Joint Naïve Bayes and LDA for Unsupervised Sentiment Analysis

Yong Zhang[1,2], Dong-Hong Ji[1], Ying Su[3], and Hongmiao Wu[4]

[1] Computer School, Wuhan University, Wuhan, P.R. China
[2] Department of Computer Science, Huazhong Normal University, Wuhan, P.R. China
[3] Department of Computer Science, Wuchang Branch, Huazhong University of Science and Technology, Wuhan, P.R. China
[4] School of Foreign Languages and Literature, Wuhan University, P.R. China
ychang.cn@gmail.com, donghong_ji2000@yahoo.com.cn, {suying929,hongmiao23}@163.com

Abstract. In this paper we proposed a hierarchical generative model based on Naïve Bayes and LDA for unsupervised sentiment analysis at sentence level and document level of granularity simultaneously. In particular, our model called NB-LDA assumes that each sentence instead of word has a latent sentiment label, and then the sentiment label generates a series of features for the sentence independently in the Naïve Bayes manner. The idea of NB assumption at sentence level makes it possible that we can use advanced NLP technologies such as dependency parsing to improve the performance for unsupervised sentiment analysis. Experiment results show that the proposed NB-LDA can obtain significantly improved results for sentiment analysis comparing to other approaches.

Keywords: Sentiment Analysis, Latent Dirichlet Allocation, Naïve Bayes, Opinion Mining.

1 Introduction

In recent years sentiment analysis or opinion mining aiming to uncover the attitude of users from the content has drawn much attention in the NLP. But most methods usually rely heavily on an annotated corpus for training the sentiment classifier. So the sentiment corpora are considered as the most valuable resources for sentiment analysis. To circumvent laborious annotation efforts, developing an unsupervised method for sentiment analysis is one of the goals of this paper.

Previous methods have tackled the problem at different levels of granularity, from document level, sentence level, phrase-level, as well as the speaker level in debates. Intuitively sentiment analysis at different level of granularity can benefit from each other [11, 15]. Though some studies have been performed to analyze the sentiment of document level and sentence level simultaneously, their approaches are usually based on structure model and try to exploit the structure information in document, and usually are supervised or semi-supervised [11, 16, 17]. So another goal of this paper is

J. Pei et al. (Eds.): PAKDD 2013, Part I, LNAI 7818, pp. 402–413, 2013.

to develop a unsupervised model that jointly classifies sentiment at sentence level and document level of granularity.

For fine-grained sentiment analysis such as for sentence, many problems must be tackled. Negation is the main problem, especially involving long-distance dependencies such as the negation of the proposition or the negation of the subject (e.g., no one thinks that it's good). In addition, contextual polarity may also be influenced by modality, word sense, the syntactic role of a word in the sentence and diminishers such as little [18]. To solve such problems, it is necessary to employ advanced NLP technologies such as dependency parsing [25]. So the third goal of this paper is to incorporate advanced NLP technologies with statistical model for sentiment analysis.

In this study, we propose an unsupervised model to incorporate the sentiment analysis at the document level and sentence level. The model is a novel and unified hierarchical generative model combining Naïve Bayes and Latent Dirichlet Allocation (LDA), which we refer to as NB-LDA. But unlike LDA, our model does not assume documents are "bags of words". Rather it assumes that each sentence has a latent sentiment label, the latent sentiment label is drawn from the distribution of sentiment over document, and the words or features in the sentence are generated by the latent sentiment label in a Naïve Bayes manner. Thus the Naïve Bayes Assumption in our model makes it easy to integrate rich features and advanced NLP technique results into it to achieve better performance. We show that this model naturally fits the task of sentiment analysis, and experimental results show the effectiveness of the proposed model.

The rest of this paper is organized as follows: Section 2 introduces related work. The proposed NB-LDA model is described in detail in Section 3. Section 4 shows the experiments setup, and experiment results are described in section 5. Lastly we conclude this paper and discuss the future work.

2 Related Work

Previous work on sentiment analysis has covered a wide range of tasks, including polarity classification [15], opinion extraction [5], and opinion source assignment [20]. And these systems have tackled the problem at different levels of granularity. Usually for short of fine-grained corpus and lack of information, the fine-grained sentiment analysis is much harder than at document level. In order to recognize the contextual polarity in phrase-level, Wilson et al. [18] have compiled a lexicon of over 8000 subjective clues and used a set of features based on dependency tree of the sentence to disambiguates the polarity of the polar expressions. But most of their approaches relay on the availability of fine-grained annotations.

Recently topic modeling has been an area of active research [3, 7, 14]. Some models extended LDA have been proposed for sentiment analysis [10, 12, 19]. These approaches jointly learn topic and sentiment to improve predictions at document level. But these approaches have a shortcoming that they take the documents as "bag of words" and each word has a latent sentiment variable. Mukherjee and Liu [14] proposed two latent variable models (TME model and ME-TME model) to simultaneously model and extract topics (aspects) and various comment expressions for online review. But only n-grams (up to 4-grams) were modelled as a commenting expression.

There are, however, obvious advantages for sentiment analysis at both document level and sentence level. Pang and Lee [15] performed minimal cuts in a sentence graph to select subjective sentences to improve the performance of document sentiment classification. But these methods try to only improve document sentiment classification via selecting useful sentences.

Some structured models have previously been used for sentiment analysis [11, 16, 17]. McDonald et al. [11] presented a structured graphical model for fine-to-coarse sentiment analysis, and adopted a sequence classification with a constrained version of Viterbi for inference of the model. Täckström and McDonald [16] used latent variable structured prediction models for fine-grained sentiment analysis in the common situation where only coarse-grained supervision is available. Similarly Täckström and McDonald [17] derive two structured conditional models, which combine coarse-grained supervision with fine-grained supervision for sentence-level sentiment analysis. But these approaches are supervised or semi-supervised.

Another very relevant model is ASUM [8], as shown in Fig. 1(a), but there are some key differences between our model and ASUM. Firstly our model can use rich feature to improve sentiment analysis. For example the sequence of the negation and polarity words can be considered in our work. But ASUM can't do so. Secondly, our model is a unified generative model. Like LDA, Naive Bayes is a generative model, and they can be naturally integrated into a unified model. Notably NB can be supervised or unsupervised while other classification algorithms such as MaxEnt, SVM are discriminative only for supervised classification and difficult to integrate into LDA. Thirdly, our model focuses on sentiment classification. However ASUM is proposed mainly to discover pairs of senti-aspects.

To our knowledge, some similar models have been proposed by taking the best of Naïve Bayes and LDA models, such as Latent Dirichlet conditional Naïve Bayes (LD-CNB) [1] and Bayesian Naive Bayes[1]. But these models are different from ours. LD-CNB assumes a Dirichlet prior with parameter α from which the mixing weights θ is sampled. Further, for an observed feature f_j , a component z_j is first sampled from θ, and x_j is sampled from the corresponding component distribution. The LD-CNB process of generating is different from ours. Also there has been some work where arbitrary features were included into the LDA model, such as Dirichlet Multinomial Regression [13], MaxEnt-LDA [20] and etc. But the purposes and methods of these works are different from ours.

[1] http://lingpipe-blog.com/2009/10/02/bayesian-naive-bayes-aka-dirichlet-multinomial-classifiers/

3 NB-LDA Model

3.1 Motivation

Table 1. Distribution of sentence labels (columns) in documents by their labels (rows) in fine-grained sentiment dataset [16]

	Pos.	Neg.	Neut.
Pos.	0.53	0.08	0.39
Neg.	0.05	0.62	0.33
Neut.	0.14	0.35	0.51

For the distribution of sentence level sentiment in each document sentiment category, Täckström and McDonald [16] have shown that the sentence level sentiment is aligned with the document sentiment, as shown in Table 1. That is, in positive documents, most of sentences are positive, and in negative documents, most of sentences are negative, and in neutral documents, most of sentences are neutral Obviously this distribution can be exploited in sentiment analysis. As we know, topic models like LDA use co-occurrence information to group similar words into a single topic. Intuitively sentence level sentiment co-occurrence in documents may be mined to achieve better performance like topic model. So we proposed a new model, NB-LDA, which can employ rich features and exploit the distribution of sentence level sentiment in each document sentiment category to improve sentiment analysis. Here Naïve Bayes is to identify the sentence sentiment polarity based on a set of features in the local context, and LDA is to exploit the sentence level sentiment co-occurrence in documents globally.

In addition, to improve the performance of unsupervised sentiment analysis, our approach employs dependency parsing and a variety of features to identify the contextual polarity of the sentence inspired by [18].

3.2 NB-LDA Model

The NB-LDA model belongs to a family of generative models for text where a document contains a fixed number of sentences, each sentence expresses a kind of sentiment polarity represented by a latent variable z, and words in the sentence are viewed as features conditionally independent given sentiment label z. As a generative classification algorithm, Naive Bayes is used to identify sentiment polarity of the sentence via a set of features. Here we assume that a sentiment polarity (positive, negative or neutral) is expressed in a sentence rather than in a word, as done in previous studies [16, 17]. As to a document, we assume that it is "bag of sentences" and is generated by a mixture distribution θ of T sentiments, which is sampled from a prior Dirichlet distribution $Dir(\alpha)$.

This generative process can be expressed more formally by defining some of variables in the model. Assume we have T sentiment labels, for example positive, negative and neutral, which play similar roles as topics. Let D be the number of documents, F be the number of features, S_d be the number of sentence in document d, S_{di} be the i-th sentence in document d, and F_{di} be the number of features in sentence S_{di}. We can parameterize the multinomial distribution over sentiment labels for each document using a matrix Θ of size D * T, where each row θ_d stand for the probabilities of sentiment labels in document d. Similarly the matrix Φ of size T * F denotes the distribution over features associated with each sentiment label, where φ_t stand for the probabilities of generating features from sentiment label t. These matrix distributions are assumed to be generated from symmetric Dirichlet priors with hyper parameters α and β respectively.

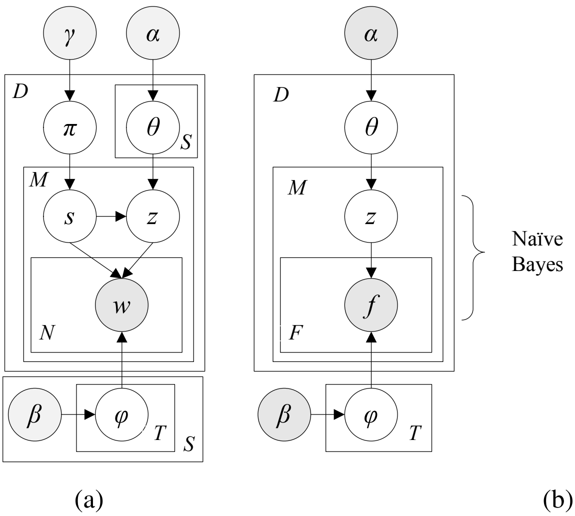

Fig. 1. (a) Graphical model of ASUM. (b) Graphical model of NB-LDA

The formal definition of NB-LDA model is the following:

- For each document d, sample $\theta_d \sim Dir(\alpha)$
- For each label t, sample $\varphi_t \sim Dir(\beta)$
- For each sentence i in S_d sentences of document d
  - Choose a sentiment label $z_{di} \sim Multi\ (\theta_d)$
  - For each feature j in F_{di} features of sentence s_{di}
  - Choose a feature value $f_{dij} \sim Multi(\varphi_{z_{di}})$.

In above generation process, the most difference from LDA is that NB-LDA chooses a latent variable z for each sentence in document d and then generates the feature values of the sentence independently with a Dirichlet prior. The graphical model corresponding to this process is shown in Figure 1(b).

Although previous studies have shown that topic and sentiment are dependent, the dependencies are usually limited in the range of one sentence. In NB-LDA, we use the NB and a set of features produced by NLP technologies such as dependency parsing to solve the local dependencies.

3.3 Inference with NB-LDA

A variety of algorithms have been used to estimate the parameters of topic models [3, 6]. In this paper, we will use Gibbs sampling [6], as it provides a simple method for obtaining parameter estimates under Dirichlet priors and allows combination of estimates from several local maxima of the posterior distribution.

Under this generative process, the joint probability of the z assignments and the features f can be factored into the following terms:

$$p(z, f) = \prod_{d=1}^{D} \prod_{i=1}^{S_d} p(z_{di}) \prod_{f=1}^{F_{di}} p(f \mid z_{di}) \tag{1}$$

By applying Gibbs sampling, we construct a Markov chain that converges to the posterior distribution on z like [6]. The transition between successive states of the Markov chain results from repeatedly drawing z from its distribution conditioned on all other variables, summing out θ and φ using standard Dirichlet integrals:

$$p(z_{d,i} = k \mid s_{d,i}, z_{d,-i}, s_{d,-i}) \propto \frac{C_{d,k}^{DT} + \alpha}{\sum_{t=1}^{T} C_{d,t}^{DT} + T\alpha} \cdot \prod_{j \in F_{d,i}} \frac{C_{j,k}^{FT} + \beta}{\sum_{f=1}^{F} C_{f,k}^{FT} + F\beta} \tag{2}$$

Here $C_{d,k}^{DT}$ is the number of times that sentences in the document d were assigned to sentiment label k, not including the current sentence. $C_{j,k}^{FT}$ is the total number of times that sentences containing feature j were assigned to sentiment label k, not including the current sentence.

For any sample from this Markov chain, being an assignment of every sentence to a sentiment label, we can estimate θ and φ using

$$\theta_{d,k} \propto \frac{C_{d,k}^{DT} + \alpha}{\sum_{t}^{T} C_{d,t}^{DT} + T\alpha} \tag{3}$$

$$\varphi_{k,j} \propto \frac{C_{j,k}^{FT} + \beta}{\sum_{f}^{F} C_{f,k}^{FT} + F\beta} \tag{4}$$

where $\varphi_{k,j}$ is the probability of using feature j in sentiment label k, and $\theta_{d,k}$ is the probability of sentiment label k in document d.

4 Experiments Setup

In this section, we introduce the data sets, prior-polarity subjectivity lexicon and feature space. In our experiments, the hyper parameters α and β are fixed at 0.3 and 0.01 respectively [6].

4.1 Corpus, Prior-Polarity Subjectivity Lexicon and Features

We used 2 corpora to evaluate our model. The first one is the Movie Review Data[2]. The second is Fine-grained Sentiment Dataset [16]. For pre-processing, we used Stanford's Suite of NLP Tools[3]. Here some tokens were removed according to the stop words list and POS, and "Fine-grained Sentiment Dataset" was adjusted slightly in order to adapt to Stanford's Suite of NLP Tools.

In the experiments, we used a public subjectivity lexicon as prior-polarity knowledge [18]. We compared the word tags to clues in the lexicon. If matched, the word token was marked with the corresponding "prior polarity". And the sentence was marked with the majority polarity voted by all matched words while taking negation into account. The prior information of sentence was only utilized during the initialization. The initialization starts by checking the "prior polarity" each sentence in corpus. If the sentence has a "prior polarity", the corresponding sentiment label is assigned to it. Otherwise, a sentiment label is randomly sampled.

Table 2. Features used in NB-LDA

Word Features
word lemma
word prior polarity: positive, negative, both, neutral
reliability class: strongsubj or weaksubj
Polarity Features
Negated: binary
negated subject: binary
modifies polarity: positive, negative, neutral, both, notmod
modified by polarity: positive, negative, neutral, both, notmod
conj polarity: positive, negative, neutral, both, notmod
Sentence Features
cardinal number in sentence: binary
pronoun in sentence: binary
modal in sentence (other than will): binary
Structure Features
Sentence position: first, mid, last

Table 2 lists the features in the experiments. Most of the features are used in [18]. Here we used word lemma as features instead of word tokens. Besides we added

[2] http://www.cs.cornell.edu/People/pabo/movie-review-data/
[3] http://nlp.stanford.edu/software/corenlp.shtml

sentence position as structure features. Note there are some differences between Stanford typed dependencies and that used in [18], and we handled them.

4.2 Baselines

In this work, we adopted four baselines to evaluate our model.

lemma-prior: In this baseline, we only use lemma as features, but didn't use any prior knowledge. In the initialization of Gibbs, all the latent variables of sentences are randomly chosen.

lemma+prior: In this baseline, we only use lemma as features, and use the prior knowledge. In the experiments, we only compared the tokens or lemmas morphologically with the words in subjectivity lexicon, and ignored the POS.

ASUM+prior: ASUM is so similar to our model. So we compared ASUM model as baseline, where we use the same prior-polarity subjectivity lexicon instead of seed words used in [8] for prior knowledge.

NB-ASUM: In the baseline, we assumed that ASUM can generate feature values just like our model. Besides the setting of ASUM+prior, we use all features described in Table 2.

5 Experiment Results

5.1 Results for Movie Review Data

In Movie Review Data, the reviews are only labeled as positive or negative at document level. So in the experiments, T is set to 2 since we only consider 2 sentiment labels: positive and negative. The document sentiment is classified based on $\theta_{d,k}$, the probability of sentiment label given document. A document d is classified as a positive sentiment document if its probability of positive sentiment label given document $\theta_{d,pos}$, is greater than its probability of negative sentiment label given document $\theta_{d,neg}$, and vice versa. As shown as Table 3, classification accuracies were averaged from 10 runs with 1000 Gibbs sampling iterations.

As can be observed from Table 3, the performance of "lemma-prior" is mediocre when no prior information was incorporated. A significant improvement, with 10.7%, is observed after incorporating prior information. It is also noted that NB-LDA with all features achieved 3.9% improvement over "lemma+prior", implying that the richer features can benefit sentiment analysis. So discovering more effective features is one of the future works.

When compared to the recently proposed unsupervised approach based on a spectral clustering algorithm [4], NB-LDA achieved better performance with about 1% overall improvement. Nevertheless, the approach proposed in [4] requires users to specify which dimensions (defined by the eigenvectors in spectral clustering) are most closely related to sentiment by inspecting a set of features derived from the reviews for each dimension, and clustering is performed again on the data to derive the final results. In our model studied here, no human judgment is required. Another recently proposed non-negative matrix tri-factorization approach in [9] also employed lexical prior knowledge for semi-supervised sentiment classification. However, when incorporating 10% of labeled documents for training, the non-negative matrix tri-factorization approach performed much worse than NB-LDA, with only around 60% accuracy achieved. Even with 35% labeled documents, it still performs worse than NB-LDA. It is worth noting that no labeled documents were used in the NB-LDA results reported here.

Table 3. Average results from 10 runs in terms of accuracy for Movie Review. Above double line: results from our model. Below double line: results from other literatures.

	Pos.	Neg.	Total
lemma -prior	55.0	59.5	57.25
lemma +prior	72.7	63.2	67.95
ASUM+prior	72.1	63.3	67.7
NB-ASUM	74.5	67.4	70.95
NB-LDA	75.3	68.4	71.85
Lin et al. [10]	74.1	66.7	70.4
Dasgupta and Ng [4]			70.9
Li et al.[9] with 10% doc. Label			60
Li et al.[9] with 35% doc. Label			about 69

Another notable result is that the prior information has different effects on positive documents and negative ones. Without prior information, the positive accuracy is less than the negative accuracy. While with prior information, the positive accuracies are better than the negative ones significantly. Manually analyzing the results reveals that there are more words matching the positive clues than negative clues in the corpus, which causes the imbalance of distribution of positive prior and negative prior during the initialization. Actually we found the similar problem in the experiments for fine-grained dataset reported in the next subsection, and in the results of [10, 16, 17].

5.2 Results for Fine-Grained Dataset

There are three sentiment categories at document level in this dataset, but there are five sentiment categories at sentence level: POS, NEG, NEU, MIX, and NR. Like [16], we considered the MIX and NR categories as belonging to the NEU category. So in the experiments, T is set to 3. The sentence sentiment polarity is classified by the latent sentiment label z after sampling.

Table 4 shows the results for our model in terms of sentence and document accuracy as well as F1-scores for each sentence sentiment category. In Table 4, the results above double line come from our experiments, and that below double line are results presented in [16] and [17] for the same dataset.

In terms of sentence accuracy, from these results it is clear that the NB-LDA model significantly outperform VoteFlip with quite a wide margin. Comparing to SaD and

DaS, the results from NB-LDA are also very competitive. However our results are still lower about 7% ~ 12% than HCRF and Interpolated in the last three rows in terms of both sentence and document accuracy. Actually the results below double line are evaluated via 294 reviews, but with 143,580 labeled reviews as supervision. Notably NB-LDA does not use any labeled data.

In our results, "lemma-prior" is mediocre. But with the prior information, "lemma+prior" performs better about 6% for sentence accuracy and about 8% for document accuracy than "lemma-prior". The NB-LDA with all features and prior information achieved much better performance than "lemma-prior" and "lemma+prior". The results suggest that both proper features and prior information are important for improving sentiment analysis. In fact, NB-LDA model can easily integrate rich features and advanced NLP technique into it to achieve better performance.

Table 4. Average results from ten runs for fine-grained dataset. Above double line: results from our model. Below double line: VoteFlip, SaD, DaS and HCRF results without observed document label in [16], and Interpolated best result in [17].

	Sentence				Document Acc
	Total Acc	Pos. F1	Neg. F1	Neut. F1	
lemma-prior	36.0	25.7	38. 6	38.8	40.5
lemma+prior	41.9	46.4	49.5	31.6	48.6
ASUM+prior	41.4	45.9	49.0	31.2	47.3
NB-ASUM	45.7	52.2	48.5	36.9	52.7
NB-LDA	46.8	53.4	49.6	38.0	54.4
VoteFlip	41.5	45.7	48.9	28.0	-
SaD	47.6	52.9	48.4	42.8	-
DaS	47.5	52.1	54.3	36.0	66.6
HCRF(soft)	53.9	57.3	58.5	47.8	65.6
HCRF(hard)	54.4	57.8	58.8	48.5	64.6
Interpolated	59.1	-	-	-	-

5.3 NB-LDA vs ASUM

Similar to our NB-LDA, ASUM is also a generative model for sentiment analysis, which is proposed mainly to discover pairs of senti-aspects. In this work, our model only focuses on sentiment classification. But ASUM incorporates aspect and sentiment together to model sentiments toward different aspects [8]. The differences between them have been discussed in motivation. Here we compared the performances between the two models, and analyzed the differences, although NB-ASUM should obtain the same results with NB-LDA in theory.

In fact as shown in Table 3, the “lemma+prior” performs better than ASUM+prior slightly. It seems that only using lemma as feature, ASUM can obtain comparable result with our model. While using all features, the total accuracy of NB-LDA is higher about 1% than that of NB-ASUM. In fine-grained dataset as shown in Table 4, we can see the similar results, where the "lemma+prior" can achieve better performances than ASUM+prior, and NB-LDA is better than NB-ASUM.

From the experimental results, we can see that NB-LDA is consistently better than ASUM. As we know, the generative process of ASUM is more complex than NB-LDA. In NB-ASUM we need compute the probabilities for each feature value conditioned on topic variable and sentiment variable. It means that in NB-ASUM, the distributions would be sparser than in NB-LDA. The sparsity might block the propagation of sentiment information during iterations, and could not conduct good results.

6 Conclusions and Future Work

In this paper we proposed a hierarchical generative model based on Naïve Bayes and Latent Dirichlet Allocation for unsupervised sentiment analysis at sentence level and document level simultaneously, only using a public subjectivity lexicon. The idea of NB assumption at sentence level makes it possible that we can use advanced NLP technologies such as dependency parsing to improve the performance of sentiment analysis. The experiments show that our model obtained better performance than VoteFlip, a rule-based approach and ASUM. However for now our model hardly reach the competitive performance to the supervised or semi-supervised approaches.

Since unsupervised approaches hardly obtain comparable performance to supervised ones. A simple extended model could be designed based on supervised LDA [2] and Naïve Bayes. Another extension of our model is to capture sentimental structures within the documents simultaneously, inspired by HTMM [7]. The third future work is to empirically investigate the effects of more features on more datasets. Finally we may split sentences not only by punctuations but also by conjunctions, since one sentence may contain multiple sentiment clauses.

Acknowledgments. The work was funded by the National Natural Science Foundation of China (No. 61133012, 61173062, 61070082) and the Major Projects of Chinese National Social Science Foundation (No.11&ZD189). Dong-Hong Ji is the corresponding author.

References

1. Banerjee, A., Shan, H.: Latent Dirichlet conditional Naive-Bayes models. In: The IEEE International Conference on Data Mining (ICDM), pp. 421–426 (2007)
2. Blei, D., McAuliffe, J.: Supervised topic models. In: Proceeding of Neural Information Processing Systems, NIPS (2007)
3. Blei, D., Ng, A., Jordan, M.: Latent Dirichlet allocation. Journal of Machine Learning Research 3(5), 993–1022 (2003)
4. Dasgupta, S., Ng, V.: Topic-wise, Sentiment-wise, or Otherwise? Identifying the Hidden Dimension for Unsupervised Text Classification. In: Proceedings of EMNLP (2009)

5. Ding, X., Liu, B., Zhang, L.: Entity discovery and assignment for opinion mining applications. In: Proceedings of KDD 2009, pp. 1125–1134 (2009)
6. Griffiths, T., Steyvers, M.: Finding scientific topics. Proceedings of the National Academy of Sciences 101(90001), 5228–5235 (2004)
7. Gruber, A., Rosen-Zvi, M., Weiss, Y.: Hidden Topic Markov Models. In: Artificial Intelligence and Statistics (AISTATS), San Juan, Puerto Rico (2007)
8. Jo, Y., Oh, A.H.: Aspect and sentiment unification model for online review analysis. In: Proceedings of WSDM 2011, pp. 815–824 (2011)
9. Li, T., Zhang, Y., Sindhwani, V.: A nonnegative matrix tri-factorization approach to sentiment classification with lexical prior knowledge. In: Proceedings of ACL-IJCNLP 2009, pp. 244–252 (2009)
10. Lin, C., He, Y.: Joint sentiment/topic model for sentiment analysis. In: Proceeding of the Conference on Information and Knowledge Management, CIKM (2009)
11. McDonald, R., Hannan, K., Neylon, T., Wells, M., Reynar, J.: Structured models for fine-to-coarse sentiment analysis. In: Proceedings of ACL (2007)
12. Mei, Q., Ling, X., Wondra, M., Su, H., Zhai, C.X.: Topic sentiment mixture: modeling facets and opinions in weblogs. In: Proceedings of WWW (2007)
13. Mimno, D.M., McCallum, A.: Topic Models Conditioned on Arbitrary Features with Dirichlet-multinomial Regression. In: Proceedings of the 24th Conference in Uncertainty in Artificial Intelligence, pp. 411–418 (2008)
14. Mukherjee, A.: Liu, Bing: Modeling Review Comments. In: Proceedings of ACL 2012, Jeju, Republic of Korea, July 8-14 (2012)
15. Pang, B., Lee, L.: A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In: Proceedings of ACL (2004)
16. Täckström, O., McDonald, R.: Discovering fine-grained sentiment with latent variable structured prediction models. In: Clough, P., Foley, C., Gurrin, C., Jones, G.J.F., Kraaij, W., Lee, H., Mudoch, V. (eds.) ECIR 2011. LNCS, vol. 6611, pp. 368–374. Springer, Heidelberg (2011)
17. Täckström, O., McDonald, R.: Semi-supervised Latent Variable Models for Sentence-level Sentiment Analysis. In: Proceedings of ACL (2011)
18. Wilson, T., Wiebe, J., Hoffmann, P.: Recognizing contextual polarity in phrase-level sentiment analysis. In: Proceedings of EMNLP (2005)
19. Zhang, Y., Ji, D.-H., Su, Y., Sun, C.: Sentiment Analysis for Online Reviews Using an Author-Review-Object Model. In: Salem, M.V.M., Shaalan, K., Oroumchian, F., Shakery, A., Khelalfa, H. (eds.) AIRS 2011. LNCS, vol. 7097, pp. 362–371. Springer, Heidelberg (2011)
20. Zhao, W.X., Jiang, J., Yan, H., Li, X.: Jointly modeling aspects and opinions with a MaxEnt-LDA hybrid. In: Proceedings of EMNLP 2010, Stroudsburg, PA, USA, pp. 56–65 (2010)