

Sentence Subjectivity Detection with Weakly-Supervised Learning

Chenghua Lin^{1,2}, Yulan He² and Richard Everson¹

¹Department of Computer Science, University of Exeter, Exeter, EX4 4QF, UK

²Knowledge Media Institute, The Open University, Milton Keynes, MK7 6AA, UK

{C.Lin, Y.He}@open.ac.uk, R.M.Everson@exeter.ac.uk

Abstract

This paper presents a hierarchical Bayesian model based on latent Dirichlet allocation (LDA), called subjLDA, for sentence-level subjectivity detection, which automatically identifies whether a given sentence expresses opinion or states facts. In contrast to most of the existing methods relying on either labelled corpora for classifier training or linguistic pattern extraction for subjectivity classification, we view the problem as weakly-supervised generative model learning, where the only input to the model is a small set of domain independent subjectivity lexical clues. A mechanism is introduced to incorporate the prior information about the subjectivity lexical clues into model learning by modifying the Dirichlet priors of topic-word distributions. The subjLDA model has been evaluated on the Multi-Perspective Question Answering (MPQA) dataset and promising results have been observed in the preliminary experiments. We have also explored adding neutral words as prior information for model learning. It was found that while incorporating subjectivity clues bearing positive or negative polarity can achieve a significant performance gain, the prior lexical information from neutral words is less effective.

1 Introduction

Subjectivity detection seeks to identify whether the given text expresses opinions (subjective) or reports facts (objective). Such a task of distinguishing subjective information from objective is useful for many natural language processing applications. For instance, sentiment classification

often assumes that the input documents are opinionated, and ideally only contain subjective statements. Document summarization systems need to summarize different perspectives and opinions. For question answering systems, extracting and presenting information of the appropriate type, i.e., opinions or facts, is imperative according to the specific question being asked (Yu and Hatzivassiloglou, 2003; Wiebe and Riloff, 2005; Pang and Lee, 2008).

Work on sentence-level subjectivity detection is relatively sparse compared to document-level sentiment classification. Early work used a bootstrapping algorithm to learn subjective (Riloff and Wiebe, 2003) or both subjective and objective (Wiebe and Riloff, 2005) expressions for sentence-level subjectivity detection. In contrast to bootstrapping, there has been some recent attempts exploring various n -gram features and different level of lexical instantiation for detecting subjective utterance from conversation data (Wilson and Raaijmakers, 2008; Raaijmakers et al., 2008; Murray and Carenini, 2009).

However, the aforementioned line of work tackled subjectivity detection either as supervised or semi-supervised learning, requiring labelled data and extensive knowledge which are expensive to acquire. On the other hand, both subjectivity and sentiment are context sensitive and in general quite domain dependent (Pang and Lee, 2008), so that classifiers trained on one domain often fail to produce satisfactory performance when shifted to new domains (Gamon et al., 2005; Blitzer et al., 2007). Moreover, user generated content from web are often massive and evolve rapidly over time, which imposes more challenges to the subjectivity detection task. These observations have thus motivated us to develop a subjectivity detection algorithm that is relatively simple compared to existing methods (e.g., based on bootstrapping or n -gram features), and yet can easily be trans-

ferred between domains through unsupervised or weakly-supervised learning without using any labelled data.

In this paper, we focus on the problem of weakly-supervised sentence-level subjectivity detection. Instead of learning subjective extraction patterns or exploring various n -gram features, we view the problem as generative model learning with the proposed subjectivity detection LDA (subjLDA) model. In this model, the generative process involves: (1) three subjectivity labels for sentences (i.e., sentence expresses subjective opinions as being positive/negative, or states facts as being objective); (2) a sentiment label for each word in the sentence (either positive, negative, or neutral), and (3) the words in the sentences.

We test the subjLDA model on the publicly available Multi-Perspective Question Answering (MPQA) dataset. Two lists of domain independent subjectivity lexicons, namely the subjClue and SentiWordNet lexicons (Esuli and Sebastiani, 2006), were incorporated as prior knowledge for subjLDA model learning. Preliminary results show that the weakly-supervised subjLDA model is able to significantly outperform baseline. Furthermore, it was found that while incorporating subjectivity clues bearing positive or negative polarity can achieve a significant performance gain, the prior lexical information from neutral words is less effective for improving the classification accuracy.

The rest of the paper is organized as follows. Section 2 reviews the previous work on subjectivity classification. Section 3 presents the subjLDA model. Experimental setup and results on the MPQA dataset are discussed in Sections 4 and 5, respectively. Finally, Section 6 concludes the paper and outlines the future work.

2 Related Work

2.1 Subjectivity Detection

While sentiment classification and subjectivity detection are closely related to each other, it has been reported that separating subjective and objective instances from text is more difficult than sentiment classification, and the improvement of subjectivity detection can benefit the latter as well (Mihalcea et al., 2007).

Early work by Riloff and Wiebe (2003) focused on a bootstrapping method for sentence-level subjectivity detection. They started with

high-precision subjectivity classifiers which automatically identified subjective and objective sentences in un-annotated texts. The subjective expression patterns were learned from syntactic structure output from the previously labelled high confidence texts. The learned patterns were used to automatically identify additional subjective sentences, which enlarged the training set, and the entire process was then iterated. Wiebe and Riloff (2005) used very similar method for subjectivity detection as Riloff and Wiebe (2003). But they moved one step forward that they also learned objective expressions apart from subjective expressions. As the subjective/objective expression patterns are based on syntactic structures, they are more flexible than single words or n -grams.

Wilson and Raaijmakers (2008) compared the performance of classifiers trained using word n -grams, character n -grams, and phoneme n -grams for recognizing subjective utterances in multiparty conversation. Raaijmakers et al. (2008) extended the work in (Wilson and Raaijmakers, 2008) by further analyzing the performance of detecting subjectivity in meeting speech by combining a variety of multimodal features including additional prosodic features. More recently, Murray and Carenini (2009) proposed to learn subjective expression patterns from both labeled and unlabeled data using n -gram word sequences. Their approach for learning subjective expression patterns is similar to (Raaijmakers et al., 2008) which relies on n -grams, but goes beyond fixed sequences of words by varying levels of lexical instantiation as in (Riloff and Wiebe, 2003).

2.2 Weakly-supervised Sentiment Classification

In this section, we first review some work in sentiment analysis using generative models as it partly inspires our work of viewing subjectivity detection as generative model learning. We then discuss other weakly-supervised sentiment classification approaches which also use prior word knowledge.

Intuitively, sentiment or subjectivity are context dependent. Therefore, modelling topic coupled with sentiment should serve a critical function in sentiment analysis. There has seen several lines of work pursuing this direction. Eguchi and Lavrenko (2006) considered the topic dependence of sentiment and combined sentiment mod-

els with topic models for sentiment retrieval. Mei et al. (2007) proposed the topic-sentiment mixture (TSM) model for capturing mixture of topics and sentiment simultaneously on Weblogs. The multi-aspect sentiment (MAS) model by Titov and McDonald (2008) focused on aggregating sentiment text for sentiment summary of rating aspects.

The more recently proposed joint sentiment-topic (JST) model (Lin and He, 2009; Lin et al., 2010) holds the closest paradigm to the proposed subjLDA model. They targeted document-level sentiment detection with weakly-supervised generative model learning, where the only knowledge being incorporated was from generic sentiment lexicons. In the JST model, topics are assumed to be generated dependent on sentiment distributions and then words are generated conditioned on sentiment-topic pairs. However, there are several intrinsic differences between JST and subjLDA: (1) JST mainly focused on document-level sentiment classification, while, in contrast, subjLDA has a different scope of targeting sentence-level subjectivity detection; (2) in JST the prior knowledge was encoded during the Gibbs sampling by assigning a word with its prior sentiment label if that word appears in the sentiment lexicon. This essentially places hard sentiment label to words and can not resort the situation when words have ambiguous sentiment polarity. Our proposed approach incorporates sentiment prior knowledge in a more principled way, in that we use sentiment lexicons to modify the topic-word Dirichlet priors and essentially create an informed prior distribution for the sentiment labels.

Another common solution to weakly-supervised sentiment classification is to make use of prior word polarity knowledge, where one uses a small number of seed words with known polarity to infer the polarity of a large set of unidentified terms. Turney and Littman (2002) classified the sentiment orientation of other terms in the corpus through mutual information, based on a small set of positive/negative paradigm words. Starting with a single seed word meaning “good” and a negation check, Zagibalov and Carroll (2008) derived a classifier through iteratively retraining, and treated sentiment and subjectivity as a continuum rather than distinct classes.

3 The SubjLDA Model

As shown in Figure 1(b), subjLDA is essentially a four-layer Bayesian model. In order to generate a word $w_{d,m,t}$ (i.e., the t^{th} word token of sentence m within document d), one first chooses a subjectivity label¹ $s_{d,m} \in [1, K]$ for each sentence in document d from the per-document subjectivity distribution π_d . Following that, one chooses a sentiment label $l_{d,m,t} \in [1, S]$ for each word in the sentences from the per-sentence sentiment distribution $\theta_{s_{d,m}}$. Finally, one draws a word from the per-corpus word distribution $\varphi_{l_{d,m,t}}$ conditioned on the corresponding sentiment label. The classification of sentence subjectivity in subjLDA is determined directly from the sentence subjectivity label $s_{d,m}$. The formal definition of the subjLDA generative process is as follows:

- For each sentiment label $l \in [1, S]$
 - Draw $\varphi_l \sim \text{Dir}(\lambda_l \times \beta_l^T)$.
- For each document $d \in [1, D]$, choose distributions $\pi_d \sim \text{Dir}(\gamma)$.
- For each sentence $m \in [1, M_d]$ in document d ,
 - Sample a subjectivity label $s_{d,m} \sim \text{Mult}(\pi_d)$,
 - Choose a distribution $\theta_{s_{d,m}} \sim \text{Dir}(\alpha_{s_{d,m}})$,
 - For each of the $N_{d,m}$ word position,
 - * Choose a sentiment label $l_{d,m,t} \sim \text{Mult}(\theta_{s_{d,m}})$,
 - * Choose a word $w_{d,m,t} \sim \text{Mult}(\varphi_{l_{d,m,t}})$.

In practice, it is quite intuitive that one classifies a sentence as subjective if it contains one or more strongly subjective clues (Riloff and Wiebe, 2003). However, the criterion for classifying objective sentences could be rather different, because a sentence is likely to be objective if there are no strongly subjective clues. In order to encode this knowledge into the subjLDA model learning, during the model initialization step, we initialized sentence subjectivity label s based on the aforementioned criterion with prior knowledge input from the sentiment lexicon. If a sentence does not match any sentiment words, its subjectivity label will be randomly sampled.

¹We have conducted another set of experiments modelling only subjective and objective labels. It was found that subjLDA performed slightly better with 3 subjectivity labels than with binary labels. We do not report the binary label results here due to page limit.

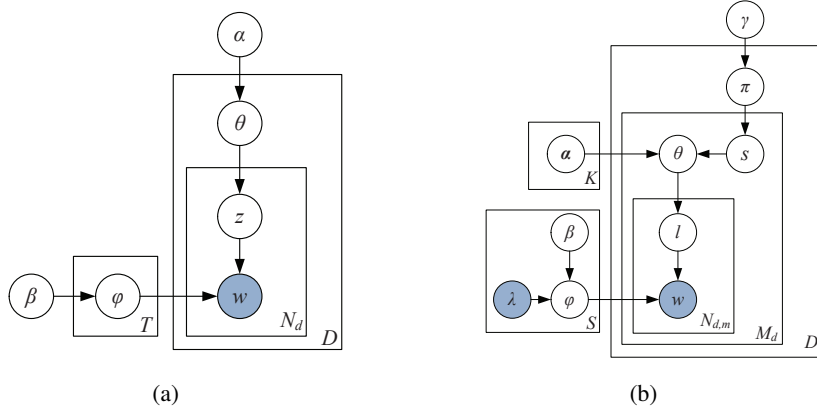


Figure 1: (a) LDA model; (b) subjLDA model.

3.1 Incorporating Model Prior

Compared to the LDA model, besides adding a sentence-level subjectivity label generation layer, we also add an additional dependency link of φ on the matrix λ of size $S \times V$ which we use to encode word prior sentiment information. The matrix λ can be considered as a transformation matrix which modifies the Dirichlet priors β so that the word prior sentiment polarity can be captured.

Intuitively, λ is firstly initialized as a matrix with all the elements taking a value of 1. Given a sentiment lexicon, for each term $w \in \{1, \dots, V\}$ in the corpus vocabulary, if w is found in the sentiment lexicon, then for each $l \in \{1, \dots, S\}$, the element λ_{lw} is updated as follows

$$\lambda_{lw} = \begin{cases} 0.9 & \text{if } S(w) = l \\ 0.05 & \text{otherwise} \end{cases}, \quad (1)$$

where the function $S(w)$ returns the prior sentiment label of w in a sentiment lexicon, i.e., positive, negative or neutral. For example, the word “*excellent*” with index w_t has a positive sentiment polarity. The corresponding row vector λ_{wt} is $[0.05, 0.9, 0.05]$ with its elements representing neutral, positive, and negative prior polarity. Multiplying β with λ , we can enforce that the word “*excellent*” has much higher probability of being drawn from the positive topic word distributions generated from a Dirichlet distribution with parameter $\beta_{l_{pos}w_t}$.

The previously proposed DiscLDA (Lacoste-Julien et al., 2008) and Labeled LDA (Ramage et al., 2009) also utilize a transformation matrix to modify Dirichlet priors by assuming the availability of document class labels. In contrast, we use word prior sentiment as supervised information to modify the topic-word Dirichlet priors.

3.2 Model Inference

The total probability of the model is

$$P(\mathbf{w}, \mathbf{l}, \mathbf{s}, \boldsymbol{\theta}, \boldsymbol{\varphi}, \boldsymbol{\pi}; \alpha, \beta, \gamma) = \prod_{j=1}^S P(\varphi_j; \boldsymbol{\lambda} \times \boldsymbol{\beta}) \cdot \prod_{d=1}^D P(\pi_d; \gamma) \prod_{m=1}^{M_d} P(s_{d,m} | \pi_d) P(\boldsymbol{\theta}_{d,m}; \alpha_{s_{d,m}}) \cdot \prod_{t=1}^{N_{d,m}} P(l_{d,m,t} | \boldsymbol{\theta}_{d,m}) P(w_{d,m,t} | \varphi_{l_{d,m,t}}), \quad (2)$$

where the bold-font variables denote vectors.

We use Gibbs sampling to estimate the posterior of subjLDA by sequentially sampling each variable of interest, $l_{d,m,t}$ and $s_{d,m}$ here, from the distribution over that variable given the current values of all other variables and the data. Letting the index $x = (d, m)$ and the subscript $-x$ denote a quantity that excludes counts in sentence m of document d , the conditional posterior for s_x is

$$P(s_x = k | \mathbf{s}_{-x}, \mathbf{l}, \mathbf{w}, \alpha, \beta, \gamma) \propto \frac{(N_{d,k} + \gamma_k) - 1}{(N_d + \sum_{k=1}^K \gamma_k) - 1} \cdot \frac{\prod_{j=1}^S \prod_{b=0}^{N_{d,m,j}-1} (b + \alpha_{s_x,j})}{\prod_{b=0}^{N_{d,m}-1} (b + \sum_{j=1}^3 \alpha_{s_x,j})}, \quad (3)$$

where $N_{d,k}$ denotes the frequency of sentences assigned to subjectivity label k in document d ; N_d is the total number of sentences in document d ; $N_{d,m,j}$ is the total number of words in sentence m of document d associated with sentiment label j ; $N_{d,m}$ is the total number of words in sentence m of document d .

In terms of the sentiment label, letting the index $y = (d, m, t)$ denote t^{th} word in sentence m of document d and the subscript $-y$ denote a quantity that excludes data from t^{th} word position, the

conditional posterior for l_t is

$$P(l_y = j | \mathbf{s}, \mathbf{l}_{-y}, \mathbf{w}, \alpha, \beta, \gamma) \propto \frac{N_{d,m,j} + \alpha_{s_{d,m,j}} - 1}{N_{d,m} + \sum_{j=1}^S \alpha_{s_{d,m,j}} - 1} \cdot \frac{Y_{j,w_t} + \lambda_{j,w_t} \beta_{j,w_t} - 1}{Y_j + \sum_{r=1}^V \lambda_{j,r} \beta_{j,r} - 1}, \quad (4)$$

where Y_{j,w_t} denotes the frequency of word w_t associated with sentiment label j in the document collection; Y_j is the total number of words associated with sentiment label j in the document collection.

Equations 3-4 are the conditional probabilities derived by marginalizing out the random variables π , θ , and φ . Samples obtained from the Markov chain are used to approximate the per-document subjectivity distribution

$$\pi_{d,k} = \frac{N_{d,k} + \gamma_k}{N_d + \sum_{k=1}^K \gamma_k}. \quad (5)$$

The approximated per-sentence sentiment distribution is

$$\theta_{d,m,j} = \frac{N_{d,m,j} + \alpha_{s_{d,m,j}}}{N_{d,m} + \sum_{j=1}^S \alpha_{s_{d,m,j}}}. \quad (6)$$

Finally, the per-corpus sentiment-word distribution is

$$\varphi_{j,r} = \frac{Y_{j,r} + \lambda_{j,r} \beta_{j,r}}{Y_j + \sum_{r=1}^V \lambda_{j,r} \beta_{j,r}}. \quad (7)$$

4 Experimental Setup

4.1 Dataset

We tested our model on the MPQA dataset² version 1.2, which is derived from a variety of foreign news documents. The whole corpus consists of 535 documents with a total number of 6,111 subjective and 5,001 objective sentences. We performed a two-stage preprocessing on the dataset by first removing stop words and non-word characters, followed by standard stemming for reducing vocabulary size and minimizing sparse data problems. After preprocessing, the MPQA dataset contains 131,220 words with 10,511 distinct terms (cf. the original dataset with 264,808 words and a vocabulary size of 31,201 without any preprocessing).

²<http://www.cs.pitt.edu/mpqa/databaserelease/>

4.2 Lexical Prior Knowledge

We explored incorporating two subjectivity lexicons as prior knowledge for subjLDA model learning, namely, the subjClue³ and SentiWordNet⁴ lexicons. We point out that the subjClue lexicon is not related to the MPQA dataset as it was collected from a number of sources, where some were culled from manually developed resources and others were identified automatically using both annotated and unannotated data (Wiebe and Riloff, 2005). We only extract the lexical clues that are considered strongly subjective, with the weakly subjective clues being discarded. The rationale behind the filtering is that while a strongly subjective clue is seldom used without a subjective meaning, weakly subjective clues are ambiguous, often having both subjective and objective uses. After stemming, removing the duplicated lexical terms and retaining those that have appeared in the corpus, we finally obtained a lexicon subset of 477 positive and 917 negative words.

SentiWordNet provides a wide coverage of lexical terms by tagging all the synsets of WordNet with three sentiment labels, i.e., positive, negative and neutral. In our experiment, we only use the neutral words from SentiWordNet for investigating how neutral words would affect the subjLDA model performance. After the same preprocessing as performed on the subjClue lexicon, a total of 193,871 neutral words were extracted. Further mapping the extracted neutral words with the corpus results in 6,457 neutral words.

5 Experimental Results

In this section, we first present the experimental results of sentence-level subjectivity classification on the MPQA dataset, and subsequently evaluate the impact on the classification performance by varying the proportion of prior information being incorporated. All the results reported here are averaged over 5 runs with 800 Gibbs sampling iterations.

5.1 Overall Results

The baseline is calculated by counting the overlap of the prior lexicon with the dataset. We classify a sentence as subjective if it contains one or more positive/negative sentiment words; if there is no matching, the sentence will be classified as

³<http://www.cs.pitt.edu/mpqa/>

⁴<http://sentiwordnet.isti.cnr.it/>

Table 1: Subjectivity classification results. (Boldface indicates the best results.)

Model	Objective (%)			Subjective (%)			Overall (%)
	Recall	Precision	F-measure	Recall	Precision	F-measure	Accuracy
Baseline	46.5	74.1	57.1	76.7	63.7	69.6	63.1
subjLDA	59.7	71.6	65.1	80.9	71.0	75.6	71.2
LDA (Sent.)	60.5	65.7	63.0	74.2	69.7	72.0	68.1
LDA (Doc.)	51.4	68.7	58.8	80.6	67.0	73.2	67.6
Wiebe 05	77.6	68.4	72.7	70.6	79.4	74.7	73.8

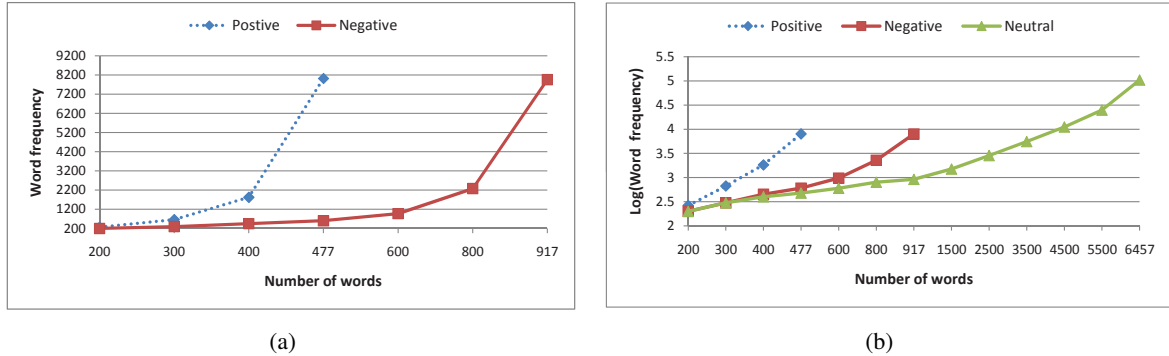


Figure 2: (a) Positive and negative lexicon statistics; (b) Positive, negative and neutral lexicon statistics.

objective. The improvement over this baseline will reflect how much subjLDA can learn from data. The LDA model (Blei et al., 2003), as shown in Figure 1(a), has been used as baseline in document-level sentiment classification in previous research (Lin et al., 2010). Thus, we also evaluated LDA on the sentence-level subjectivity detection task by modelling a mixture of three sentiment topics, i.e., positive, negative and neutral. For fair comparison, we encoded prior knowledge of sentiment lexicon into LDA as identical to subjLDA. Thus the LDA model here can be considered as a weakly-supervised version. Moreover, we tested LDA under two different modes, i.e., modelling a normal document vs. treating each individual sentence as a separate document. The sentence sentiment is determined as follows.

(a) **LDA in document mode:** sentiment of sentence m in document D is calculated using

$$P(l|m) \propto P(m|l)P(l|d) = \prod_{w_t \in m} P(w_t|l)P(l|d). \quad (8)$$

We define that sentence m is classified as an objective sentence if its probability of neutral label given sentence $P(l = neu.|m)$, is greater than both $P(l = pos.|m)$ and $P(l = neg.|m)$. Otherwise, the sentence is classified as subjective.

(b) **LDA in sentence mode:** sentence subjectivity is directly determined based on the per-

sentence sentiment distribution θ , using identical classification metrics to the document mode.

As can be seen from Table 1, a significant performance gain was observed for both subjLDA and LDA over the baseline. Particularly, more than 8% gain was observed for subjLDA, giving the best overall accuracy of 71.2% which is 3.1% and 3.6% higher than LDA(Sent.) and LDA(Doc.), respectively. In addition, except for objective recall, subjLDA outperforms LDA in both the sentence and document modes for all the other evaluation metrics, with more balanced objective and subjective F-measures being attained compared to the other two models. On the other hand, it was observed that while LDA(Doc.) can achieve a comparable subjective F-measure to LDA(Sent.), its objective F-measure is nearly 5% lower, resulting in worse overall performance. This is probably due to the fact that by treating each individual sentence as a document, LDA(Sent.) can avoid inferencing global sentiment topics and thus capture more accurate sentiment information from local topics. We measured the overall accuracy significance with paired t-Test (critical $P=0.01$). Results show that the improvements of subjLDA over both LDA(sent.) and LDA(doc.) are highly statistically significant. Thus, we conclude that subjLDA is superior than LDA in the subjectivity detection task.

When compared to the previous proposed boot-

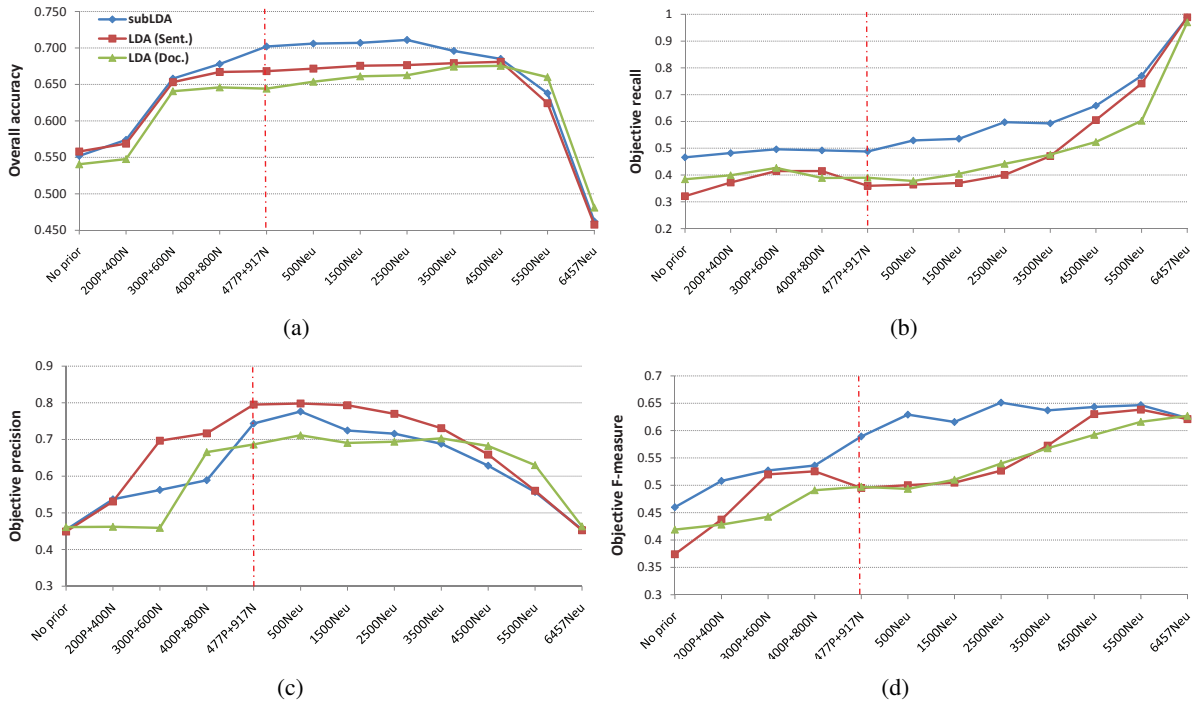


Figure 3: Subjectivity classification performance vs. different prior information by gradually adding the subjective and neutral words. The vertical dashed line denotes the point where all the positive and negative words have been incorporated into the model; 200P+400N denotes adding the least frequent 200 positive and 400 negative words; 500Neu denotes adding the least frequent 500 neutral words in addition to all the positive and negative words.

strapping approach (Wiebe and Riloff, 2005), subJLDA is about 2% lower in terms of overall accuracy. However, it should be noted that, their approach used a much larger training set for self-training which consists of more than 100,000 sentences. Moreover, apart from subjectivity clues, they also used additional features such as subjective/objective pattern and POS for the Naive Bayes sentence classifier training. In contrast, the proposed subJLDA model is relatively simple with only a small set of subjectivity clues being incorporated as prior knowledge.

5.2 Classification Results with Different Priors

While positive/negative sentiment lexicon is commonly used in lexical approaches to sentiment classification, the impact of incorporating neutral words remains relatively unexplored. In this experiment, we investigated the impact on the model performance by incorporating additional knowledge from neutral words. We started by first considering the positive and negative words only and gradually increased the number of words starting with the lowest frequency words. After all the pos-

itive and negative words have been incorporated, we then gradually added additional neutral words into the model also from the lowest frequency to the highest. Figure 2 shows the lexicon statistics of all the positive, negative and neutral words, where the value on the x-axis represents the number of words sorted by word frequency and the corresponding y-axis value indicates the total number of times those words appear in the corpus. For instance, the 400 least frequent positive words appear a total of 1,826 times in the corpus as shown in Figure 2(a).

Figure 3 depicts the subjectivity classification results of subJLDA and LDA by varying the proportion of lexical terms being incorporated. It is quite obvious from the overall accuracy shown in the figure that both subJLDA and LDA benefit from incorporating the information of subjective words, and in general, the more lexical items the better the results. Without using any neutral words, all the three models achieved the best results when all the subjective words were incorporated. It was noted that subJLDA performed similar to LDA when only a small number of low frequency subjective words were used. However,

with more higher frequency subjective words being incorporated, subjLDA shows stronger performance boosting over LDA and gives the best accuracy of 70.2% when all the subjective words were incorporated, being 3.4% and 5.8% better than LDA(Sent.) and LDA(Doc.), respectively, as indicated by the vertical dashed line in the figure.

On the other hand, adding neutral words is also beneficial, where performance gain was observed for all the 3 models in addition to the best results using subjective words only (i.e., subjLDA by 1%, LDA(Sent.) by 1.6% and LDA(Doc.) by 2.9%). Analyzing the objective recall and precision shown in Figure 3(b) and 3(c) reveals that, while incorporating the 4,500 least frequent neutral words considerably increases the objective recall, the objective precision does not drop much which eventually leads to the overall improvement of all the three models.

However, compared to the subjective words, the classification improvements by incorporating additional neutral words are less significant. This is probably due to the fact that while the presence of positive/negative words conveys clear subjective meanings, neutral words are relatively vague which could bear objective or subjective sense under different contexts. Furthermore, all three models experience a significant performance drop after the point of (4500Neu). Examining Figure 2 reveals that, while the 4,500 least frequent neutral words appear 11,142 times in the corpus, the 1,957 most frequent words (i.e., from 4500 to 6457) appear 93,036 times, nearly 10 times as much as the former. Thus, the high frequency neutral words become dominant in the model and result in severe classification bias towards the objective class. Therefore, appropriate filtering of neutral words is necessary in order to avoid introducing bias into model learning.

5.3 Sentiment Topics

In subjLDA, we model three topics in the per-corpus word distribution, each of which corresponds to neutral, positive and negative sentiment. Figure 4 shows the top 15 topic words of the sentiment topics extracted from the MPQA dataset by subjLDA. It can be easily observed that while the positive and negative sentiment topics consist of clear sentiment bearing words, the neutral topic contains mostly theme words with no sentiment, which illustrates the effectiveness of subjLDA in

Sentiment topics		
Neutral	Positive	Negative
countri	state	terror
presid	right	opposit
unit	support	concern
govern	gener	evil
intern	want	critic
bush	interest	question
report	posit	mean
elect	move	protest
china	remark	violat
militari	hope	accus
war	alli	refus
prison	agre	despit
taiwan	live	reject
minist	provid	fail
foreign	consent	impos

Figure 4: Sentiment topics extracted by subjLDA.

extracting sentiment bearing topics from text.

6 Conclusions and Future Work

This paper presents the subjectivity detection LDA Model (subjLDA) for sentence-level subjectivity classification. In contrast to most of the existing approaches requiring labelled corpora or linguistic pattern extraction, we view this problem as weakly-supervised generative model learning where the only input to the model is a small amount of domain independent subjective/neutral words. The subjLDA model has been evaluated on the MPQA dataset. Preliminary results show that except slightly lower in objective recall, subjLDA outperformed LDA over all other evaluation metrics, and is comparable to the previously proposed bootstrapping approach using a much larger training set. Moreover, it was found that while incorporating more subjective words can generally yield better results, the performance gain by employing extra neutral words is less significant.

There are several directions we would like to pursue in the future. While word lexical prior information is incorporated by modifying the Dirichlet prior for topic-word distributions here, it is also possible to explore other mechanisms to define expectation or posterior constraints. In addition, the current subjLDA model only models bag-of-words features, another future step would be extending subjLDA to include higher order information such as bigrams for improving model performance.

Acknowledgements

This work was supported in part by the EC-FP7 projects ROBUST (grant number 257859) and VPH-Share (grant number 269978).

References

- David M. Blei, Andrew Y. Ng, and Michael I. Jordan. 2003. Latent Dirichlet allocation. *J. Mach. Learn. Res.*, 3:993–1022.
- John Blitzer, Mark Dredze, and Fernando Pereira. 2007. Biographies, bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification. In *Proceedings of the Association for Computational Linguistics (ACL)*, pages 440–447.
- Koji Eguchi and Victor Lavrenko. 2006. Sentiment retrieval using generative models. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 345–354.
- A. Esuli and F. Sebastiani. 2006. SentiWordNet: A publicly available lexical resource for opinion mining. In *Proceedings of LREC*.
- M. Gamon, A. Aue, S. Corston-Oliver, and E. Ringger. 2005. Pulse: Mining customer opinions from free text. *Lecture Notes in Computer Science*, 3646:121–132.
- S. Lacoste-Julien, F. Sha, and M.I. Jordan. 2008. DiscLDA: Discriminative learning for dimensionality reduction and classification. In *NIPS*.
- Chenghua Lin and Yulan He. 2009. Joint sentiment/topic model for sentiment analysis. In *Proceedings of the ACM conference on Information and knowledge management (CIKM)*.
- Chenghua Lin, Yulan He, and Richard Everson. 2010. A comparative study of Bayesian models for unsupervised sentiment detection. In *Conference on Computational Natural Language Learning (CoNLL)*.
- Qiaozhu Mei, Xu Ling, Matthew Wondra, Hang Su, and ChengXiang Zhai. 2007. Topic sentiment mixture: modeling facets and opinions in weblogs. In *Proceedings of the 16th international conference on World Wide Web (WWW)*, pages 171–180.
- R. Mihalcea, C. Banea, and J. Wiebe. 2007. Learning multilingual subjective language via cross-lingual projections. In *In Proceedings of the Association for Computational Linguistics (ACL)*, page 976.
- G. Murray and G. Carenini. 2009. Predicting subjectivity in multimodal conversations. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1348–1357.
- Bo Pang and Lillian Lee. 2008. Opinion mining and sentiment analysis. *Found. Trends Inf. Retr.*, 2(1-2):1–135.
- S. Raaijmakers, K. Truong, and T. Wilson. 2008. Multimodal subjectivity analysis of multiparty conversation. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 466–474.
- D. Ramage, D. Hall, R. Nallapati, and C.D. Manning. 2009. Labeled LDA: A supervised topic model for credit attribution in multi-labeled corpora. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 248–256.
- E. Riloff and J. Wiebe. 2003. Learning extraction patterns for subjective expressions. In *Proceedings of the conference on Empirical methods in natural language processing (EMNLP)*, pages 105–112.
- Ivan Titov and Ryan McDonald. 2008. A joint model of text and aspect ratings for sentiment summarization. In *Proceedings of the Annual Meeting on Association for Computational Linguistics and the Human Language Technology Conference (ACL-HLT)*, pages 308–316.
- Peter D. Turney and Michael L. Littman. 2002. Unsupervised learning of semantic orientation from a hundred-billion-word corpus. *ArXiv Computer Science e-prints*, cs.LG/0212012.
- J. Wiebe and E. Riloff. 2005. Creating subjective and objective sentence classifiers from unannotated texts. In *Proceedings of the Conference on Computational Linguistics and Intelligent Text Processing (CICLing)*, volume 3406, pages 486–497. Springer.
- T. Wilson and S. Raaijmakers. 2008. Comparing word, character, and phoneme n-grams for subjective utterance recognition. In *Proceedings of INTERSPEECH*, pages 1614–1617.
- Hong Yu and Vasileios Hatzivassiloglou. 2003. Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 129–136.
- T. Zagibalov and J. Carroll. 2008. Automatic seed word selection for unsupervised sentiment classification of Chinese text. In *Proceedings of the International Conference on Computational Linguistics (COLING)*, pages 1073–1080.