

Joint Sentiment-Topic Detection from Text Document

Gauri Nivrutti Tuplondhe

Department of Computer Engineering,
KKWIEER, University of Pune,
Nashik- 422003, India

Abstract: Automated tools are used to detect subjective information like attitudes, opinions and feelings. Such process is called as sentiment analysis. The Joint Sentiment-Detection (JST) model is the probabilistic model which is extension of Latent Dirichlet Allocation (LDA) model that detects sentiment and topic simultaneously from text. Supervised approaches to sentiment classification often fail to produce satisfactory results when applied to other domains while the JST model is weakly supervised in nature where supervision only comes from domain independent sentiment lexicon. Thus, makes JST model portable to other domains. The proposed system incorporates a small amount of domain independent prior knowledge which is sentiment lexicon to further improve the sentiment classification accuracy. It also carry out experiments and evaluates the model performance on different datasets.

Keywords: Joint sentiment-topic (JST) model, Latent Dirichlet Allocation (LDA) , semi-supervised approach, sentiment analysis.

1. INTRODUCTION

Companies and consumers have the greater impact of opinion reach resources like online reviews and social networks compared to traditional media. The demand of gleaning insights into such vast amount of user-generated data, work on developing new algorithms for automated sentiment analysis has bloomed in the past few years.

Sentiment classification is the major task of sentiment analysis. A large portion of work concentrates on classifying a sentiment-bearing document according to its sentiment polarity, i.e. either positive or negative as a binary classification like [1], [2], [3], [9]. Most of this work rely on labeled corpora where documents are labeled as positive, negative prior to the training. In real world applications such labeled corpora may not be easily available. Also, sentiment classification models trained in one domain might not work well when moving to another domain. Furthermore, topic/feature detection and sentiment classification are mostly performed separately. But sentiments are context dependent, so that sentiment expressions can be quite different for different topics or domains. For instance, when appearing under different topics within movie review data, the adjective “complicated” may have negative sentiment orientation as “complicated role” in one topic, and positive orientation as “complicated plot” in another topic. This suggests that modeling sentiment and topic simultaneously may help find better feature representations for sentiment classification. Therefore, these problems motivated the need of using weakly supervised or unsupervised approaches for domain-independent sentiment classification.

Sentiment and topic of sentiment are simultaneously detected from text at document level by Joint Sentiment-Topic (JST) which is weakly supervised in nature. A mechanism is introduced to incorporate prior information about the sentiment lexicons into model learning by modifying the Dirichlet priors of the topic-word distributions. . This model extends the topic model latent dirichlet allocation (LDA) [6] by adding sentiment layer. It is different from other sentiment-

topic model in that: 1) It is weakly supervised. 2) It can detect topics and sentiment simultaneously. Unlike supervised approaches to sentiment classification, which often fail to produce satisfactory performance when applied to other domains, the weakly-supervised nature of JST makes it highly portable to other domains, as will be verified by the experimental results on datasets from different domains.

2. RELATED WORK

2.1 Sentiment Classification

Standard machine learning techniques such as support vector machines (SVMs) and Naive Bayes (NB) classifiers are used for sentiment classification approaches. These approaches are corpus-based, in which a domain-specific classifier is trained with labeled training data. The work in [3] employed machine learning techniques including SVMs, NB and Maximum Entropy to determine whether the sentiment expressed in a movie review was “thumbs up” or “thumbs down”. In subsequent work [4], they further improved sentiment classification accuracy on the movie review dataset using a cascaded approach. The work [2], [3], [4] only focus on sentiment classification in one domain while the work in [5] addresses the issue of cross-domain sentiment classification. Four strategies have been explored for customizing sentiment classifiers to new domains [5] like small number of labeled examples can be used as training set or it can combine labeled data with large amount of unlabeled data from target domain. All the above work has some similar limitations: 1) the mixture of topics is ignored while doing sentiment classification, 2) They consider supervised learning approach by using labeled corpora for training which is not suitable for cross-domain work.

2.2 Sentiment-Topic Models

The work related to jointly determine sentiment and topic simultaneously from text is relatively sparse. Most closely related to our work is [7], [8], [9]. . Topic-sentiment model (TSM) [7] models the mixture of topics and sentiments simultaneously from web-blogs. TSM is based on the

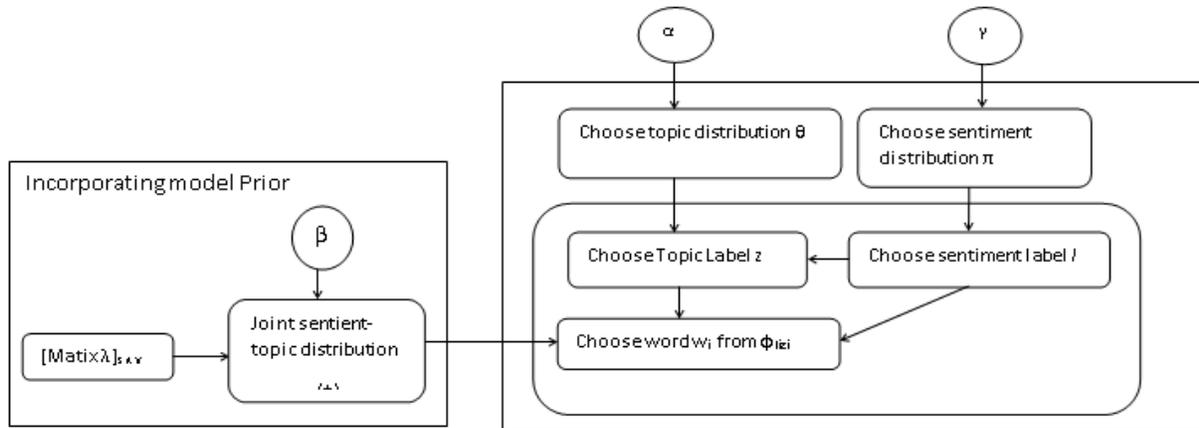


Figure 1. Block Diagram

probabilistic latent semantic indexing (pLSI). It finds the latent topics in a Weblog collection, sentiments and the subtopics in the results of query. If the word is common English then it samples a word from background component model. Else, a word is sampled from a topical model or sentiment model. Thus, the word generation for sentiment is independent of topic. While in JST, a word is drawn from the joint distribution of sentiment and topic label. To obtain the sentiment coverage, TSM performs postprocessing. JST gives the document sentiment by using probability distribution of sentiment label for a given document.

The Multi-Grain Latent Dirichlet Allocation (MG-LDA) [8] is more appropriate to build topics in which a customer provide a rating for each aspect that is customer will annotate every sentence and phrase in a review as being relevant to some aspect. Each word is generated from either a global topic or a local topic. The model uses a topic model in that it assigns words to a set of induced topics, each of which may represent one particular aspect. The limitation of MG-LDA is that it does not considers the associations between sentiments and topics.

The MG-LDA model is extended to Multi-Aspect Sentiment [MAS] model [9]. The model extracts the ratable aspects of an object and cluster them into coherent topics. Then model uses various techniques to classify and aggregate sentiment over each of these aspects. Thus limitation of MG-LDA is overcome by MAS. It differs from JST in that it is a supervised model because it requires that every aspect should be rated which may not be possible in real world applications. While JST is a weakly supervised model which only requires minimum prior information.

3. METHODOLOGY

3.1 Joint Sentiment-Topic Model

JST model is the extension of existing LDA framework which has three hierarchical layers, where topics are associated with documents, and words are associated with topics. JST [10] introduces fourth layer to the LDA model called sentiment layer in order to consider sentiment of the document. Hence, JST becomes four-layer model, where sentiment labels are associated with documents, under which topics are associated with sentiment labels and words are

associated with both sentiment labels and topics. The graphical model of JST is given in figure 1.

Consider a corpus with a collection of D documents denoted by $C = \{d_1, d_2, d_3, \dots, d_D\}$, each document in the corpus is a sequence of N_d words denoted by $d = (w_1, w_2, \dots, w_{n_d})$, and each word in the document is an item from a vocabulary index with V distinct terms denoted by $\{1, 2, \dots, V\}$. S be the number of distinct sentiment labels, and T be the total number of topics. The procedure for generating a word w_i in document d under JST can be given as: 1) Choose a sentiment label l from the per-document sentiment distribution π_d . 2) Choose a topic from the topic distribution $\theta_{d,l}$, where $\theta_{d,l}$ is conditioned on the sampled sentiment label l . Each document is associated with S topic distributions, each of which corresponds to a sentiment label l with the same number of topics. Thus, JST model can predict the sentiment associated with the extracted topics. 3) Draw a word from the per-corpus word distribution conditioned on both topic and sentiment label.

The graphical model of JST approach as shown in figure 1 can be defined as follows:

- 1) For every l (sentiment label) $l \in \{1, \dots, S\}$
 - For every topic $j \in \{1, \dots, T\}$, draw $\phi_{lj} \sim \text{Dir}(\lambda_l \times \beta_{lj}^T)$.
- 2) For every document d , choose a distribution $\pi_d \sim \text{Dir}(\gamma)$.
- 3) For every $l \in \{1, \dots, S\}$ under document d , choose a distribution $\theta_{d,l} \sim \text{Dir}(\alpha)$.
- 4) For every word w_i in document d
 - choose $l_i \sim \text{Mult}(\pi_d)$,
 - choose $z_i \sim \text{Mult}(\theta_{d,l_i})$,

- choose a word w_i from ϕ_{li} which is a multinomial distribution over words conditioned on both sentiment label l_i and topic z_i .

The hyperparameters α and β in JST is the number of times topic j associated with sentiment label l is sampled from a document and the number of times words sampled from topic j are associated with sentiment label l , respectively. The hyperparameter γ is number of times sentiment label l sampled from a document before any word from the corpus is observed. π is per-document sentiment distribution, θ is per-document sentiment label specific topic distribution, and ϕ is per corpus joint sentiment-topic word distribution .

3.2 Incorporating Model Priors

JST model is the extension of LDA model in which additional dependency link of ϕ on the matrix λ of size $S \times V$ is used to encode word prior sentiment information into the JST model. A transformation matrix λ modifies the Dirichlet priors β of size $S \times T \times V$, so that the word prior sentiment polarity can be captured. The process of incorporating prior knowledge into the JST model is as follows: first, λ is initialized with all the elements equal to 1. For every sentiment label $l \in \{1, \dots, S\}$ and every word $w \in \{1, \dots, V\}$ in the corpus vocabulary, if word w is also available in the sentiment lexicons used, the element λ_{lw} is updated as follows:

$$\lambda_{lw} = \begin{cases} 1, & \text{if } S(w) = 1 \\ 0, & \text{otherwise.} \end{cases}$$

where $S(w)$ is the function which returns the prior sentiment label of w found in a sentiment lexicon (neutral, positive, or negative). Suppose, a word 'Bad' have polarity negative which is from vocabulary with index i . The corresponding row vector of λ is given by $[1, 0, 0]$ which corresponds to negative, positive, neutral prior polarity. Now, for each topic $j \in \{1, \dots, T\}$, multiply λ_{ij} with β_{ij} . Here, the value of β_{inegij} is retained only and β_{iposji} and β_{ineuji} becomes 0.

3.3 Model Inference

To obtain the distributions of π , θ , and γ , first estimate the posterior distribution over z and l , i.e., the assignment of word tokens to topics and sentiment labels for a corpus. The sampling distribution for a word given remaining topics and sentiment labels is given by, $P(z_i=j, l_i=k | \alpha, \beta, \gamma)$. All words in the collection except for the word at location ' t ' in document D are given by z^{-t} and l^{-t} which are vectors of assignment of topics and sentiment labels.

The joint probability of the words, topics and sentiment label assignments can be given by

$$P(w, z, l) = P(w|z, l) P(z, l) = P(w|z, l) P(z|l) P(l) \quad (1)$$

To estimate the posterior distribution by sampling the variables z_t and l_t , the process of Gibbs sampling is used. Let, the superscript $-t$ denote a quantity that excludes word from t^{th} position. By marginalizing out random variables ϕ , θ and π , the conditional posterior for variables of interest z_t and l_t is given as

$$P(z(t) = j, l(t) = k | w, z(-t), l(-t), \alpha, \beta, \gamma) \propto$$

$$\frac{N(k, j, wt) - t + \beta}{N(k, j) - t + V\beta} \frac{N(d, k, j) - t + \alpha(k, j)}{N(d, k) - t + \sum_j \alpha(k, j)} \frac{N(d, k) - t + \gamma}{N(d) - t + S\gamma}$$

Samples obtained from the Gibbs sampling are used to approximate the per-corpus sentiment-topic word distribution which can be given as:

$$\phi(k, j, t) = \frac{N(k, j, wt) + \beta}{N(k, j) + V\beta}$$

The approximate per-document topic distribution specific to the sentiment label can be given as:

$$\theta(d, k, j) = \frac{N(d, k, j) + \alpha(k, j)}{N(d, k) + \sum_j \alpha(k, j)}$$

And the approximate per-document sentiment distribution can be given as

$$\pi(d, k) = \frac{N(d, k) + \gamma}{N(d) + S\gamma}$$

3.4 Algorithm

Algorithm : Procedure of Gibbs sampling for JST model.

Input: corpus, α, β, γ

Output : sentiment and topic label assignment for all word tokens in the corpus.

- 1: Initialize $S \times T \times V$ matrix Φ , $D \times S \times T$ matrix Θ , $D \times S$ matrix Π .
- 2: for $i = 1$ to maximum Gibbs sampling iterations do
- 3: for all documents $d = [1, D]$ do
- 4: for all terms $t = [1, Nd]$ do
- 5: Exclude term t associated with topic label z and sentiment label l from variables $N_d, N_{d,k}, N_{d,k,j}, N_{k,j}$ and $N_{k,j,l}$;
- 6: Sample a new sentiment-topic pair \tilde{l} and \tilde{z} using above equation 2;
- 7: Update variables $N_d, N_{d,k}, N_{d,k,j}, N_{k,j}$ and $N_{k,j,l}$ using the new sentiment label \tilde{l} and topic label \tilde{z} ;
- 8: end for
- 9: end for
- 10: for every 25 iterations do
- 11: Using Maximum Likelihood Estimation Update hyperparameter α ;
- 12: end for
- 13: for every 100 iterations do
- 14: Update matrices Θ, Φ , and Π with new Sampling results;
- 15: end for
- 16: end for

3.5 Hyperparameter Setting and Classifying Document Sentiment

In the JST model implementation, set the symmetric prior $\beta = 0.01$, the symmetric prior $\gamma = (0.05 \times L) / S$, where L is the average document length, S is the total number of sentiment labels. The asymmetric prior α is learned directly from data using maximum-likelihood estimation [11] and updated every 25 iterations during the Gibbs sampling procedure.

3.6 Classifying Document Sentiment

The document sentiment is classified as the probability of a sentiment label given a document $P(l|d)$. Experiments only consider the probability of positive and negative labels for a given document, while the neutral label probability is ignored. A document d is classified as a positive if the probability of a positive sentiment label $P(l_{pos}|d)$ is greater than its probability of negative sentiment label $P(l_{neg}|d)$, and vice versa.

4. RESULTS AND DISCUSSION

4.1 Datasets

Two easily available data sets, movie review (MR) data set (<http://www.cs.cornell.edu/people/pabo/movie-review-data>) and Multi-domain sentiment (MDS) data set <http://www.cs.jhu.edu/~mdredze/datasets/sentiment/index2.html> are used in the experiments. The MR data set contains 1,000 positive and 1,000 negative movie reviews with average of 30 sentences each document. MDS data set is crawled from Amazon.com which includes reviews of four different products. Both data sets are first preprocessed in which punctuation, non-alphabet characters, numbers and stop words are removed. Two subjectivity lexicons, appraisal lexicon (http://lingcog.iit.edu/arc/appraisal_lexicon_2007b.tar.gz) and MPQA lexicon (<http://www.cs.pitt.edu/mpqa/>) are combined and incorporated as model prior information. Stemming is performed on both data sets and both lexicons in the preprocessing. The two lexicons used in work are fully domain independent and do not bear any supervised information related to the MR and MDS data set.

4.2 Performance Analysis

4.2.1 Sentiment Classification Results versus Different Number of Topics

As JST models sentiment and topic mixtures simultaneously, it is therefore worth exploring how the sentiment classification and topic extraction tasks affect/benefit each other and in addition, the model behave with different topic number settings on different data sets when prior information is incorporated.

Modeling sentiment and topics simultaneously help to improve sentiment classification. For the cases where a single topic performs the best, it is observed that the drop in sentiment classification accuracy by additionally modeling mixtures of topics is only marginal, but it is able to extract sentiment-oriented topics in addition to document-level sentiment detection.

4.2.3 Topic Extraction

Manually examining the data reveals that the terms that seem not convey sentiments under the topic in fact appear in the context of expressing positive sentiments.

5. CONCLUSION

JST model detects sentiment and topic simultaneously from a text at document level in a weakly supervised fashion. Only sentiment prior knowledge is incorporated which is

independent of the domain. For general domain sentiment classification, by incorporating a small amount of domain independent prior knowledge, JST model achieves either better or comparable performance compared to existing semi-supervised approaches without using labeled documents. Thus, JST is flexible in the sentiment classification task. Weakly supervised nature of JST makes it highly portable to other domains. Moreover, the topics and topic sentiments detected by JST are indeed coherent and informative.

In future, incremental learning of the JST parameters can be done when facing with new data. Also, the modification of the JST model can be achieved by incorporating other supervised information into JST model learning, such as some known topic knowledge for certain product reviews or document labels derived automatically from the user supplied review ratings.

6. REFERENCES

- [1] C. Lin, Yulan He, R. Everson “Weakly Supervised Joint Sentiment-Topic Detection from Text” IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 24, NO. 6, JUNE 2012.
- [2] P.D. Turney, “Thumbs Up Or Thumbs Down?: Semantic Orientation Applied to Unsupervised Classification of Reviews,” Proc. Assoc. for Computational Linguistics (ACL '01), pp. 417-424, 2001.
- [3] B. Pang, L. Lee, and S. Vaithyanathan, “Thumbs Up?: Sentiment Classification Using Machine Learning Techniques,” Proc. ACL Conf. Empirical Methods in Natural Language Processing (EMNLP) pp. 79-86, 2002.
- [4] B. Pang and L. Lee, “A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts,” Proc. 42th Ann. Meeting on Assoc. for Computational Linguistics (ACL), pp. 271-278, 2004.
- [5] A. Aue and M. Gamon, “Customizing Sentiment Classifiers to New Domains: A Case Study,” Proc. Recent Advances in Natural Language Processing (RANLP), 2005.
- [6] D.M. Blei, A.Y. Ng, and M.I. Jordan, “Latent Dirichlet Allocation,” J. Machine Learning Research, vol. 3, pp. 993-1022, 2003.
- [7] Q. Mei, X. Ling, M. Wondra, H. Su, and C. Zhai, “Topic Sentiment Mixture: Modeling Facets and Opinions in Weblogs,” Proc. 16th Int'l Conf. World Wide Web (WWW), pp. 171-180, 2007.
- [8] I. Titov and R. McDonald, “Modeling Online Reviews with Multi-Grain Topic Models,” Proc. 17th Int'l Conf. World Wide Web, pp. 111-120, 2008.
- [9] I. Titov and R. McDonald, “A Joint Model of Text and Aspect Ratings for Sentiment Summarization,” Proc. Assoc. Computational Linguistics—Human Language Technology (ACL-HLT), pp. 308-316 2008.

[10] C. Lin and Y. He, “Joint Sentiment/Topic Model for Sentiment Analysis,” Proc. 18th ACM Conf. Information and Knowledge Management (CIKM), pp. 375-384, 2009.

[11] T. Minka, “Estimating a Dirichlet Distribution,” technical report, MIT, 2003.

[12] S. Li and C. Zong, “Multi-Domain Sentiment Classification,” Proc. Assoc. Computational Linguistics—Human Language Technology (ACL-HLT), pp. 257-260, 2008.

[13] T. Hofmann, “Probabilistic Latent Semantic Indexing,” Proc. 22nd Ann. Int’l ACM SIGIR Conf. Research and Development in Information Retrieval, pp. 50-57, 1999.