# Enhancement Bag-of-Words Model for Solving the Challenges of Sentiment Analysis

Doaa Mohey El-Din
Information Systems Department
Faculty of Computers and Information, CU
Cairo, Egypt

***Abstract*—Sentiment analysis is a branch of natural language processing, or machine learning methods. It becomes one of the most important sources in decision making. It can extract, identify, evaluate or otherwise characterizes from the online sentiments reviews. Although Bag-Of-Words model is the most widely used technique for sentiment analysis, it has two major weaknesses: using a manual evaluation for a lexicon in determining the evaluation of words and analyzing sentiments with low accuracy because of neglecting the language grammar effects of the words and ignore semantics of the words. In this paper, we propose a new technique to evaluate online sentiments in one topic domain and produce a solution for some significant sentiment analysis challenges that improves the accuracy of sentiment analysis performed. The proposed technique relies on the enhancement bag-of-words model for evaluating sentiment polarity and score automatically by using the words weight instead of term frequency. This technique also can classify the reviews based on features and keywords of the scientific topic domain. This paper introduces solutions for essential sentiment analysis challenges that are suitable for the review structure. It also examines the effects by the proposed enhancement model to reach higher accuracy.**

***Keywords—Sentiment analysis; Bag-Of-Words; sentiment analysis challenges; text analysis; Reviews***

## I. Introduction

Sentiment Analysis (SA) [1] is a Natural Language Processing and Information Extraction task that aims to obtain researcher's feelings expressed in positive or negative reviews or opinion by analyzing a big numbers of documents and papers [2]. The important issue [3] in sentiment analysis is to recognize how sentiments are expressed in reviews and whether the expressions refer (acceptable) or negative (unacceptable) reviews toward the subject. Sentiment analysis is a laborious [4] task for performing with computers and algorithms. Identifying some patterns is hard for machines or even impossible while it is easy for human beings. There are some intractable situations for computers such as: Can't deal with pronouns and what they refer to. It also can't understand the different meaning of words such as ironies or sarcasm. The difficulty in putting a standard in some punctuation marks such as (!!!). It's very hard to evaluate the world knowledge information. Sentiment analysis requires to generate a big lexicon which takes a big time to search for required words. The Hardness in evaluating words within two polarities: positive and negative. Sometimes we need to understand some features or keywords for each topic or deal with multi-language. Other problems face sentiment analysis in understanding reviews meaning and grammar for one language or multi-languages. Especially, there are several words that have different meaning and polarities. Thus, sentiment analysis involves identification of sentiment meaning, expressions [5], Polarity and expressions strength, and their relationship to the subject. The volume of linguistics resources is enormous. The bag-of-words (BOW) [6,7] models evaluation uses many techniques such as Naive Bayes (NB), Support Vector Machine (SVM) and Maximum Entropy (ME) classifiers [8] that have been exhibited to go well in the binary positive negative sentiment classification tasks on document-level datasets like movie reviews. There are three essential classification levels in sentiment analysis. The differences between document-level, sentence-level, aspect level and word-level SA declares in the level of evaluation the sentiments and classification polarity based on document, sentence, aspect/entity or word/term. The sequence of sentiment analysis process of each one of them as followed in the figure 1:

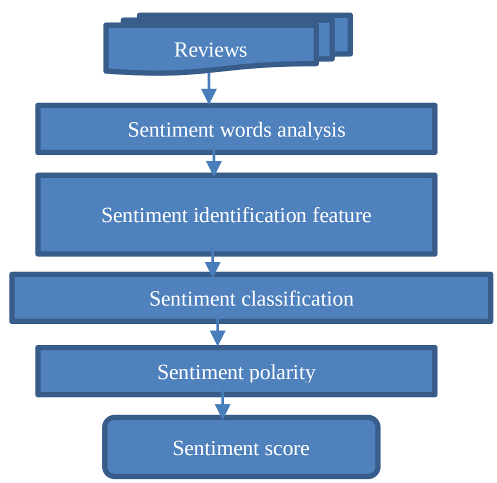


Fig. 1. Sentiment analysis approach

The difference between the four levels of sentiment classification declares in sentiment classification polarity level on document as a whole [9, 10] or on each sentence in a document or text [11] or on aspect or entity level [12] or on each word in text. In this paper, we discuss the sentiment classification and evaluation on word level in one topic domain. We present the sentiment classification regard to word polarity with the specific topic of features or keywords.

The first step is to identify the features and entities the scientific topic. The review writers can give different reviews for different features of the same feature like this sentence "the journal of publication of this paper is not good, but it has a big number of citation". The sentence is subjective, Word-level SA [13] will determine whether the review expresses positive or negative based on evaluate each word polarity related with the feature.

In this paper, we present a new technique which called Sentiment Analysis Of Online Papers "SAOOP". It can evaluate online sentiment reviews for research paper domain. This technique is a new technique which introduces an enhancement of Bag-of-words model to solve major weaknesses of the Bag-Of-Words model in sentiment analysis evaluation. It depends on the word level of sentiment analysis in one topic domain. Additionally, we can extract features and keywords of the domain to classify sentiments reviews and reach the accurately meaning of each review. The proposed technique also can introduce solutions for sentiment analysis challenges to improve accuracy.

The rest of this paper is organized as follows: Section 2 represents related works. Section 3, the presentation of the proposed technique “SAOOP”. In Section 4, outlines of the Experiment comparison between the standard and enhancement proposed models. Section 5 highlights the evaluation and discussion results. Finally, Section 6 conclusion and proposes directions for future work.

## II. Related Works

The target of this paper evaluates sentiment analysis and classifies the sentiment polarity automatically and more accurately. Although the sentiment analysis is a hot area to research, No research finds enough in this field till now.

The authors [14] discusses that sentiment analysis becomes the most motivating research area among natural language processing (NLP) community. There are many tools and applications for opinion Mining or sentiment analysis. They also face many research challenges. There are some innovative and effective techniques required to be invented which should overcome the current challenges faced by Opinion Mining and Sentiment Analysis.

In movie domain, the research works on analyzing sentiments on the document level. There is the most distinguished work [15] by using “Poor” and “Excellent” seed words to compute the semantic orientation, point wise the mutual information approach used to calculate the semantic orientation. The sentiment orientation (SO) of the document was computed as the average SO of all such phrases. The accuracy results achieved to 66%.

But when using a lexicon-based approach [16] on also movie review domain, the approach is used to implement the sentiment polarity classification. For this sentiment classification positive and negative words lexicons is used and semantic orientation calculator (SO-CAL) is built that combine intensifiers and negative words. The approach's results showed to have 59.6% to 76.4% accuracy on 1900 documents.

Further research of machine learning [17], which develops a new approach for extracting product features and opinions from a group of free customers;' sentiment reviews about a service or product. This approach depends on a language-modeling framework that, using a seed set of opinion words, can be compatible with any domain reviews and language. The proposed approach combined both a statistical mapping between words and a kernel-based model of opinion words learned from the seed set to approximate a model of product features from which the retrieval is performed

## III. Proposed Technique

In this paper, we propose the enhancement bag-of-words (BOW) model on sentiment analysis by presenting a new technique is called Sentiment analysis of online papers “SAOOP” [18]. The proposed technique SAOOP is based on a word weight instead of the term frequency of each word in the standard BOW. The standard BOW model uses a huge lexicon [19] which has duplications of word and repetition. This lexicon is built manually which requires to create a 'positive' and 'negative' words list [20] by recognize the sentiment polarities based on the personal observation. This approach takes a big time and efforts to compute the total score of sentiments reviews. Another problem of BOW is low accuracy [21] because the standard BOW model neglects text grammatically and ordering of words. So we introduce a new miniature lexicon to reduce the standard lexicon of BOW and deal with adjectives, nouns, verbs, adverbs, adjectives, prefixes, suffixes or other grammatical classes as a word by similarity [22] and differences algorithms. The proposed lexicon is constructed automatically which is based on hierarchical database model [23] to give the correct scores with respect a topic features and keywords. The new lexical approach uses for saving time and ease searching process for each word.

The target 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. With caring topic features with word level of sentiment analysis, the proposed technique also can classify reviews into some categorizations. These categorizations are based on scientific papers topic features and keywords parameters as (place of publication, citation number, topic, publishing paper date, and authors). Although SAOOP aims at the evaluation of words but it can handle some cases of expressions and phrases with respect the order of each word. SAOOP also computes the total score of each review by calculating the aggregate score of review words. For measure accuracy, we make the comparison between our proposed enhancement BOW technique and the standard BOW model on the scientific domain.

### A. *SAOOP Architecture*

he SAOOP technique is presented here which illustrated in fig.2. The architecture has five phases for reaching to sentiment score. The input is online reviews. The phase one called "Analyze Data" which includes some functions as web scraping and extracting data, text analysis, NLP linguistics, and Enhancement BOW.

Web Scarping and extracting data which extracts the papers data and their parameters data and creates rows records. Text analysis contains that splits sentiment reviews into sentences and tokenizes each review into some words. Natural language processing (NLP) Linguistics makes normalization functions and reformat data. Our proposed technique introduce an Enhancement Bag-of-words (BOW) algorithm which is based on a word weight.

The phase two called "Lexicon data" illustrates that creating a lexicon, topic features and keywords, finding names, and classify sentiment reviews. This phase explains that classify reviews from one or more class in assuming five classes (Topic, citation number, the publishing date of paper, authors, and place of publications). In the second function, extract the features and keywords of scientific domain as the names of authors, the names or shortcuts of conferences and journals. Finding names declares that how to recognize some names of each class. And the last function is the classification reviews by using the previous two functions [24]. In the third phase entitled "sentiment analysis score and polarity", the proposed technique can detect the polarity based on one of sentiment classification (very negative, negative, neutral, positive, and very positive). Although the proposed technique is based on the word-by-word evaluation, it can handle expressions, wish words, some special cases, with caring grammar by using a newly constructed lexicon. The fourth phase is the solutions of sentiment challenges. This depends on the word level, it contains proposed solutions to deal with some challenges "Spam and fake detections", "Implicit and Explicit Negation", and "World knowledge" based on topic features. Additionally creating a map guide which is based on the sentiment scores related to the most related papers according to keywords and fields classification. The output declares in the total sentiment score and polarity of each paper.

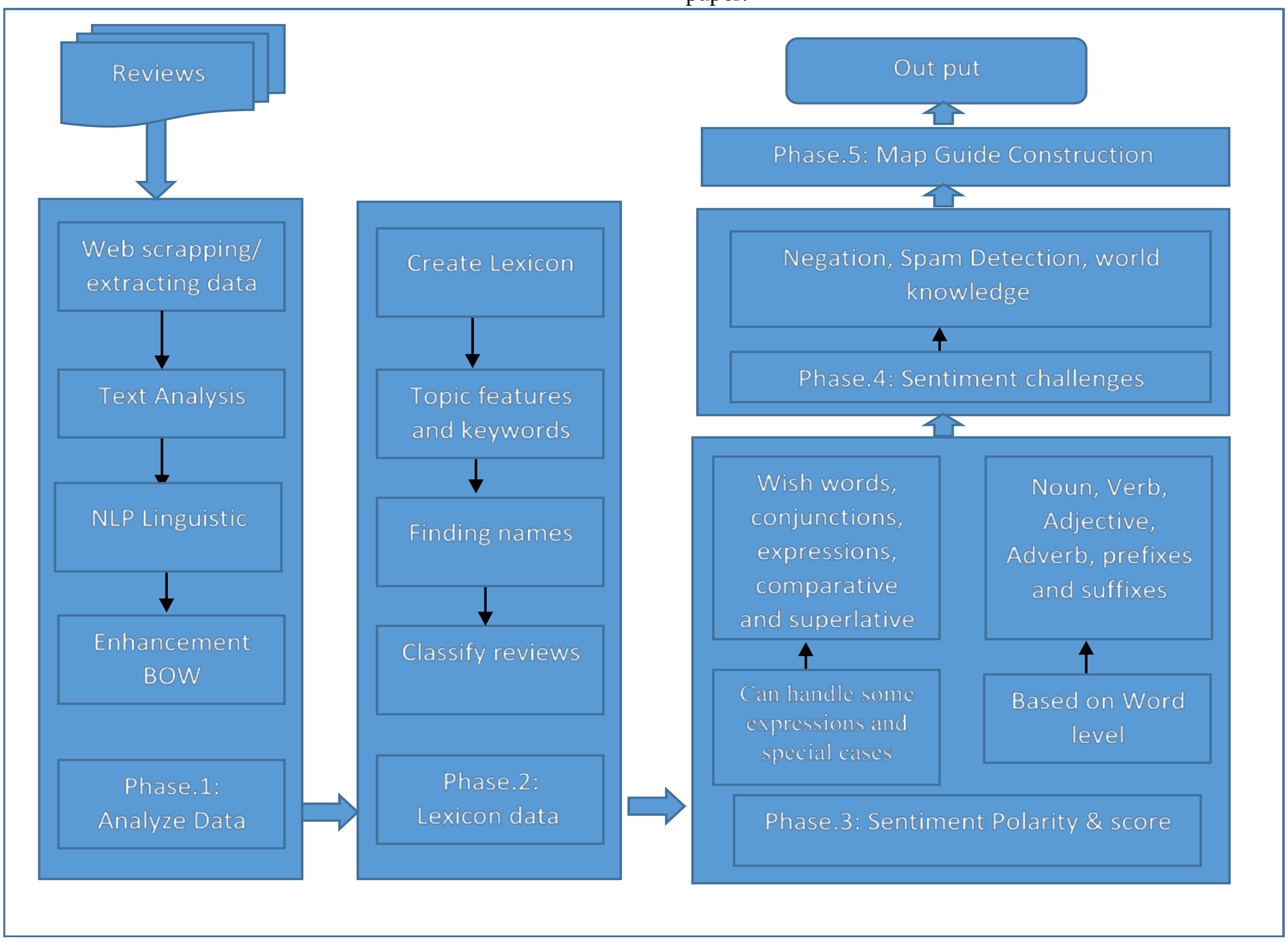


Fig. 2. Proposed technique (saoop) architecture

## B. SAOOP Lexicon

The proposed technique creates a new lexicon for sentiments reviews which is based on the hierarchal database model. This model is a data model where the data is organized like a tree. The structure allows repeating information using parent/child relationships: each parent can have many children but each child only has one parent. All attributes of a specific record are listed under a feature type. The lexicon has words, prefixes and suffixes and hierarchal nouns to produce a solution for bi-polar words and evaluate topic features of reviews. We use Part-of-speech (POS) tagging [24] model to recognize nouns in constructing the nouns tree in lexicon. The advantage of this hierarchical model of nouns is each parent can hold the same name of child with different value that supports us in evaluating bi-polar and fuzzy words or the topic features which declares in Figure.3:

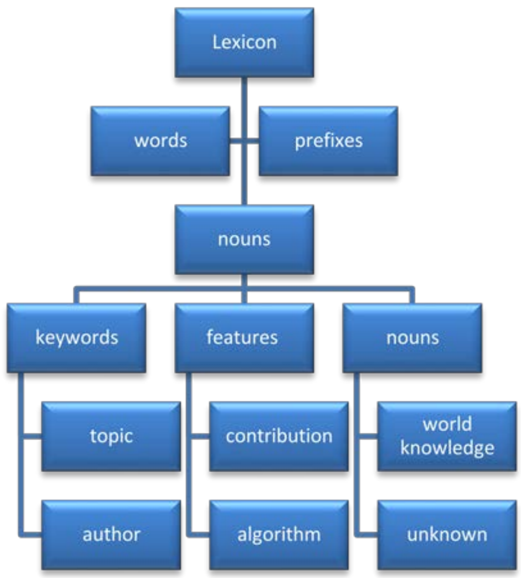


Fig. 3. Proposed lexicon structure

For example if keyword is "topic" and the before word is [old] the polarity will be [negative] but if the keyword is "author" the polarity to the same word [old] will be [positive]. The same case in features but other nouns we splits to identify the world knowledge words and polarities such as a name of famous scientist [Einstein] holds a positive polarity or unknown words.

The proposed lexicon saves a positive *Wp* and negative *Wn* score for each word with the recommended polarity, by assuming the equation,

$$\sum Wp + Wn = 1 \tag{1}$$

This model expresses words into positive or negative polarities with saving two scores to support the accurate meaning such as [not great] polarity does not equal [bad] polarity but it equals [good] polarity. SAOOP proposed technique evaluate sentiments with a word weight. It also assumes a desirable state of each word based on the meaning (e.g., "great" and "good") have a positive polarity, while words that encode an undesirable state have a negative polarity (e.g., "bad" and "worst"). Although sentiment polarity normally applies to adjectives and adverbs, there are verbs, expressions, conjunctions, prefixes or suffixes and noun sentiment words as well. We can compile sets of sentiment words and phrases for adjectives, adverbs, verbs, expressions, prefixes, suffixes and nouns respectively.

### *C. SAOOP Enhancement BOW*

In this section, we will explain the comparison between the standard and enhancement bag of words model. We also can compare between the methodology and challenges, they can face.

#### *1) Standard BOW:*

The input of the first technique is some documents, and the output is the sentiment scores for each word. Each document is a bag of words, meaning: Assumes order of words has no significance (the word “home made” has the same probability as “made home”). The first standard BOW algorithm follows the next steps: Bag of words representation (or vector space representation) [25] is the main methodology proposed by information retrieval researchers to represent text corpus, which is an easy approach to converts unstructured text to structured data based on word by word , and neglecting the grammar. This algorithm declares the relationship between documents and evaluates the words based on the term frequency in these documents. There are many algorithms to calculate term weight, we apply here the term frequency–inverse document frequency (tf-idf) which is a numerical statistic that aim at reflecting the importance word is to a text in a groups or corpus.

There are some goals for this algorithm and it can't only to evaluate sentiment score. But it depends on supervised or unsupervised algorithms to compute the score. It can't deal with the words order. We declare the standard algorithm in the following:

Algorithm 1: Pseudo code of Standard BOW

1. The input: Given a corpus of K documents, comprising a dictionary of M words, find the “relations” of words and documents (usually cluster the documents
2. Get the number of reviews.
3. Create dictionary for all words for all reviews.
4. Calculate the number of frequencies (occurrences) in each review manually.
5. This following formula declares the word frequency in the text.
6. $tfi,j = \frac{ni,j}{\sum_k nk,j}$
7. Where n i,j is the number of occurrences of the considered term in document dj, and the denominator is the number of occurrences of all terms in document dj.
8. With |D| is a total number of documents in the corpus and |{dj: ti € dj}| is the number of documents where the term ti appears.
9. Then,
10. $tfidf\ i,j = tf\ i,j * idfi$
11. Create vector of frequencies for each review.
12. Create dictionary of positive, negative and neutral words in three files manually.
13. Check on each word manually positive or negative or neutral.
14. It also neglects the grammar and the meaning of the grammar.
15. It can't evaluate the score separately, but it must apply one of the supervised or unsupervised algorithm as K-Nearest Neighbor or Naïve Base Classifier,
16. Compute the sentiment score for each word.
17. Then try to classify the total score polarity class between positive or negative classes.

*2) Enhancement BOW*

By study of analyzing the online scientific reviews [26], we identify the nature of review structure which is short and formal. So the BOW is the most suitable model to deal with these reviews but we need to improve the accuracy and avoid its weaknesses. Hence, we present an enhancement BOW model in the proposed technique (SAOOP) which is explained steps in the following: The input is online sentiment reviews, and the output is the sentiment scores of each word. Algorithms.2 discusses the pseudo code of the enhancement.

**Algorithm 2: Pseudo code of Enhancement BOW**

1. For each paper P do
2. Web scrapping data
3. Reformat data
4. Calculate number of sentiment S
5. For each review R in P
6. Check and delete fake or spam sentiments reviews.
7. Get the real number R or reviews.
8. For sentence sent ∈ classification reviews data do.
9. for review category a ∈ A do, class='Topic', and Score= 0.
10. For word w ∈ s do.
11. If O (w) > 0 then.
12. Remove stop and punctuation lists.
13. Convert all w into UPPER case.
14. Create a new lexicon for all positive and negative words.
15. Assume each word w has two values (positive and negative), and The total score of w equal 1,

$$\sum Wp + Wn = 1$$

Assuming each word has 2 values (W p=positive value, W n=negative value)

16. If having explicit negative words (such as not)?
17. Check on the next word w to detect score.
18. We assume the negative value for positive word, And assume the positive value for the negative value for the negative word,

$$V(W) = W(N) - 0.2.$$

19. Each w has class from five sentiment classification levels (very negative, negative, neutral, positive, and very positive).

$$V(W) = W(N) + 0.2.$$

20. Detect sentiment score and polarity.
21. End If.
22. Else if having a word from second negative list such as never.
23. Convert the polarity of the sentence by,

$$V(Sent) = S(Sent) * -1.$$

24. End if.
25. Else if check on future words such as (wish, hope).
26. Check on next word and detect polarity and score (go to step17).
27. Else If.
28. Use POS tagging to check on nouns.
29. If w is noun?
30. If w is feature?
31. Detect sentiment score and polarity
32. End if
33. Else if w is keyword?
34. Detect sentiment score and polarity
35. Else if w is world knowledge
36. Detect sentiment score and polarity
37. Else if go to step 17.
38. End for.
39. Assign review classification of each S in R.
40. If s € review classes
41. Determine class.
42. End if
43. Else If class = 'topic'
44. Compute sent score and polarity.
45. Calculate R (SA) is a total sentiment score of each review r.
46. End for
47. End for
48. Calculate T (SA) is a total sentiment score of each paper p. Calculate AVG (SA) is an average sentiment score of each p. T (SA): is the real total score of all reviews. r: is the number of real reviews without spam or fake reviews.
49. End For.

*D. Sentiment Polarity Detection*

The main goal of this word level reaches the accurate polarity for the sentiment review. The proposed technique can detect the polarity values for each input word of the evaluation dataset (with the summation of positive and negative is equal 1). But the polarity depends on the meaning of the review characteristics. SAOOP can assign a 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.1. 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 division is suitable for the five sentiment class's levels [18]. There is two scores of sentiment analysis, a real sentiment and the average sentiment scores. The r (SA): is a total score of sentiment score of all reviews on each paper with caring of the number of positive reviews. In the next equation,

$$r(SA) = \sum_{i=1}^{n} \frac{P(SA(R))}{n} \tag{3}$$

The AVG (SA): is a total score of sentiment score of all reviews on each paper. In the next equation,

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

The calculation of the total score of all reviews depends on the score of each review. There is a difficult problem between a large number of reviews and evaluating sentiment polarity of each one, this problem is improper the most review number having assessment higher score.

*E. Sentiment Analysis Challenges Scope*

With scanning a data set of CiteuLike [27] scientific reviews for papers, we can detect also the most essential challenges [28] in evaluating sentiments and opinions that are implicit and explicit negative, world knowledge and spam or fake reviews.

The discussion of the solutions in the following:

*1) Implict and Explicit Negative challenge*

Negation is the biggest challenge in sentiment analysis [29]. 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. First: explicitly is deliberately formed and are easy to self-report and by keywords. Second implicitly 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.

*2) World knowledge requirement Challenge*

The proposed technique presents a solution for Knowledge [30] 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. Hierarchal model between nouns to achieve the polarity, score and meaning. Also to differ between them and keywords or features. In a hierarchical model, data is organized into a tree-like structure, implying a single parent for each record. A sort field keeps sibling records in a particular order. Hierarchical structures were widely used in the early mainframe database management systems. This structure allows one one-to-many relationship between two types of data. This structure is very efficient to describe many relationships in the real world; recipes, table of contents, ordering of paragraphs/verses, any nested and sorted information.

*3) Spam and Fake Reviews Challenge:*

The Internet includes both realistic and spam contents [31]. For effective Sentiment classification, this spam content should be eliminated before processing. The proposed technique 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. The proposed SAOOP can avoid and cure the most of them by deleting empty reviews and removing the duplicate sentiment reviews by considering the same reviewer for computing the real number of reviews.

## IV. Experiment

The discussion of this experiment explains the comparison between the proposed technique and the standard BOW technique in online scientific papers domain. This comparison shows the accuracy results based on real dataset. A real set (1000 reviews) from the CiteULike website in computer science branch. This comparison also discusses the challenges solutions impact on evaluating sentiment analysis. We compare between accuracy [32] with the next metrics [33]:

TABLE I. Measurement Metrics

| | Classified as positive | Classified as Negative |
|---|---|---|
| Class Positive | TP | FN |
| Class Negative | FP | TN |

Let in a group of reviews have positive sentiment (belong to class positive) and reviews have negative sentiment (belong to class negative). N p N n N After classifying these sentiment reviews, class positive had reviews correctly classified under it and reviews wrongly classified under it, while class negative had documents correctly classified under it and documents wrongly classified under it. Then, in relation to class positive: TP FP TN FN.

$$precision = \frac{TP}{TP+FP} \quad (5),$$

$$Recall = \frac{TP}{TP+FN} \quad (6),$$

The accuracy equation declares in the next equation,

$$Accuracy\ (acc.) = \frac{TP+TN}{TP+TN+FP+FN} \quad (7),$$

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).

## V. Evaluation & Discussion

With the examination of the percentage degree of different techniques accuracy [33] on text reviews content. For computing the accuracy of each model, by calculating the intersections of the positive or negative proportion given by each technique. Table.4 presents the percentage of accuracy for the two compared models. Table 4 shows techniques recall, precision, and accuracy.

TABLE II. Average Results for All Datasets

| Metric | **BOW** | **SAOOP** |
|---|---|---|
| **Precision** | 0.834 | 0.856 |
| **Recall** | 0.560 | 0.867 |
| **Accuracy** | 0.618 | 0.817 |

SAOOP gets a better results of accuracy (82%) than standard (62%). So the enhancement bag of words increases the accuracy with around 20%, as figure.

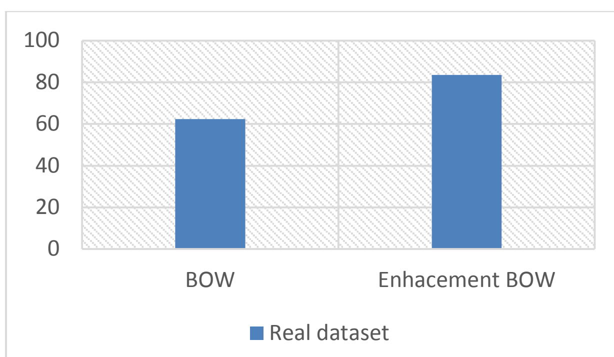


Fig. 4. differences between standard BOW algorithm and proposed SAOOP based on real dataset

In the next table, we discuss the comparison between the two algorithms standard BOW and proposed Enhanced BOW with SAOOP technique. This table comparison relies on several features, the goal of the algorithm, sentiment classification levels, and the input is reviews, the data size of the input data, data set scope, clarity, efficiency, memorability, simplicity [34]. The definition of memorability is the quality or state of being easy to remember or worth remembering, which can help to declare the relationship between the data and its features. The Algorithm will be clear if it is familiar and easy to use. The efficiency of algorithms depends on the accuracy results in figure.4. The last issue simplicity of algorithm which is simple if it is concise to write down and easy to grasp. Simplicity" of an algorithm is affected by "cultural" factors: Means of presentation (notation, assumptions…etc.) and Previous knowledge of the reader.

TABLE III. COMPARISON BETWEEN STANDARD BOW AND ENHANCEMENT BOW

| Algorithm | Standard BOW | Enhancement BOW |
|---|---|---|
| **Goal** | Text analysis and give polarity for words in text | Evaluate sentiment score for reviews |
| **Sentiment classification** | 2 or three classes | 5 classes |
| **Input type** | Documents, text or images | Reviews |
| **Data size** | Small number of texts or review | Large number of reviews |
| **Data set** | Any scope, refer topic domain to minimize dictionary | Topic domain is the best to minimize dictionary and can extraction features and entities |
| **Clarity** | No | Yes |
| **Efficiency** | No , less accuracy and manually dictionaries | Yes, high accuracy |
| **Memorability** | No | Yes |
| **Simplicity** | Yes | Yes |

Although the last comparison is illustrated, we find the SWOT [35] analysis comparison is very useful to show strengths, weaknesses, opportunities and threats. SWOT [37] analysis [36] became one of the most popular tools for strategic planning or making a decision. It can help in improving our models. Strengths are those features of the business which allow you to operate more effectively than your competitors. Weaknesses are areas capable of improvement. Opportunities identify any new opportunities for techniques. Threats can be external or internal, and are anything which can adversely affect the techniques. With applying SWOT analysis on the compared two algorithms, the results presents in table that discuss the weaknesses of the bow and how can handle them in the strengths of the proposed enhanced BOW model in a new SAOOP technique.

TABLE IV. COMPARISON BETWEEN STANDARD BOW AND ENHANCEMENT BOW IN SWOT ANALYSIS

| Algorithm / SWOT | BOW | SAOOP |
|---|---|---|
| **Strengths** | - Ease to use<br>- Using for small reviews<br>- Topic domain<br>- Deal with images, text, and documents | -Improve Bag of words and combine with POS tagging algorithm<br>-categorize reviews<br>- extract features<br>- identify objects and evaluate it<br>Applied KNN- Naïve base classifiers to measure accuracy.<br>Graphic reports<br>-Handle some sentiment analysis challenges<br>Easy<br>Clarity<br>High accuracy<br>Topic domain or any domain based on dictionary<br>Memory ability<br>Scale classification -1,0,1 |
| **Weaknesses** | - Less accuracy<br>- Manually dictionary<br>- neglect grammar<br>- neglect ordering<br>- Don't deal with Numbers<br>Questions<br>- Fake or spam review<br>-World knowledge<br>User mention<br>-Hash tags<br>-Emotions | - Not fast enough<br>-Don't deal with<br>-Numbers<br>-Questions |
| **Opportunities** | Automate algorithm<br>High accuracy | -More fast<br>-Arabic sentiment analysis for scientific papers<br>-Create some lexicons to suitable with some domains |
| **Threats** | Binary words<br>World knowledge<br>Numbers<br>Questions<br>User mention<br>Hash tags<br>Emotions | Numbers (10/10, or 100%)<br>Questions<br>Words not splitting<br>Emotions<br>Deal with hash tags, user mentions and emotions. |

Enhancement BOW model in the proposed SAOOP technique has been shown to extremely effective, since it captures more contextual meaning based on word weight, resulting a classification accuracy of 83.5%. Our observation the precision and recall for each sentiment category separately, since the effect of our proposed technique has a significantly different impact on the negative and positive class. Although, the proposed technique increases the precision and recall of both the classes, we could observe a significantly higher improvement of precision and recall in dealing with sentiment challenges. This is a clear indication of the effectiveness of incorporating the impact of world knowledge, spam detection, and negation, by interesting the topic domain features and keywords and constructing the newly miniature lexicon. Although the proposed technique is based on the word-by-word model, it can understand some phrases as do not directly through caring with the classification of reviews.

## VI. Conclusion & Future Work

The technique described in this paper proposes an approach to evaluate sentiment score at the word level. Our contributions include the enhancement of Bag-of-Words model on online scientific papers reviews and the incorporate contextual polarity and effect of sentiment analysis challenges to improve the sentiment accuracy. SAOOP aims at evaluating for reviews of scientific papers and from scientific papers is called CiteULike website, analyzes and classifies the textual content of the sentiment reviews of each paper. The proposed SAOOP can classify sentiment reviews and visualize the relationships between them based on extract features and keywords of scientific domain.

This paper makes a comparison between the standard bag-of-words model and our proposed enhancement bag-of-words and test the impact on sentiment analysis challenges and the accuracy. The experimental results show that our technique obtains sentiment classification accuracy with (83.5 %) that significantly better than the standard BOW (62%). Further, the efficiency of the proposed algorithm improves over standard BOW algorithm. Future research will focus on enhancing the proposed technique further by working on phrases in order to have sufficient local information to determine the polarity. Further, working on the proposed technique (SAOOP) to apply on the Arabic language in scientific paper research.

## References

[1] Bing, L., Sentiment Analysis and Opinion Mining. Morgan & Claypool Publishers, 2012.

[2] 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.

[3] Erik, B., Pieter, H., & Koen, D., "Automatic Sentiment Analysis in On-line Text", Proceedings ELPUB2007 Conference on Electronic Publishing – Vienna, Austria, June 2007.

[4] Raisa, V., `& Jayasreee, M., "A SURVEY ON SENTIMENT ANALYSIS AND OPINION MINING", IJRET: International Journal of Research in Engineering and Technology, eISSN: 2319-1163 | pISSN: 2321-7308, 2013.

[5] Tetsuya, N., & Jeonghee, Y., "Sentiment Analysis: Capturing Favorability Using Natural Language Processing", IBM Research, Tokyo Research Laboratory, K-CAP'03, October 23–25, 2003, Sanibel Island, Florida, USA, 2003.

[6] 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.

[7] Yin, Z., Rong, J., & Zhi-Hua, Z., "Understanding Bag-of-Words Model: A Statistical Framework", International Journal of Machine Learning and Cybernetics, 2010.

[8] Steven, B., Ewan, K., and Edward, L., Natural Language Processing with Python, 1st Edition, O'Reilly Media publisher, 2009.

[9] Wilson, T., Wiebe, J., & Hoffman, P., "Recognizing contextual polarity in phrase-level sentiment analysis", In Proceedings of HLT/ EMNLP; 2005.

[10] Bing, L., "Sentiment Analysis: A Multi-Faceted Problem", IEEE Intelligent Systems, 2010.

[11] Bishan , Y., & Claire, C., "Context-aware Learning for Sentence-level Sentiment Analysis with Posterior Regularization", Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, pages 325–335, Baltimore, Maryland, USA, June 23-25 2014. c 2014 Association for Computational Linguistics.

[12] Haseena, R.P, "Opinion Mining and Sentiment Analysis - Challenges and Applications", International Journal of Application or Innovation in Engineering & Management (IJAIEM), Volume 3, Issue 5, May 2014, ISSN 2319 – 4847, 2014.

[13] ELS: A word-level method for entity-level sentiment analysis, proceeding WIMS '11 proceedings of the international conference on web intelligence, mining and semantics, ACM, USA, 2011.

[14] P. Turney, "Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews", Proceedings of the Association for Computational Linguistics, July 2002.

[15] Richa, S., Shweta, N., & Rekha, J., "OPINION MINING OF MOVIE REVIEWS AT DOCUMENT LEVEL", International Journal on Information Theory (IJIT), Vol.3, No.3, July 2014.

[16] Zagibalov, Taras, and John Carroll,(2008), "Unsupervised Classification of Sentiment and Objectivity in Chinese Text", In Proceedings of the Third International Joint Conference on Natural Language Processing,304–311.

[17] Yuya,Y., Tomoharu, I., & Hiroshi, S., "Latent Support Measure Machines for Bag-of-Words Data Classification", Advances in Neural Information Processing Systems 27, 2014.

[18] Doaa Mohey El-Din, Hoda M.O. Mokhtar and Osama Ismael, "Online Paper Review Analysis", International Journal of Advanced Computer Science and Applications(IJACSA), 6(9), 2015.

[19] Lukasz, A., Tomasz, K., Piotr, S., and Włodzimierz, T.,"Simpler is Better? Lexicon-based Ensemble Sentiment Classification Beats Supervised Methods", International Workshop on Curbing Collusive Cyber-gossips in Social Networks (C3-2014), Proc. IEEE/ACM Int. Conf. Advances in Social Network Analysis and Mining, ASONAM, Beijing, China,August 17, 2014

[20] Ilia, C., & Natalia, L., "Extraction and Use of Opinion Words for Three-Way Review Classification Task", CDUD'11–Concept Discovery in unstructured data, 2011

[21] Georgios, P., & Mike, T., "More than Bag-of-Words: Sentence-based Document Representation for Sentiment Analysis", Proceedings of Recent Advances in Natural Language Processing, pages 546–552, Hissar, Bulgaria, September 2013.

[22] Rugved, D., Ketan, V., Suratsingh, R., & Tushar, J., "Comparative Study of Document Similarity Algorithms and Clustering Algorithms for Sentiment Analysis", International Journal of Emerging Trends & Technology in Computer Science (IJETTCS), Volume 3, Issue 5, September-October 2014 ISSN 2278-6856,2014

[23] Joseph, V. H., and Paul, J. K., "A COMPARISON OF THE RELATIONAL DATABASE MODEL AND THE ASSOCIATIVE DATABASE MODEL", Issues in Information Systems, Volume X, No. 1, 2009.

[24] Efthymios, K., Theresa, W., and Johanna, M.,"Twitter Sentiment Analysis: The Good the Bad and the OMG!", Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media, Association for the Advancement of Artificial Intelligence,2011.

[25] D.McCarthy, J.Carroll ,"Disambiguating Nouns, verbs, and adjectives using automatically acquired selectional preferences", Association for Computational Linguistics, 2003.

[26] John, M. B., "Counting publications - Journals vs. Conferences in Computer Science", in NIK-2013 conference, 2013.

[27] Bas, H., Paul, v.I., Alexander, H., Flavius, F., & Uzay, K., "Accounting for Negation in sentiment analysis", DIR, Amsterdam, the Netherlands, 2011.

[28] Bas, H., Paul, v.I., Alexander, H., Flavius, F., & Uzay, K., "Accounting for Negation in sentiment analysis", DIR, Amsterdam, the Netherlands, 2011.

[29] 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.

[30] Sujata, R. & Patreek, K., "Challenges of Sentiment Analysis and Existing State of Art", International Journal of Innovation and Research in Computer Science (IJIRCS), 2014.

[31] John, M. B., "Counting publications - Journals vs. Conferences in Computer Science", in NIK-2013 conference, 2013.

[32] 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. 1st Edition, Series title: Communications in Computer and Information Science, Springer International Publishing, 2014.

[33] Subhabrata, M., & Pushpak, B., "WikiSent : Weakly Supervised Sentiment Analysis Through Extractive Summarization With Wikipedia", Lecture Notes in Computer Science Volume 7523, 2012, pp 774-793, Information Retrieval (cs.IR); Computation and Language (cs.CL,September 2012.

[34] Salvatore, G., Bernadette, B. M., Giulianella, C., Mario, F., Benedetto, M., & Ronald, R. Y., "Advances in computational intelligence", 14th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems, IPMU 2012, Catania, Italy, July 9-13, 2012.

[35] Lu, Y., & Harold, B., "A WEIGHTED-TREE SIMPLICITY ALGORITHM FOR SIMILARITY MATCHING OF PARTIAL PRODUCT DESCRIPTIONS", in Proc. of ISCA 14 th International Conference on Intelligent and Adaptive Systems and Software Engineering. Toronto, 2005.

[36] Tamlin M. P., "RECENT PROGRESS IN DEVELOPMENT OF SWOT RIVER DISCHARGE ALGORITHMS", Department of Geological Sciences, University of North Carolina and participants in the 2012 SWOT Discharge Algorithms Workshop, June 18-20, Chapel Hill, NC, 2012.

[37] Boleslaw, G., Daniel, P., "SWOT ANALYSIS IN THE FORMULATION OF TOURISM DEVELOPMENT STRATEGIES FOR DESTINATIONS", Tourism 2010, 20/2, 2011.