Topic Modeling with Sentiment Clues and Relaxed Labeling Schema

Yasuhide Miura
Fuji Xerox Co., Ltd., Japan
`yasuhide.miura`
`@fujixerox.co.jp`

Keigo Hattori
Fuji Xerox Co., Ltd., Japan
`keigo.hattori`
`@fujixerox.co.jp`

Tomoko Ohkuma
Fuji Xerox Co., Ltd., Japan
`ohkuma.tomoko`
`@fujixerox.co.jp`

Hiroshi Masuichi
Fuji Xerox Co., Ltd., Japan
`hiroshi.masuichi`
`@fujixerox.co.jp`

Abstract

This paper proposes a method to extract sentiment topics from a text collection. The method utilizes *sentiment clues* and a *relaxed labeling schema* to extract sentiment topics. Experiments with a quantitative and a qualitative evaluations was done to confirm the performance of the method. The quantitative evaluation with a polarity classification marked the accuracy of 0.701 in tweets and 0.691 in newswire texts. These performances are comparable to support vector machine baselines. The qualitative evaluation of polarity topic extraction showed an overall accuracy of 0.729, and a higher accuracy of 0.889 for positive topic extraction. The result indicates the efficacy of our method in extracting sentiment topics.

1 Introduction

Continuous increase of text data arose an interest to develop a method to automatically analyze a large collection of texts. Topic modeling methods such as Latent Dirichlet Allocation (LDA)(Blei et al., 2003) are popular methods for such analysis. For example, they have been applied to analyze newswire topics (Blei et al., 2003; Rajagopal et al., 2013), scientific topics (Griffiths and Steyvers, 2004), weblogs (Mei et al., 2007), online reviews (Titov and McDonald, 2008b), and microblogs (Ramage et al., 2010; Zhao et al., 2011). Topic modeling methods generally extract probability distributions of word as *topics* of a given text collection. Note that this definition is quite different from the definitions in sentiment analysis or opinion mining literatures (Yi et al., 2003; Kim and Hovy, 2006; Stoyanov and Cardie, 2008; Das and Bandyopadhyay, 2010a; Das and Bandy-

opadhyay, 2010b) which basically define *topic* as an object of an opinion. Extracted topics are useful as a summary to catch a broad image of a text collection, but they are not always intuitively interpretable by humans. Typical methods for estimating topic modeling parameters aim to maximize a likelihood of training data (Blei et al., 2003; Griffiths and Steyvers, 2004). This objective is known to form topics that are not always most semantically meaningful (Chang et al., 2009).

Approaches to extract more explicit topics using observed labels are being proposed. Supervised LDA (Blei and McAuliffe, 2007), Labeled LDA (Ramage et al., 2009), and Partially Labeled Dirichlet Allocation (PLDA)(Ramage et al., 2011) are such supervised topic models. Labels of these supervised topic models are not required to be *strictly* designed. Strictly designed labels here mean organized and controlled labels like the categories of Reuters Corpora (Lewis et al., 2004). Ramage et al. (2009) and Ramage et al. (2010) showed the effectiveness of using labels like del.icio.us tags, Twitter hashtags, and emoticons. The use of these non-strict labels can avoid cost-intensive manual annotations of labels. However, available labels completely depend on a community that provides them. This is problematic when a text collection to analyze is already specified since we may not find labels that are suitable for an analysis.

Sentiment labels such as a product rating and a service rating are widely used labels that are community dependent. For example, a hotel may be positively rated for food but be negatively rated for room. These labels have been used successfully to extract sentiments of various aspects (Blei and McAuliffe, 2007; Titov and McDonald, 2008a). However, these kind of rating labels can not be expected to exist in communities other than review

sites.

This paper presents a method to extract sentiment topics from a text collection. A noticeable characteristic of our method is that it does not require strictly designed sentiment labels. The method uses *sentiment clues* and a *relaxed labeling schema* to extract sentiment topics. Sentiment clue here denotes meta data or a lexical characteristic that strongly relates to a certain sentiment. Some examples of sentiment clues are: a happy face emoticon that usually expresses a positive sentiment and a social tag[1] of a disaster that tends to bear negative sentiment. Sentiment label here is expected to be label that expresses a general sentiment like positive, neutral, or negative. Relaxed labeling schema is a schema that defines a process of setting labels to a text using the given sentiment clues. The key feature of this schema is that a text with a sentiment clue gets a sentiment-clue-specific label and a sentiment label. This assumes that words that co-occur with a sentiment clue tend to hold the same sentiment as the sentiment clue. The assumption follows an idea from supervised sentiment classification methods of Go et al. (2009), Read (2005), and Davidov et al. (2010) which presume strong relationships between certain emoticons and certain sentiments.

Our contributions in this paper are two-fold: (1) we propose a method that does not require strictly designed sentiment labels to extract sentiment topics from a text collection, (2) we show the effectiveness of our method by performing experiments with a quantitative and a qualitative evaluations. The rest of this paper is organized as follows. Section 2 describes our method in detail. Section 3 explains data that are used in the experiment of the method. Section 4 demonstrates the effectiveness of the method with an experiment. Section 5 indicates related works of the method. Section 6 concludes the paper with some future extensions to the method.

2 Methods

2.1 Partially Labeled Dirichlet Allocation

Our method utilizes PLDA (Ramage et al., 2011) as a supervised topic modeling method. PLDA is an extension of LDA (Blei et al., 2003) which is an unsupervised machine learning method that models topics of a document collection. LDA as-

---

[1]Social tag here means a non-strict tag that is defined in a web community (e.g. a del.icio.us tag or a Twitter hashtag).
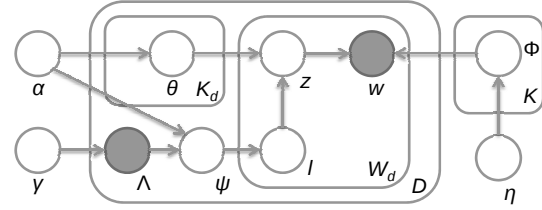

Figure 1: The graphical model of PLDA. Shaded elements represent observed elements.

sumes that documents can be expressed as an mixture of topics, where a topic is a distribution over words. PLDA incorporates supervision to LDA by constraining the use of topic with observed labels. The generative process of PLDA shown in Figure 1 is as follows:

> For each topic $k \in \{1 \ldots K\}$
> $\quad$ Pick $\Phi_k \sim \mathrm{Dir}(\eta)$
> For each document $d \in \{1 \ldots D\}$
> $\quad$ For each document label $j \in \Lambda_d$ (observed labels)
> $\quad\quad$ Pick $\theta_{d,j} \sim \mathrm{Dir}(\alpha)$
> $\quad$ Pick $|\Lambda_d|$ size $\psi_d \sim \mathrm{Dir}(\alpha)$
> $\quad$ For each word $w \in W_d$
> $\quad\quad$ Pick label $l \sim \mathrm{Mult}(\psi_d)$
> $\quad\quad$ Pick topic $z \sim \mathrm{Mult}(\theta_{d,l})$
> $\quad\quad$ Pick word $w \sim \mathrm{Mult}(\Phi_z)$

In the process, $\mathrm{Dir}(\cdot)$ represents a Dirichlet distribution and $\mathrm{Mult}(\cdot)$ represents a multinomial distribution.

The learning process of PLDA will be a problem to estimate parameters Φ, ψ, θ that maximizes the joint likelihood $P(\mathbf{w}, \mathbf{z}, \mathbf{l} | \mathbf{\Lambda}, \alpha, \eta, \gamma)$ of a given document collection. An efficient method for estimating these parameters are presented in Ramage et al. (2011).

2.2 Proposed Method

We propose a simple three step method to extract sentiment topics from a text collection.

Step 1: Preparation of Sentiment Clues

Firstly, a set of sentiment clue is prepared. Typical examples of sentiment clues are emoticons and social tags. Table 1 shows an example of a sentiment clue set.

Step 2: Relaxed Labeling Schema

Secondly, labels are set to texts using the sentiment clue set defined in Step 1. Labels are set

Sentiment Clue	Clue Name	Sentiment Label
:-)	happy face emoticon	positive
:-(	sad face emoticon	negative

Table 1: An example of a sentiment clue set.

to text differently in condition of sentiment clue existence. A text with a sentiment clue gets a sentiment-clue-specific label and a sentiment label that corresponds to it. For example, with the sentiment clue set of Table 1, a text including :-) gets a *happy face emoticon* label and a *positive* label. A text without any sentiment clue gets all sentiment labels that are defined in Step 1. For example, with the sentiment clue set of Table 1, a text that does not include :-) and :-(gets a *positive* and a *negative* labels. Table 2 summarizes how labels are set to texts. The basic policy of this process is to label texts with all possible labels. We call this schema *relaxed labeling schema* because this all-possible policy is non-strict, thus *relaxed*.

Step 3: Supervised Topic Modeling

Thirdly, a supervised topic modeling using PLDA is processed to the labeled texts of Step 2. Sentiment topics will be extracted as the topics that are labeled by the sentiment labels of Step 1. Note that our method is not fully dependent to PLDA. An alternate supervised topic modeling method that allows multiple labels to a text can be used instead of PLDA.

3 Data

We performed an experiment to confirm the effectiveness of the proposed method. Prior to explaining the details of the experiment, we will describe data that we used in it.

3.1 Emoticon Polarity List

We have done a preliminary investigation of emoticons to define sentiment clues. Firstly, we picked up six emoticons that are widely used in Japanese. Secondly, 300 tweets, 50 per emoticon, that include one of the six emoticons were annotated by three annotators with one of the following four polarities: positive, negative, positive and negative, and neutral. Thirdly, the number of positive annotations and negative annotations that two annotators or more agreed were counted for each emoticons. Table 3 shows polarity annotations that

Emoticon	Polarity
(´ ▽ `)ノ	positive
＼(ˆoˆ)／	positive
(ˆ-ˆ)	positive
orz	negative
(´・ω・`)	negative
(>﹏<)	negative

Table 3: The six emoticons and their largest vote polarities.

Criterion	Tweets
HAPPY	10,000
SAD	10,000
NO-EMO	200,000
total	220,000

Table 4: The summary of the topic modeling data.

each of the emoticons got the largest vote.

3.2 Topic Modeling Data

Tweets are used as the topic modeling data of the proposed method. *Public streams* tweets in Japanese during the period of May 2011– August 2011 are collected using the Twitter streaming API [2]. From there, we sampled total of 220,000 tweets that satisfy one of the following three criteria:

HAPPY 10,000 tweets that contain (´ ▽ `)ノ (a happy emoticon in Japanese, here on EMO-HAPPY).

SAD 10,000 tweets that contain orz (a sad emoticon in Japanese, here on EMO-SAD).

NO-EMO 200,000 tweets that do not contain any emoticon[3]. For this criterion, following two conditions were also considered: a tweet consists of five words or more and a tweet is not a retweet. These conditions are set to reduce the number of uninformative tweets and duplicate tweets.

In the sampling of NO-EMO, a Japanese morphology analyzer Kuromoji[4] is used for word segmentation. Table 4 shows the summary of the sampled tweets.

---

[2]https://dev.twitter.com/docs/streaming-apis
[3]10,924 Japanese emoticons which we collected from several web sites are used in this process.
[4]http://www.atilika.org/

Text Type	HFE label	positive label	negative label	SFE label
Texts with the happy face emoticon	√	√		
Texts without any emoticon		√	√	
Texts with the sad face emoticon			√	√

Table 2: The summary of how labels are set to texts with the sentiment clues of Table 1. In the table HFE is "happy face emoticon" and SFE is "sad face emoticon".

3.3 Polarity Classification Evaluation Data

Two data sets, *Tweet* and *Newswire*, are used to evaluate the performance of polarity classification. Tweet is an evaluation set of general tweets whose domain is same as the topic modeling data. Newswire is an evaluation set of newswire texts whose domain is quite different from the topic modeling data. The details of these sets are described in the following subsections.

3.3.1 Tweet

3,000 tweets satisfying the following three conditions are sampled from the May 2011–August 2011 tweets of Section 3.2:

  a. A tweet consists of five words or more (same as NO-EMO).

  b. A tweet includes an adjective, an adverb, an adnominal, or a noun-adverbial. This condition expects to increase the number of tweets that include evaluative content.

  c. A tweet does not have a POS tag that composes more than 80% of its words. This condition is set to exclude tweets such as a list of nouns or an interjection that includes a repeated character.

Note that words and their POS tags are extracted using Kuromoji like in NO-EMO.

The sampled 3,000 tweets were annotated with one of the following six polarity labels: positive, negative, positive and negative, neutral, advertisement, and uninterpretable. Label *advertisement* is defined to avoid annotating an advertising tweet to positive. Label *uninterpretable* is defined to prevent annotating a tweet that requires its accompanying context to determine a polarity.

Eighteen annotators formed ten pairs[5] and each pair annotated 300 tweets. The annotation agreement was 0.417 in Cohen's Kappa. We extracted

---

[5]Two annotators participated in two pairs.

Type	Polarity	Number
Tweet	Positive	384
	Negative	339
Newswire	Positive	107
	Negative	327

Table 5: The compositions of the polarity classification evaluation data.

723 tweets that two annotators agreed with positive or negative as polarity classification evaluation data. Tweet in table 5 shows the composition of the data.

3.3.2 Newswire

434 sentences of the Japanese section of NTCIR-7 Multilingual Opinion Analysis Task (MOAT) (Seki et al., 2008) that satisfies the following condition is extracted:

  a. A sentence with a positive or a negative polarity that two or more annotators agreed.

The Japanese section consists of 7,163 sentences from Mainichi Newspaper. Polarities are annotated to these sentences by three annotators. Note that the sentences are newswire texts, and are mostly non-subjective or neutral polarity. Newswire in table 5 shows the composition of the data.

4 Experiment

We performed an experiment and two evaluations to confirm the effectiveness of the proposed method.

4.1 Sentiment Clues

The sentiment clue set of Table 6 was used in the experiment. Note that the topic modeling data include 10,000 tweets that contain EMO-HAPPY and 10,000 tweets that contain EMO-SAD since they are used in the sampling process of them (Section 3.2).

Sentiment Clue	Sentiment Label
EMO-HAPPY	positive
EMO-SAD	negative

Table 6: The sentiment clues used in the experiment.

4.2 Preprocesses

Number of preprocesses were done to the topic modeling data to extract words from them.

1. Following normalizations are applied to the texts: Unicode normalization in form NFKC[6], repeated 'w's (a character used to express laugh in casual Japanese) are replaced with 'ww', a Twitter user name (e.g. @user) is replaced with 'USER', a hashtag (e.g. #hashtag) is replaced with 'HASH-TAG', and a URL (e.g. http://example.org) is replaced with 'URL'.

2. Words and their POS tags are extracted from the texts using Kuromoji.

3. Words that do not belong to the following POS tags are removed (stop POS tag process): noun[7], verb, adjective, adverb, adnominal, interjection, filler, symbol-alphabet, and unknown.

4. Six very common stop words such as *suru* "do" and *naru* "become" are removed.

5. The words are replaced with their base forms to reduce conjugational variations.

6. Words that appeared twice or less in the data are removed.

4.3 Supervised Topic Modeling

Stanford Topic Modeling Toolbox[8] is used as an implementation of PLDA. For the priors of PLDA, symmetric topic prior α and symmetric word prior η were set to 0.01. Number of topics for each labels were set to the numbers listed in Table 7. *Background* in the table is a special topic that can be used to generate words in any documents (tweets) regardless of their sentiment labels. In supervised topic modeling, this kind of topic can be used to extract label independent topic (Ramage et al., 2010).

Label	Number
positive	50
negative	50
EMO-HAPPY	1
EMO-SAD	1
background	1

Table 7: The number of topics set to each labels.

The parameter estimation of PLDA is done to the preprocessed data using CVB0 variation approximation (Asuncion et al., 2009) with max iteration set to 1000. Table 8 shows some examples of the extracted topics.

4.4 Evaluations

4.4.1 Quantitative Evaluation of Topics

A discriminative polarity classification was performed as a quantitative evaluation. Note that this evaluation dose not directly evaluate the performance of a sentiment topic extraction. However, following the previous works that jointly modeled sentiment and topic (Lin et al., 2012; Jo and Oh, 2011), we perform a sentiment classification evaluation. A more direct evaluation will be presented in Section 4.4.2.

Using the parameter estimated topic model, document-topic distribution inferences were conducted to the polarity classification evaluation data described in Section 3.3. From there, a positive and a negative score were calculated for each tweet with the following equation:

$$score(d, l) = \sum_{t_l} P(t_l|d) \qquad (1)$$

In the equation, d is a document (tweet), l is a label (either positive or negative), t_l is a topic of l, and $P(t_l|d)$ is the posterior probability of t_l given d. For each tweet, a label that maximizes Equation 1 was set as a classification label.

We also prepared a baseline support vector machine (SVM) based polarity detector similar to Go et al. (2009) for a comparison. HAPPY criterion tweets and SAD criterion tweets of Section 3.2 are used as the positive samples and the negative samples of SVM respectively. Following the best accuracy setting of Go et al. (2009), only bag-of-word unigrams were used as the features of SVM. For preprocesses, same preprocesses as the proposed method (Section 4.2) with EMO-HAPPY and EMO-SAD emoticons added to the

Label	Probable Words (Top 10)
EMO-HAPPY	(´▽`)ノ [EMO-HAPPY], USER [normalized user name], ない "no", ん [interjection], ?, の "thing", w [laugh expression], ww [laugh expression], 笑 "laugh", ... [ellipsis]
EMO-SAD	orz [EMO-SAD], USER [normalized user name], !, ー [macron], (,), ... [ellipsis], ゜ [degree symbol], д [a character often used in Japanese emoticons], 行く "go"
positive #11	USER [normalized user name], 食べる "eat", 美味しい "delicious", 飲む "drink", 屋 "shop", 料理 "meal", ラーメン "ramen", 店 "shop", コーヒー "coffee", 肉 "meat"
positive #30	!, USER [normalized user name], ありがとう "thank you", よろしく "please", お願い "please", くださる [honorific word], イイ "good", これから "from now", 楽しむ "enjoy", できる "can"
negative #2	さ [suffix similar to -ness], 暑い "hot", 夏 "summer", この "this", そう [reply word], 中 "inside", 今日 "today", 風 "wind", 外 "outside", 汗 "sweat"
negative #48	くる "happen", 目 "eye", 痛い "hurt", 入る "enter", 風呂 "bath", 寝る "sleep", 頭 "head", お腹 "stomach", すぎる "too", ない "no"

Table 8: Examples of extracted labeled topics with Table 7 setting. Bracketed expressions in the table are English explanations of preceding Japanese words that can not be directly translated.

Type	Method	Accuracy
Tweet	Majority Baseline	0.531
	Proposed	0.701
	SVM	0.705
Newswire	Majority Baseline	0.753
	Proposed	0.691
	SVM	0.712

Table 9: The polarity classification results. The majority baseline is the case when all predictions were same. This is positive for Tweet and negative for Newswire.

stop words. These two emoticons are added to stop words since they are used as the labels of this SVM baseline. As an implementation of SVM, LIBLINEAR[9] was used with L2-loss linear SVM and the cost parameter C set to 1.0.

Table 9 shows the results of polarity classifications. The proposed method marked an accuracy of 0.701 in Tweet, which is comparable to 0.705 of the SVM baseline. An accuracy was 0.691 for Newswire which is also comparable to 0.712 of the SVM baseline. However, the simple majority baseline has the highest accuracy of 0.753 in Newswire.

4.4.2 Qualitative Evaluation of Topics

The quantitative evaluation evaluated the performance of the sentiment topic extraction indirectly

with the sentiment classification. As a more direct qualitative evaluation, two persons manually evaluated the extracted 50 positive and 50 negative topics.

The evaluators were presented with *top 40 probable words* and *top 20 probable tweets* for each topic. Top 40 probable words of topic t_l were simply the top 40 words of the topic-word distribution $P(w|t_l)$. The extraction of top 20 probable tweets were more complex compared to the extraction of words. Document-topic distribution inferences were run to the training data using the parameter estimated topic model. For each topic t_l, top 20 tweets of document-topic distribution $P(t_l|d)$ were extracted as the top 20 probable tweets of t_l.

The evaluators labeled positive, negative, or uninterpretable to each of the topics by examining the presented information. The evaluators are instructed to label positive, negative, or uninterpretable. Label *uninterpretable* is an exceptional label. Topics with probable words and tweets that satisfy one of the following conditions were labeled uninterpretable: (a) majority of them are not in Japanese (b) majority of them are interjections or onomatopoeias, and (c) majority of them are neutral.

The agreement of the two evaluations was 0.406 in Cohen's Kappa. We extracted 59 topics that the two evaluators agreed with positive or negative, and measured the accuracies of the 50 positive and 50 negative topics. Table 10 shows the detail of the evaluation result. The overall accuracy was 0.729,

11

Label	#P	#N	Accuracy
positive	24	3	0.889
negative	13	19	0.594
overall			0.729

Table 10: The evaluation result of the 50 positive topics and the 50 negative topics. #P and #N are the number of topics that the two evaluators agreed as positive and negative respectively.

which indicates the success of the sentiment topics extraction.

5 Related Works

There are several works that simultaneously modeled topic and sentiment. Mei et al. (2007) proposed Topic Sentiment Mixture (TSM) model which is a multinomial mixture model that mixes topic models and a sentiment model. Lin et al. (2012) proposed joint sentiment-topic model (JSTM) that extends LDA to jointly model topic and sentiment. Jo and Oh (2011) proposed Aspect and Sentiment Unification Model (ASUM) that adapts LDA to model aspect and sentiment pairs. Titov and McDonald (2008a) proposed Multi-Aspect Sentiment (MAS) model that models topic with observed aspect ratings and latent overall sentiment ratings. Blei and McAuliffe (2007) proposed supervised LDA (sLDA) that can handle sentiments as observed labels. Our method is different from TSM model, JSTM, and ASUM since these models handle sentiments as latent variables. MAS model and sLDA utilize sentiments explicitly like in our method. However, not like in the relaxed labeling schema of our method, they have not presented a technique specialized for non-strict labels.

Sentiment analysis (Pang and Lee, 2008) also has a close relationship with our method. We borrowed the idea of using sentiment clues from sentiment analysis methods of Go et al. (2009), Read (2005), and Davidov et al. (2010). Our method is different from these method in the objective that the method aims to extract sentiment topics, not sentiments, from a text collection.

6 Conclusion

We proposed a method to extract sentiment topics using sentiment clues and the relaxed labeling schema. The quantitative evaluation with the polarity classification marked the accuracy of 0.701

in tweets and the accuracy of 0.691 in newswire texts. These performances are comparable to the SVM baselines 0.705 and 0.712 respectively. The qualitative evaluation of sentiment topics showed the overall accuracy of 0.729. The result indicates the success in the extraction of sentiment topics. However, compared to the high accuracy of 0.889 achieved in the extraction of positive topics, the extraction of negative topics showed the moderate accuracy of 0.594.

One characteristic of our method is that the method only requires a small set of sentiment clues to extract sentiment topics. Even though the method has its basis on a supervised topic modeling method, cost-intensive manual annotations of labels are not necessary. Despite the weakness of extracting negative topics shown in the qualitative evaluation, we think this highly applicable nature makes our method a convenient method. For future extensions of the method, we are planning the following two works:

Extraction of Aspect Topics

In this paper, we proposed a method that extracts sentiment topics using sentiment clues. Similar approach can be taken to extract non-sentiment topics if there are *clues* for them. For example, Twitter communities use hashtags to group variety of topics (Ramage et al., 2010). As a future work, we are planning to perform an aspect topic extraction using social tags as aspect clues.

Introduction of Non-parametric Bayesian Methods

In the experiment of our method, we set the equal number of topics to a positive and a negative labels. How polarities distribute should differ among domains, and this equal number setting may not work well on some domains. We are planning to introduce a non-parametric Bayesian method (Blei and Jordan, 2005; Ramage et al., 2011) to our method so that the number of topics can be decided automatically.

References

Arthur Asuncion, Max Welling, Padhraic Smyth, and Yee Whye Teh. 2009. On smoothing and inference for topic models. In *Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence*, pages 27–34.

David M. Blei and Michael I. Jordan. 2005. Vari-

ational inference for Dirichlet process mixtures. *Bayesian Analysis*, 1:121–144.

David M. Blei and Jon D. McAuliffe. 2007. Supervised topic models. In *Neural Information Processing Systems 20*, pages 121–128.

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

Jonathan Chang, Jordan Boyd-Graber, Sean Gerrish, Chong Wang, and David M. Blei. 2009. Reading tea leaves: how humans interpret topic models. In *Neural Information Processing Systems 22*, pages 288–296.

Dipankar Das and Sivaji Bandyopadhyay. 2010a. Extracting emotion topics from blog sentences: use of voting from multi-engine supervised classifiers. In *Proceedings of the 2nd international workshop on Search and mining user-generated contents*, pages 119–126.

Dipankar Das and Sivaji Bandyopadhyay. 2010b. Identifying emotion topic - an unsupervised hybrid approach with rhetorical structure and heuristic classifier. In *Proceedings of The 6th International Conference on Natural Language Processing and Knowledge Engineering*.

Dmitry Davidov, Oren Tsur, and Ari Rappoport. 2010. Enhanced sentiment learning using twitter hashtags and smileys. In *Proceedings of the 23rd International Conference on Computational Linguistics: Posters*, pages 241–249.

Alec Go, Richa Bhayani, and Lei Huang. 2009. Twitter sentiment classification using distant supervision. Technical report, Stanford University.

Thomas L. Griffiths and Mark Steyvers. 2004. Finding scientific topics. *Proceedings of the National Academy of Sciences*, 101(Suppl 1):5228–5235.

Yohan Jo and Alice Oh. 2011. Aspect and sentiment unification model for online review analysis. In *Proceedings of the fourth ACM international conference on Web search and data mining*, pages 815–824.

Soo-Min Kim and Eduard Hovy. 2006. Extracting opinions, opinion holders, and topics expressed in online news media text. In *Proceedings of the Workshop on Sentiment and Subjectivity in Text*, pages 1–8.

David D. Lewis, Yiming Yang, Tony G. Rose, and Fan Li. 2004. RCV1: A new benchmark collection for text categorization research. *Journal of Machine Learning Research*, 5:361–397.

Chenghua Lin, Yulan He, Richard Everson, and Stefan Rùger. 2012. Weakly supervised joint sentiment-topic detection from text. *IEEE Transactions on Knowledge and Data Engineering*, 24(Issue 6):1134–1145.

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*, pages 171–180.

Bo Pang and Lillian Lee. 2008. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1-2):1–135.

Dheeraj Rajagopal, Daniel Olsher, Erik Cambria, and Kenneth Kwok. 2013. Commonsense-based topic modeling. In *Proceedings of the Second International Workshop on Issues of Sentiment Discovery and Opinion Mining*, pages 6:1–6:8.

Daniel Ramage, David Hall, Rameshi Nallapati, and Christopher D. Manning. 2009. Labeled LDA: A supervised topic model for credit attribution in multi-labeled corpora. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, pages 248–256.

Daniel Ramage, Susan Dumais, and Dan Liebling. 2010. Characterizing microblogs with topic models. In *Proceedings of the fourth international AAAI conference on Weblogs and Social Media*, pages 130–137.

Daniel Ramage, Chistopher D. Manning, and Susan Dumais. 2011. Partially labeled topic models for interpretable text mining. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 457–465.

Jonathon Read. 2005. Using emoticons to reduce dependency in machine learning techniques for sentiment classification. In *Proceedings of the ACL Student Research Workshop*, pages 43–48.

Yohei Seki, David Kirk Evans, Lun-Wei Ku, Le Sun, Hsin-Hsi Chen, and Noriko Kando. 2008. Overview of multilingual opinion analysis task at NTCIR-7. In *Proceedings of the 7th NTCIR Workshop Meeting on Evaluation of Information Access Technologies: Information Retrieval, Question Answering, and Cross-Lingual Information Access*, pages 185–203.

Veselin Stoyanov and Claire Cardie. 2008. Annotating topics of opinions. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation*, pages 3213–3217.

Ivan Titov and Ryan McDonald. 2008a. A joint model of text and aspect ratings for sentiment summarization. In *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 308–316.

Ivan Titov and Ryan McDonald. 2008b. Modeling online reviews with multi-grain topic models. In *Proceedings of the 17th international conference on World Wide Web*, pages 111–120.

Jeonghee Yi, Tetsuya Nasukawa, Razvan Bunescu, and Wayne Niblack. 2003. Sentiment analyzer: Extracting sentiments about a given topic using natural language processing techniques. In *Proceedings of the Third IEEE International Conference on Data Mining*, pages 427–434.

Wayne Xin Zhao, Jing Jiang, Jianshu Weng, Jing He, Ee-Peng Lim, Hongfei Yan, and Xiaoming Li. 2011. Comparing Twitter and traditional media using topic models. In *Proceedings of the 33rd European conference on Advances in information retrieval*, pages 338–349.