 1.d), each document has a distribution over topics; and each topic has a distribution over words being drawn from classes, whose transition follows a distribution having a markov dependency. In the Author-Topic Model [18] (Figure 1.b), each author is associated with a multinomial distribution over topics. Each topic is assumed to have a multinomial distribution over words.

An approach to capture author-specific topic preference is described in [12]. The work considers seed facets like 'food', 'ambience', 'service' *etc.* and uses dependency parsing with a lexicon to find the sentiment about each facet. A WordNet similarity metric is used to assign each facet to a seed facet. Thereafter, they use linear regression to learn author preference for the seed facets from review ratings. The work is restrictive as it considers only manually given seed facets, topic-ratings are subjected to the lexicon coverage, and it does not incorporate review coherence.

2.1 Motivating JAST from Related Work:

While writing a review, an author has some topics in mind and for each topic there are facets. For example, an important topic in a movie review is 'acting', which has facets like 'semiotics', 'character', 'facial expression', 'vocabulary' *etc.* For each topic, the author decides on a topic-rating based on his actual experience. Thereafer, the author writes a word based on the topic, topic-rating and the *class*. Classes can be visualized as part-of-speech or semantic-syntactic categories for a word. For example, to frame the sentence "The acting is awesome" the author chooses a syntactic word from the 'Article' class, followed by a topic word from the class 'Nouns', a syntactic word from the class 'Verbs' and finally, a sentiment word from the class 'Adjectives'. The probability of a word drawn from a class not only depends on the class but also on the previous class. This

places a restriction on the way the classes are chosen, so that an invalid class distribution like "Article Noun Adjective Verb" cannot be formed. In this work, we further say that the class distribution is author-specific as each author has his own writing style. For example, the distribution of content words and function words in the writing of males and females is different [17]. In another experiment [3], women cognition for words in 'emotional' category are found to be more than men's. Even within the same gender, some authors are verbose whereas others make abrupt topic switches. In this work, we propose an author-specific generative process of a review that captures all aforementioned facets.

3 Joint Author Sentiment Topic Model

The JAST model describes a generative process of writing a review by an author.

3.1 Generative Process for a Review by an Author:

Consider a corpus with a set of D documents denoted by $\{d_1, d_2, \ldots d_D\}$. Each document has a sequence of N_d words denoted by $\{d = w_1, w_2, \ldots w_{N_d}\}$. Each word is drawn from a vocabulary V having unique words indexed by $\{1, 2, \ldots V\}$. Consider a set of A authors involved in writing the documents in the corpus, where a_d is the author of document d. Each document has an *author-specific overall review rating distribution* $\Omega_d^{a_d}$. Consider a set of R distinct sentiment ratings for any document. The author a_d draws a rating r_d from the multinomial distribution $\Omega_d^{a_d}$ for document d. Consider a sequence of topic assignments $\{z = z_1, z_2, \ldots z_T\}$, where each topic z_i can be from a set of T possible topics. Consider a sequence of sentiment rating assignments for topics $\{l = l_1, l_2, \ldots l_L\}$, where each topic z_i can have a sentiment rating l_i from a set of L possible topic-ratings. Consider a sequence of class assignments $\{c = c_1, c_2, \ldots c_C\}$, where each class c_i is from the set of C possible classes. The words in the document can now be generated as follows. The author a_d draws a rating r_d for document d from $\Omega_d^{a_d}$. The author then chooses a class c_i from the set of classes, where the *author-specific class transition* follows a distribution $\pi^{c_{i-1}}$. If $c_i = 1$, the author decides to write on a new topic. The author chooses a topic z_i and its sentiment rating l_i from the topic-rating distribution ϕ_{a_d, r_d} conditioned on the overall rating r_d chosen by a_d for the document. If $c_i = 2$, the author decides to continue writing on the previous topic. However, he chooses a new sentiment rating l_i for the topic from ϕ_{a_d, r_d}. Once the topic and its label are decided, the author draws a word from the *per-corpus word distribution* ξ_{z_i, l_i}. If $c_i \neq 1, 2$, the author decides to draw a background word from the syntactic class distribution ξ_{c_i, l_i}, where $l_i = 0$ is the objective po-

larity of the function word drawn. Once all the latent topics and topic-ratings are learnt, the author revises his estimate of the overall document rating distribution over the learnt parameters. Figure 2 shows the graphical model of JAST. Algorithm 3.1 shows the generative process for JAST.
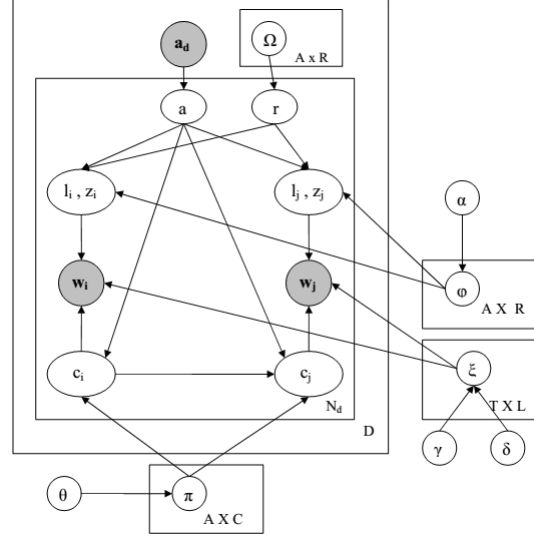

Figure 2: Graphical Model of JAST

ALGORITHM 3.1. Generative Process for a Review by an Author

1. For each document d, the author a_d chooses an overall rating $r_d \sim Mult(\Omega_d^{a_d})$ from the author-specific overall document rating distribution

2. For each topic $z_i \in z$ and each sentiment label $l_i \in l$, draw $\xi_{z_i, l_i} \sim Dir(\gamma)$

3. For each class $c_i \in c$ and each sentiment label $l_i = 0 \in l$, draw $\xi_{c_i, l_i} \sim Dir(\delta)$

4. Choose the author-specific class transition distribution π_{a_d}

5. The author a_d chooses the author-rating specific topic-label distribution $\varphi_{a_d, r_d} \sim Dir(\alpha)$

6. For each word w_i in the document

   (a) Draw $c_i \sim Mult(\pi_{a_d}^{c_{i-1}})$.

   (b) If $c_i = 1$, Draw $z_i, l_i \sim Mult(\varphi_{a_d, r_d})$. Draw $w_i \sim Mult(\xi_{z_i, l_i})$.

   (c) If $c_i = 2$, Draw $z_{i-1}, l_i \sim Mult(\varphi_{a_d, r_d})$. Draw $w_i \sim Mult(\xi_{z_{i-1}, l_i})$.

   (d) If $c_i \neq 1, 2$, Draw $w_i \sim Mult(\xi_{c_i, l_i})$.

Some authors are more inclined to give average ratings to reviews than extreme ones; whereas some authors are more likely to assign extreme ratings to reviews than moderate ones. Correspondingly, the topic-label distribution also differs for an author across the review ratings. The hyper-parameter $\alpha_{z_i,l_i}^{a,r}$ is the prior observation on the number of times topic z_i is associated to label l_i in a document rated r by an author a.

The hyper-parameters γ and δ are the prior observation on the number of times the word w_i is associated to topic z_i and class c_j with sentiment labels l_i and $l_j = 0$ respectively. The transition between classes is influenced by an author's stylistic features. The hyper-parameter θ_a is the prior observation on the number of class transitions for an author a, that form the rows of the transition matrix π_a in the Hidden Markov Model.

3.2 Model Inferencing:

In this section, we discuss the inferencing algorithm to estimate the distributions Ω, ϕ, ξ and π in JAST. For each author, we compute the conditional distribution over the set of hidden variables l, z and c for all the words in a review and r for the overall review. The exact computation of this distribution is intractable. EM algorithm can also be used to estimate the parameters, but it is shown to perform poorly for topical models with many parameters and multiple local maxima. So we use collapsed Gibbs Sampling [4] to estimate the conditional distribution for each hidden variable, which is computed over the current assignment over all other hidden variables and integrating out other parameters in the model.

Let A, R, Z, L, C and W be the set of all authors, ratings, topics, topic ratings, classes and words in the corpus. The joint probability distribution of JAST is given by:

$$
\begin{aligned}
&P(A, R, Z, L, C, W, \Omega, \phi, \xi, \pi; \alpha, \gamma, \theta, \delta) = \\
&\prod_{x=1}^{A} \prod_{i=1}^{D} P(\Omega_i^x) \prod_{y=1}^{R} P(\phi_{x,y}; \alpha) \times \\
&\prod_{k=1}^{T} \prod_{u=1}^{L} P(\xi_{k,u}; \gamma, \delta) \times \\
&\prod_{s=1}^{C} P(\pi_{x,s}; \theta) \times P(r_i|\Omega_i^x) \prod_{j=1}^{N_d} P(z_{i,j}, l_{i,j}|\phi_{x,r_i}) \times \\
&P(c_{i,j}|\pi_{x,s}) \times P(w_{i,j}|\xi_{z_{i,j},l_{i,j}}, \xi_{c_{i,j},l_{i,j}} \pi_{x,c_{i,j}}) \cdots (1)
\end{aligned}
$$

Let $n_{d,v,t,l,c}^{a,r}$ be the number of times the word w, indexed by the vth word in the vocabulary, appears in the document d with rating r, written by author a, in the topic t with topic rating l and class c. $z_{i,j}$ denotes the topic of the jth word of the ith document written by an author.

Integrating out ϕ, $P(Z, L|A, R; \alpha)$ is given by,

$$
\prod_{x=1}^{A} \prod_{y=1}^{R} \frac{\Gamma(\sum_{k,u} \alpha_{k,u}^{x,y}) \prod_{k,u} \Gamma(n_{(.),(.),k,u,(.)}^{x,y} + \alpha_{k,u}^{x,y})}{\prod_{k,u} \Gamma(\alpha_{k,u}^{x,y}) \Gamma(\sum_{k,u} n_{(.),(.),k,u,(.)}^{x,y} + \sum_{k,u} \alpha_{k,u}^{x,y})}
$$

Integrating out ξ, $P(W|Z, L, C; \gamma)$ is given by,

$$
\underbrace{\prod_{k=1}^{K} \prod_{u=1}^{L} \frac{\Gamma(\sum_v \gamma_v) \prod_v \Gamma(n_{(.),v,k,u,1}^{(.),(.)} + \gamma_v)}{\prod_v \Gamma(\gamma_v) \Gamma(\sum_v n_{(.),v,k,u,1}^{(.),(.)} + V\gamma)}}_{g_1}, c = 1
$$

$$
\underbrace{\prod_{u=1}^{L} \frac{\Gamma(\sum_v \gamma_v) \prod_v \Gamma(n_{(.),v,k=k^*,u,2}^{(.),(.)} + \gamma_v)}{\prod_v \Gamma(\gamma_v) \Gamma(\sum_v n_{(.),v,k=k^*,u,2}^{(.),(.)} + V\gamma)}}_{g_2}, c = 2, k^* = z_{i,j-1}
$$

$$
\underbrace{\prod_{c=1}^{C} \frac{\Gamma(\sum_v \delta_v) \prod_v \Gamma(n_{(.),v,(.),u=0,c}^{(.),(.)} + \delta_v)}{\prod_v \Gamma(\delta_v) \Gamma(\sum_v n_{(.),v,(.),u=0,c}^{(.),(.)} + \delta\gamma)}}_{g_3}, c \neq 1, 2
$$

Let m_{a,c_i}^{a,c_i-1} denote the number of class transitions from the $(i-1)^{th}$ class to the ith class for the author a. The class transition probability is $P(c_i|c_{i-1}, a) = \frac{(m_{a,c_i}^{a,c_i-1} + \theta_a)(m_{a,c_{i+1}}^{a,c_i} + I_a(c_{i-1}=c_i) \times I_a(c_{i+1}=c_i+\theta_a))}{m_{a,(.)}^{a,c_i} + I_a(c_{i-1}=c_i) + C\theta_a}$.

The conditional distribution for the class transition

$$
P(C|A, Z, L, W) \propto P(W|Z, L, C) \times P(C|A)
$$

$$
\propto \begin{cases} \prod_{a=1}^{A} g_1 \times P(c_i|c_{i-1}, a), c_i = 1 \\ \prod_{a=1}^{A} g_2 \times P(c_i|c_{i-1}, a), c_i = 2 \qquad \cdots (2) \\ \prod_{a=1}^{A} g_3 \times P(c_i|c_{i-1}, a), c_i \neq 1, 2 \end{cases}
$$

In case of a supervised model, the ratings are observables and hence the distribution Ω_i^x is known. However, in our case JAST is unsupervised and we estimate the overall rating distribution as follows. At any step of the review generation process, the overall review rating of the document influences the topic and topic-rating selection for individual words in the review. Once all the topics and topic-ratings are determined for all the words in a review, the review rating can now be visualized as a response variable with a probability distribution over all the latent topic and topic-ratings in the review. Such a kind of updation model is used in [7]. The overall review rating distribution can be updated now as:

$$
\Omega_d^{a,r} = \frac{\sum_{k,u} I(r = \arg\max_{r^*} \phi_{a,r^*}[k,u]) \times \phi_{a,r}[k,u]}{K} \cdots (3)
$$

For each topic k with topic-rating u, the above equation finds the rating that maximizes the author-specific topic-rating preference given by $\phi_{a,r}$. In Gibbs sampling, the conditional distribution is computed for each hidden variable, based on the current assignment of other hidden variables. The values for the latent variables are sampled repeatedly from this conditional distribution until convergence. Let the subscript $-i$ denote the value of the variable excluding the data at the ith

position. The conditional distributions for Gibbs sampling for JAST are given by:

$$P(z_i = k, l_i = u | a_d = a, r_d = r, z_{-i}, l_{-i}) \propto$$
$$\frac{n^{a,r}_{(.),(.),k,u,(.)} + \alpha^{a,r}}{\sum_{k,u} n^{a,r}_{(.),(.),k,u,(.)} + \sum_{k,u} \alpha^{a,r}_{k,u}} \cdots (4)$$

$$P(w_i = w | z_i = k, l_i = u, c_i = c, w_{-i}) \propto$$

$$\underbrace{\frac{n^{(.),(.)}_{(.),w,k,u,1} + \gamma}{\sum_w n^{(.),(.)}_{(.),w,k,u,1} + V\gamma}, c = 1}_{h_1}$$

$$\underbrace{\frac{n^{(.),(.)}_{(.),w,k=k^*,u,2} + \gamma}{\sum_w n^{(.),(.)}_{(.),w,k=k^*,u,2} + V\gamma}, c = 2, k^* = z_{i-1}}_{h_2}$$

$$\underbrace{\frac{n^{(.),(.)}_{(.),w,(.),u=0,c} + \delta}{\sum_w n^{(.),(.)}_{(.),w,(.),u=0,c} + V\delta}, c \neq 1, 2 \cdots (5)}_{h_3}$$

$$P(c_i = c | a_d = a, z_i = k, l_i = u, c_{-i}, w_i = w) \propto$$
$$h_1 \times P(c_i | c_{i-1}, a), c = 1$$
$$h_2 \times P(c_i | c_{i-1}, a), c = 2$$
$$h_3 \times P(c_i | c_{i-1}, a), c \neq 1, 2 \cdots (6)$$

The conditional distribution for the joint updation of the latent variables is given by,

$$P(z_i = k, l_i = u, c_i = c | a_d = a, r_d = r, z_{-i}, l_{-i}, c_{-i}, w_i = w) \propto$$

$$\frac{n^{a,r}_{(.),(.),k,u,(.)} + \alpha}{\sum_{k,u} n^{a,r}_{(.),(.),k,u,(.)} + \sum_{k,u} \alpha^{a,r}_{k,u}} \times \frac{n^{(.),(.)}_{(.),w,k,u,1} + \gamma}{\sum_w n^{(.),(.)}_{(.),w,k,u,1} + V\gamma} \times \Omega^{a,r}_d,$$
$$c = 1$$

$$\frac{n^{a,r}_{(.),(.),k,u,(.)} + \alpha}{\sum_{k,u} n^{a,r}_{(.),(.),k,u,(.)} + \sum_{k,u} \alpha^{a,r}_{k,u}} \times \frac{n^{(.),(.)}_{(.),w,k=z_{i-1},u,2} + \gamma}{\sum_w n^{(.),(.)}_{(.),w,k=z_{i-1},u,2} + V\gamma} \times \Omega^{a,r}_d$$
$$c = 2$$

$$\frac{n^{a,r}_{(.),(.),k,u,(.)} + \alpha}{\sum_{k,u} n^{a,r}_{(.),(.),k,u,(.)} + \sum_{k,u} \alpha^{a,r}_{k,u}} \times \frac{n^{(.),(.)}_{(.),w,(.),u=0,c} + \delta}{\sum_w n^{(.),(.)}_{(.),w,(.),u=0,c} + V\delta} \times \Omega^{a,r}_d,$$
$$c \neq 1, 2 \cdots (7)$$

The model parameters Ω, ϕ, ξ and π are updated according to equations $3, 4, 5$ and 6 respectively.

3.3 Rating Prediction of Reviews:
JAST model assumes the identity of the author to be known. Once the model parameters are learnt, for each of the words in the given review its topic and topic-rating (k, u) are extracted from $\xi_{T \times L}[w]$. The overall review rating of document d is given by $\arg\max_r \Omega^{a,r}_d$. For an unseen document refer to equation 3 to estimate $\Omega^{a,r}_d$.

4 Experimental Evaluation

We performed experiments in movie review and restaurant review domains. The first one is the widely used IMDB movie review dataset [16] which serves as a benchmark in sentiment analysis. The second dataset consists of restaurant reviews from Tripadvisor [12].

4.1 Dataset Pre-Processing:
The movie review dataset contains 2000 reviews and 312 authors with at least 1 review per author. In order to have sufficient data per author, we retained only those authors with at least 10 reviews. This reduced the number of reviews to 1467 with 65 authors. The number of reviews for the 2 ratings (pos and neg) is balanced in this dataset.

The restaurant review dataset contains 1526 reviews and 9 authors. Each review has been rated by an author on a scale of 1 to 5. However, the number of reviews per rating is highly skewed towards the mid ratings. In order to have sufficient data for learning per review rating, oversampling is done to increase the number of reviews per rating. JAST uses 80% (unlabeled) data per author to learn parameters during inferencing. Table 1 shows the dataset statistics.

4.2 Incorporating Model Prior:
Bing Liu sentiment lexicon [5] is used to initialize the polarity of the words as *positive, negative* or *objective* in the matrix $\xi_{T \times L}[W]$ prior to inferencing. The lexicon contains 2006 positive terms and 4783 negative terms. The review ratings, topic and class labels are initialized randomly. The dirichlet priors are taken to be symmetric.

JAST requires the initialization of 2 important model parameters, *i.e.* the number of topics (T) and the number of classes (C). The number of authors (A), topic ratings (L) and review ratings (R) are pre-defined according to the dataset in hand. We use the *model perplexity* to initialize T and C, which is an important measure used for language and topic modeling. A higher value of perplexity indicates a lesser model likelihood and hence lesser generative power of the model. We analyze the change in model perplexity with the change in parameters $(T$ and $C)$, by keeping one constant and varying the other. Finally, the values are chosen at which the model perplexity is the minimum. Table 2 shows the model initialization parameters.

4.3 Model Baselines:
Lexical classification is taken as the first baseline for our work. A sentiment lexicon [5] is used to find the polarity of the words in the review. The final review polarity is taken to be the majority polarity of the opinion words in the review. Negation handling is done in this baseline. The same sentiment lexicon is also used in the JAST model for incorporating prior information.

Joint Sentiment Topic Model [9] is considered to be the second baseline for JAST. It does not incorporate

Dataset	Authors	Avg Rev/ Author	Rev/ Rating							Avg Rev Length	Avg Words/ Rev
Movie Review*	312	7	**Pos** 1000		**Neg** 1000			**Total** 2000		32	746
Movie Review⊥	65	23	**Pos** 705		**Neg** 762			**Total** 1467		32	711
Restaurant Review*	9	170	**R 1** 43	**R2** 134	**R 3** 501	**R 4** 612	**R 5** 237	**Total** 1526		16	71
Restaurant Review⊥	9	340	**R 1** 514	**R 2** 532	**R 3** 680	**R 4** 700	**R 5** 626	**Total** 3052		20	81

Table 1: Movie Review and Restaurant Review Dataset Statistics (R denotes review rating) (* denotes the original dataset, ⊥ indicates processed dataset)

Model Parameters	Movie Review	Restaurant Review
A	65	9
R	2	5
T	50	25
L	3	3
C	20	15
$\alpha = 1/T \times L$	0.007	0.013
$\gamma = 1/T \times L$	0.007	0.013
$\delta = 1/C \times L$	0.017	0.022
$\theta = 1/A \times C$	0.0007	0.007

Table 2: Model Initialization Parameters

Models	Accuracy
Lexical Baseline	65
JST [9]	82.8
Mukherjee *et al.* (2013) [12]	84.39
JAST	**87.69**

Table 3: Accuracy Comparison of JAST with Baselines in IMDB Movie Review Dataset

author information or syntactic/semantic dependencies. Since we do not perform subjectivity detection (which is a supervised classification task) before inferencing, we compare our work with JST model performance, without subjectivity analysis, using only unigrams. It is notable here that JAST, like JST, is unsupervised, apart from the prior lexicon used.

The third baseline considered for our work is [12]. The authors consider a set of seed facets and use dependency parsing with a lexicon to find facet ratings. They use linear regression to find author-facet preferences from the overall review ratings and facet ratings.

A large number of works have been reported on the IMDB movie review dataset [16]. We compare the performance of our approach to all the existing state-of-the-art systems in the domain on that dataset.

4.4 Results: For the IMDB movie review dataset, if $|\Omega_d^{a,r=+1} - \Omega_d^{a,r=-1}| < \epsilon$, the lexical baseline rating is taken instead of the JAST rating. Such cases are indicative of the fact that either of the 2 ratings may be possible for the review and JAST cannot make the decision. Table 3 shows the accuracy comparison of JAST with the baselines in the movie review dataset. Table 4 shows the accuracy comparison of JAST with the existing models in the domain on the same dataset.

The restaurant reviews are rated on a scale of 1 to 5, compared to the binary rating in movie reviews. In the aspect rating problem, a low value of mean absolute error (MAE) between predicted and ground ratings is taken as a good performance indicator. Table 5 shows the MAE comparison of the baselines with JAST on the restaurant reviews. Table 6 shows a snapshot of the extracted words, given topic and topic ratings. Graph 1 and 2 shows the variation in the probability of an author liking a specific facet with overall review rating in the 2 datasets respectively.

5 Discussions

1. Review Rating Distribution: In the movie review dataset, pre-processing filtered out authors having less than 10 reviews each. This decreased the number of authors by 79%. It is observed that the average number of positive reviews per author (13) is less than the number of negative reviews per author (18) in the processed dataset, which means movie critics can be difficult to impress. On the contrary, in restaurant reviews number of good ratings (R 4) > average ratings (R 3) ≫ excellent ratings (R 5) ≫ bad ratings (R 2) ≫ worse ratings (R 1). This means food critics are more likely to write positive reviews than negative ones.

2. Comparison with Lexical and Author-Specific Baselines: JAST performs much better than lexical baseline in both domains with an accuracy improvement of 22% in movie reviews and mean absolute error (MAE) reduction of 0.63 in restaurant reviews.

JAST achieves an accuracy improvement of 3.3% over the supervised author-specific baseline in [12] in
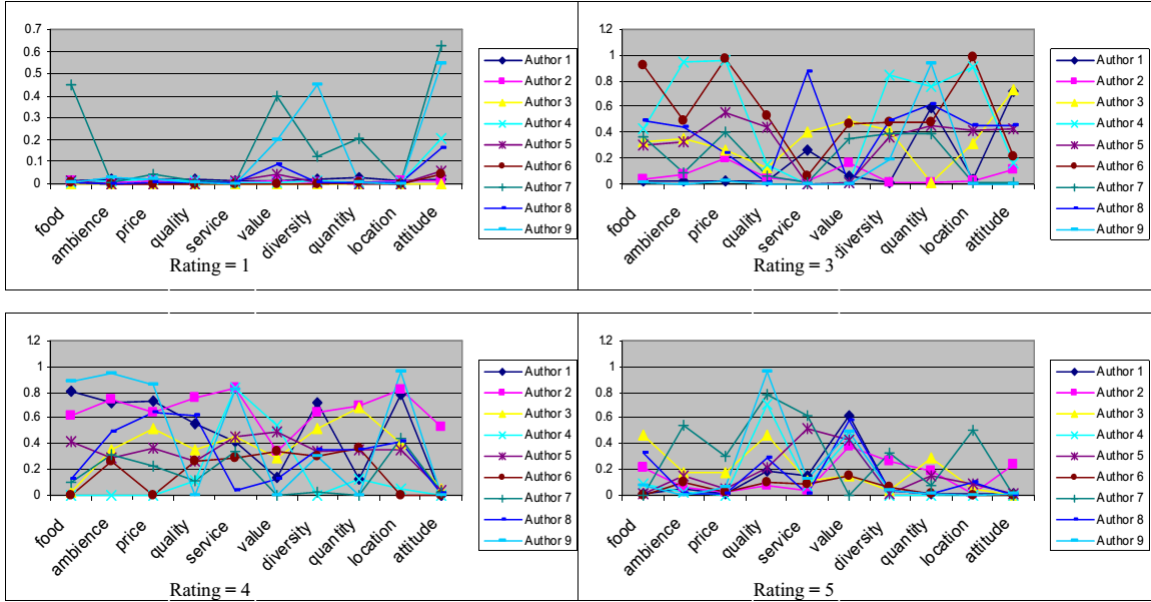
Figure 3: Variation in Author Satisfaction for Facets with Overall Review Rating in Restaurant Review Dataset
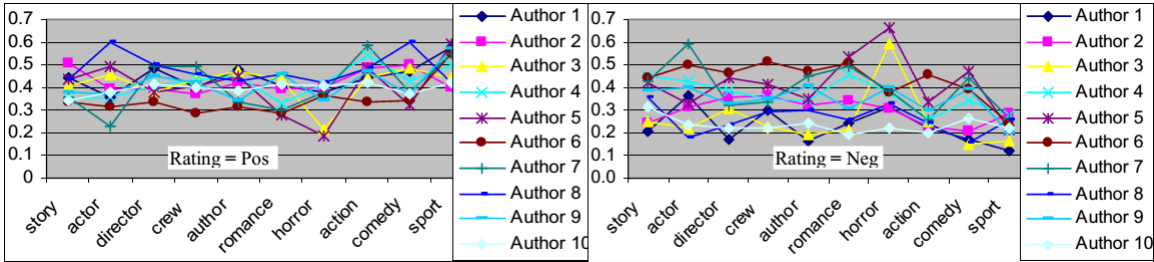

Figure 4: Variation in Author Satisfaction for Facets with Overall Review Rating in Movie Review Dataset

movie reviews. In restaurant reviews, JAST attains an MAE reduction of 0.14 over facet-specific topic rating model, and 0.10 over facet-specific author-specific topic rating model. Unlike the baseline model using a handful of seed facets, JAST discovers all latent facets, facet-ratings and author-facet preferences.

3. Comparison with Joint Sentiment Topic Model: JAST achieves an accuracy improvement of 5% over JST [9] without subjectivity analysis in IMDB dataset, and MAE reduction of 0.40 in restaurant reviews. Unlike JST, JAST incorporates authorship information to find author-specific topic preferences and author writing style to maintain review coherence.

JAST has less data requirement than JST as it uses 80% data to learn parameters during inferencing, compared to JST which uses the entire dataset. However, JAST has an overhead of author identity requirement. Another distinction is that JST learns all the document ratings during inferencing itself. Unlike JAST, it does not say how to find the rating of an unseen document.

Subjectivity analysis has been shown to improve

classifier performance in sentiment analysis. We did not incorporate subjectivity detection in our model as the task is fully supervised requiring another set of labeled data with sentences tagged as subjective or objective. But even with subjectivity detection JST fares poorly compared to JAST.

4. Comparison with Other Models in Movie Review Dataset: The proposed JAST model is unsupervised requiring no labeled data for training (only authorship information). However, it attains a much better performance than the supervised version of the Recursive Auto Encoders [20] and Tree-CRF [14], both of which report 10% cross-validation accuracy. It also performs better than the supervised classifiers used in [16], [15] and [6]. JAST performs much better than all the other unsupervised and semi-supervised works in the movie review domain.

5. Topic Label Word Extraction: In the JAST model, the author first chooses a syntactic or a semantic class from the author-class distribution. Given that semantic class is chosen, the author gets to choose a

Movie Review Dataset					Restaurant Review Dataset				
T=bad	T=good	T=actor	T=actor	T= actor	T=food	T=food	T=food	T=service	T=bad
L=neg	L=pos	L=neg	L=pos	L=obj	L=obj	L=neg	L=pos	L=pos	L=neg
bad	good	kevin	funny	cruise	food	bad	dish	ambience	average
suppose	great	violence	comedy	name	diner	awful	price	face	noth
bore	sometimes	comic	laugh	run	customer	seem	din	hearty	wasn
unfortunate	different	early	joke	ship	sweet	just	first	pretty	bad
stupid	hunt	someth	fun	group	kitchen	cheap	beautiful	exceptional	basic
waste	truman	not	eye	patch	feel	wasn	chicken	diner	nor
ridiculous	sean	long	talk	creature	meal	stop	quality	friendly	didn
half	excellent	every	hour	tribe	front	cold	recommend	perfection	don
terrible	relationship	support	act	big	home	quite	lovely	help	last
lame	amaze	type	moment	rise	serve	small	taste	worth	probably
dull	damon	somewhat	close	board	warm	loud	fun	extra	slow
poorly	martin	question	scene	studio	waitress	no	available	effort	sometimes
attempt	chemistry	fall	picture	sink	treat	common	definitely	warm	serious

Table 6: Extracted Words given Topic (T) and Label (L) in Movie and Restaurant Review Dataset

Models	Acc.
Eigen Vector Clustering [2]	70.9
Semi Supervised, 40% doc. Label [8]	73.5
LSM Unsupervised with prior info [10]	74.1
SO-CAL Full Lexicon [21]	76.37
RAE Semi Supervised Recursive Auto Encoders with random word initialization [20]	76.8
WikiSent: Extractive Summarization with Wikipedia + Lexicon [13]	76.85
Supervised Tree-CRF [14]	77.3
RAE: Supervised Recursive Auto Encoders with 10% cross-validation [20]	77.7
JST: Without Subjectivity Detection using LDA [9]	82.8
JST: With Subjectivity Detection [9]	84.6
Pang *et al.* (2002): Supervised SVM [16]	82.9
Supervised Subjective MR, SVM [15]	87.2
Kennedy *et al.* (2006): Supervised SVM [6]	86.2
Appraisal Group: Supervised [25]	90.2
JAST: Unsupervised HMM-LDA	**87.69**

Table 4: Comparison of Existing Models with JAST in the IMDB Dataset

Models	MAE
Lexical Baseline (Hu et. al 2004)	1.24
JST [9]	1.01
Facet Specific General Author Preference [12]	0.75
Facet and Author Specific Preference [12]	0.71
JAST	**0.61**

Table 5: Mean Absolute Error Comparison of JAST with Baselines in the Restaurant Dataset

topic labels manually assigned to the word clusters. It is observed that the extracted words are meaningful and coherent which serve as a qualitative evaluation of the effectiveness of JAST in extracting topic-label-words.

6. Author Rating Topic Label Extraction: In the JAST model, an author draws an overall rating for the review from the author-specific review rating distribution. The author chooses a syntactic/semantic class from the author-class distribution. Given that semantic class is chosen, the author chooses a topic and topic-rating conditioned on the overall rating of the review. Graph 1 and 2 show the probability of an author liking a specific facet in a review with his chosen rating in the 2 domains. The graph traces out the reason why an author assigns a specific rating to a given review. Let us consider Author 1 and Graph 1. The author talks highly of 'value' in reviews rated 5 by him, which is thus a very important facet to him. The author assigns a rating 4 to reviews where he finds the facets 'location', 'diversity', 'price', 'ambience', 'food' satisfactory. However, the author does not find much 'value' in these reviews probably due to a poor 'attitude' (which has a very low probability in such reviews). In the reviews rated 3, the author finds only the facets 'attitude' and 'quantity' of food interesting while the rest does not attract him much, and consequently the review gets an average rating. For obvious reasons, the

new topic to write on (or continues writing on the previous topic), as well as the topic rating based on the overall rating chosen by him for the review. Once the topic and topic-rating are decided, the author chooses a word from the topic and topic-rating distribution of the corpus. The last distribution is author-independent and depends only on the per corpus word distribution.

Table 6 shows a snapshot of the extracted words, given the chosen topic and topic-rating. Given a seed word (T) and the desired topic rating L, the topic that maximizes the corresponding word and topic-rating distribution is chosen. The words in the corresponding distribution are shown in the column corresponding to (T,L), in descending order of their probabilities, with

author does not like any facet in reviews rated 1.

In Graph 2, Author 1 gives positive rating to movies with a good 'story' (and hence 'author') and 'director' ('actor' may be poor), and negative rating to those with a good 'actor' but poor 'story' and 'director'; whereas the preferences of Author 8 is the contrary, which validates our claim in Example 1 of introduction.

6 Conclusions and Future Work

In this work, we have shown that sentiment classification and aspect rating prediction models can be improved a lot, if the author identity is known. Authorship information is required to extract author-specific topic preferences and ratings, as well as to maintain the review coherence by exploiting the author's writing style, reflected from the author-specific semantic-syntactic class transition and topic switch. The proposed JAST model is unsupervised (except the sentiment lexicon used to incorporate prior information), although it bears the overhead of requiring the author identity. Experiments are conducted in movie review and restaurant review domains, where JAST is found to perform much better than other existing models. It is found to perform even better than the supervised classification models in the movie review domain in the benchmark IMDB dataset.

In future work, we would like to experiment with other features like incorporating higher order information in the form of bigrams and trigrams, subjectivity detection (for movie reviews) etc. It would also be interesting to use JAST for author-ship attribution task and predict the author for a review, given the overall rating and the learnt model parameters.

References

[1] DAVID M. BLEI, ANDREW Y. NG, AND MICHAEL I. JORDAN, *Latent dirichlet allocation*, J. Mach. Learn. Res., 3 (2003), pp. 993–1022.

[2] SAJIB DASGUPTA AND VINCENT NG, *Topic-wise, sentiment-wise, or otherwise?: Identifying the hidden dimension for unsupervised text classification*, EMNLP '09, 2009, pp. 580–589.

[3] BREMNER ET AL., *Gender differences in cognitive and neural correlates of remembrance of emotional words.*, Psychopharmacol Bull, 35 (2001), pp. 55–78.

[4] THOMAS L GRIFFITHS ET AL., *Integrating topics and syntax*, Advances in NIPS, 17 (2005), pp. 537–544.

[5] MINQING HU AND BING LIU, *Mining and summarizing customer reviews*, KDD '04, pp. 168–177.

[6] ALISTAIR KENNEDY AND DIANA INKPEN, *Sentiment classification of movie reviews using contextual valence shifters*, Computational Intelligence, 22 (2006).

[7] HIMABINDU LAKKARAJU ET AL., *Exploiting coherence in reviews for discovering latent facets and associated sentiments*, in SDM '11, 2011, pp. 28–30.

[8] TAO LI, YI ZHANG, AND VIKAS SINDHWANI, *A non-negative matrix tri-factorization approach to sentiment classification with lexical prior knowledge*, in ACL/IJCNLP, 2009, pp. 244–252.

[9] CHENGHUA LIN AND YULAN HE, *Joint sentiment/topic model for sentiment analysis*, CIKM '09, pp. 375–384.

[10] CHENGHUA LIN, YULAN HE, AND RICHARD EVERSON, *A comparative study of bayesian models for unsupervised sentiment detection*, CoNLL '10, pp. 144–152.

[11] YUE LU, CHENGXIANG ZHAI, AND NEEL SUNDARESAN, *Rated aspect summarization of short comments.*, in WWW, 2009, pp. 131–140.

[12] SUBHABRATA MUKHERJEE, GAURAB BASU, AND SACHINDRA JOSHI, *Incorporating author preference in sentiment rating prediction of reviews*, 2013.

[13] SUBHABRATA MUKHERJEE AND PUSHPAK BHATTACHARYYA, *Wikisent: weakly supervised sentiment analysis through extractive summarization with wikipedia*, ECML PKDD'12, 2012, pp. 774–793.

[14] TETSUJI NAKAGAWA, KENTARO INUI, AND SADAO KUROHASHI, *Dependency tree-based sentiment classification using crfs with hidden variables*, HLT '10, 2010.

[15] BO PANG AND LILLIAN LEE, *A sentimental education: sentiment analysis using subjectivity summarization based on minimum cuts*, ACL '04, 2004.

[16] BO PANG AND VAITHYANATHAN SHIVAKUMAR LEE, LILLIAN, *Thumbs up?: sentiment classification using machine learning techniques*, EMNLP '02, 2002.

[17] JAMES W. PENNEBAKER ET AL., *Linguistic Inquiry and Word Count*, Mahwah, NJ, 2001.

[18] MICHAL ROSEN-ZVI ET AL., *The author-topic model for authors and documents*, UAI '04, 2004, pp. 487–494.

[19] BENJAMIN SNYDER AND REGINA BARZILAY, *Multiple aspect ranking using the good grief algorithm*, in HLT 2007, ACL, April 2007, pp. 300–307.

[20] RICHARD SOCHER ET AL., *Semi-supervised recursive autoencoders for predicting sentiment distributions*, EMNLP '11, 2011, pp. 151–161.

[21] MAITE TABOADA ET AL., *Lexicon-based methods for sentiment analysis*, Computational linguistics, 37 (2011), pp. 267–307.

[22] IVAN TITOV AND RYAN T. MCDONALD, *A joint model of text and aspect ratings for sentiment summarization.*, ACL, 2008, pp. 308–316.

[23] HONGNING WANG, YUE LU, AND CHENGXIANG ZHAI, *Latent aspect rating analysis on review text data: a rating regression approach*, in KDD, 2010, pp. 783–792.

[24] HONGNING WANG ET AL., *Latent aspect rating analysis without aspect keyword supervision*, KDD '11, 2011.

[25] CASEY WHITELAW, NAVENDU GARG, AND SHLOMO ARGAMON, *Using appraisal groups for sentiment analysis*, CIKM '05, 2005, pp. 625–631.

[26] JIANXING YU ET AL., *Aspect ranking: Identifying important product aspects from online consumer reviews.*, ACL, 2011, pp. 1496–1505.