International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-7 Issue-5, June 2018

Analysis on Automatic Opinion Extraction from Product Reviews

Sint Sint Aung, Myat Su Wai

Abstract: ***Opinion mining also known as sentiment analysis is the computational study of subjective information towards different entities. Entities usually refer to products, organizations, services or/and their features, functions, components and attributes. Opinion mining is a major task of Natural Language Processing (NLP) that studies methods for identifying and extracting opinions from written text, such as product reviews, discussion groups, forums and blogs. Natural Language Processing techniques and lexicon-based approaches for opinion mining are used to extract aspects and customer opinions. Extracting opinion words and product features is an important task in many sentiment analysis applications. Opinion lexicon also plays a very important role because it is very useful for a wide range of tasks. Although there are several opinion lexicons available, it is hard to maintain a universal opinion lexicon to cover all domains. So, it is necessary to expand a known opinion lexicon that is useful for some domains. The aim of this system is to automatically expand opinion lexicon and to extract product features based on the dependency relations. Stanford Core NLP dependency parser is used to identify the dependency relations between features and opinions. Extraction rules are predefined according to these dependency relations. This work proposed an algorithm based on double Propagation to extract feature and opinions. The polarity orientation is annotated by using Vader lexicon. Unlike the existing approaches, this system contributes verbs opinions and verb product features. In order to increase the precision and recall, the system also proposes additional patterns besides 8 rules in Double Propagation. And, general words that are not features and adjectives that are not opinions are filtered in the proposed system. According to experimental studies, our approach is better than the existing state of the art approach.***

Index Terms: Opinion Mining, Opinions, Aspects

I. INTRODUCTION

Opinion mining also known as sentiment analysis is the computational study of subjective information towards different entities. Entities usually refer to products, organizations, services or/and their features, functions, components and attributes. Opinion mining is a major task of Natural Language Processing (NLP) that studies methods for identifying and extracting opinions from written text, such as product reviews, discussion groups, forums and blogs. It makes the Web an extensive and excellent source of information to gather opinions about a specific object. With the undeniable growth of the Web, individuals and organizations are using online content for their buying and manufacturing decision-making.

Every time someone attempts to discover what other people think about something on the Web, the response is an enormous amount of data, which makes it difficult to find useful information easily. For organizations, tracking customer feedback can help to measure the level of satisfaction and make optimal manufacturing and selling decisions. Due to human mental and physical limitations, it is difficult to manually gather and analyze the massive amount of information on the Web. Therefore, a system that can automatically summaries documents is increasingly desirable. Such a system extracts relevant information and presents it in a manner that is easy to read and understand in order to make informed decisions.

This work focuses on analyzing a large corpus of online reviews about products. Opinion lexicon expansion and features extraction tasks are performed simultaneously based on propagation using the relations. Firstly, word tokenization, part-of speech tagging and syntax or dependency parsing are done to process the sentences of the input datasets. Stanford ford Core NLP dependency parser is used to identify dependency relations. To process the propagation, the system only requires a seed opinion lexicon. The idea of the propagation approach is first to extract opinion words and features using the seed opinion lexicon and then use the newly extracted opinion words and features for new features and opinion words extraction.

The propagation ends until no more opinion words and features can be identified. In this way, even if the seed opinion lexicon is small, features can still be extracted with high recall and at the same time the opinion lexicon is also expanded. Incorrect features are removed by using general word set obtained from WordNet and NLTK. Finally, input sentences are classified their polarity orientations such as positive or negative by using Vader lexicon. By this way, opinion lexicon is expanded automatically [20] [28].

Feature-based opinion mining from customer reviews is a challenging problem for opinion mining and sentiment analysis. This research contributes to methods to identify and extract product features and sentiment from customer reviews by employing natural language processing (NLP) in unsupervised learning techniques. The contributions of the proposed system are as follows;

- This system contributes an algorithm to extract opinions and product features simultaneously and iteratively.
- This system also considers verb features, verb opinions and context dependent opinion words by using only dependency relations.

Manuscript published on 30 June 2018.
***** Correspondence Author (s)
Sint Sint Aung, Department of Academic Affairs, University of Computer Studies, Mandalay (UCSM), Mandalay, Myanmar, E-mail: ssaung@gamil.com
Myat Su Wai, Web Mining Lab, University of Computer Studies, Mandalay (UCSM), Mandalay, Myanmar, E-mail: missmyatsuwai@gmail.com

The rest of this paper is organized as follows. Section 2 describes about some related works of this work. Section 3 explains about the research method, Section 4 displays the experimental results and analysis. Finally, section 5 concludes this work.

II. RELATED WORKS

Many researches have been done about opinion word extraction. In general, the existing work can be categorized as corpora-based and dictionary-based approaches. Our work falls into the corpora-based approaches. According to Banitaan et al., (2010), Binali et al., (2009), and Glance et al., (2004), there are different categories of entities [11]. A broad overview organizes them into four entity categories that represent different types of words in a review text. These four categories are components, functions, features, and opinions. Some entities may not fit in any category. Therefore, a fifth category, "other" is formed and left open for any suggested categories [9]

Qian Liu (2013) proposed a logic programming approach for aspect extraction. In their system, they implemented double propagation in Answer Set Programming using 8 ASP rules. The recall is low because correct aspects were pruned as incorrect features and they considered only direct relations. Moreover, their approach may miss some infrequent features because this method extracted frequent noun or noun phrases as product features [14]. Yahui Xi (2013) developed an approach for extracting Chinese product features from Chinese product re-views. The authors also emphasize only on product features not on opinions [17 [28]].

Then, graph propagation algorithms, such as presented in Esuli and Sebastiani, (2007), label propagation Rao and Ravichandran, (2009) or random walk Baccianella et al., (2010), are utilized to iteratively calculate the sentiment score of each item. Under this direction, parsing results, syntactic contexts or linguistic clues in thesaurus are mostly explored to calculate the similarity between items. Wiebe, (2000) utilize the dependency triples from an existing parser Lin, (1994) [9]. Qiu et al. (2009) adopt dependency relations between sentiment words and feature words.

Esuli and Sebastiani (2005) exploit the glosses information from Wordnet [7] [12]. The lexicon-based approach Hu and Liu, (2004), Kim and Hovy, (2004) Ding et al., (2008), and Taboada, et al., (2010) determines the sentiment or polarity of opinion via some function of opinion words in the document or the sentence. As discussed earlier, this method can result in low recall for our entity-level sentiment analysis [18].

Qian Liu (2013) proposed a logic programming approach for aspect extraction. In their system, they implemented double propagation in Answer Set Programming using 8 ASP rules. The recall is low because correct aspects were pruned as incorrect features and they considered only direct relations. Moreover, their approach may miss some infrequent features because this method extracted frequent noun or noun phrases as product features [14] [35] [39].

To construct the Twitter-specific sentiment lexicon, Mohammad et al. (2013) use pointwise mutual information (PMI) between each phrase and hashtag/emoticon seed words, such as #good, #bad, :) and :(. Chen et al. (2012) utilize the Urban Dictionary and extract the target-dependent sentiment expressions from Twitter. Unlike Mohammad et al. (2013) that only capture the relations between phrases and sentiment seeds, we exploit the semantic and sentimental connections between phrases through phrase embedding and propose a representation learning approach to build sentiment lexicon [31].

In fact, the Arabic language in social media presents several challenges for sentiment mining as detailed by El-Beltagy and Ali (2013). First, the unavailability of colloquial Arabic parsers makes the morphological analysis task harder. Moreover, there is no publicly available sentiment lexicon for Arabic. Entity name recognition and handling idiomatic Arabic expressions in different dialects are also additional challenges for Arabic sentiment mining. For more information on Arabic morphological complexity and dialectal variations, see Habash (2010 [42] [44].

As another attempt to create a lexicon-based approach for sentiment mining, Alhazmi et al. (2013) linked the Arabic WordNet to ESWN through the provided synset offset information. The efficiency of the lexicon for sentiment mining was not evaluated [43].

Using ESWN, Mukherjee et al., (2012) developed TwiSent which collects tweets and classifies them as positive, negative or objective. Besides detecting the sentiment of the tweet, TwiSent addresses four known problems for tweets: spams, structural anomalies, entity specifications and pragmatics [47]. Addressing these inputs improved sentiment classification by 10 % compared to other sentiment mining applications that this work is only limited to the English language.

Davidov et al., (2010) describe a technique that transforms hashtags and smileys in tweets into sentiments. The described process is divided into two parts: identifying sentiment expressions, and determining the polarity of the identified expressions. Each tweet is divided into 4 different groups: words, punctuation, n-grams, and patterns [13] [21] [29]. Then for each group a separate technique is applied to detect a positive or a negative sentiment. Although this approach analyzes hashtags and smileys which are multilingual, it is still mainly designed for the English language.

Last but not least, Aly & Atia, (2013) presents a LABR: Large Arabic Book Reviews dataset consisting of 63K book reviews with rating from 1 to 5. The authors present baseline approaches for performing sentiment mining and set benchmarks for future research and approaches in sentiment mining [36] [45].

Agarwal et al., (2015) employed dependency relations between words to extract features from text based on ConceptNet ontology. Afterwards they used a method called "mRMR", which works as a feature selection scheme to eliminate redundant information. [Somprasertsri and Lalitrojwong, (2010) presented a method that extracts opinions and product features considering the syntactic and semantic information and based on dependency relations and ontology knowledge [4] [15] 17] [29].

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

Zhao et al (2015) presented a new method called joint propagation and refinement for mining opinion words and targets. The authors used frequency based threshold to prune incorrect targets. So, targets that are not occurred frequently, i.e. infrequent features are removed in their system. Threshold need to be raised to improve the precision which will affect the recall [18].

III. METHOD

Automatic aspects and opinions extraction method was proposed to extract both opinion words and features simultaneously by exploiting certain syntactic relations of opinion words and features. Although it was originally designed to work with product reviews, a reimplementation and extension of it has been applied on Twitter data, forum discussions, and blog postings. It has also been successfully used to analyze Chinese online discussions. The method needs only an initial set of opinion word seeds as the input and no seed features are required. It is based on the observation that opinions almost always have targets, and there are natural relations connecting opinion words and targets in a sentence due to the fact that opinion words are used to modify targets [16] [20] [28].

Furthermore, it was found that opinion words have relations among themselves and so do targets. The opinion targets are usually features. Thus, opinion words can be recognized by identified features, and features can be identified by known opinion words. The extracted opinion words and features are utilized to identify new opinion words and new features, which are used again to extract more opinion words and features [29] [46].

This propagation process ends when no more opinion words or features can be found. As the process involves propagation through both opinion words and features, the method is called double propagation. Extraction rules are designed based on different relations between opinion words and features and also opinion words and features themselves. Dependency grammar was adopted to describe these relations [39] [44].

The algorithm uses only a simple type of dependencies called direct dependencies to model useful relations. A direct dependency indicates that one word depends on the other word without any additional words in their dependency path or they both depend on a third word directly. Some constraints are also imposed. Opinion words are considered to be adjectives and features nouns or noun phrases. Thus, the potential POS tags for opinion words are JJ (adjectives), JJR (comparative adjectives), and JJS (superlative adjectives), while those for features are NN (singular nouns) and NNS (plural nouns). The dependency relations describing relations be- tween opinion words and features include mod, pnmod, subj, s, obj, obj2, and desc, while the relations for opinion words and features themselves contain only the conjunction relation conj. We use OA-Rel to denote the relations between opinion words and features, OO-Rel between opinion words themselves, and AA-Rel between features. Each relation in OA-Rel, OO-Rel, or AA-Rel can be formulated as a triple <POS (wi), R, POS (wj)> where POS (wi) is the POS tag of word wi and R is the relation. The values of POS (wi) and R were listed above.

The extraction process uses a rule-based approach with the relations defined above. For example, in an opinion sentence "Canon G3 produces great pictures," the adjective "great" is parsed as directly depending on the noun "pictures" through mod, formulated as an OA-Rel <JJ, mod, NNS>. If we know "great" is an opinion word and are given the rule "a noun on which an opinion word directly depends through mod is taken as a feature," we can extract "pictures" as a feature. Similarly, if we know "pictures" is a feature, we can extract "great" as an opinion word using a similar rule. The propagation performs four subtasks:

1. Extracting features using opinion words
2. Extracting features using the extracted features
3. Extracting opinion words using the extracted features
4. Extracting opinion words using both the given and the extracted opinion words.

Table 1 Dependency Relations used In this System

Rule	Observation	Constraint	Output
R_{11}	O →O-Dep →F F→F-Dep→O	O∈{O} O-Dep ∈{DR} F-Dep ∈ {DR} POS(F)∈ {NN, VB}	F=Fature
R_{12}	O→O-Dep→H←O-Dep←F	O∈{O} O-Dep∈{DR} F-Dep∈{DR} POS(F)∈ {NN, VB}	F=Feature
R_{21}	Fi →Fi -Fep→Fj	Fj ∈{F} Fi-Dep=Fj-Dep POS(Fi)∈ {NN}	Fi=Feature
R_{22}	Fi →Fi -Dep→H←Fj Dep←Fj	Fj ∈{F} Fi-Dep=Fj-Dep POS (Fi) ∈{NN}	Fi=Feature
R_{31}	O → O-Dep → F F→F-Dep→O	F∈{F} O-Dep∈{DR} F-Dep∈{DR} POS(O) ∈{JJ}	O=Opinion
R_{32}	O→O-Dep→H←O-Dep←F	F∈{F} O-Dep∈{DR} F-Dep∈{DR} POS(O) ∈ {JJ, VB}	O=Opinion
R_{41}	Oi →Oi-Dep →Oj	Oj ∈{O} Oi-ep∈{CONJ} POS (Oj) ∈{JJ}	Oi=Opinion
R_{42}	Oi→Oi-Dep→H←Oj-Dep←Oj	Oi ∈{O} Oi-Dep=Oj-ep, POS (Oj) ∈{JJ}	Oi=Opinion

Column 1 is the rule ID, column 2 is the observed relation (line 1) and the constraints that it must satisfy (lines 2 – 4), column 3 is the output. OA-Rels are used for tasks (1) and (3), AA-Rels are used for task (2), and OO-Rels are used for task (4).

Four types of rules are defined, respectively, for these four subtasks and the details are given in Table 3.3. In the table, o (or a) stands for the output (or extracted) opinion word (or feature). {O} (or {A}) is the set of known opinion words (or the set of features) either given or extracted. H means any word.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

POS (O (or A)) and O (or A) - Dep stand for the POS tag and dependency relation of the word O (or A) respectively. {JJ} and {NN} are sets of POS tags of potential opinion words and features respectively. {JJ} contains JJ, JJR and JJS; {NN} contains NN and NNS. {MR} consists of dependency relations describing relations between opinion words and features (mod, pnmod, subj, s, obj, obj2, and desc). {CONJ} contains conj only. The arrows mean dependency. For example, O --> O-Dep --> A means O depends on A through a syntactic relation O-Dep. Specifically, we employ R1i to extract features (a) using opinion words (O), R2i to extract opinion words (o) using features (A), R3i to extract features (a) using extracted features (Ai) and R4i to extract opinion words (o) using known opinion words (Oi). Take R11 as an example. Given the opinion word O, the word with the POS tag NN and satisfying the relation O-Dep is extracted as a feature.

It should be noted that although this method only finds noun feature words and adjective opinion words, it can be extended to features and opinion words of other parts-of-speech by adding more dependency relations. It also has a method to join words to form feature phrases and a method to determine the opinion orientations of the extracted opinion words.

IV. EXPERIMENT AND ANALYSIS

The performance of the proposed system is evaluated by measuring the evaluation criteria as shown in the above section according to the extracted features and their corresponding opinions from customer reviews. For experiment, we use core i7 processor with 2.20 GHz speed (2 gen), 8GB RAM with 1333 MHz speed and 64-bit Ubuntu OS, and, the proposed system is implemented with python programming language (Py Charm IDE for python).

In this research, 10 reviews datasets are chosen to test the proposed system as resources for experiment. 8 product review datasets are collected from https://www.cs.uic.edu/~liub/FBS/sentiment-analyssas.

Table 2 Datasets Used in the Proposed System

Dataset	No of Sentences	No of Features	No of Opinions
Restaurant	1083	1193	463
Hotel	266	212	185
Router	245	304	207
Speaker	291	435	227
Computer	239	346	215
Ipod	161	293	214
Linksys Router	192	375	193
Nokia 6000	363	633	277
Norton	210	302	185
Diaper Champ	212	239	169

Table 2 shows the information of dataset according to their names, the number of sentences and the number of features Amog them, three datasets; computer, router and speaker are annotated by: Qian Liu, Bing Liu, 2015, School of Computer Science and Engineering, Southeast University, China and Department of Computer Science, University of Illinois at Chicago, USA. And, another 2 datasets; restaurant and hotel are obtained from SemEval research group. SemEval (Semantic Evaluation) is an ongoing series of evaluations of computational semantic analysis systems, organized under the umbrella of SIGLEX, the Special Interest Group on the Lexicon of the Association for Computational Linguistics (ACL). The rests are annotated and used in (Ding, Liu and Yu, WSDM-2008), which improves the lexicon-based method proposed in (Hu and Liu, KDD-2004).

The proposed approach and state of the art approach are evaluated on newly extracted opinion words without considering the seeds in the evaluation. The proposed system used two lexicons as the seed. 10 different seeds number such as 100, 200, 300, 400 and 500 words from NRC Affect Intensity lexicon. In this case, S means seed. The results are shown in Tables 3 to 7.

Table 3 Comparison of Newly Extracted Opinion Words with Seed 100

Dataset	Proposed Approach			State of the art Approach		
	P	R	F1	P	R	F1
Speaker	0.9052	0.7577	0.8249	0.9375	0.5286	0.6760
Router	0.7489	0.8357	0.7899	0.7187	0.5582	0.6284
Hotel	0.6758	0.5297	0.5939	0.6764	0.375	0.4825
Restaurant	0.7951	0.8484	0.8209	0.7847	0.6220	0.6939
Nokia 6600	0.7444	0.8519	0.7946	0.7068	0.5899	0.6431
Linksys Router	0.8105	0.8020	0.8062	0.7633	0.5181	0.6172
iPod	0.8855	0.8317	0.8578	0.8731	0.5467	0.6724
Diaper Champ	0.8300	0.7470	0.7863	0.7692	0.4705	0.5839
Computer	0.8232	0.8271	0.8251	0.8475	0.6495	0.7354
Norton	0.7621	0.8486	0.8030	0.7368	0.5326	0.6182

Precision, recall, and F-score are computed on the newly extracted opinion words. This is an important point because only the new extractions are meaningful. Using all the extracted words to compute precision and recall is not appropriate as they can include many words that are already in the seed list.

According to the experimental results of precision, the proposed approach gets 92% average highest precision in speaker dataset and 65% average lowest precision in ABSA Hotel dataset with larger seed number 3500.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

According to the experimental result of recall, the proposed approach gets 85% average highest recall in Restaurant, Nokia 6600 and Norton datasets and 51% average lowest recall in ABSA Hotel dataset.

According to the results of f1-score, the proposed approach gets 85% average highest f1-score in iPod dataset and 57% average lowest f1-score in ABSA Hotel dataset. One reason for dropping in precision, recall and f1-score of ABSA Hotel dataset is that the nature of this dataset is quite different with the rest of product review datasets used in the proposed system.

Table 4 Comparison of Newly Extracted Opinion Words with Seed 200

Dataset	Proposed Approach			State of the art Approach		
	P	R	F1	P	R	F1
Speaker	0.9139	0.7555	0.8272	0.9365	0.5244	0.6723
Router	0.7467	0.8382	0.7898	0.7232	0.5582	0.6301
Hotel	0.6690	0.5219	0.5864	0.6764	0.375	0.4825
Restaurant	0.7922	0.8474	0.8189	0.7823	0.6187	0.6909
Nokia 6600	0.7412	0.8498	0.7918	0.7056	0.5884	0.6417
Linksys Router	0.8095	0.8052	0.8073	0.7615	0.5156	0.6149
iPod	0.8855	0.8317	0.5783	0.8740	0.5488	0.6742
Diaper Champ	0.3006	0.7470	078637	0.7669	0.4702	0.5830
Computer	0.1775	0.8293	0.8235	0.8518	0.6478	0.736
Norton	0.7598	0.8469	0.8010	0.7368	0.5355	0.6202

Table 5 Comparison of Newly Extracted Opinion Words with Seed 300

Dataset	Proposed Approach			State of the art Approach		
	P	R	F1	P	R	F1
Speaker	0.9052	0.7610	0.8269	0.9370	0.5288	0.6761
Router	0.7489	0.8357	0.7899	0.7151	0.5566	0.6260
Hotel	0.6690	0.5248	0.5882	0.6796	0.3783	0.4861
Restaurant	0.7914	0.8468	0.8181	0.7829	0.6195	0.6917
Nokia 6600	0.7428	0.8509	0.7932	0.7056	0.5905	0.6429
Linksys Router	0.8085	0.8	804232	0.7596	0.5130	0.6125
iPod	0.8855	0.8317	0.8578	0.8740	0.5539	0.6781
Diaper Champ	0.83	0.7470	0.7863	0.7647	0.4642	0.5777
Computer	0.8186	0.8262	0.8224	0.4720	0.6476	0.7331
Norton	0.7586	0.8555	0.804	0.7368	0.5326	0.6182

Table 6 Comparison of Newly Extracted Opinion Words with Seed 400

Dataset	Proposed Approach			State of the art Approach		
	P	R	F1	P	R	F1
Speaker	0.9071	0.7511	0.8217	0.9435	0.5223	0.6724
Router	0.7330	0.8341	0.7916	0.7133	0.5517	0.6222
Hotel	0.6805	0.5326	0.5975	0.6732	0.3736	0.4805
Restaurant	0.7938	0.8480	0.8200	0.7817	0.6165	0.6894
Nokia 6600	0.7444	0.8550	0.7959	0.7056	0.5884	0.6417
Linksys Router	0.8095	0.8010	0.8052	0.7596	0.5157	0.6144
iPod	0.8826	0.8277	0.8543	0.8731	0.5467	067241
Diaper Champ	0.8278	0.7440	0.7836	0.7647	0.4670	0.5799
Computer	0.8215	0.8293	0.8254	0.8447	0.6476	0.7331
Norton	0.7574	0.8453	0.7989	0.7328	0.5303	0.6153

Table 7 Comparison of Newly Extracted Opinion Words with Seed 500

Dataset	Proposed Approach			State of the art Approach		
	P	R	F1	P	R	F1
Router	0.7433	0.8316	0.7850	0.725	0.5631	0.6338
Restaurant	0.7942	0.8502	0.8212	0.7787	0.6123	0.6855
Computer	0.8181	0.8300	0.8240	0.8466	0.6509	0.736
Speaker	0.9100	0.7577	0.8269	0.9370	0.5360	0.6819
iPod	0.8844	0.8301	0.8564	0.8721	0.5471	0.6724
Linksys Router	0.8074	0.7947	0.8010	0.7596	0.5157	0.6144
Nokia 6000	0.7420	0.8472	0.7911	0.7030	0.5897	0.6414
Norton	0.7538	0.8448	0.7967	0.7307	0.5337	0.6168
Hotel	0.6783	0.5271	0.5932	0.7	0.3804	0.4929
Diaper Champ	0.8289	0.7590	0.7924	0.7669	0.4730	0.5851

The comparative results of precision on newly extracted opinion words between the proposed approach and state of the art approach with different seed numbers and seed lexicons are described.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

The proposed approach outperforms in precision for all different seed numbers although the seed lexicons are changed. The proposed approach gets 79% average precision while states of the art approach 76% average precision. The average recall of the proposed approach and state of the art approach using two seed lexicons and different seed numbers. The proposed approach performs better than the state of the art approach for all seed numbers in both seed lexicons. The proposed approach has 78% average recall whereas state of the art approach gets only 52% average recall. This is because state of the art approach only considers adjective as opinion words. So, their recall is very low.

The average f1-score of the proposed approach and and state of the art approach using two seed lexicons and different seed numbers are described in figure. The proposed approach performs better than the state of the art approach for all seed numbers in both seed lexicons. The proposed approach has 78% average f1-score whereas, and state of the art approach gets only 61% average f1-score. Finally, we can conclude tha, the proposed approach has generally higher recall and less FNR rate than state of the art approach in all experiments.

V. CONCLUSION

Sentiment analysis also has a great potential as a sub-component technology for other systems. In this work, an unsupervised lexicon-based approach to opinion lexicon expansion and feature extraction is proposed. Features and opinions words are extracted simultaneously by using the proposed algorithm. The biggest advantage of the method is that it requires no additional resources except an initial seed opinion lexicon. According to experimental results, the proposed system works well in feature extraction, opinion extraction and polarity classification. As the future works, when the new version of Stanford Core NLP is released, more dependency relation between words will be handled and improve the performance of this work.

REFERENCES

1. http://sentiment.christopherpotts.net/lexicon, 27. 10. 2016.
2. http://mpqa.cs.pitt.edu/corpora/product_debates/18.9.2016.
3. Agarwal, B., Poria, S., Mittal, N., Gelbukh, A., and Hussain, A. “Concept-level sentiment analysis with dependency-based semantic parsing: A novel approach”. Cognitive Computation, pages 1–13, 2015.
4. Asghar M, Khan A, Ahmad S, Kundi F, “A Review of Feature Extraction in Sentiment Analysis”, Journal of Basic and Applied Scientific Research, 4(3): 181–186, 2014.
5. Broß, J. “Aspect-Oriented Sentiment Analysis of Customer Reviews Using Distant Supervision Techniques”. PhD thesis, Freie Universität Berlin, 2013.
6. Bing Liu, (2011) “Web Data Mining, Springer”, Second Edition, Department of Computer Science, University of Illinois, Chicago, USA.
7. Chen, L., Wang, W., Nagarajan, M., Wang, S., Sheth, A. P. “Extracting Diverse Sentiment Expressions with Target-Dependent Polarity from Twitter”, In the Proceedings of the Sixth International AAAI Conference on Weblogs and Social Media(ICWSM), 50–57, 2012.
8. Chinsha, T. and Joseph, S. “A syntactic approach for aspect-based opinion mining”. In Semantic Computing (ICSC), 2015 IEEE International Conference on, pages 24–31. IEEE.
9. De Marneffe, M.-C. and Manning, C. D. “Stanford typed dependencies manual. Report”, Technical report, Stanford University, 2016.
10. Fabbrizio G, Aker A, Gaizauskas R012) “Summarizing Online Reviews Using Aspect Rating Distributions and Language Modeling”, IEEE Intelligent Systems, 28(3): 28–37, 2012.
11. G. Qiu, B. Liu, J. Bu, and C. Chen, “Opinion word expansion and target extraction through double propagation,” Computational Linguistics 37, no.1, 2011, pp. 9–27.
12. Jiang, M. Yu, M. Zhou, X. Liu, and T. Zhao, “Target-dependent twitter sentiment classification,” In Proc. of the 49th Ann. Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1, pp. 151-160. Association for Computational Linguistics, 2011.
13. Khan, K., Baharudin, B., Khan, A., and Ullah, A. “Mining opinion components from unstructured reviews: A review”. Journal of King Saud University-Computer and Information Sciences, 26(3):258–275, 2014.
14. Kundi F, Ahmad S, Khan A, Asghar, “Detection and Scoring of Internet Slangs for Sentiment Analysis Using SentiWordNet”, Life Science Journal, 11(9):66–72, 2014.
15. Liu, B. “Web data mining: exploring hyperlinks, contents, and usage data”. Springer Science & Business Media, 2011.
16. Liu, B. “Sentiment analysis and opinion mining”. Synthesis Lectures on Human Language Technologies, 5(1):1–167, 2015.
17. Liu, B. “Sentiment analysis mining opinions, sentiments, and emotions”. 1:1–386, 2015.
18. Liu, L. Xu and J. Zhao, “Opinion target extraction using word-based translation model”, Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, pp. 1346-1356, 2012.
19. Moghaddam, A. S., “Aspect-based opinion mining in online reviews”. PhD thesis, Applied Sciences: School of Computing Science, 2014.
20. Myat Su Wai and Sint Sint Aung, “Simultaneous Opinion Lexicon Expansion and Product Feature Extraction”, Proceeding of 16th IEEE/ACIS International Conference on Computer and Information Science (ICIS 2017), ISBN: 978–1–5090–5506–7.
21. Neviarouskaya A, Aono M, “Sentiment Word Relations with Affect, Judgment, and Appreciation”, IEEE Transactions on Affective Computing, 4(4): 425–438, 2014.
22. Pontiki, M., Galanis, D., Papageorgiou, H., Manandhar, S., and Androutsopoulos, I., “Semeval-2015 Task 12: Aspect Based Sentiment Analysis”. In Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015), pages 486–495, Denver, Colorado. Association for Computational Linguistics.
23. Pontiki, M., Galanis, D., Pavlopoulos, J., Papageorgiou, H., Androutsopoulos, I., and Manandhar, S.. “Semeval-2014 Task 4: Aspect Based Sentiment Analysis”. In Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014), pages 27–35, Dublin, Ireland. Association for Computational Linguistics.
24. S. Poria, B. Agarwal, Gelbukh, A. Hussain, N. Howard, “Dependency-based semantic parsing for concept-level text analysis”, In Gelbukh, A. (ed.) Computational Linguistics and Intelligent Text Proceeding, Part I. LNCS, vol. 8403, pp. 113–127. Springer, Heidelberg, 2014.
25. S. Poria, E. Cambria, G. Winterstein, G. B. Huang, “Sentic patterns: Dependency-based rules for concept-level sentiment analysis,”. Knowledge-Based Systems, vol.69, pp. 45-63, 2014.
26. S. Poria, E. Cambria, L.W. Ku, C. Gui, A. Gelbukh, “A rule-based approach to aspect extraction from product reviews,” In Workshop Proc. of the 25th International Conference on Computational Linguistics, COLING’14, pp. 28-37, 2014.
27. Shariaty, S. and Moghaddam, S.. “Fine-grained opinion mining using conditional random fields”. In Data Mining Workshops (ICDMW), 2011 IEEE 11th International Conference on, pages 109–114. IEEE..
28. Sint Sint Aung and Myat Su Wai, Domain Independent Feature Extraction Using Rule Based Approach, Advances in Science, Technology and Engineering Systems Journal (ASTESJ), volume 3, issue 1, 2018.
29. Souza M, Vieira R, Busetti D, Chishman R, Alves I, “Construction of a Portuguese Opinion Lexicon from Multiple Resources”, in the proceedings of the 8th Brazilian Symposium in Information and Human Language Technology (STIL’ 2011), 59–66. Springer, pp.411-448, 2007.
30. Maria Pontiki et al. (2016) “Aspect Based Sentiment Analysis”, Association for Computational Linguistics (ACL), San Diego, California, June 16-17, 2016.
31. Saif M. Mohammad. arXiv preprint arXiv, Word Affect Intensities , April 2017.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

32. Xu, T. J. Zhao, D. Q. Zheng, S. Y. Wang, "Product features mining based on Conditional Random Fields model", Proceedings of the 2010 International Conference on Machine Learning and Cybernetics, pp. 3353-3357, 2010.
33. Xu, X., Cheng, X., Tan, S., Liu, Y., and Shen, H. "Aspect-level opinion mining of online reviews". China Communications, 10(3):25–41.
34. Y. Wu, Q. Zhang, X. Huang, and L. Wu, "Phrase dependency parsing for opinion mining,". In EMNLP'09, Volume 3-Volume 3, pp. 1533-1541. Association for Computational Linguistics, 2009.
35. Y. Zheng, L. Ye, G. Wu and X. Li, "Extracting product features from Chinese customer reviews," Proceedings of 3rd International Conference on Intelligent System and Knowledge Engineering, pp. 285-290, 2008.
36. Z. Hai, K. Chang, G. Cong, "One seed to find them all: mining opinion features via association", Proceedings of the 21st ACM international conference on Information and knowledge management, pp. 255-264, 2012
37. Zhai, Z., Liu, B., Xu, H., and Jia, P., "Clustering product features for opinion mining". In Proceedings of the fourth ACM international conference on Web search and data mining, pages 347–354. ACM.
38. Zhang D, Dong H, Yi J, Song L, "Opinion summarization of customer reviews", In proceedings of the International Conference on Automatic Control and Artificial Intelligence (ACAI 2012), 1476–1479.
39. Zhang, L. and Liu, B. "Aspect and entity extraction for opinion mining". In Data mining and knowledge discovery for big data, pages 1–40. Springer.
40. Zhang, Y. and Zhu, W. "Extracting implicit features in online customer reviews", In Proceedings of the 15th ACM international conference on Information and knowledge management, pages 43–50. ACM.
41. Guang Qiu, Bing Liu, Jiajun Bu and Chun Chen. "Opinion Word Expansion and Target Extraction through Double Propagation." Computational Linguistics, March 2011, Vol. 37, No. 1: 9.27.
42. Qian Liu, Zhiqiang Gao, Bing Liu and Yuanlin Zhang. A Logic Programming Approach to Aspect Extraction in Opinion Mining. Proceedings of IEEE/WIC/ACM International Confernece on Web Intelligence (WI-2013), 2013.
43. Qian Liu, Zhiqiang Gao, Bing Liu and Yuanlin Zhang. "Automated Rule Selection for Aspect Extraction in Opinion Mining." Proceedings of International Joint Conference on Artificial Intelligence (IJCAI-2015), July 25-31, 2015.
44. Yahui Xi (2013), "Extracting Product Features from Chinese Product Reviews", Journal of Multimedia, 2013. C.J. Hutto and Eric Gilbert, 2014, "VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Texts", Association for the Advancement of Artificial Intelligent AAAI, , 2014.
45. Zhao et al. (2014) "Joint Propagation and Refinement for Mining Opinion Words and Targets", IEEE Data Mining Workshop, 2014, pp.417-424.
46. Zhiyuan Chen, Arjun Mukherjee, and Bing Liu. Aspect Extraction with Automated Prior Knowledge Learning. Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (ACL 2014), June 22-27, Baltimore, USA.

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)