Identifying Malicious Reviews Using NLP and Bayesian Technique on Ecommerce Historical Data

Sagar Ashokrao Mahajan

Government College of Engineering, Aurangabad, Dept. of Computer Science & Engineering, Aurangabad

---

Abstract: These days, online item audits assume an essential part in the buy choice of shoppers. A high extent of positive audits will bring significant deals development, while negative surveys will cause deals misfortune. Driven by the huge monetary benefits, numerous spammers attempt to advance their items or downgrade their rivals' items by posting phony and one-sided online surveys. By enlisting various records or delivering assignments in publicly supporting stages, numerous individual spammers could be coordinated as spammer gatherings to control the item audits together and can be additionally harming. Existing deals with spammer bunch discovery separate spammer bunch applicants from audit information and distinguish the genuine spammer bunches utilizing unaided spamicity positioning strategies. As a matter of fact, as per the past examination, marking few spammer bunches is simpler than one expects, nonetheless, hardly any techniques attempt to utilize this significant named information. In this paper, we propose a halfway administered learning model (PSGD) to distinguish spammer gatherings. By naming some spammer bunches as certain examples, PSGD applies positive unlabeled learning (PU-Learning) to examine a classifier as spammer bunch indicator from positive occasions (marked spammer gatherings) and unlabeled cases (unlabeled gatherings). In particular, we remove solid negative set regarding the positive occasions and the unmistakable highlights. By joining the positive examples, extricated negative occasions and unlabeled occurrences, we convert the PU-Learning issue into the notable semi supervised learning issue, and afterward utilize a Naive Bayesian model and an EM calculation to prepare a classifier for spammer bunch discovery. Examinations on genuine Amazon.cn informational index show that the proposed PSGD is viable and outflanks the cutting edge spammer bunch location techniques.

Keywords- Information search and retrieval, NLP, opinion mining, opinion feature.

I Introduction

Opinion mining can likewise be alluded to as estimation investigation whose objective is to break down individuals' opinions, mentalities, and feelings toward substances, functions, and their characteristics. An opinion assumes a significant part in dynamic. Slants or opinions explained in audits are analyzed at a scope of goals. This is material to people just as for associations moreover. At the point when a few associations need to investigate the opinions of clients about their items and administrations, they can direct overviews.

Clients post their perspectives and opinions on the seller's webpage or on their websites, discussions, and social destinations. In the present life, profoundly accessible CGM for example shopper produced media like message sheets, wikis, gatherings, online journals, and news stories board huge accommodation however they are liable for some presentation too. For the formation of some new creative open doors which are favorable to buyers, endeavors can examine purchaser created media to understand client's assessment about their items and administrations. At the point when certain issues are not settled quickly and effectively, obliviousness of such shopper created media can influence and produce hazards in brand picture and undertaking impact on the lookout, in light of the fact that the transmission speed of CGM data could reestablish obstinate consideration over the web. Ground-breaking diagnostic models are required which are favorable in the evaluation of customer conclusions.

Document level opinion mining distinguishes the general subjectivity or assessment communicated on an element in a survey report, yet it doesn't connect opinions with explicit parts of the substance. Purchasers of the item are perpetually discontent with the opinion rating of that item. People groups are more intrigued to know why it gets the rating, positive just as contrary ascribes that impacts on conclusive rating of item. In this way, it is basic to mine the exact opinionated features from text surveys and partner them to opinions. In opinion mining, an opinion features shows an element or a property of a substance on which clients expresses their opinions.

Opinion mining incorporates opinion include whose errand is to determine an element or a quality of a substance on which purchasers express their perspectives and opinions. Proposed framework perceives such features from unstructured literary audits. In opinion mining, a lot more methodologies have been now proposed which unique opinion features . To remove opinion include from surveys, regulated learning model work in given domain just yet the model must be retrained in the event that it is applied to various domains [1], [4].

Unsupervised learning approaches incorporate regular language handling (NLP) which utilizes domain-free syntactic formats or laws. These layouts and rules are utilized to catch the reliance jobs and nearby setting of the element terms. In any case, these standards are not material to genuine surveys since they need legitimate plan. Rules which are not in legitimate structure can't function admirably on informal genuine surveys. Point displaying approaches can extricate coarse-grained and nonexclusive subjects, which are really semantic element groups of the exact features remarked on explicitly in reviews [3].

A spammer bunch comprises of a bunch of commentators who co-audits a bunch of normal items. Accordingly, the assessment mining procedure could be used utilizing NLP to extricate the gatherings [12], [13]. Be that as it may, since numerous clients might be fortuitously assembled due to the comparative interest, the gatherings removed by FIM are just the spammer bunch applicants and should be additionally checked to distinguish the genuine spammer gatherings. Consequently, the recognition of spammer bunches for the most part contains two stages: (I) Discover spammer bunch competitors, (ii) Identify the genuine spammer bunches from the applicants.

II Literature Review

Opinion mining, which is likewise alluded as notion investigation, incorporates advancement of a framework which ready to amass and characterize opinions of a buyers about an item. Mechanized opinion mining regularly utilizes AI, a sort of man-made brainpower (AI), for the reason to dig text for slant. Data accessible in text configuration can be classified into realities and opinions. A reality speaks to the target explanations about substances and functions where as opinions represent abstract proclamations. Opinions emulate individuals' feelings about the elements and functions. Opinions gave by buyers in text audits are inspected from archive, sentences remembered for that report and word and expressions remembered for that record [11].

Objective of such sort of report level (sentence-level) opinion mining is to arrange the general subjectivity or notion communicated in an individual audit archive (sentence). Assessment of writings at the record or the sentence level does represents the opinions of clients, for example, different preferences. A positive archive doesn't speak to the all sure opinions of customers on features of specific article. Also, a negative report doesn't represent all negative opinions of clients on features of specific item [7]. Text record which incorporates assessments holds both positive and negative parts of specific article or element as per client's perspectives.

For the most part, generally assessment on the article may contain some sure viewpoints and some negative perspectives. Solid examination of features level is needed to discover total viewpoints about item or element.

For this reason three significant errands are as per the following:

1) Identifying object features

2) Determining opinion directions

3) Grouping equivalent words Identify object features search out for intermittent things and thing phrases as features , which are generally true features.

Existing data extraction techniques which are appropriate for recognizing object highlights are as restrictive irregular fields (CRF), shrouded Markov models (HMM). Determining opinion directions close whether the opinions given by buyer on the highlights of article or element are positive, negative or impartial. Existing vocabulary based methodology utilizes opinion words and expressions in a sentence to choose the direction of an opinion on a component. One article highlights can be communicated with various words or expressions, gathering equivalents task gathering's equivalent words together.

To compute sentence subjectivity Hatzivassiloglou and Wiebe [16] presents supervised grouping strategy to figure sentence subjectivity. Hatzivassiloglou and Wiebe proposed the general impacts of dynamic modifiers, semantically situated descriptors, and gradable descriptors on anticipating subjectivity of the content report holding audits.

Ache and Lee [11] proposed a sentence-level subjectivity finder for the reason to discover the sentences in a record as either emotional or objective. This strategy holds abstract sentences and disposes of the goal sentences. After then they applied opinion classifier. Errand of assumption classifier is to digest come about subjectivity with improved outcomes.

To order whole film surveys into positive or negative slants, Pang et al. [14] presented AI model as guileless Bayes, greatest entropy, and backing vector machines. They finish up results created by standard AI techniques are better than result by human-produced baselines. In any case, AI strategy performs well on just customary subject based order and need usefulness on assessment arrangement. An unsupervised learning strategy was proposed to order audit records into positive or negative in which as approval spoke to inspiration of report and disapproval speaks to antagonism of archive [8]. Past work indicated that customary assessment examination approaches can be very successful.

To mechanize the investigation of estimation materials, various methodologies were utilized for the forecast for the notions of words, articulations and furthermore archives [7], which incorporate Natural Language Processing (NLP) and example based [8]–[11], AI calculations, for instance NB, ME, SVM [12], and unsupervised learning [13].

Kim and Hovy [14] first created an equivalent arrangement of competitor words with obscure feelings.

Govindarajan [15] proposed a strategy for slant examination on eatery audits utilizing half and half order innovation. While most analysts center around AI based opinion investigation, others center around extremity vocabularies based strategies [16].

Kamps et al. [17] decided word assessment direction in the wake of figuring their semantic separation with their benchmarks in the WordNet [18] equivalent structure graph.

Wang et al. [19] first examined the characters about the assumption phrases in the NTUSD extremity word bank to get their polarities and qualities dependent on their characters. Cambria [20] received human-PC communication, data recovery and multi-modular sign handling advances to extricate individuals' assumptions among the ever-developing on the web social information base. Since every one of the above investigations had restricted inclusion [21] and deficiencies in forecast, we should think about semantic fluffiness when building slant vocabulary. This paper proposed another methodology, for example Multi-Strategy estimation examination dependent on semantic fluffiness, which is a blend of AI and opinion vocabularies based methodology.

Normal assessment directions of expressions and words are determined of each audit record to envision estimation of survey report. To register conclusions of expressions in audit record, domain-subordinate relevant data is utilized however this method has impediment as it relies upon outside web crawler. Zhang et al. [6] presented a standard based semantic investigation method to arranged notions for text surveys. Word reliance structures are utilized to arrange the supposition of a sentence.

Zhang et al. anticipated record level assessments by totaling notions of sentence. This procedure has restriction as rule-based techniques experience helpless presentation as they don't hold completeness in their guidelines.

To stay away from this, Maas et al. [15] introduced technique for both report level and sentence-level slant grouping. This proposed technique utilizes blend of unsupervised and supervised ways to deal with learn vectors. For learning measure, they catch semantic term-report data just as rich conclusion content. It is fundamental to take note of that opinion mining of the report, sentence, or expression (word) level doesn't figure out what precisely individuals loved and hated in audits. It neglects to consolidate the recognized suppositions and comparable highlights remarked on in the audits. Obviously, a removed opinion without the comparing features (opinionated objective) is of restricted an incentive in all actuality [2].

Opinion Feature Extraction Opinion features extraction is a subproblem of opinion mining. Existing methods of opinion include extraction can be classified into two classifications as, supervised and unsupervised. To check highlights or parts of watched elements, supervised learning consolidates concealed Markov models and contingent arbitrary fields. This is otherwise called a joint auxiliary labeling issue. Despite the fact that supervised models perform well on given domain, they required broad retraining when utilized in a few domains. To utilize supervised models in various domains, move learning measure is required.

Unsupervised Natural language Processing NLP techniques use mining of syntactic examples of highlights to extract opinion highlights. Unsupervised methodologies decide syntactic relations between include terms and opinion words in sentences. To decide relations unsupervised methodologies utilize created syntactic guidelines or semantic job marking [10]. This connection helps to find highlights related with opinion words just as mine huge number of invalid highlights of online surveys. With the end goal of extraction of successive itemsets .

Hu and Liu [12] presented an affiliation rule mining (ARM) method which depends on recurrence of itemsets. Regular itemset comprises of potential opinion highlights, which are things and thing phrases with high sentence-level recurrence. Yet, this strategy has limitations as: 1) successive yet invalid highlights are separated erroneously, and 2) uncommon yet legitimate highlights might be disregarded. Su et al. [8] proposed a common support grouping (MRC) procedure to handle include based opinion mining issues. Common support bunching techniques are utilized to mine the relations between features classes and opinion word gatherings. Extraction measure relies upon a co event weight lattice produced from the given survey corpus. MRC additionally ready to separate inconsistent highlights if the shared connections among features and opinion bunches found through the grouping stage is precise. MRC's exactness is low as it experiences difficulties in acquiring great groups on genuine surveys.

Latent Dirichlet portion (LDA) [13] approaches have been used to settle perspective based opinion mining undertakings. These LDA's are developed for extraction of dormant subjects which may not be opinion highlights communicated explicitly in surveys. Regardless of whether these methodologies are useful in determining of fundamental structures of survey information, they might be less productive in determining exact element terms remarked in audits. In our proposed framework, we group some syntactic reliance arrangements to mine up-and-comer highlights, as in NLP approaches and we use the SPAM AND NON SPAM FEATURES strategies to classify the favored domain-explicit opinion highlights. The vital differentiation of SPAM AND NON SPAM FEATURES contrasted with existing techniques lies in its shrewd combination of domain-ward and domain free data sources.

III Proposed System

NLP consolidates Syntactic assessment and Semantic examination. Feature that discontinuously appears in the given review domain, and routinely appears outside the domain, for instance, in a domain-self-governing corpus known as Explicit features. Subsequently, domain-explicit opinion features will imply even more on and on in the domain corpus of studies when stood out from a domain-independent corpus.

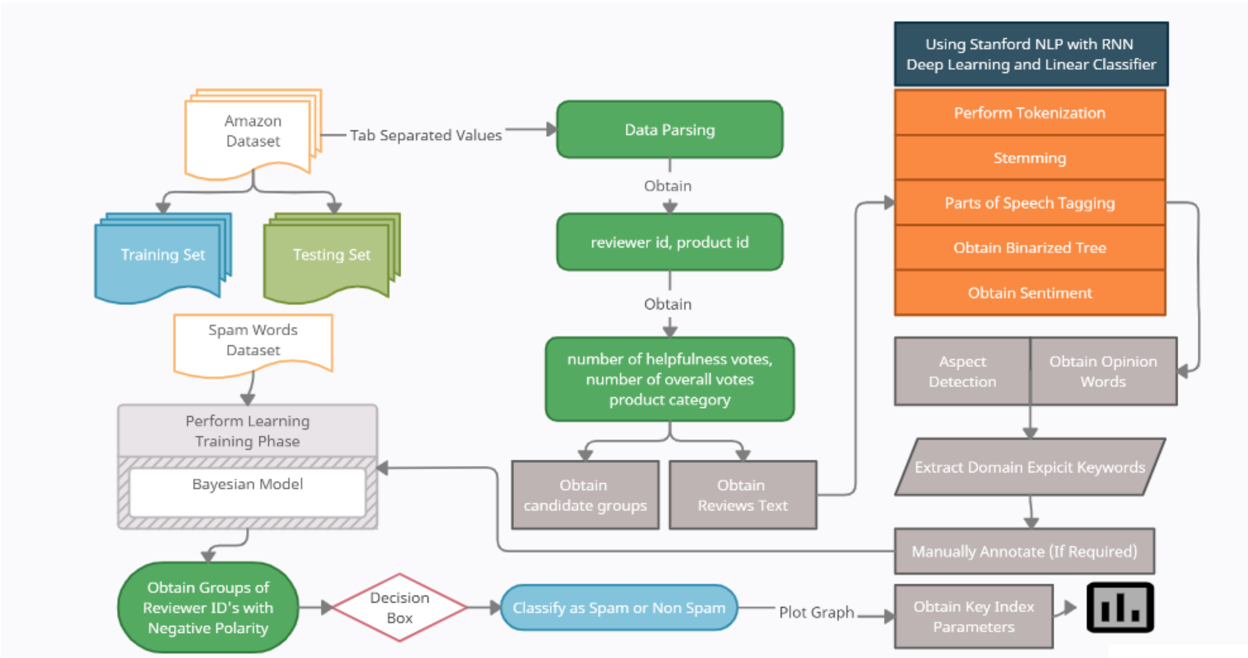


Figure 1 Proposed Architecture Flow

Fig. 1 portrays the progression of proposed technique. By utilizing physically expressed syntactic principles, from the audit corpus we first mine top notch of up-and-comer highlights. Later we process SPAM FEATURES and NON SPAM FEATURES for each preoccupied applicant highlight. SPAM FEATURES portrays the factual relationship of the possibility to the given domain corpus and EDR, which reproduce the measurable significance of the contender to the domainindependent corpus. Competitor highlights with SPAM FEATURES increases more noteworthy than a predefined Implicit significance edge and NON SPAM FEATURES scores not exactly Explicit pertinence limit are affirmed as legitimate opinion highlights.

A. Candidate Feature Extraction

Opinion highlights are comprised of things or thing phrases, by and large which are develop as the subject or object of an audit sentence. On account of reliance language structure, the subject opinion features has a syntactic relationship of type subject action word (SBV) with the sentence predicate. The article opinion include has a reliance relationship of verbobject (VOB) on the predicate. Moreover, it additionally has a reliance relationship of relational word object (POB) on the prepositional word in the sentence. From the previously mentioned reliance relations, i.e., SBV, VOB, and POB, we present three syntactic guidelines as follows:

Table 1.0 Syntactic Rules (Heuristic Rules)

Rules	Interpretation
NN+SBV	Identify NN as CF, If NN has a SBV dependency relation
NN+VOB	Identify NN as CF, If NN has a VOB dependency relation
NN+POB	Identify NN as CF, If NN has a POB dependency relation

Component of the technique of competitor features extraction is as per the following: 1) First of all, to perceive the syntactic association of each sentence in the given audit corpus, reliance parsing (DP) is utilized. 2) Later, the three principles in depicted in Table 1 are applied to the perceived reliance structures, and when a standard is terminated, the relating things or thing phrases are mined as set of up-and-comer highlights. Proposed competitor include extraction technique is language dependent.

B. Opinion Feature Identification

Domain Relevance Domain importance depicts the manner by which a term is identified with a specific domain dependent on two kinds of measurements as scattering and deviation. Scattering checks the number of quantities of times a term is alluded across records by figuring the distributional importance of that term across various archives in the total domain which is commonly known as level centrality. Deviation reflects how consistently a term is uncover in a specific record by ascertaining its distributional essentialness in the archive which is commonly known as vertical centrality. Scattering and deviation are registered by using the recurrence converse record recurrence (TF-IDF) term loads. Proposed domain reliance measure has two distinct perspectives as:

• Actually, domain reliance was acquainted with track news functions by choosing theme words from function words verbalized in the reports. Proposed work doesn't recognize point and functions As another option, proposed strategy use the proposed domain significance as a measure to characterize opinion highlights from unstructured content audits.

• As proposed framework doesn't separate among subject and function words. Domain significance recipe is modified for evaluating between corpus measurements uniqueness; especially, it is tuned to bind the distributional incongruities of opinion includes crossways two domain.

C. Implicit and Explicit Domain Relevance

Implicit Domain Relevance represents domain importance of specific opinion includes determined on given domain specific survey corpus. SPAM FEATURES duplicate the exact substance of the component to the domain audit corpus.

Explicit-domain importance is estimated by domain significance of specific opinion features on given domain independent survey corpus. NON SPAM FEATURES shows the factual relationship of the element to the domain-free corpus. As applicant terms are identified with possibly one corpus or other. They never identified with both simultaneously. In such case, NON SPAM FEATURES likewise shows the immateriality of an element to the given domain audit corpus. There some generally normal terms that are used all over the place and furthermore in an audit corpus as highlights.

IV Performance Analysis

The experiments were conducted using the customer reviews using Amazon dataset.

A system based on the proposed techniques has been implemented in Java using Stanford NLP classifier. The proposed system evaluated from three perspectives:

- The effectiveness of feature extraction.

- The effectiveness of opinion sentence extraction.
- The accuracy of orientation prediction of opinion sentences.

For each sentence in a review, if it shows user's opinions, all the features on which the reviewer has expressed his/her opinion are tagged. Whether the opinion is positive or negative (i.e., the orientation) is also identified. If the user gives no opinion in a sentence, the sentence is not tagged as we are only interested in sentences with opinions in this work. For each product, we produced a feature list. All the results generated by our system are compared with the manually tagged results. Tagging is fairly straightforward for both product features and opinions. A minor complication regarding feature tagging is that features can be extrinsic or intrinsic in a sentence. Most features appear in opinion sentences, e.g., *pictures* in *"the pictures are absolutely amazing".* Both extrinsic and intrinsic features are easy to identify by the human tagger.

Another issue is that judging opinions in reviews can be somewhat subjective. It is usually easy to judge whether an opinion is positive or negative if a sentence clearly expresses an opinion. However, deciding whether a sentence offers an opinion or not can be debatable.

Table 1.0 Confusion Matrix

	Not predicted as aspect	Predicted as aspect
Wrong aspects	TN	FP
Correct aspects	FN	TP

Tables given below shows the TP, TN, FP, FN, Precision, Recall, F-score for Amazon Dataset. Based on the results generated of confusion matrices we computed precision, recall l. Table contains recall and precision of frequent feature generation for each product, which uses association mining.

The precision is improved significantly by this pruning. The recall stays steady. The recall level almost does not change. The results from clearly demonstrate the effectiveness of these two pruning techniques. The precision drops a few percents on average.

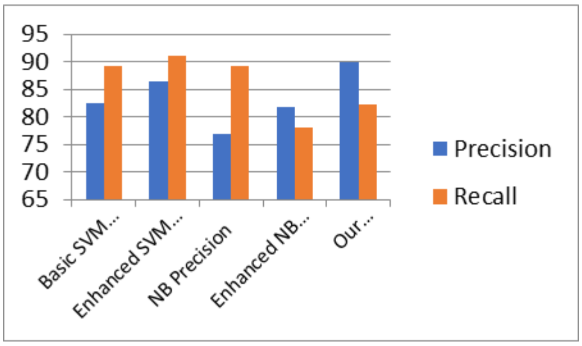
The reviews represent typical user generated content; they are all written by customers/users instead of being authored by any professional editor. Texts exhibit a rather informal style they often lack correct English grammar, exhibit the use of slang words, and contain an above average amount of misspellings. Web crawls are mono lingual, all crawled documents are written in English.

Table 1.0 NLP Classification Results

Table Data	Precision	Recall	F-score
Test Data	97.51	100	98.74

Table 2.0 Naïve Bayes

Table Data	Precision	Recall	F-score
Apparel	82.79	99.57	90.45



Graphical Result and Comparison Chart

When precision, recall, and F-measure are applied to aspect term occurrences, TP is the number of aspect term occurrences tagged (each term occurrence) both by the method being evaluated, FP is the number of aspect term occurrences tagged by the method and FN is the number of aspect term occurrences tagged by the method. The three measures are then defined as above. They now assign more importance to frequently occurring distinct aspect terms, but they can produce misleadingly high scores when only a few, but very frequent distinct aspect terms are handled correctly. Furthermore, the occurrence-based definitions do not take into account that missing several aspect term occurrences or wrongly tagging expressions as aspect term occurrences may not actually matter, as long as the most frequent distinct aspect terms can be correctly reported.

When precision, recall, and F-measure are computed over aspect term occurrences, all three scores appear to be very high, mostly because the system performs well with the occurrences of 'design', which is a very frequent aspect term, even though it performs poorly with all the other aspect terms. This also does not seem right. In the case of distinct aspect terms, precision and recall do not consider term frequencies at all, and in the case of aspect term occurrences the two measures are too sensitive to high-frequency terms.

V Conclusion

Proposed framework set up approach for opinion features extraction which used SPAM AND NON SPAM FEATURES include sifting standard. SPAM AND NON SPAM FEATURES features sifting measure utilizes the differences in distributional qualities of highlights across two corpora, out of which one is domain-explicit and another is domain-autonomous. SPAM AND NON SPAM FEATURES perceive up-and-comer highlights as domain-explicit and domain-free. Proposed IDER technique prompts detectable improvement in both element extraction execution and features based opinion mining results when contrasted with existing IDR, NON SPAM FEATURES LDA, ARM, MRC, and DP. We figure the effect of corpus size and subject choice on include extraction execution. To figure this, great nature of domain autonomous corpus is fundamental. We see that using a domain-autonomous corpus with comparable size as yet topically unique in relation to the given survey domain will create predominant opinion features extraction results.

REFERENCES

1) ZHIANG WU et al.: Detecting Spammer Groups From Product Reviews: A Partially Supervised Learning Model 10.1109/ACCESS.2018.2820025
2) B. Liu, "Sentiment Analysis and Opinion Mining," Synthesis Lectures on Human Language Technologies, vol. 5, no. 1, pp. 1-167, May 2012.
3) Y. Jo and A.H. Oh, "Aspect and Sentiment Unification Model for Online Review Analysis," Proc. Fourth ACM Int'l Conf. Web Search and Data Mining, pp. 815-824, 2011.
4) N. Jakob and I. Gurevych, "Extracting Opinion Targets in a Singleand Cross-Domain Setting with Conditional Random Fields," Proc. Conf. Empirical Methods in Natural Language Processing, pp. 1035- 1045, 2010.
5) F. Li, C. Han, M. Huang, X. Zhu, Y.-J. Xia, S. Zhang, and H. Yu, "Structure-Aware Review Mining and Summarization," Proc. 23rd Int'l Conf. Computational Linguistics, pp. 653-661, 2010.
6) C. Zhang, D. Zeng, J. Li, F.-Y. Wang, and W. Zuo, "Sentiment Analysis of Chinese Documents: From Sentence to Document Level," J. Am. Soc. Information Science and Technology, vol. 60, no. 12, pp. 2474-2487, Dec. 2009.
7) G. Qiu, C. Wang, J. Bu, K. Liu, and C. Chen, "Incorporate the Syntactic Knowledge in Opinion Mining in User-Generated Content," Proc. WWW 2008 Workshop NLP Challenges in the Information Explosion Era, 2008.
8) Q. Su, X. Xu, H. Guo, Z. Guo, X. Wu, X. Zhang, B. Swen, and Z. Su, "Hidden Sentiment Association in Chinese Web Opinion Mining," Proc. 17th Int'l Conf. World Wide Web, pp. 959-968, 2008.

9) R. Mcdonald, K. Hannan, T. Neylon, M. Wells, and J. Reynar, "Structured Models for Fine-to-Coarse Sentiment Analysis," Proc. 45th Ann. Meeting of the Assoc. of Computational Linguistics, pp. 432- 439, 2007.
10) S.-M. Kim and E. Hovy, "Extracting Opinions, Opinion Holders, and Topics Expressed in Online News Media Text," Proc. ACL/COLING Workshop Sentiment and Subjectivity in Text, 2006.
11) B. Pang and L. Lee, "A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts," Proc. 42nd Ann. Meeting on Assoc. for Computational Linguistics, 2004.
12) M. Hu and B. Liu, "Mining and Summarizing Customer Re- views," Proc. 10th ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining, pp. 168-177, 2004.
13) D.M. Blei, A.Y. Ng, and M.I. Jordan, "Latent Dirichlet Allocation," J. Machine Learning Research, vol. 3, pp. 993-1022, Mar. 2003.
14) B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up?: Sentiment Classification Using Machine Learning Techniques," Proc. Conf. Empirical Methods in Natural Language Processing, pp. 79-86, 2002.
15) P.D. Turney, "Thumbs Up or Thumbs Down?: Semantic Orientation Applied to Unsupervised Classification of Reviews," Proc. 40th Ann. Meeting on Assoc. for Computational Linguistics, pp. 417- 424, 2002.
16) V. Hatzivassiloglou and J.M. Wiebe, "Effects of Adjective Orientation and Gradability on Sentence Subjectivity," Proc. 18th Conf. Computational Linguistics, pp. 299-305, 2000.
17) J. Kamps, ``Using wordnet to measure semantic orientation of adjectives,'' in Proc. Int. Conf. Lang. Resour. Eval., 2004, pp. 1115_1118. [Online]. Available: http://www.baidu.com
18) D. Lin, ``WordNet: An electronic lexical database,'' Comput. Linguistics, vol. 25, no. 2, pp. 292_296, 1999.
19) M. Wang and H. Shi, ``Research on sentiment analysis technology and polarity computation of sentiment words,'' in Proc. IEEE Int. Conf. Prog. Informat. Comput. (PIC), vol. 1. Dec. 2010, pp. 331_334.
20) E. Cambria, ``Affective computing and sentiment analysis,'' IEEE Intell. Syst., vol. 31, no. 2, pp. 102_107, Mar./Apr. 2016.
21) F. Zhou, R. J. Jiao, and J. S. Linsey, ``Latent customer needs elicitation by use case analogical reasoning from sentiment analysis of online product reviews,'' J. Mech. Des., vol. 137, no. 7, p. 071401,2015.
22) R. Li, S. Shi, H. Huang, C. Su, and T. Wang, ``A method of polarity computation of chinese sentiment words based on Gaussian distribution,'' in Proc. 15th Comput. Linguistics Intell. Text Process. (CICLing), vol. 8404.