# Sentiment Analysis on Social Media Using Morphological Sentence Pattern Model

Youngsub Han
Department of Computer and Information Sciences
Towson University
Towson, USA
yhan3@students.towson.edu

Kwangmi Ko Kim
Department of Mass Communication
Towson University
Towson, USA
kkim@towson.edu

***Abstract*—Social media became popular than ever as people are willing to share their emotions and opinions or to participate in social networking. Accordingly, the understanding of social media usage became important. The sentiment analysis is emerged as one of useful methods to analyze emotional stats expressed in textual data including social media data. However, this method still presents some limitations, particularly with an accuracy issue. For example, our previous sentiment analysis used a probability model and needed to adopt human-coded train-sets to maintain an acceptable accuracy level (89%). To overcome and improve this weakness, we propose an automated sentiment analysis in this paper by using the morphological sentence pattern model. We found that this new approach presented in this paper allowed us to achieve a higher level of accuracy (91.2%). The movie reviews were used for this analysis from IMDb, Rotten Tomatoes, Metacritic, YouTube and Twitter.**

***Keywords— sentiment analysis; natural language processing, social media, morphological patterns, aspect based approach***

## I. Introduction

In recent years, social networking sites, such as Twitter, Facebook, Instagram and YouTube is one of the most popular online media for sharing information, emotions, news and promotions [1]. People express not only their personal feelings but also share their pleasant/unpleasant, or dissatisfied experiences with brands or companies. Also, those data provide relevant information for understanding trends, issues, and various aspects of human behavior. Furthermore, these data can be used for identifying influential opinion leaders [2]. Various industries have spared no efforts to build advantages on social media data, which helps to reach out to their target audiences and customers efficiently. Sentiment analysis aimed to observe and summarize people's opinions or emotional states through textual data. Despite the demands of sentiment analysis methods for analyzing social media data, fundamental challenges still remain mainly because user-generated data is unstructured, unlabeled, and "noisy." Also, how to match sentiment lexicon on a lexicon-based approach is one of the biggest challenges because that doesn't mean that all matched lexicon are meaningful [3, 4, 5]. In our previous research, we proposed a probability model-based sentiment analyzer [6]. It yielded a relatively higher accuracy level (89%) than other approaches, and it can be broadly used to analyze text-based data. However, there are some limitations on this model. The first one is that all words have polarity including articles ("a" and "the") and prepositions ("to" and "on"), which are meaningless words in a sentence. This caused inaccurate results. Table I shows examples of the problem.

TABLE I. Examples of limitation of probability model

| Tweet | Positivity (%) |
|---|---|
| Freddie Gray Not Alone: 1997 Baltimore Police Case Raised Same Issue.#ColorOfHeart WakeUpAmerica | Positive (100) |
| RT @samswey: #FreddieGray isn't an isolated incident. Baltimore police dept killed 5 people last year. All were black. Every. Last. One. | Positive (87) |

The second limitation is that a probability-based sentiment analysis required human-coded train-set to maintain a higher level of accuracy. This means that the train-set must be rebuild continuously because this method cannot analyze un-trained words such as new-words [6]. It also requires a new train-set for a new topic. Therefore, we propose an automated and aspect-based approach using a morphological sentence pattern model to minimize the limitations discussed earlier [7, 8, 9]. It is expected that this new approach would yield a higher level of accuracy than the previous one.

## II. Related Works

### A. Natural Language Processing

To analyze text based data, we used a natural language processing (NLP) tool, the "Stanford Core NLP" made by The Stanford Natural Language Processing Group [10]. This tool is well-known for text mining because this tool provides refined and sophisticated results based on English grammar. In this research, we used the base forms of words, the part of speech (POS), and the structure of sentence [10]. There are two main advantages to use this tool for our method. The first advantage is that our system requires refined and tokenized words because user generated online textual data

978-1-5090-5756-6/17/$31.00 ©2017 IEEE
SERA 2017, June 7-9, 2017, London, UK

has a lot of linguistic problems such as spacing errors, idioms, and jargons. The second advantage is that our system requires part of speech tags to extract morphological sentence patterns for sentiment analysis.

### *B. Sentiment Analysis using User's Tendency*

Guerra et al. proposed a sentiment analysis algorithm to measure the bias of social media users toward a topic [11]. They hypothesized that users tend to express their opinion multiple times and a user's bias tends to be more consistent over time as a basic nature of human behavior. Thus, the bias toward a topic of social media users was measured and the sentiment was analyzed by transferring user's biases into textual features. Kucuktunc et al. also proposed a method to analyze sentiment based on demographic characteristics of users, such as gender, age and education [12]. However, these methods cannot be broadly used because they required users' relationship data and previous messages that the users posted, which are not always available in social networks due to the privacy laws.

### *C. Sentiment Analysis using Graph Representation*

Speriosu et al. applied a label propagation (LPROP) approach based on graph representation to analyze the sentiment of messages in Twitter [13]. Their assumption was that each tweet written by a user is linked to other tweets written by the same user, and each user is influenced by tweets written by other users that he or she follows. They represented such a relationship using a graph where the features of the message, such as words, emoticon and authors, are inter-related to each other. Those features determine the positivity or negativity of the message in the graph. Also, the results were examined based on the accuracy of the LPROP approach with messages in four different topics. The accuracy of the proposed LPROP approach is the highest among other sentiment analysis approaches as its level ranges from 65.7% to 84.7%, depending on the topics. However, we think this accuracy could be further improved because its average accuracy is still 72.08%.

### *D.* Lexicon based Sentiment Analysis

Lexicon-based approach is a traditional sentiment analysis approach. O'Connor et al. analyzed political opinions using this approach [1]. They collected tweets related to political issues from 2008 to 2009. Then, they built a lexicon where each word was categorized as either positive or negative keywords based on OpinionFinder [14]. Once positive and negative keywords were counted for all the messages, each message is classified as positive or negative . As a result, the ratio of positive messages versus negative messages was compared with survey results, which showed a strong correlation (80%) between the results of the sentiment analysis and survey results. The finding suggested that the lexicon-based method can be used as a supplement for traditional survey. However, this lexicon-based approach also has a weakness in that a message including positive keywords does not necessarily yield a positive opinion all the time. For instance, a word "like" is categorized as a positive word in the lexicon, and hence, if a message includes the word "like," it is categorized as a positive message. However, , if the message includes the word "don't" right before "like," the actual opinion of message should be categorized as negative. Such limitation should be improved with lexicon-based approach.

*Probability Model based Sentiment Analysis:* Lee et al. proposed a sentiment analysis method using a probability model as we did with our previous research [6]. This method requires a train-set with human coders to build a sentiment lexicon that contains the list of words that appeared in the text messages. For example, when a word occurs in 7 positive messages out of 10 messages, the word would have 0.7 (70%, 7/10) probability. Then, it computes the positivity score of text messages in a test-set using the list of words in a message and sentiment lexicon. Each message is categorized as either positive or negative, depending on threshold value calculated using a train set. As we mentioned in Section I, this model also has limitations as all words can have positivity even though words are not meaningful. It causes over-analysis to extract opinion from the data. Also, the train-set must be updated frequently as online data and messages change.

*Aspect based Sentiment Analysis:* The aspect-based sentiment analysis is a lexicon-based approach because this approach uses the lexicon as a measurement. The major difference between the two is that the aspect-based analysis provides a more in-depth analysis because all results are categorized into each aspect. In this approach, an aspect seems an attribute of objects. For example, when an object is a mobile phone, its aspects can be "display," "size," "price," "camera," or "battery" to describe the mobile phone. Thus, expected results are paired with an aspect and an expression such as "display-clean", "price-good", or "camera-awesome" [7, 8, 9]. Therefore, we will adopt this approach for more in-depth analysis.

### *E. Morphological Sentence Pattern Model*

This model was developed for building aspect-based sentimental lexicon, which can be used for sentiment analysis in our previous research [15, 16]. In this model, the recognizer extracts which parts of speeches (POS) are surrounding aspects or expressions as shown in Fig. 1. This model also used the "Stanford Core NLP" to recognize patterns and extract lexicon [10]. To analyze multi-social media data including YouTube and Twitter, the system builds morphological sentence patterns for each media separately because people share their opinions and emotions differently depending on the sources [16]. It provides relatively higher accuracy than existing approaches without involving a human coding stage.

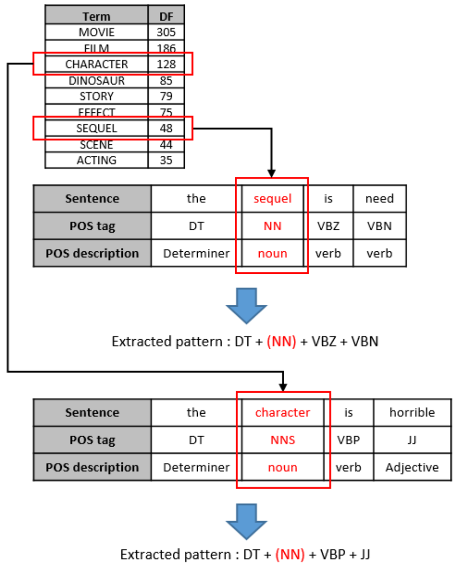


Fig. 1. Example of extracting morphological sentence patterns.

## III. METHODOLOGY

The system consists of four main phases; collector, aspects and expressions extractor, sentiment pattern extractor, and sentiment analyzer. In the first phase, the collector crawls movie reviews, tweets and YouTube comments. In the second phase, the aspect-expression extractor discovers aspects and expressions using the MSP model. In third phase, sentiment pattern extractor builds all candidate morphological sentence patterns for sentiment analysis. In the last phase, the sentiment analyzer matches the patterns and sentimental lexicon with the collected data. Fig. 2 shows the system architecture and flow.

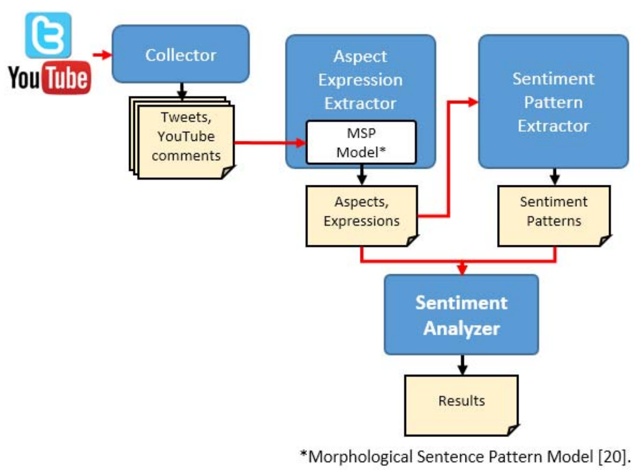


Fig. 2. System Architecture and flow

### *A. Data Collecting*

To collect movie reviews from IMDb, Rotten Tomatoes, and Metacritic, we used a collecting tool, which was developed by Y. Han et al. [15]. The tool collects movie reviews with their ratings generated by users. The system uses the "Jsoup HTML parser" which is an open-source Java library of methods and is designed to extract and manipulate data stored in a hypertext markup language (HTML)[1]. The collector automatically collects reviews using movie names as seeds such as "Jurassic World", "Avengers: Age of Ultron". Each site has different rating systems. Rotten Tomatoes let writers indicate their opinions whether "Fresh" as a positive, or "Rotten" as a negative. IMDb and the Metacritic have writers indicate their opinions on scale of 1 to 10 points. The bigger number means more positive evaluation. Thus, we decided the ratings of 8 to 10 as positive opinions and the ratings of 1 to 3 are as negative opinions to calculate positivity of expressions.

To collect tweets from Twitter, we used a Twitter-collecting tool, which was developed by Y. Han et al in our previous research [16]. The crawler retrieves tweets by keywords related to the target objects such as companies, products, politicians or movies. The crawler requests tweets with the keywords with application key provided from Twitter, then Twitter gives tweets including the keyword. All keywords are stored in our database with the object information. Twitter allows to search the recent 9 days of tweets by each keyword. With this, the crawler collected tweets repeatedly because Twitter allows 180 requests per key in 15 minutes, and a request includes up to 100 tweets.

To collect YouTube comments, we used a YouTube-collecting tool, which was developed by Lee et al [6]. YouTube provides APIs to collect data such as video information, user profiles, and user comments. The crawler collects comments posted on movies that are retrieved by keywords related to the target objects such as companies, products, politicians or movies. It also collects the data repeatedly within a scheduled time based on user requests.

### *B. Extracting Aspects and Expressions*

II. EXAMPLES OF MATCHED PATTERNS

| Rank | Aspect | | Expression | |
|---|---|---|---|---|
| | *Word* | *Count* | *Word* | *Count* |
| 1 | MOVIE | 397 | GOOD | 149 |
| 2 | FILM | 248 | GREAT | 100 |
| 3 | DINOSAUR | 185 | AMAZING | 47 |
| 4 | CHARACTER | 153 | BAD | 47 |
| 5 | PARK | 106 | PREDICTABLE | 43 |
| 6 | STORY | 106 | AWESOME | 37 |
| 7 | ACTING | 95 | MAKE | 35 |
| 8 | PLOT | 73 | OPEN | 34 |
| 9 | CGI | 65 | BETTER | 30 |
| 10 | PEOPLE | 50 | MORE | 28 |

Using MSP Model, the system extracted aspects and expressions because the sentiment pattern extractor uses aspects and expressions to build sentiment patterns [15, 16].

[1] Jonathan Hedley, Jsoup HTML parser, http://jsoup.org/

As we mentioned in section II, we used this model to analyze multi-source online data  Table II shows the examples of extracted aspects and expressions.

### A. *Extracting Sentiment Patterns*

TABLE II. EXAMPLES OF SNTIMENT PATTERNS

| | Pattern | | | Text |
|---|---|---|---|---|
| | *Aspect* | *Infix* | *Expression* | |
| Words | character | - | best | the /DT/ best /JJS/ character /NN/ |
| Pattern (POS) | NN | adjacent | JJS | |
| Words | chris pratt | is | fine | chris /NNP/ pratt /NNP/ is /VBP/ fine /JJ/ |
| Pattern (POS) | NNP+NNP | VBP | JJ | |
| Words | movie | is not | bad | this /DT/ movie /NN/ is /VBZ/ not /RB/ bad /JJ/ |
| Pattern (POS) | NN | VBZ+RB | JJ | |

The MSP Model [15] focuses on which part of speech exists between aspects and expressions. Table III shows a basic idea with examples. A pattern consists of three parts, which are aspect, expression and Infix. Our system matches aspects and expressions with sentimental lexicon when sentence contains the pattern. When matching the patterns, the system considers their sequence of POS(s) because it affects to their meaning [18]. Table IV shows examples of sentiment analysis results.

TABLE III. EXAMPLES OF SENTIMENT ANALYSIS RESULT

| Aspect : Expression | Sentiment | Correct | Text | Pattern |
|---|---|---|---|---|
| STORYLINE : GREAT | P | O | The action is intense, and the storyline is great. | (STORYLINE) /VBZ/ (GREAT ) |
| ACTION : INTENSE | P | O | The action is intense, and the storyline is great. | (ACTION) /VBZ/ (INTENSE) |
| ACTING : GREAT | P | O | Movie also has great acting. | (GREAT) (ACTING) |
| MOVIE : GREAT | P | O | Movie also has great acting. | (MOVIE) /RB/ /VBZ/ (GREAT) |
| TREVORRO W : GREAT | P | O | Colin trevorrow is a great director. | (TREVORROW ) /VBZ/ /DT/ (GREAT) |
| SEQUEL : BANAL | N | O | Spielberg and company found a way to destroy its flavor with two banal sequels. | (BANAL) (SEQUEL) |
| MOVIE : SUCK | N | O | You'll most likely see title's following the words, "this movie sucked" or "i'm gonna go on a rant here". | (MOVIE) (SUCK) |
| ANYTHING : TERRIBLY | N | X | To be honest there wasn't anything terribly new in this film, | (ANYTHING) (TERRIBLY) |
| EFFECT : SPECIAL | P | X | This movie has great special effects. | (SPECIAL) (EFFECT) |
| SPECIAL EFFECT : GREAT | P | O | This movie has great special effects. | (GREAT ) (SPECIAL EFFECT) |
| CHARACTER : LIKABLE | P | O | Sure chris pratt's character is the only likable person in the film but the dinosaurs are epic. | (CHARACTER) /VBZ/ /DT/ /JJ/ (LIKABLE) |
| ACTING : BELIEVABLE | P | O | Aside from a few overdramatic scenes by claire, the acting is really believable. | (ACTING) /VBZ/ /RB/ (BELIEVABLE) |

### B. *Sentiment Analysis*

To examine the patterns, we collected 200 documents for each data source. Then, we randomly selected 100 matched results for each data source and compared the results with a hand-coded answer-set labeled by two graduate students. At the initial testing, we just matched the patterns without any other processing. As shown in Table V, the accuracy of the sentiment analysis is about 86.5 % on average. This is reasonable compared with related approaches however we found 2 main problems, which are partial matching problems and mismatching problems. The portions of the partial matching problems are about 8% and the mismatch problems are about 7.5% on average. If we solve these problems, we can expect a higher accuracy. In the next section, we will discuss the problems in detail.

TABLE IV. SENTIMENT ANALYSIS RESULTS OF INITIAL TESTING

| | Movie Reviews | YouTube Comments | Twitter Tweets |
|---|---|---|---|
| **Accuracy** | 87.5% | 89% | 83% |
| **Partial Matching** | 8.5% | 5% | 10.5% |
| **Mismatch** | 11% | 1.5% | 10% |

### C. *Matching Problem Solving*

*Partial Matching Problem:* In Table IV, the ninth and tenth rows show an example of a partial matching problem. Our system can extract both 'EFFECT' and 'SPECIAL EFFECT' as aspects in a sentence because our system matches all possible patterns even though the 'EFFECT' is a part of 'SPECIAL EFFECT'. The portion of the problem is about 8% on average. To solve this problem, we set the system to extract the longest matched word(s) because "SPECIAL EFFECT" seems more meaningful as an aspect in the example sentence, "THIS MOVIE HAS GREAT SPECIAL EFFECTS."

*Mismatching Problem:* In Table IV, the third and fourth rows show an example of a mismatching problem. The analyzer paired an expression ‘GREAT’ with ‘ACTING’ and ‘MOVIE’ in a sentence because the system matches all possible patterns. The portion of the problem is about 7.5%. To solve this problem, we set the system matches to the nearest pair when this situation happens because the expression, “GREAT” targets “ACTING” in the example sentence, “MOVIE ALSO HAS GREAT ACTING.” Therefore the system should extract “ACTING” as an aspect and “GREAT” as an expression.

## IV. Experiments

To experiment our method, we collected 1,000 documents from movie review sites (IMDb, Rotten Tomatoes, and Metacritic), Twitter and YouTube related to a movie, ‘Jurassic world’. Then, the pattern extractor generates the sentimental sentence patterns using aspects and expressions, which are extracted from MSP model. From all extracted patterns, we selected 165 patterns for movie reviews, 113 patterns for tweets, and 102 patterns for YouTube comments when their frequency was 2 or greater.

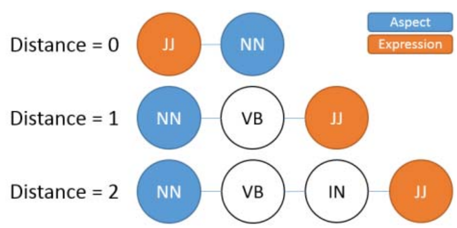


Fig. 3. Calculating Word Proximity between aspect and expression.

We calculated a distance of aspects and expressions (word proximities) from all extracted patterns as shown in Fig. 3. The average distance was 1.17 (movie reviews: 1.81, YouTube: 1.41, Twitter: 1.91). This means that most of the extracted patterns have one or two part of speech(s) between aspects and expressions on average.

102 patterns from YouTube

| YouTube | count | % |
|---|---|---|
| None(adjacent) | 121 | 36.1% |
| DT | 45 | 13.4% |
| VBD | 26 | 7.8% |
| IN | 11 | 3.3% |
| IN DT | 8 | 2.4% |
| VBZ | 8 | 2.4% |
| VBD RB | 6 | 1.8% |
| -LRB- | 4 | 1.2% |
| VBZ RB | 4 | 1.2% |
| '' | 3 | 0.9% |
| : | 3 | 0.9% |
| JJ | 2 | 0.6% |
| MD VB DT | 2 | 0.6% |
| PRP VBZ RB | 2 | 0.6% |
| RB | 2 | 0.6% |

165 patterns from movie

| Movie | count | % |
|---|---|---|
| None(adjacent) | 165 | 35.6% |
| DT | 21 | 4.5% |
| VBZ | 19 | 4.1% |
| IN | 16 | 3.5% |
| VBD | 11 | 2.4% |
| VBZ RB | 10 | 2.2% |
| IN DT | 9 | 1.9% |
| RB | 9 | 1.9% |
| VBZ DT | 8 | 1.7% |
| VBP | 6 | 1.3% |
| '' VBZ DT | 4 | 0.9% |
| NN | 4 | 0.9% |
| JJ | 3 | 0.7% |
| MD VB | 3 | 0.7% |
| POS | 3 | 0.7% |

113 patterns from tweets

| YouTube | count | % |
|---|---|---|
| None(adjacent) | 66 | 26.6% |
| NNP | 37 | 14.9% |
| JJ | 12 | 4.8% |
| RB | 10 | 4.0% |
| DT | 8 | 3.2% |
| VBZ | 6 | 2.4% |
| IN | 4 | 1.6% |
| NNP NNP | 4 | 1.6% |
| IN DT | 3 | 1.2% |
| TO VB | 3 | 1.2% |
| VBG | 3 | 1.2% |
| CD | 2 | 0.8% |
| DT JJ | 2 | 0.8% |
| DT NN | 2 | 0.8% |
| IN NNP NN VBZ | 2 | 0.8% |

Fig. 4. Example of Extracted Sentimental Sentence Patterns

Fig. 4 shows examples of extracted sentence patterns in order of their frequency. In this result, about 30% of aspects and expressions are adjacent (distance is 0) and one or two POS(s) (distance is 1 or 2) exists between aspects and expressions. This means most of the expressions affect to near aspects on all three sources through these experiments.

To examine the sentiment analyzer, we used a sentiment lexicon generated by Bing Liu [19], which contains 2,003 positive words and 4,782 negative words. We also added 102 positive words and 98 negative words extracted from the MSP model. One hundred two stop-words were further included, such as “HTTP”, “RT”, and “@”. Then, a total of 1,000 documents collected from each data source were analyzed. Then, we also randomly selected 200 sentences, which contain one or more lexicon for each data source and we compared the results with hand-coded answer-set labeled by two graduate students.

TABLE V. Results of Sentiment Analysis

| | Accuracy |
|---|---|
| **Movie Reviews** | 92.5% |
| **YouTube Comments** | 93% |
| **Twitter Tweets** | 88% |
| **Average** | **91.2%** |

As shown in Table VI, the accuracy of the sentiment analysis was at the level of about 91.2% with the proposed the MSP model, which improved the partial matching problem and the mismatching problem. The number is improved by 4.7% from the results of the initial testing from 86.5 %.

## V. Conclusion

TABLE VI. Comparison with exist methods

| Method | F-Score |
|---|---|
| PANAS-t | 0.737 |
| Emoticons | 0.948 |
| SASA | 0.754 |
| SenticNet | 0.810 |
| SentiWordNet | 0.789 |
| SentiStrength | 0.894 |
| Happiness Index | 0.821 |
| LIWC | 0.731 |
| Probability Model | 0.890 |
| **Our approach** | **0.912** |

In this research, we proposed a sentiment analysis method using the morphological sentence pattern model. This model improved the accuracy (91.2%) than what previous studies used as shown in Table VII. Through experiments, we found a characteristic that most of the meaningful aspects and expressions are adjacent or located within 1 to 2 distances in terms of part of speech. Also, our approach was useful in solving the partial matching problem and mismatching problem. The main advantage of our method was that no human-coded train-sets are required to achieve a higher level of accuracy. On the other hand, our approach has a limitation. The problem is that if the pattern is not detected, this model cannot analyze the data. Therefore, the future work might consider applying a convolutional neural network (CNN) algorithm in the sentiment analysis which is widely used in the artificial intelligence field. We expect that the CNN algorithm helps to solve the limitation in terms of the diversity of pattern matching.

# REFERENCES


[1] B. O'Connor, R. Balasubramanyan, B. R. Routledge and N. A. Smith, "From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series", Proceedings of the International AAAI Conference on Weblogs and Social Media, (2010), pp. 122-129.

[2] A. M. Kaplan and M. Heinlein. "Users of the world, unite! The challenges and opportunities of social media", Business Horizons, vol. 53, (2010), pp. 59-68.

[3] A. Sharma and S. Dey, "A comparative study of feature selection and machine learning techniques for sentiment analysis", Proceedings of the 2012 ACM Research in Applied Computation Symposium, (2012) October, pp. 1-7.

[4] P. Goncalves, M. Araújo, F. Benevenuto and M. Cha, "Comparing and combining sentiment analysis methods", Proceedings of the first ACM conference on online social networks, (2013) October, pp. 27-38.

[5] P. Melville, W. Gryc and R. D. Lawrence, "Sentiment analysis of blogs by combining lexical knowledge with text classification", Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining, (2009) June, pp. 1275-1284.

[6] H. Lee, Y. Han and K, Kim, "Sentiment Analysis on Online Social Network Using Probability Model", In Proceedings of the Sixth International Conference on Advances in Future Internet, (2014), pp. 14-19.

[7] J. Bross, H. Ehrig, "Automatic Construction of Domain and Aspect Specific Sentiment Lexicons for Customer Review Mining", CIKM'13, Oct. 27–Nov. 1, 2013, San Francisco, CA, USA, ISBN: 978-1-4503-2263-8

[8] T. T. Thet, J.-C. Na, and C. S. G. Khoo. "Aspect-based sentiment analysis of movie reviews on discussion boards", Journal of Information Science, 2010, 36:823-848.

[9] F. Wogenstein, J. Drescher, D. Reinel, S. Rill, J. Scheidt , "Evaluation of an Algorithm for Aspect-Based Opinion Mining Using a Lexicon-Based Approach", WISDOM '13, August 11 2013, Chicago, USA. ISBN : 978-1-4503-2332-1

[10] Christopher D. Manning, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven J. Bethard and David McClosky, "The Stanford CoreNLP Natural Language Processing Toolkit", Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations, June 2014, pp. 55-60.

[11] P. H. Guerra, A. Veloso, W. Meira and V. Almeida, "From Bias to Opinion: A Transfer-Learning Apporach to Real-Time Sentiment Analysis", Proceddings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining(KDD 11), August 2011, pp. 150-158, ISBN:978-7-4503-0813-7

[12] O. Kucuktunc, B. B. Cambazoglu, I. Weber and H. Ferhatosmanoglu, "A Larege Scale Sentiment Analysis for Yahoo! Answers", Proceedings of the fifth ACM international conference on Web search and data mining, February 2012, pp. 633-642, ISBN: 978-1-4503-0747-5

[13] M. Speriosu, N. Sudan, S. Upadhyay and J. Baldridge, "Twitter Polarity Classification with Label Propagation over Lexical Links and the Follower Graph", Proceedings of EMNLP 2011, Conference on Empirical Methods in Natural Language Processing, July 2011, pp. 53-64, ISBN: 978-1-937284-13-8

[14] T. Winson, P. Hoffman, S. Somasundaran, J. Kessler, J. Wiebe, Y, Choi, C. Cardie, E. Riloff and S. Patwardhan, "OpinionFinder:A System for Subjectivity Analysis", Proceedings of HLT/EMNLP 2005 Interactive Demonstrations, October

[15] Y. Han, Y. Kim, and I. Jang, A Method for Extracting Lexicon for Sentiment Analysis based on Morphological Sentence Patterns, Software Engineering Research, Management and Applications, Studies in Computational Intelligence, SCI, Springer, 654, 85-101. 2016.

[16] Y. Han, Y. Kim and J. Song, A Cross-Domain Analysis using Morphological Sentence Pattern Approach for Extracting Aspect-based Lexicon, International Journal of Multimedia and Ubiquitous Engineering (IJMUE), In-press, 2016.

[17] Y. Han, H, Lee and Y, Kim, "A Real-time Knowledge Extracting System from Social Big Data using Distributed Architecture", Proceedings of the 2015 Research in Adaptive and Convergent Systems, (2015) October, pp. 74-79.

[18] J. J. Song Word order, Cambridge University Press, 2012, ISBN:978-0-521-87214-0

[19] L. Bing, "Sentiment Analysis and Subjectivity." An chapter in Handbook of Natural Language Processing, Second Edition, (editors: N. Indurkhya and F. J. Damerau), 2010