

Study on Identification of Subjective Sentences in Product Reviews Based on Weakly Supervised Topic Model

Wei Jiang^{1,2}

1. College of Computer Science and Technology, Wuhan University of Technology, Wuhan, China

2. Department of Information Engineering, City College of Wuhan University of Science and Technology, Wuhan, China

jeew13@126.com

Abstract—Sentiment analysis or opinion mining in online product reviews is a method that can automatically detect subjective information regarding the entity such as opinions, attitudes, and feelings expressed by consumers. Online product reviews always include objective and subjective sentences; identification of subjective sentences in the given content is a very important and foundational task in the research of opinion mining. In this paper, we focus on the problem of identification of sentence-level subjective sentences, propose a weakly supervised model mixed topics based on LDA for identification of the subjective sentences, considering the impact of multiple topic factors on the identification of subjective sentences. The approach exploits semi-supervised learning method, and extended the existing basic LDA topic model for the identification of subjectivity in text. This work iterates the model prior probability by using a small domain-independent lexicon. Finally, the proposed model is applied to a online review corpus and the experimental shows that the proposed model can effectively improve the recognition effect.

Index Terms—online product reviews, opinion mining, topic model, subjective sentence

I. INTRODUCTION

With the advent of Web2.0, the rapid development of BBS, blog, e-commerce sites, as well as the major web portals are changed the way people use the Internet thoroughly. Buyers can easily publish the comments on the internet for the purchase or service; these comments are often subjective view, or objective fact of descriptions of commodities. These evaluations can help potential buyers to choose suitable goods or services for their own, can guide the business to improve their service quality, also can make the producers of goods or services to see the consumer preferences, understanding the differences to their own and competing goods or services, which can help the producers to find their own shortcomings, and can be targeted to improve the quality of their products, to make their products gain greater recognition of consumers. Unfortunately, due to a large number of comments, making a potential buyers, sellers and goods maker to cannot obtain the effective overall opinion

distribution in a short time relatively, thus effective auxiliary decision-making.

Because more and more users are willing to share their views or experience on the internet, this kind of comment information rapidly expand, only by artificial method is difficult to deal with massive amounts of online information collection and processing, so urgently we need the computers to help users quickly and sort out these related evaluation information, based on the technology of product reviews mining arises at the historic moment, which are mainly dealing with online reviews text of product. In short, it's an analysis, processing, summing up and reasoning process of the subjective text with opinion. The original sentiment analysis from the analysis of the sentiment words of predecessors [1], such as "nice" is a positive word; the "ugly" is a negative word. Due to the sentiment analysis based on product reviews can help users understand the products' reputation in the mind of the public, so favored by many consumers and business website. As a lot of subjectivity text appears on the internet, the researchers gradually from the simple analysis of the opinion words study transfer to more complex opinion mining. Based on this, according to the processing of text granularity, sentiment analysis can be divided into words, phrases, sentences and discourse level, etc. Subjective expression refers to the collection of words or phrases in the unit subjective text. In view of the comment text words is part of the subjective expression. In addition, the combination of certain words, such as the village idiot can clearly identify the subjectivity of text, although any one of the words alone may not the evaluate words. How to obtain these meaningful phrases is the key of the recognition. If anyone can detect the subjective sentences in reviews resource and make effective treatment, the performance of sentiment analysis would arise to some extend [2]. Therefore, detecting the subjective sentences in online review documents should serve a critical function in helping opinion mining task.

Online product reviews including objective and subjective sentences, identification of subjective sentences in the given content is a very important and mainly task in the research of opinion mining. Generally,

the subjective sentences have the characteristics of domain independence and context-sensitive. Wiebe and Riloff et al. [3,4] used bootstrapping to learn the subjective evaluation phrases or objective & objective evaluation phrases and then to capture the subjective sentences. Lin and He et al. [5] regarded the subjective recognition as a generative model, exploited a small subjective domain-independent lexicon to find out the tendency of sentences. But intuitively, we found that the subjective expression of sentiment tendency always depends on the specific topic information [6]. In recent years, Mei et al. [7,8,9] has proposed the probability generative model mixed with topic and sentiment, but they are only focused on sentiment analysis of document level. Considering the impact of multiple topic factors on the identification of subjective sentences, this paper aim at the problem of identification of sentence-level subjective sentences, exploit semi-supervised learning method, and extended the existing basic LDA model for the identification of subjectivity (subjLDA)[5], and finally proposed a new weekly supervised topic model(Multi-subjLDA) to recognize the subjective sentences in reviews.

In this paper, the proposed model is closely related with the JST and subjLDA model, but they are very different:(1) the JST model was considering the influence of subject factors on sentiment analysis, but it is only for the document, and the goals of the model is to discover subjective sentences, which aim to sentiment analysis of sentence level.(2) the subjLDA model only considers three kind of sentiment polarities, i.e. positive, negative and neutral as the topics of sentiment analysis , without thinking of multiple topics of document collection. But sentiment or subjectivity often relies on topic, thus this paper takes the influence of multiple topics on subjective sentences recognition into account on the basis of subjLDA. Finally, the proposed model is applied to a online review corpus and the experimental shows that the proposed model can effectively improve the recognition effect.

In summary, this paper makes two main contributions:

1) It proposes a semi-supervised topic modelling method, called Multi-subjLDA, which combining multiple topics and domain-independent subjectivity lexicon to improve subjective sentences recognition.

2) It considers the influence of multiple topics on subjective sentences detection based on subjLDA proposed by Lin et al. [5]. In the same time , the proposed model extends ability of subjLDA, which can extract latent subjective topics from document collections.

The rest of this paper is organized as follows: In Section 2, we first discuss related works and then introduce LDA model and its inference in Section 3.1. We present the proposed model and its generative process in Section 3.2 and 3.3.Experiment's setup is reported in Section4.1 and then we discussed the comparative results in Section4.2. Finally, we conclude our work and indicates future work in Section5.

II. RELATED WORK

With the development and popularization of internet and more and more consumers have left their views, suggestions and even personal preference in the web, it has very practical value for how to summarize automatically these text information which contains strong sentiment tendency. Therefore, there have been many research interests in sentiment analysis and opinion mining on review text.

Turney et al. [10] proposed points of mutual information (PMI) method to expand the benchmark vocabulary with sentiment polarity, and made use of the algorithm of latent semantic analysis (LSA) to analyze the sentiment that expressed in text. They identified appraise words by calculating the correlation value between all subjective words in WordNet and seed words "good" and "bad" that represent positive and negative polarity respectively. However, it's not all sentiment sources of other language as rich as English language, some researchers translate sentiment lexicon of rich resource language into the other language with less resource of sentiment lexicon[11], for instance ,some researchers translated English lexicon into Chinese. But the results presented by many experiments showed that many appraise words have changed their polarity after translation.

B.Pang et al. [12] used some machine leaning methods such as Bayesian, maximum entropy (EM) and SVM algorithm to classify the review of film; Liu and Hu [13] discussed the method of mining product features from product reviews so as to get the overall sentiment orientation of certain feature of the products that expressed by consumers. Most classification methods aimed to sentence level and document level is based on the sentiment polarity identification of word and phrases. Many researchers developed the supervised or semi-supervised method to achieve the classification of sentiment. Hatzivassiloglou and Wiebe [14] have researched the classification of subjectivity in sentence level, they focused on the problem that determine the sentences is subjective sentences or objective sentences. Hu and Liu [13], Hovy [16], Wiebe and Roiloff [4] carried on the thorough research about the polarity classification of opinion expressed in sentences.

In recent years, with the research of topic model [17,18] rise gradually, many researchers applied it to the sentiment analysis. As the evaluative object is contained in some topics from the opinionate text, so we can use the topic model to classify the evaluative object. Tivo et al. [19] adopted Multi-Grain LDA topic model to mine the appraise entity from product reviews and clustered similar object. Though this method improved performance of recall in extracting the appraise object theoretically, there has no experiment to compare with the other traditional methods based on noun or noun phrase.

Mei et al. [7] constructed a Topic-Sentiment Mixture (TSM) model based on pLSA to study the sentiment analysis of Micro-blog. The model they proposed consists of a background language model and a topic model.TSM model can be used in cross-domain comment

text because it does not need prior domain knowledge. However, the shortcoming of pLSA is over fitting in the process of parameter estimation.

Lin et al. [8] proposed Joint Sentiment Topic (JST), which used sentiment label to mark sentiment layer in the model and documents, topics and words are all associated with sentiment in JST model structure. The model the proposed considered different probability distribution for each sentiment polarity and employed prior polarity of words to achieve aspect extraction in cross domain.

Brody et al. [20] proposed a supervised method called Aspect-Sentiment Model. They firstly regarded each sentence in reviews as a independent document and identified ratable aspect by the basis LDA model, then exploited adjective words to recognize opinion words related with specific aspect.

Jo et al. [9] proposed Sentence-LDA (SLDA) model. SLDA assumed that each words in a sentence belongs to a same aspect. Aspect that identified by SLDA can match details in comments. Furthermore, they proposed Aspect and Sentiment Unified Model (ASUM) by modeled sentiment that corresponding to different aspects based on SLDA. ASUM can capture the important information pairs like (aspect, sentiment) from reviews.

Mukherjee and Liu et al. [21] proposed TME(Topic and Multi-Expression) model, which can model topic and different type of expression in comments simultaneously. TME also can make a distinction between topic and different kind of expression via a transition variable. Furthermore, they exploited EM prior as guide to separate topics and improved TME so as to obtain a new model called Maximum-Entropy TME.

Identification of subjective sentences is much highly difficult compared with sentiment analysis [11], at the same time, improving the performance of subjective sentences recognition is in favour of improving the accuracy of sentiment classification. McDonald et al. [22] holds a view that the identification and analysis on opinion sentences of varied granularities can significantly improve the effectiveness of the sentiment analysis. Wiebe and Riloff et al. [3] used bootstrapping method to identify the subjective sentences. They adopted the subjectivity classifier based rules to classify subjective sentences and objective sentences, and then learn subjective evaluation phrase pattern from annotated text, and finally identify subjective sentences by these patterns automatically and expand the training set simultaneously. Zhao et al. [23] used a automatic-selected syntax tree to identify comment phrases, for thinking that it is very important for these evaluation phrases to judge the sentiment polarity.

Recently, the use of topic model with unsupervised learning method for opinion mining has attracted a lot of researcher's attention. Lin and He et al. [5] puts forward a kind of weakly supervised model(subjLDA) for extracting subjective sentences. The model they proposed similar with ours proposed model, but they didn't put the influence of multiple subjects on subjective sentences into consideration. Mei et al. [7] proposed a hybrid topic

model which mixed with sentiment via introducing a background model and two independent sentiment models. Lin et al. [8] proposed a joint sentiment topic model (JST) that can find sentiment and topics simultaneously, in which they assumed that the topic generated depends on the distribution of sentiment, and the word is generated while the topics and sentiment are known. A unified sentiment model (ASUM) proposed by Jo et al. [9] is similar with JST and what's different is that the former must choose the model constraints from the distribution of the same word while the latter allows word to choose from different distributions.

III. ANALYSIS OF MODEL

A. Introduction to LDA Topic Model

In recent years, with the gradual rise of statistical topic model, many researchers applied it to the field of sentiment analysis. Statistical topic model is the current research in the field of text mining major paradigm, LDA model is a typical representative of it. Simple probabilistic model can be seen as a generative probabilistic model which indicates generative process of document

LDA model is that a three-layer generative Bayesian probabilistic model consists of documents, topics and words. Fig. 1 shows the graphical model of LDA model. Assume that we have a document collection with D documents, LDA model takes into account that each document is a probabilistic mixture of the topic set K , where K denote the number of topic set, and each topic k is a multinomial distribution on the words in vocabulary. The only observable variable is word w marked with color gray as present in Fig. 1. w_{dn} represents the n th word in d th documents, $w_{dn} \in V$; V is a vocabulary with all words in D documents; z_{dn} is the topic that generate word w_{dn} ; α is a prior hyper parameter about topic distribution in the document collection; In the document layer, θ_d is the distribution of all the topics in document d , which obeys a Dirichlet distribution $Dir(\theta_d | \alpha)$; a topic ϕ_k is the distribution on words in the vocabulary V , The model depicted in Fig. 1 contains distribution $\phi_{1:K}$ of K topics on word, N is the total number of words in a document d including duplicate words. LDA is a generative probabilistic model, assuming K is confirmed and given the parameters α and $\phi_{1:K}$, generation of document d consists of two steps: At first, choose randomly a vector θ_d with K dimensions from a Dirichlet distribution $p(\theta | \alpha)$ to generate topic distribution of document d ; then generate each word w_{dn} of document d according to $p(w_{dn} | \theta_d, \phi_{1:K})$.

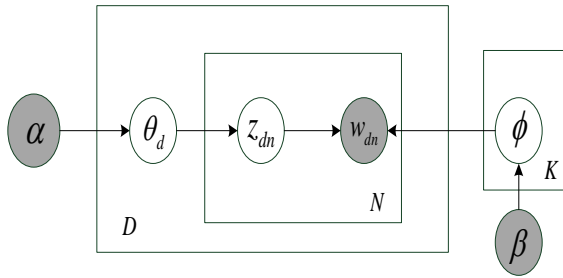


Figure 1. The graphical model of the LDA topic model.

The process of extracting implicit topic from document collections that based on LDA model is similar with generative process of document that have mentioned above. In the conditions of given word $w_{1:D,1:N}$ of document collections, parameters such as $\phi_{1:K}$, $\theta_{1:D}$ and α can be got by launching word $w_{1:D,1:N}$ corresponding hidden variables $z_{1:D,1:N}$ in reverse. Many researchers adopted method of approximate reasoning because of it is hard to make the accurate estimation in terms of more than one unknown variable exists. Blei D et al. [17] calculated it by Variational Bayes inference algorithm, Griffiths T L et al. [24] lead in hyper-parameters β , with the following Eq. (1), Eq. (2) and Eq. (3) using Gibbs Sampling method to estimate posterior distribution of the topic z_{di} for the current sample word w_{di} , then obtained the model parameters θ and ϕ .

$$p(z_{di} = j | w_{di} = m, z_{-di}, w_{-di}) \propto \frac{C_{mj}^{VK} + \beta}{\sum_{m'} C_{mj}^{VK} + V\beta} \times \frac{C_{dj}^{DK} + \alpha}{\sum_{j'} C_{dj}^{DK} + K\alpha} \quad (1)$$

$$\phi_{mj} = \frac{C_{mj}^{VK} + \beta}{\sum_{m'} C_{mj}^{VK} + V\beta} \quad (2)$$

$$\theta_{dj} = \frac{C_{dj}^{DK} + \alpha}{\sum_{j'} C_{dj}^{DK} + K\alpha} \quad (3)$$

Where ϕ_{mj} is the probability on the current word m on the topic j ; θ_{dj} is the probability on the document d that includes topic j , and z_{-di} represent all the topics assignments excluding the i th term; C_{mj}^{VK} denotes the occurrence times of probability m of word w_{di} in Statistics of word frequency matrix $K \times V$; and C_{dj}^{DK} is the number of times that topic j has occurred in document d ; $z_{di} = j$ denotes that the topic j assigned to w_{di} . The basic idea is to iterate in order to obtain all hidden variables by using the distribution ϕ

of all topic that has been sampled on the words and distribution θ_d that document d distributed on the topic inferred topics z_{di} of sampling of current word w_{di} .

The LDA model we introduced above can get all the topics discussed in reviews, which is called global topics. However, LDA model cannot save the local relation between each topic and corresponding sentiment because representation of documents is employed by bag-of-words. To address this problem, the influence of multiple topics on subjective sentences recognition into account on the basis of LDA model.

B. Generative Process of Proposed Model

From what has been discussed above, Fig. 2 depicts the proposed model named Multi-subjLDA, which is a 4-level Bayesian model. This model introduced the factors of topics on the base of subjLDA model, combining the topics with the subject sentence for identification. We use $D = \{d_1, d_2, \dots, d_D\}$ to denote the document collection, $M_d = \{1, 2, \dots, N_d\}$ is the number of sentences in each document. Let N_d denote the number of sentences in d document. Each sentence consists of $N_{d,m}$ words in a sequence of $M_d = \{w_1, w_2, \dots, w_{N_{d,m}}\}$, in which $N_{d,m}$ means the number of words in the sentence of M_d in the document.

Assume the document collection of words from the set V without repeating terms, let T and K denote the number of topics and the number of sentence-level subjectivity labels respectively. Let S denotes the number of word-level subjectivity labels. $\{z, s, l\}$ in the model are hidden variables which represent topic, sentence-level subjectivity label and word-level subjectivity label respectively. The procedure of word that generated in the document d by the model is as follows: At first, choose a sentence level subjectivity label $s_{d,m}$ for each sentence in the document d from the document subjective distribution π_d , then choose a word level subjectivity label $l_{d,m,t}$ for each word in the sentence from the sentence level subjective distribution $\mu_{s_{d,m}}$; secondly, choose a topic level subjectivity label $z_{d,m,t}$ randomly from the topic level subjective distribution $l_{d,m,t}$, using word level subjectivity label; finally, choose a word from the word level subjective distribution $\phi_{z_{d,m,t}, l_{d,m,t}}$ on the base of the word level subjectivity label $l_{d,m,t}$ and the topic $z_{d,m,t}$. Sentence-level subjectivity can be got from the sentence level subjectivity label $s_{d,m}$.

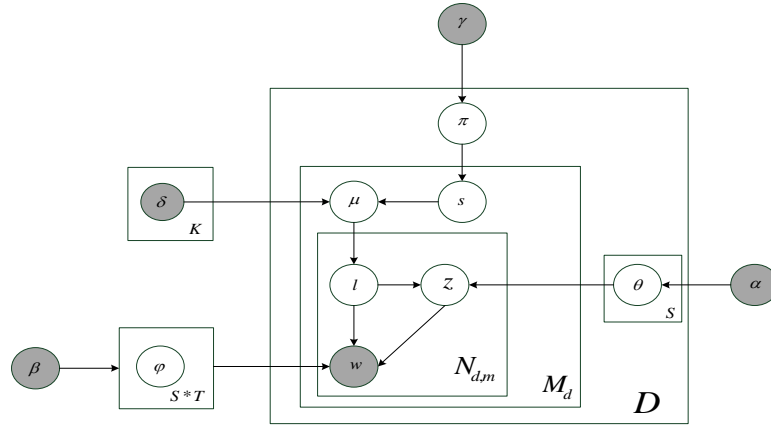


Figure 2. Multi-subjLDA model.

C. Inference

According to generation of Multi-subjLDA model, we derive the joint probability distribution as follows:

$$p(w, z, s, l | \alpha, \beta, \gamma, \delta) = p(w | z, l, \beta) p(z | l, \alpha) p(l | s, \delta) p(s | \gamma) \quad (4)$$

To evaluate 4 distributions as θ , ϕ , μ , π , we can use Gibbs Sampling to evaluate the posterior probability distribution of hidden variables: $z_{d,m,t}$, $l_{d,m,t}$, $s_{d,m}$, and then evaluate model parameters through sequences of words known. According to the Conditional posterior distribution, it can be inferred from Eq. (5) and (6).

For each sentence, let $x = (d, m)$ denotes an index in the sentence m of the document d and $-x$ denotes a set of sentences except from the current sentence m . According to the Gibbs sampling method, posterior distribution of the sentence-level subjectivity label s_x can be evaluated as follows:

$$P(s_x = i | s_{-x}, l, w, z) \propto \frac{N_{d,i}^{-x} + \gamma}{N_d^{-x} + K\gamma} \times \frac{\prod_{k=1}^S \prod_{b=0}^{N_{d,m,k}-1} (b + \delta_{s,k})}{\prod_{b=0}^{N_{d,m,k}-1} (b + \sum_{k=1}^S \delta_{s,k})} \quad (5)$$

In which, $N_{d,i}$ denotes the number of sentences that endowed with the sentence-level subjectivity label i . N_d denotes the number of all sentences in the document. $N_{d,m,k}$ denotes the number of the sentence m that endowed with the sentence level subjectivity label k . $N_{d,m}$ denotes the number of the sentence m in the document.

For each word, let $t = (d, m, n)$ denotes the set of sentences except n th word in the m th sentence, and then posterior probability distribution of z_t and l_t is represented as follows:

$$p(l_t = k, z_t = j | w, s, z_{-t}, l_{-t}) \propto \frac{N_{k,j,i}^{-t} + \beta}{N_{k,j}^{-t} + V\beta} \times \frac{N_{d,k,j}^{-t} + \alpha}{N_{d,k}^{-t} + T\alpha} \times \frac{N_{d,m,j}^{-t} + \delta_{s_{d,m},j}}{N_{d,m}^{-t} + \sum_{j=1}^S \delta_{s_{d,m},j}} \quad (6)$$

In which, $N_{k,j,i}$ denotes the number of the word i that endowed with the topic j and the sentence level subjectivity label k . $N_{k,j}$ denotes the number of all words endowed with the topic j and the sentence level subjectivity label k . $N_{d,k,j}$ denotes the number of the document d endowed with the topic j and the sentence level subjectivity label k . $N'_{d,k}$ denotes the number of the document d endowed with the sentence level subjectivity label k . With Eq. (5) and (6), sampling method of Markov Chain can be used to evaluate as follows:

1) distribution of document-level subjectivity:

$$\pi = \frac{N_{d,i} + \gamma}{N_d + K\gamma} \quad (7)$$

2) distribution of sentence-level subjectivity:

$$\mu = \frac{N_{d,m,k} + \delta_{s_{d,m},k}}{N_{d,m} + \sum_{k=1}^S \delta_{s_{d,m},k}} \quad (8)$$

3) distribution of subjectivity of topics in documents:

$$\theta = \frac{N_{d,k,j} + \alpha}{N'_{d,k} + T\alpha} \quad (9)$$

TABLE I.
DATA SET OF DIGITAL CAMERA REVIEWS

Documents	Subjective Sentences	Objective Sentences	Words	Entries of Lexicon
672	2681	591	10421	7152

TABLE II.
RESULTS OF IDENTIFICATION

Model	Subjectivity(%)			Objectivity(%)		
	Recall	Precision	F-measure	Recall	Precision	F-measure
LSM	84.1	81.5	82.8	62.1	58.5	60.2
JST	73.8	71.2	72.5	67.8	52.2	58.9
subjLDA	87.8	85.3	86.5	63.8	55.3	59.2
Multi-subjLDA	89.2	84.6	86.8	66.2	60.6	63.3

TABLE III.
SUBJECTIVE AND OBJECTIVE SENTENCES IDENTIFIED FROM CORPUS

Digital Camera Reviews	
Subjective Sentences	Objective Sentences
#Red eye is very easy to correct	#I have searched for a digital camera for 3 months
#The camera comes with an excellent easy to install software	#The camera does not have a digital zoom
#The picture are absolutely amazing	#This is my second machine
#the software that comes with it is amazing	#I purchased it a year ago and have had many problems
#This is the best digital camera on the market	#I have used the cannon500 in my role as a reporter last year

4) distribution of subjectivity of words in set:

$$\varphi = \frac{N_{k,j,i} + \beta}{N_{k,j} + V\beta} \quad (10)$$

IV. EXPERIMENT AND ANALYSIS

A. Experiment Settings

We used review data set provided in [13] to evaluate the performance of the model proposed in this paper. We select 672 documents from data set randomly about digital camera review, containing 3272 sentences in all. With the method of manual annotation, get 2681 words of subjective sentences and 591 objective sentences. Eventually obtained 10421 words and a lexicon contains 7152 words after pre-processing. Specific data is shown in Table I.

In the process of learning the model, combined with the subjective index (subjClue) dictionary is as a prior knowledge. We extract subjective emotional words that are marked as strong, and then removed the repeating words, and eventually get a emotional dictionary consists of 1012 subjective words. In the experiment, the prior

knowledge of emotional dictionary is only used when posterior distribution is initialized, if a item in the collection of lexical matches with the word in dictionary, the word will be marked as a emotional word, if a sentence contains more than one emotion word, the sentence will be annotated as a subjective sentence; Otherwise, the subjectivity label of the words or sentences was randomly generated.

B. Experimental Results and Analysis

α , β , λ , δ are 4 hyper parameters in this model, all of them are setted in generative process by using the method that is similar with [5,6,8]. In this experimental, we assigned the value $\gamma = (0.05 \times L) / K$, $\beta = 0.01$ and $\alpha = 50 / T$ respectively, in which, parameter L represent the average length of documents and δ is learned from corpus by using maximum likelihood estimate(MLE).

In order to compare the influence of recognition with the different topics in the Multi-subjLDA model and observe the recognition results in the above product reviews data set, some topic models we have mentioned above used to compare with the proposed model (Multi-subjLDA) in our experiment, the results shows in the following Table II.

As showing in Table II, the model of Multi-subjLDA performance certainly ascends a little compared with the other three models. Although Multi-subjLDA recognition accuracy is lower than subjLDA in the task of subjective words recognition, its recall and F-measure are better and obtained accuracy 89.2% and 86.8% respectively. In addition, the experiment evaluates the model of objective sentence recognition, which performance is generally low, but still higher than the other three models compared with the subjective sentences identification, the recall, precision and F-measure are respectively reached 66.2%, 60.6% and 66.2%. The following Table III shows the five subjective and objective sentences extracted from corpus, we can find that subjective sentences contains some obvious emotional words, while objective topics are mostly some neutral ones.

Due to the influence of different topics, the subjectivity sentence recognition may be influenced in the experiment. We always design different topics to analysis the subjective sentences. Fig. 3 depicts that the F-measure of Multi-subjLDA model is higher than other methods while the topics number is 40 to 80, but there are little performance degradation when the number of topic increased, the reason perhaps is that subjective word distribution becomes sparse with the increasing of topic quantity.

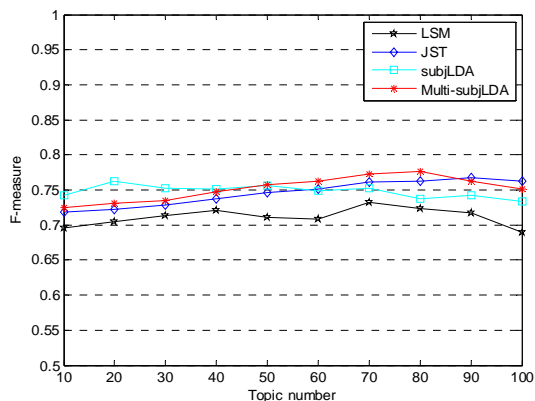


Figure 3. Performance of different models with different topic number

V. CONCLUSION

This paper has proposed the sentence recognition model aiming at multiple topics, and the identification of subjective sentences only extracted from the text collections. By combining multiple topics and domain-independent subjectivity lexicon to improve subjective sentences recognition. The experiment of the comments corpora of digital camera user show that the subjectivity of sentence recognition model can effectively improve the performance of subjective sentences recognition and extract meaningful topics. Future, we will find other ways to solve aiming to the problem of subjective word distribution sparse.

REFERENCES

- [1] Hatzivassiloglou, Vasileios, Kathleen R. McKeown. "Predicting the semantic orientation of adjectives," In Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and Eighth Conference of the European Chapter of the Association for Computational Linguistics. Association for Computational Linguistics, 1997, pp. 174-181.
- [2] Riloff, Ellen, Janyce Wiebe, William Phillips. "Exploiting subjectivity classification to improve information extraction," In Proceedings of the National Conference On Artificial Intelligence. Vol. 20(3). Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999, 2005, pp. 1106-1111.
- [3] Riloff, Ellen, and Janyce Wiebe. "Learning extraction patterns for subjective expressions," In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), 2003.
- [4] J. Wiebe, E. Riloff. "Creating subjective and objective sentence classifiers from unannotated texts," In Proceedings of the Conference on Computational Linguistics and Intelligent Text Processing (CICLing), 2005.
- [5] Chenghua Lin, Yulan He, Richard Everson. "Sentence subjectivity detection with weakly-supervised learning," In Proceedings of International Joint Conference on Natural Language Processing (IJCNLP), 2011.
- [6] Chenghua Lin, Yulan He, Richard Everson. "A comparative study of bayesian models for unsupervised sentiment detection," In Proceedings of Conference on Computational Natural Language Learning (CoNLL), 2010.
- [7] Qiao Zhu Mei, Xu Ling, Matthew Wondra, et al. "Topic sentiment mixture: modeling facets and opinions in Weblogs," In Proceedings of the 16th International Conference on World Wide Web(WWW), 2007, pp. 171-180.
- [8] Chenghua Lin, Yulan He. "Joint sentiment/topic model for sentiment analysis," In Proceedings of the 18th ACM Conference on Information and Knowledge Management, 2009, pp. 375-384.
- [9] Yohan Jo, Alice Oh. "Aspect and sentiment unification model for online review analysis," In Proceedings of the Fourth ACM International Conference on Web Search and Data Mining, 2011, pp. 815-824.
- [10] Turney, Peter D., Michael L. Littman. "Measuring praise and criticism: Inference of semantic orientation from association," ACM Transactions on Information Systems (TOIS) 21(4), 2003, pp. 315-346.
- [11] Mihalcea Rada, Carmen Banea, Janyce Wiebe. "Learning multilingual subjective language via cross-lingual projections," In proceedings of the Association for Computational Linguistics (ACL). Vol. 45(1). 2007, pp. 976-983.
- [12] Pang, Bo, Lillian Lee, and Shivakumar Vaithyanathan. "Thumbs up? sentiment classification using machine learning techniques," In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), 2002, pp. 79-86.
- [13] Hu, Mingqing, and Bing Liu. "Mining and summarizing customer reviews," In Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2004, pp. 168-177.
- [14] Hatzivassiloglou, Vasileios, Janyce M. Wiebe. "Effects of adjective orientation and gradability on sentence subjectivity," In Proceedings of the 18th conference on Computational Linguistics (COLING'00), 2000, pp. 299-305.