

# Online Paper Review Analysis

Doaa Mohey El-Din

Information Systems Department  
Faculty of Computers and  
Information, CU  
Cairo, Egypt

Hoda M.O. Mokhtar

Information Systems Department  
Faculty of Computers and  
Information, CU  
Cairo, Egypt

Osama Ismael

Information Systems Department  
Faculty of Computers and  
Information, CU  
Cairo, Egypt

**Abstract**—Sentiment analysis or opinion mining is used to automate the detection of subjective information such as opinions, attitudes, emotions, and feelings. Hundreds of thousands care about scientific research and take a long time to select suitable papers for their research. Online reviews on papers are the essential source to help them. The reviews save reading time and save papers cost. This paper proposes a new technique to analyze online reviews. It is called sentiment analysis of online papers (SAOOP). SAOOP is a new technique used for enhancing bag-of-words model, improving the accuracy and performance. SAOOP is useful in increasing the understanding rate of review's sentences through higher language coverage cases. SAOOP introduces solutions for some sentiment analysis challenges and uses them to achieve higher accuracy. This paper also presents a measure of topic domain attributes, which provides a ranking of total judging on each text review for assessing and comparing results across different sentiment techniques for a given text review. Finally, showing the efficiency of the proposed approach by comparing the proposed technique with two sentiment analysis techniques. The comparison terms are based on measuring accuracy, performance and understanding rate of sentences.

**Keywords**—Sentiment analysis; Opinion Mining; Reviews; Text analysis; Bag of words; sentiment analysis challenges

## I. INTRODUCTION

World Wide Web (www) has become the most popular communication platforms to the public reviews, opinions, comments and sentiments about products, places, scientific books or papers and to daily text reviews. The number of active user bases and the size of their reviews created daily on online websites are massive. There are 2.4 billion active online users, who write and read online and Internet usage around the world [1]. Scientific research domain has a big world in journals and conferences, there are more than 4000 rated conferences and 5000 ranked journals [2]. Each one of them has thousand number of papers such as ACM, Springer and Science direct. Notably, a large fragment of WWW researchers makes their content public, allowing researchers, societies, universities, corporations to use and analyze data. According to a new survey conducted by Dimensional Research, April 2013: 90% of customer's decisions depends on Online Reviews [3]. According to 2013 Study [4]: 79% of customer's confidence is based on online personal recommendation reviews. As the result, a large number of studies and research have monitored the trending new research increasing year by year. In this work, trying to achieve trusted scientific reviews evaluation to be useful for researchers and facilitate the selection of the suitable papers.

Recently, several websites encourage researchers to express and exchange their views, suggestions and opinions related to scientific papers. Sentiment analysis [5] depends on two issues sentiment polarity and sentiment score. Sentiment polarity [6] is a binary value either positive or negative. On the other hand, sentiment score which relies on one of three models [7]. Those models are Bag-of-words model (BOW) [8], part of speech (POS) [9], and semantic relationships [10]. BOW [11] is the most popular for researchers and based on the representation of word but BOW neglects language grammar. POS [12] which is grammatically tagging especially verbs, adjectives and adverbs [13]. For example, (The research is not good.) declaring in (The/DT research/NN is/VBZ not/RB good/JJ. /.). In the example DT refers to "Determiner", NN refers to "Noun", singular or mass, VBZ refers to "Verb", RB refers to "Adverb", and JJ refers to "Adjective". But a semantic relationship method is the most complex method, which is based on the relationship between concepts or meanings for example antonym, synonym, homonym etc.

There is a research gap the sentiment analysis accuracy because of sentiment evaluation drawbacks and sentiment analysis challenges [14]. The evaluation sentiment drawbacks that Reflected in language coverage. This paper focuses on understanding text reviews and introduces solutions for some sentiment challenges. The sentiment analysis challenges summarized in ten challenges [15]. They are spam and fake reviews Detection, Limitation of classification filtering, Asymmetry in availability of opinion mining software Incorporation of opinion with implicit and behavior data, Incorporation of opinion with implicit and behavior data, and Natural language processing overheads (ambiguity), Generation of highly content lexicon database, handling of bipolar sentiments, dealing with short Sentence like abbreviations, Requirement of World Knowledge, Negation. All challenges have a bad effect on the understanding of reviews.

In this paper, the research aims to fill this research gap by proposing the new technique for sentiment analysis of online scientific papers reviews (SAOOP). The technique also measures efficiency by making a comparison between SAOOP, and other two sentiment analysis techniques [16]. Namely "Natural Language Toolkit-Text processing" (NLTK) and "recursive deep models for semantic" (NLPS). The results depend on comparing accuracy, performance and rate of coverage of language through two datasets.

The rest of this paper is organized as follows: Section 2 represents related works. Section 3, the presentation of the

new technique “SAOOP”. In Section 4, outlines of the Experiment as well as the sample used for comparison. Section 5 highlights the comparison results. Finally, Section 6 concludes and proposes directions for future work.

## II. RELATED WORKS

The purpose of this paper is sentiment evaluation which means to find the sentiment polarity (positive, negative, or neutral) of a text reviews data and evaluate the sentiment score of the text review. Generally a text review is divided into single sentences (“sentence-based”) and words (“words-based”) or very short texts from a single source.

### A. Sentiment Analysis: An Overview

The author in (Sentiment analysis of document based on annotation) presented a tool which judges the quality of text based on annotations on scientific papers [17]. The authors' methodology declares in collective's sentiment of annotations in two approaches. This methodology counts all the annotation produces the documents and calculates total sentiment scores. The problem of this methodology appears in a relationship between annotations that is complex. The technique needs to have a big query knowledge base containing metadata. The notion declares in that the values are not accurate enough such as the value of “Good=0.875” has greater value than the value of “Best=0.75” although the result is wrong in logical meaning. Nevertheless, believing that collecting metadata and evaluating them could be useful to achieve higher analysis quality.

The researchers proposed a “Web Based Opinion Mining system” for hotel reviews [18]. They introduced an evaluation system for online user's reviews and comments to support quality controls into hotel management. The research is capable of detecting and retrieving reviews on the web and deals with German reviews. The multi-topic/multi-polarity is the method of this research; the system would recognize the neutral e.g., “don't know” to “classify sentiment polarity that as neutral” and the multi-topic cases identified in their corpus. The major weakness illustrates in not handling some cases in multi-topic segments. The authors [19] analyzed sentiments reviews of mobile devices products. Their Machine learning (ML) [20] system investigates the classification accuracy of Naïve Bayes algorithm. In addition to Judge the product quality and status in the market is advantageous. They use three machine learning algorithms (Naïve Base Classifier, K-nearest neighbor [21], and random forest [22] to calculate the sentiments accuracy. The random forest improves the performance of the classifier.

### B. Sentiment Analysis Techniques

This section provides a brief description of the two sentiment analysis techniques investigated in this paper. These techniques are the most popular in the literature and they cover diverse techniques such as the use of Natural Language Processing (NLP) [23] in assigning polarity and sentiment score.

1) *Natural Language Toolkit*: The authors aim at an evaluation sentiment scores and polarity. They produce the Natural Language Toolkit (NLTK) [12]. NLTK is a text analysis technique that evaluates cognitive and constitutional components of a given text reviews based on using lexicon including words. They use hierarchal sentiment classification level with two levels (Neutral, Positive, and Negative). The drawback of this technique illustrated in low accuracy and some logical errors. Because the technique needs to increase handling of language coverage [24].

2) *NLP Stanford sentiment (NLPS)*: The researchers introduce recursive neural models have in common: word vector representations and classification [25]. The authors released a tool named “NLP Stanford” NLPS [26], which develops an integration of learning techniques that produces better results and higher accuracy training model empirically. Their goal is based on Semantic word spaces have been very beneficial but NLPS cannot express the meaning of longer phrases in a primary way. So they improve this technique by detection the sentiment requires wider supervised training and evaluation resources.

## III. SENTIMENT ANALYSIS OF ONLINE PAPERS (SAOOP)

In this section, Sentiment analysis of online papers “SAOOP” will be presented. SAOOP is used in opinion mining [27] and based on a new English lexical dictionary [28]. This lexical dictionary groups adjectives, nouns, verbs, adverbs, adjectives, prefixes, suffixes and other grammatical classes into synonym. The proposed technique is an enhancement on Bag-Of-Words (BOW) model [29] in sentiment analysis to achieve high accuracy, which depends on word weight replacing term frequency of each word. The proposed technique solves two important Bag-of-words weaknesses.

The standard bag of words is not automatic in classification and creating polarity lexicons because BOW model needs to create manual lists of 'positive' and 'negative' words [30]. That means the review judgment is based on the probability of positive or negative words. The second is low accuracy because the standard BOW model neglects text grammatically. Sentiment classification levels will be divided into five classes (very positive, positive, negative, very negative and Neutral).

The proposed technique makes the sentiment classification levels are more detailed and easy by word percentage of each class. The goal of SAOOP is for inferring the polarity of common meaning and polarity concepts from natural language text at a word level, rather than at the syntactic level. SAOOP also classifies reviews into some categorizations based on papers parameters. In addition, the estimation rank of each paper based on evaluation some parameters.

### A. SAOOP Overview

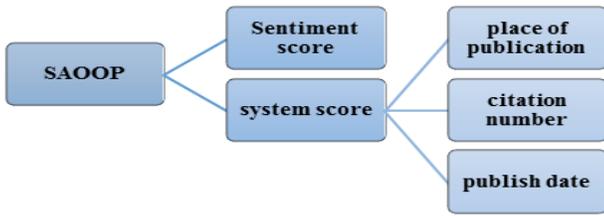


Fig. 1. SAOOP Overview

“Fig.1” shows that SAOOP model consists of two components sentiment score and system score. SAOOP can evaluate any paper based on the components. Sentiment score depends on total reviews evaluation score. And system score which depends on the sum of total scores of three parameters of paper (place of publication), citation number of paper and paper publishing date. SAOOP technique helps researchers to select the suitable paper with the total paper score.

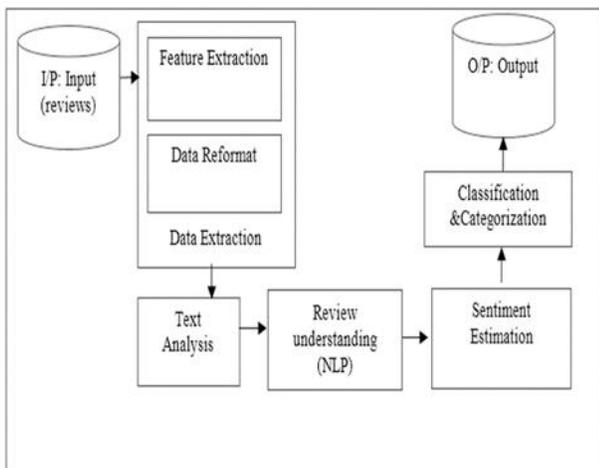


Fig. 2. SAOOP Technique overview

“Fig.2” declares SAOOP Technique overview. The input is scientific paper website link. In data extraction [31] level two parts: first, using Easy web extract tool which is web scraping tool to extract data of paper from scientific papers website online. Part two is data reformat from Excel sheet which is one output of EasWebExtract tool [32] suitable with SAOOP database format. In text analysis level, SAOOP applies some functions of text analysis on reviews of each paper. In the first, applying the splitting sentences function, tokenizing words function, and checking of stop list and removing them [33].

In review understanding (NLP) level, the proposed technique understands the sentences meaning and check words in vocabulary lexicon with similarity and differences algorithms. In estimation phase, showing the evaluation sentiment score for each word into text review and the polarity

detection for each one and each sentence and calculate the total score of sentiment review score. In classification phase, that’s splitting into two parts, first the reviews classification into five sentiment classification levels (very positive, positive, negative, very negative and objective (neutral), also having degrees of each sentiment level with scale from [-1, 0, 1]. There is also another classification which declares each review categorization based on five meaning classes (topic, date, author, citation, and place of publication). The benefit from the extracted data to memorize them and make relationships between evaluated papers and reviews and categorize reviews logically based on topic domain parameters. Output is the sentiment evaluation score of all reviews with all papers with caring of number of reviews parameter, and evaluation of scientific paper parameters score which is based on metadata of each paper (place of publication, publishing date, and number of citation). So the consequent result is ranking to each paper with the total score of sentiment and system scores with graphical reports of results.

### B. SAOOP Methodology

SAOOP can assign polarity based on this approach, considering the words weight replacing term frequency, by assuming each word has two values and polarity with this assumption equation,

$$V(w) = \sum(W(p) + W(n)) = 1. \tag{1}$$

V (w) is value of word, W (p) refers to positive value and W (n) refers to negative word, the selection between positive or negative polarity Influenced by the meaning of words and each other polarity. But the sentence contains negative that differs in the word value. If the word is positive, convert to negative polarity and the negative score will be as in the equation,

$$V(w) = W(p) - 0.2. \tag{2}$$

And if the word is negative, the score will be calculated by  $V(w) = W(n) + 0.2$ . The selection of 0.2 because this disison is suitable for the five sentiment class’s levels [18]. The proposed technique also creates papers ranks with calculating sentiment and measuring domain parameters. By assuming,

$$P(TS) = \sum_{R=0}^n (T(SA) + T(SS)). \tag{3}$$

In the equation, P (TS) refers to a total score of each paper, T (SA): is a total score of sentiment score of all reviews on each paper with caring of number of positive reviews. In the next equation,

$$(SA) = \sum_{i=1}^n \frac{P(SA(R))}{n}. \tag{4}$$

The calculation of the total score of all reviews depends on the score of each review. There is a difficult problem between large number of reviews and evaluating sentiment polarity of each one, this problem is improper the most review number having assessment higher score. For example, one paper publishing in 2013 that’s mean from 2 years and this paper has twenty reviews, not equal evaluation one paper publishing in 2005, that’s mean from 10 years and the second paper has twenty reviews. The first one is the top rated because the evaluation number of reviews in short time. In other example, one paper publishing from 2 years and having twenty negative

reviews, not equal evaluation other one publishing from 10 years and having positive twenty reviews. The second one has maximum rated because the evaluation numbers of positive reviews is larger than the one, although the second is the oldest. As mentioned before double trouble with reviews number and the relationship between date and other relation between sentiment polarity of reviews and number of reviews. That interprets difficulty of evaluation domain parameters.

The proposed technique faces these challenges and evaluates the percentage of positive reviews over total scores. But still there is a problem in relationship between date and number of reviews, for example: one paper publishing from 2 years which has twenty positive reviews, not equal evaluation other one publishing from 10 years which has positive twenty reviews. Actually that is not equal their selves because the recent has bigger reviews number. So SAOOP presents a solution for date relation with reviews number, according with two parameters number of positive reviews and the recent paper. T (SS): is a total score of system score parameters that are evaluated logically of paper parameters according to this equation,

$$V(SS) = \sum \left( \frac{S(pp)}{\lambda} + \frac{S(C)}{2\lambda} + \frac{S(D)}{2\lambda} \right). \quad (5)$$

V (SS) expresses the value of systems score. S (PP) means the score of publication place, S(C) refers to the score of paper citation number, and S (D) means the score of paper publish date. Assuming  $\lambda$  is a constant equal 2, dividing into  $\lambda$  and  $2\lambda$  to determine the priority of evaluation of the parameter. The evaluation topic parameters process does not ease because of depending on the logical meaning of each one. So the research focuses on scientific papers domain to put the foundation of evaluation parameters to achieve the fact value of each paper to support researcher with sentiment analysis by ranking papers based on total score of them. There is inverse relationship between publishing date and number of citation of the paper, which declare in this equation,

$$S(C) = \frac{\text{current year} - \text{publish date year}}{\text{number of Citation}}. \quad (6)$$

The result is not true the highest citation number having the highest evaluation score of it. For example, one paper publishing in 2013 that's mean from 2 years which has ten citations, not equal evaluation one paper publishing in 2005, that's mean from 10 years which has ten citations. The first one is the highest score because number of citations in shortly is high, this first paper will be predicated if the paper has the same time 10 years, it mostly has 50 reviews not 10 reviews such as the second paper. In other words, the first paper has 5 papers into each year but the second has 1 into each year. To evaluate score of publishing place conference which depends on ACM conferences tiers with a sample into computer science conferences, such as "VLDB: Very Large Data Bases is in the top tier: tier 1", "ER: Intl Conf. on Conceptual Modeling (Conf. on the entity Relationship Approach)" is in next tier which is in lower tier: Tier 2, and "IDEAS: Intl Database Engineering and Application Symposium" is in a lower tier: Tier 3" [34].

### C. SAOOP & Sentiment Challenges

SAOOP enables to make solutions to most significance sentiment analysis challenges [35]. The proposed technique

can produce some solutions for main challenges to reach to higher accuracy. The discussion of the solutions in the following:

#### 1) Topic domain independence

Domain-dependent [36] is a difficult challenge to recognize topic nature. There are some words have many meanings and different sentiment values relevant to the topic. There is also a problem shows in extracting keyword or features and how to evaluate words based on each topic. One feature set may give very good performance in one domain, at the same time it performs very poor in some other domain. The produced solution suitable with a small scale by applying the proposed technique on one topic domain and examine domain parameters evaluation by categorization reviews because they also give different meaning with the same word. This research presents a technique to recognize topic nature automatically. The proposed technique is based on extracting keywords and relevant features of each topic. In addition, to produce a solution for some words have many meanings and different sentiment values relevant to the topic. The proposed technique is based on Classification review of each domain features and keywords.

For example, "IEEE is [great +] publication for your paper", SAOOP can put IEEE is in a place of publication classification (based on feature name of publication) and the polarity is positive. "The publishing conference is [great+]", this review refers to the place of publication classification (based on keywords) and the polarity refers to positive. In other example, "The paper publishing date is [old-]", this reviews refers date classification (based on date attribute) and "Old" having the negative score. "The author is [old] in this field", but SAOOP can categorize the last review in author classification that is meaning the author is expert in this field so "Old" will be had positive score.

SAOOP improves the sentiment score to be more accurate and fair. By assuming some words have 0 value because of depending on classifications of each sentence of each review, there are some groups of words having a polarity and score to relate with the detected classification.

#### 2) Negation

Negation is the biggest challenge in sentiment analysis [37]. The new technique produces a solution to improve evaluation negative with the enhanced bag of words technique. This research handles the two techniques: explicitly and implicitly negative [38]. First: explicitly is deliberately formed and are easy to self-report and by keywords. Second implicitly [5] is the unconscious level, are involuntarily formed and are typically unknown to us without any keywords of negative. In addition, the handling the negative meaning of some conjunctions such as "not only", and "But". The dual negative is the most important case which cares to achieve the total sentiment polarity. Reverses polarity of mid-level terms: great V.S not great.

A method often followed in handling negation explicitly in sentences like:

"I do [not like+] ^ the paper", is to detect the negative polarity because the word (not) and convert the sentence

operator to negative. But this does not work for “I do [not like<sup>+</sup>] this research but I [like<sup>+</sup>] the field”. But still there can be problems.

Other example, “I find the functionality of this new methodology [less <sup>-</sup>] practical”, this review refers to explicit comparative negative. “This algorithm is [not great<sup>+</sup>]”, the proposed technique handles in this review the positive and negative evaluation which declares in [not great! = bad] but [not great = good]. Implicitly negative such as “This research is [very [complex <sup>-</sup>] ]” this example does not have any negative keyword, but the meaning has negative and the polarity will be negative of this sentence.

There are sentences having keywords of negation and they don't have the negative polarity such as “[Not only<sup>+</sup>] I [like<sup>-</sup>] this algorithm, but also [easy<sup>+</sup>] to understand and apply.” the polarity is not reversed after “not” due to the presence of “only”. So this type of combinations of “not” with other words like “only” has to be kept in mind while designing the algorithm.

There is a difference between “not only” and not because not only strengths the meaning (more positive or less negative) based on the polarity of the sentence. In this example other case of implicit negative, I [wish <sup>-</sup>] to work [harder <sup>-</sup>]. In the last review, the new technique presents future words e.g. wish refers to the negative polarity but first must check the polarity of the next sentence polarity because maybe changed the polarity depends on meaning.

### 3) Creation lexicon

The proposed SAOOP yields an improvement over prior published bag of words built lexicons. This technique also provides an improvement in calculation technique used in reviews sentiment analysis. SAOOP technique presents a solution to take care of grammar (which is one of limitation of Bag-Of-Words) and to save time took is N-gram algorithm to create subsequences of terms. There are two phases that will be produced:

- Phase 1. Data Preparation Phase

Less number of words in vocabulary lexicon to fast search based on similarity and differences algorithms. SAOOP neglects verbs tenses or word formula (singular or plural), that's meaning neglecting English grammar and syntax because of the comparison and differentiation with the infinitive verbs, and singular words with most letters similarity.

- Phase 2. Lexicon Development Phase

Evaluation words /terms: is based on enhanced bag of words: the proposed technique doesn't depend on term frequency. This phase is based on assuming each word has two values and the total of them equal 1. Each term has 2 polarities (+/-).

### Negative Algorithm

1. For each review R in paper P sets
2. For each sentence Sent in Review R
3. Apply Pre-Processing: Remove the stop words
4. Convert all words to Upper case
5. Check on expressions have “No or Not” e.g Not Only
6. Check on first Negation keywords list, which effects on the polarity of the words e.g “Not” I don't like this paper.
7. By assuming, the negative value for positive word ,  
 $V(w) = W(p) - 0.2$
8. And assume the positive value for the negative word,  
 $V(w) = W(N) + 0.2$
9. Check on the next word after negative e.g “like” has positive value and polarity but here it will take negative polarity and value.
10. Detect the polarity and value.
11. Check on the second list of negation keywords, which effects on the polarity of sentence e.g. never, yet, neither.
12. Convert polarity of the sentence by multiplication with(-1),  
 $V(Sent) = S(Sent) * -1$
13. Check on future words e.g “wish/hope”.
14. Check on the next sentence polarity.
15. End for
16. Detect sentence value and polarity.
17. End For
18. calculate review value and polarity  
(Note: knowing our attention of review classification.)

### 4) World knowledge requirement

SAOOP technique produces a solution for Knowledge about worlds' facts, events, people are often required to correctly classify the text. Trying to achieve higher accuracy and get the evaluation for some neutral reviews. The World knowledge challenge solution is based on the hierarchical database of nouns. Semantic (hierarchal) relationships between nouns to achieve the polarity, score and meaning. Also to differ between them and keywords or features. Consider the following example, “the author is a [lion<sup>-</sup>] in this field”, the previous review present negative polarity because lion is a name of animal but in real evaluation in the review refers to a positive polarity. In the next review, “Bing is really [Einstein<sup>+</sup>]” evaluation sentiment analysis without world knowledge classifies above sentence as neutral, but this review is an objective sentence because Einstein is the name of the famous scientist, so it refers a positive polarity also. This review is very hard for software to understand that automatically. SAOOP creates a huge lexicon database to contain the world knowledge especially related to researchers and the most common in the reviews. The solution of world knowledge also assumes values of the words based on the most common meaning. The evaluation of these world knowledge depends on keywords and classification of reviews.

### 5) Spam and Fake Reviews:

The WWW contains both realistic and spam contents. For effective Sentiment classification, this spam content should be eliminated before processing.

SAOOP can be done by empty or identifying duplicates, by detecting outliers and by considering the reviewer reputation. The proposed Technique enhances reviews spam and fake. SAOOP technique can avoid and cure the most of them by:

- Remove empty reviews: To calculate the real number of reviews.
- Delete duplicate reviews by considering the same reviewer: To Calculate the real sentiment score of the paper.

For example, one paper has 10 reviews, 3 of them for the same text review and with the same user, and 2 is empty reviews, in most sentiment application, if having 10 reviews number and the same repeated reviews will calculate together, the sentiment score is not real because having fake reviews and the results became fake also. And also there are some reviews are general are not related to the paper actually. SAOOP can produce solution for the case study on citeulike.com website, through making quaternary relationship between a set of paper parameters “paper name”, “author names”, “review” and “Username” (who is a review writer) with taking into consideration review written time, if the review is repeated by the same review writer with ensuring if the review is fake by all parameters and time, the proposed technique will delete the spam review before calculating the sentiment analysis. SAOOP can also deal with fake reviews if it empty and deleted.

In this paper, showing the implementation of SAOOP technique using C# programming language working on Microsoft visual studio 2010 platform. The newly created lexicon is based on SQL Server Management Studio 2008.

#### IV. EXPERIMENT

In this section, the discussion of the comparison between the proposed technique and two sentiment analysis techniques. This comparison shows the accuracy and performance results based on two datasets. This comparison also compares with the effects and solutions of sentiment analysis challenges.

##### A. Datasets

The comparison uses two different datasets: 1) real data set: which splits into two data sets with training set (1000 text reviews) and test set (5000 text reviews), 2) verified data set: which is a real set with unknown evaluation around 10.000 text reviews (including more than 5.000 positive words, 5.000 negative words).

##### 1) Real dataset

The first sample set is a sample of WWW.citeulike.com papers reviews and Metadata posted by computer science papers branch [39, 34]. The comparison in real data set in computer science scope including two parts: training data and test data [40]. Training data is a set of data to evaluate sentiment around 1000 reviews, knowing the values before.

The second part is a test data: which is a set of data to evaluate sentiment with hide class label around 5000 text reviews. Citeulike receives in excess of 200,000 distinct visits (defined by Google Analytics as a group of page views by a unique user with timeout after 30 minutes inactive) monthly, with each visit originating an average of 2.77 page views [41]. Of that 200,000 around distinct users who have previously visited the site on multiple occasions.

There are currently 505,402 items posted in the database (counting n people post the same article); 1,676,130 tags (counting n if there are 'n' tags applied to an article); and 130,548 distinct words used these numbers are growing exponentially. This sample set allows us to study the responses to noticeable past texts. In addition, to evaluate the improved levels of techniques, methodologies in sentiment analysis. SAOOP can handle ten cases to ease to understand text review accurately by CiteULike users they illustrated in table 1. SAOOP can care and evaluate of some English grammar to improve BOW model.

##### 2) Verified dataset

The second dataset which is called verified data set is a real data set but they can't be known the evaluation before. The dataset has around 10.000 text reviews in this sample. This data set is splitting into two parts of verified data reviews as positive and negative. These datasets include a wide range of online papers texts reviews: general reviews. In Table 2, the sample reviews of online scientific papers. SAOOP technique can evaluate sentiment score with the relationship of reviews categorization. With applying on this human-verified sample set [29], by fitting to quantify the range with different sentiment analysis techniques can accurately evaluate polarity of text reviews.

TABLE I. TABLE ENGLISH LANGUAGE COVERAGE HANDLING BY SAOOP

Cases	Definition & Examples
1. Expressions 2. Topic objects [3] 3. Negation	Based on syntax (e.g., Not Only, No one.) Features (e.g., lot of contributions) Implicit (e.g., Independent)) and explicit (e.g. not bad, does not very good) meaning
4. Suffixes & prefixes	The beginning or end letters of word to have different meaning (e.g., dislike, opposed to, useless) Converting verbs tenses into infinitive
5. Verbs 6. Adjective & Adverb 7. Nouns 8. Comparative [4] comparisons 9. Phrases 10. Some special (Need, Wish)	e.g., Well, improved, highly e.g., algorithms, improving, enhancing e.g., easier convert to → easy. (“More”; “higher”) and (“most”, “highest”). e.g., very good, the professional work (e.g., hope, wish): in the most times, they have negative polarity.

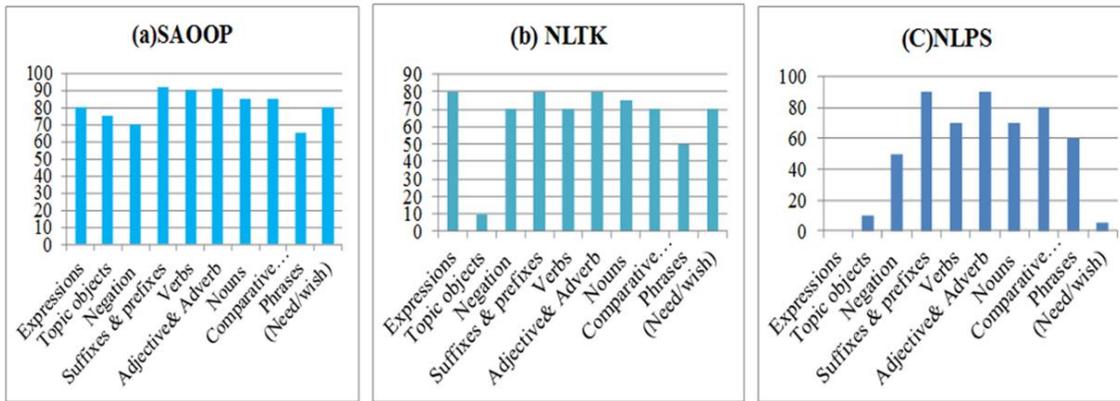


Fig. 3. Coverage Rate of ten cases understanding cases with the three techniques

TABLE II. SAMPLE OF REVIEWS

Sample of reviews
This paper is very well.
It's not great
The best web point
I am interesting in this field
Extremely good
It's not only hot research area but also having new scientific contributions.
This point research is more affected in web mining than using in neural network.
high accuracy
It's not have good value enough
Citation is valuable

B. Comparison Measures

In order to define the evaluation of accuracy and performance of the three techniques, which will consider in the following table.3:

TABLE III. MEASUREMENT TABLE

		Actual observation	
		Positive	Negative
Predicted expectation	Positive	x	y
	Negative	z	w

Let present True positive (x) was defined when a text was correctly classified as positive, False Positive (y) is a negative text which was classified as positive, False Negative (z) is a positive text but was classified as negative, and the last one True Negative (w) is a correctly classified as negative [42]. In order to compare and evaluate the techniques, by considering the following metrics, commonly used in information retrieval: true positive rate or recall:  $R = x/(x + z)$ , false positive rate or precision:  $P = x/(x + y)$ , accuracy:  $A = (x + w)/(x + y + z + w)$ , and F-measure (performance):  $F = 2 \cdot (P \cdot R)/(P + R)$ . In many cases simply use the F-measure, as it is a measure of a test's accuracy and relies on both the precision and recall [10]. By reporting, all the measurement mentioned above by practical interpretation. The true positive rate or recall can be understood as the rate at

which positive reviews are predicted to be positive (R), whereas the true negative rate is the rate at which negative reviews are predicted to be negative.

The accuracy represents the rate at which the method predicts results correctly (A). The precision also called the positive predictive rate, calculates how close the measured values are to each other (P). The comparison also provides the F-measure results, since it is a standard way of summarizing precision and recall (F). Ideally, a polarity identification method reaches the maximum value of the F-measure, which is 1, meaning that its polarity classification is perfect. The y-axis is a percentage of the understanding sentence rate.

V. COMPARISON RESULTS

In order to facilitate understanding the advantages, disadvantages, and limitations of the various sentiment analysis techniques [43]. This section also presents the comparison results among them.

Understanding of word coverage: in the beginning, the comparison of the coverage of English grammar cases across the representative scientific reviews from CiteULike website. Then examination the intersection of the covered reviews cases across the techniques were in table 1. "Fig. 3 (a)" shows the result for the proposed technique SAOOP, which explain in section 4. "Fig. 3 (b)" declares the NLTK technique. NLTK which is a teaching tool works in, computational linguistics using Python [44]. And "Fig.3 (c)" shows NLP technique. NLPS technique which is predicting the sentiment of reviews based on a recursive model.

As shown in the figure, SAOOP has the highest understanding sentence coverage with 82.5 % with two data sets with three data sets samples, respectively, followed by NLPS which can't evaluate the total sentiment score but with detecting word by word polarity its percentage is 72%.

NLTK can interpret less than 10% of all relevant reviews. In addition, we compare with the percentage of handling sentiment analysis challenges to high accuracy and performance of sentiment analysis of the three techniques of the text reviews depicted in "Fig. 3".

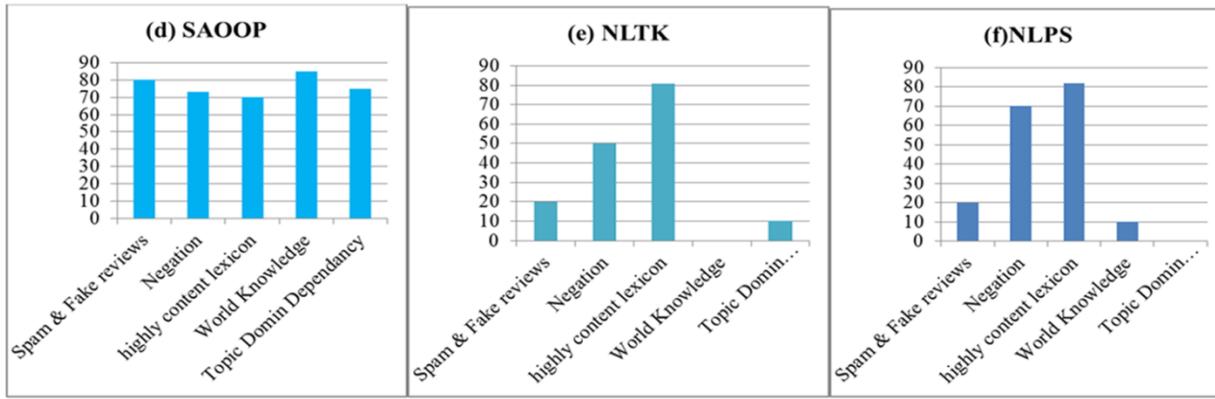


Fig. 4. Percentage of handling sentiment analysis with the three techniques

According to “Fig.4 (d)” in fact, SAOOP had a new solution for some sentiment challenges but NLPS and NLTK, they can’t produce methodology to solve them expect some cases in negation but they have many logical errors, that shown by “Fig.4 (E) and (F)”. The analysis results in table 3, refers to the: Percentage of accuracies between techniques based on different data set size. Also we examine the average result analysis of the two big data set that spirited into three data sets, that illustrate the highest average results with sentiment score of the proposed SAOOP technique then NLPS and the lowest one is a NLTK Technique. Finally, the summarization the results with the average of the three data sets (real and verified sets), we find the average of sentiment score of the proposed technique improve the results. Because of working binary analysis solutions of some important challenges and evaluate some technical cases in the text which have a problem in evaluation to be more accurate. In next section, we discuss the accuracy results of the comparison.

a) *Accuracy:* With the examination of the percentage degree of different techniques accuracy on text reviews content. In order to compute the accuracy of each technique, by calculating the intersections of the positive or negative proportion given by each technique. Table.4 presents the percentage of accuracy for the three compared techniques. For each technique in the first column, showing the estimation from the two data sets of reviews. Finding that some techniques have a high coefficient as in the case of SAOOP (82.5%), while others have least overlap such as NLTK (62%) and NLPS (70.2%).

The last “column” of the table shows on average to what extent each technique agrees with the other two samples. The last “row” quantifies how other methods agree with a certain technique, on average.

With the results of table 4, they illustrate differences between accuracy and performance of the three techniques. Table 4 shows techniques recall, precision, accuracy and performance.

“Fig.5” is shown the accuracy results of them. In a summary, the result indicates that existing tools vary widely in terms of accuracy about sentiment score, with scores ranging from 60% to 80%.

TABLE IV. AVERAGE RESULTS FOR ALL DATASETS

Metric	SAOOP	NLTK	NLPS
Recall	0.856	0.571	0.253
Precision	0.867	0.845	0.846
Accuracy	0.817	0.629	0.715
F-measure	0.846	0.665	0.729

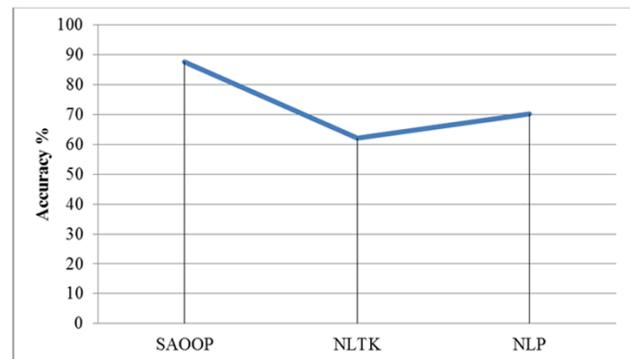


Fig. 5. Differences between Accuracies of three techniques

TABLE V. PERCENTAGE OF ACCURACY BETWEEN TECHNIQUES

Dataset	SAOOP	NLTK	NLPS	Average
Training set 1.000	83.5%	62%	72%	72%
Test set 5.000	81.99%	61%	70%	70.99%
Teal set 10.000	82.5%	60%	71.56%	71.186%
Average	82.5%	61.514%	71.604%	-

b) *Perfromance:* In this section, showing an evaluation of the performance of the three compared techniques. For comparing the performance results, Table.5 which gives the average of the results obtained for all datasets. For the F-measure, a score of 1 is ideal and 0 is the worst possible. The technique with the highest F-measure was faced sentiment analysis challenges and cover ten cases of each text review (0.846), which had the highest sentiment accurate and understanding text coverage. The second rated technique in the understanding of F-measure is NLPS, which obtained a much higher

coverage than understanding and challenges. It is important to note this problem that it can't be interpreted into of total score of the text review. For observation better performance on data sets that contain more expressed sentiment, such as text reviews (e.g., papers online) and the lowest performance is NLTK technique.

## VI. CONCLUSION

Sentiment analysis is the most important source in decision making. Almost people becomes depends on it to achieve the efficient product. Thousands of researchers rapidly year by year that focuses on scientific online reviews for papers to help them. So the researchers introduce a new sentiment technique. In this paper, the researchers create a new technique is called sentiment analysis of online papers "SAOOP". The proposed technique will be a suitable and efficient solution to analyze online reviews. The target of technique to improve accuracy and achieve to accurate review meaning. The proposed SAOOP approach is based on two methods: evaluation and analysis reviews (sentiment analysis) and solve some sentiment analysis challenges. In order to serve researchers in selecting efficient papers. In addition, it evaluates topic domain parameters of scientific papers (place of publication, publishing date, and a number of citation paper) to evaluate the total score of papers. To evaluate SAOOP efficiency, making a comparison between it and two famous techniques. The results have a comparison between the accuracy and performance between the three techniques when the researchers apply the techniques on three data sets (training, test and verified). The comparison results illustrate how proposed technique can increase accuracy and performance with facing many language coverage cases and solving some sentiment analysis challenges. The accuracy results show in NLTK (62%) and NLPS (70%) to 82% (SAOOP) with the proposed technique.

## REFERENCES

- [1] Suman, D.R, Wenjun, Z., Social Multimedia Signals: A Signal Processing Approach to Social Network Phenomena, ISBN-13: 978-3319091167, Springer International Publishing Switzerland, 2015.
- [2] Larsen, Peder Olesen, and Markus von Ins. "The Rate of Growth in Scientific Publication and the Decline in Coverage Provided by Science Citation Index." *Scientometrics* 84.3 (2010): 575–603. PMC. Web. 25 Sept. 2015.
- [3] Peng, L., Cui, G., Zhuang, M., & Li, C. (2014). What do seller manipulations of online product reviews mean to consumers? (HKIBS Working Paper Series 070-1314). Hong Kong: Hong Kong Institute of Business Studies, Lingnan University.
- [4] Thomas, B., Keep Social Honest , "What Consumers Think about brands on social media, and what businesses need to do about it" Report, 2013.
- [5] Namita, M., Basant, A., Garvit, C., Nitin, B., & Prateek, P., "Sentiment analysis of Hindi Review based on Negation and Discourse Relation", International Joint Conference on Natural Language Processing, Nagoya, Japan, 2013.
- [6] Theresa, W., Janyce, W. .& Paul, H., "Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis", proceeding HLT'05 proceedings of the conference of Human Language Technology and Empirical Methods in Natural Language Processing, 2006.
- [7] Ian, H.W., Eibe, F., & Mark A. H., Data Mining: Practical machine learning tools and techniques. The Morgan Kaufmann series in data management systems, 3<sup>rd</sup> Edition, Morgan Kaufmann Publishers is an imprint of Elsevier, 2011.
- [8] Yin, Z., Rong, J., & Zhi-Hua, Z., "Understanding Bag-of-Words Model: A Statistical Framework", International Journal of Machine Learning and Cybernetics, 2010.
- [9] Diana, M., & John.C., "Disambiguating Nouns, verbs, and adjectives using automatically acquired Selectional preferences", Association for Computational Linguistics , Vol. 29, 2003.
- [10] Presutti, V., Stankovic, M., Cambria, E., Cantador, I., Di Iorio, A., Di Noia, T., Lange, C., Reforgiato Recupero, D., Tordai, A., "Semantic web evaluation challenge. 1<sup>st</sup> Edition, Series title: Communications in Computer and Information Science, Springer International Publishing, 2014.
- [11] Bing, L., "Sentiment Analysis and Subjectivity". In Nitin, I. & Fred, J. (eds). Handbook of Natural Language Processing. 2nd Ed, Machine Learning & pattern recognition series, Chapman& Hall/CRC, 2010.
- [12] Samih, Y., Erdogan, Y., & Halife, K., "Tagging Accuracy Analysis on Part-of-Speech Taggers", Journal of Computer and Communications, 2014.
- [13] Mitchell, P.M., Mary A.M., and Beatrice, S., "Building a Large Annotated Corpus of English: The Penn Treebank", Computational Linguistics Journal , Voul. 19, Number 21993.
- [14] Saifee, V., & Jay, T., "Applications and Challenges for Sentiment Analysis: A Survey", International Journal of Engineering Research & Technology (IJERT), Vol. 2 Issue2, 2013.
- [15] Hassena, R.P., "Opinion Mining and Sentiment Analysis Challenges and Applications", International Journal of Application or Innovation in Engineering & Management (IJAIEM), Volume 3, Issue 5, 2014.
- [16] Bing, L., Sentiment Analysis and Opinion Mining. Morgan & Claypool Publishers, 2012.
- [17] Archana, S., "Sentiment analysis of document based on annotation", CORR Journal, Vol. abs/1111.1648, 2011.
- [18] Walter, K., & Mihaela, V., "Sentiment analysis for hotel reviews", proceedings of the computational linguistics-applications, Jacharanka Conference, 2011.
- [19] A.Tamilsevi, & M.Parveen, T., "Sentiment analysis of micro blogs using opinion mining classification algorithm", International Journal of Science and Research (IJSR), Vol. 2 Issue 10, 2013.
- [20] Nisha, J., & Dr.E. Kirubakaran, "M-Learning sentiment analysis with Data Mining Techniques", International Journal of Computer Science and Telecommunications, Volume 3, Issue 8, 2012.
- [21] G'unes, E., Ahamed, H., Qian, D., & Dragomir, R., "Improved Nearest Neighbor Methods for Text", Technical Report CSE-TR-576-11, University of Michigan, Department of Electrical Engineering and Computer Science, 2011.
- [22] Frederick, L., "Implementation of Breiman's Random Forest Machine Learning Algorithm", ECE591Q Machine Learning Journal Paper, 2005.
- [23] Christof, M., "Machine learning for query formulation in question and answering", Natural Language Engineering, Cambridge University, 2011.
- [24] Steven, B., Ewan, K., and Edward, L., Natural Language Processing with Python, 1<sup>st</sup> Edition, O'Reilly Media publisher, 2009.
- [25] Richard, S., Alex, P., Jean, Y.W., Jason, C., Christopher, D.M., Andrew, Y.N. & Christopher, P., "Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank", Conference on Empirical Methods in Natural Language Processing, 2013.
- [26] I. Hemalatha , Dr. G. P Saradhi V., & Dr. A. Govardhan, "Preprocessing the Informal Text for efficient Sentiment Analysis" , International Journal of Emerging Trends & Technology in Computer Science (IJETCS), Volume 1, Issue 2, 2012.
- [27] Dushyant, B.R., & Samrat, K., "A Review on Emerging Trends of Web Mining and IT's Application", International Journal of Engineering Development and Research | IJEDR, 2013.
- [28] Bo, P., & Lillian, L., "Opinion mining and sentiment analysis", Journal Foundations and Trends in Information Retrieval, Vol. 2, 2008.
- [29] Bing, L., "Sentiment Analysis: A Multi-Faceted Problem", IEEE Intelligent Systems, 2010.

- [30] Hang, C., Vibhu, M., & Mayur, D., "Comparative experiments on sentiment classification for online product reviews", American Association for Artificial Intelligence (AAAI Conference), 2006.
- [31] Mr. Akshay, A. Adsod, & Prof. Nitin R.C., "A review on web mining", International Journal of Engineering Trends and Technology (IJETT)", Volume 10 Number 3, 2014.
- [32] Alberto, H.F.L., Berthier, A.R-N, Altigran, S. d. S., and Juliana, S. T., "A brief survey of web data extraction tools", ACM SIGMOD, Vol.31, Issue 2., New York , USA, June 2002
- [33] Farah, B., Carmine, C., & Diego.R., "Sentiment analysis: Adjectives and Adverbs are better than Adjectives Alone", International Conference on Weblogs and Social Media -ICWSM , Boulder, CO USA, 2007.
- [34] Utkarsh, G., Clemente, I., and Mike P. W., Understanding Factors Influencing the Citation Count of Networking Conference Papers, in proc. CHI '09 proceedings of the SIGCHI Conference on human factors in computing systems, 2009.
- [35] Sujata, R. & Patreek, K., "Challenges of Sentiment Analysis and Existing State of Art", International Journal of Innovation and Research in Computer Science (IJIRCS), 2014.
- [36] Chenghua, L., and Yulan, H., "Joint Sentiment/Topic Model for Sentiment Analysis", CIKM '09 Proceedings of the 18th ACM conference on Information and knowledge management, Hong Kong, China, ACM 978-1-60558-512-3/09/11, 2009.
- [37] Andreas, H., Gerhard, P., and Andreas, N., "A Brief Survey of Text Mining", LDV Forum - GLDV Journal for Computational Linguistics and Language Technology, 2005.
- [38] Bas, H., Paul, v.L, Alexander, H., Flavius, F., & Uzay, K., "Accounting for Negation in sentiment analysis", DIR, Amsterdam, the Netherlands, 2011.
- [39] John, M. B., "Counting publications - Journals vs. Conferences in Computer Science", in NIK-2013 conference, 2013.
- [40] Quoc, L., & Tomas, M., "Distributed Representations of Sentences and Documents", Proceedings of the 31 st International Conference on Machine Learning, Beijing, China. JMLR: W&CP volume 32, 2014.
- [41] Umer, F., Yang, S., John, M. C., and Lee, G., "Social Bookmarking for Scholarly Digital Libraries", Published by the IEEE Computer Society, Pennsylvania State University, December 2007.
- [42] Pollyanna, G., Matheus, A., Fabrico, B., & Meeyoung, C., "Comparing and Combining Sentiment Analysis Methods", ACM Association for Computing Machinery, 2013.
- [43] Mark, C., Oliver, D., & Fatih, U., "Potential and Limitations of Commercial Sentiment Detection Tools", Emotion and Sentiment in Social and Expressive Media-ESSEM, 2013.
- [44] Steven, B., Ewan, K., Edward, L., & Jason, B., "Multidisciplinary Instruction with the Natural Language Toolkit", Proceedings of the Third Workshop on Issues in Teaching Computational Linguistics (TeachCL-08), Columbus, Ohio, USA, 2008.