International Journal on Emerging Technologies **10**(3): 238-243(2019)

ISSN No. (Print) : 0975-8364
ISSN No. (Online) : 2249-3255

Semantic Rules based linguistic Computational Model of Sentiment Analysis

***Y. Sahu*[1], *G.S. Thakur*[2] *and S. Dhyani*[1]**
[1]*Research Scholar, Department of Computer Application, MANIT, Bhopal, (Madhya Pardesh), India.*
[2]*Assistant Professor, Department of Computer Application, MANIT, Bhopal, (Madhya Pardesh), India.*

(Corresponding author: Yatendra Sahu)
(Received 26 June 2019, Revised 29 August 2019 Accepted 25 September 2019)
(Published by Research Trend, Website: www.researchtrend.net)

ABSTRACT: Online web interface provides an efficient source of consumers reviews to analyze and understand the performance of the system. Using online review system, any web user may submit their experience about the service and that experience may be utilized to understand the satisfaction level of service by other users. In a system where number of users is large, a computational sentiment analysis and opinion mining can summarizes all the reviews about a particular service, product or event. Sentiment analysis and opinion mining generate the text polarity to evaluate positivity or negativity of the text of review submitted by the users. In this paper we have presented a computational model based on semantic rule based linguistic pattern to generate more improved text polarity for understanding the satisfaction level of the user about the service. The proposed computational model performs better in comparison of existing techniques on some well known dataset with an Accuracy of 79.8%.

Keywords: Concept of Words, Emotion Detection, Features Extraction, Semantic Rules, Textual Sentiment Analysis, NLP.

I. INTRODUCTION

Now day's people have started to analyze the service performance of any system on the basis of online reviews of people available on the internet [1]. For pickup any kind of service like product shopping, hotel booking, couching institute and doctor etc, people always take advice and recommendation from their family members and friends [2]. Now maximum services are available online, so online reviews are very popular. People post reviews to share their own experience and these experiences are utilizes by other person to predict the service quality to understand the actual service of the service between the customers [3]. That would help the people to decide that whether service is satisfactory or not. Over last few years, online review became most famous and popular due to its easiness. Online review system are attractive with following key feature like 24×7 hours availability, easy access through internet, sharing from anywhere without any restriction of location. Service providers also analyze these reviews to understand the performance of their product and to understand the need of any up gradation [4]. Many retail company like Flipkart, Amazon have started to work on these concept to analyze the feedback of their customers about the product they sell. Some company like Yelp has also offered online details review system of local shops, hotels, restaurants, tourist places and other business. The reviews of customers about any business are more valuable for service providers. These all types of text analysis based are many handle by sentiment analysis and opinion mining. Sentiment analysis is mainly categorized into machine learning based sentiment analysis and lexical based sentiment analysis. Machine learning based sentiment analysis has two types [5]: supervised and unsupervised.

Sentiment Analysis Classification

1. Machine Learning Approach
   - A. Supervised Learning
     - I. Linear Classifiers
       - o Support Vector Machine
       - o Neural Network
     - II. Probabilistic Classifiers
       - o Naïve Bayes
       - o Bayesian Network
       - o Maximum Entropy
     - III. Decision Tree Classifiers
     - IV. Rule Based Classifiers
   - B. Unsupervised Learning
2. Lexical Based Approach
   - A. Dictionary Based Approach
   - B. Corpus Based Approach
     - I. Statistical
     - II. Semantic

Fig. 1. Sentiment classification techniques.

Both techniques have a number of approaches as given in Fig. 1. The lexical based approach also has two types: dictionary based sentiment analysis and corpus based analysis.
Various researchers have already worked on the above listed approach. In section II, we discuss some of the recent work published by different authors. Section III explains the basic idea of semantic rule based approach useful for sentiment analysis. In section IV, we elaborate on the proposed computational model with its architecture. Section V presents our experimental result analysis and discussion. Section VI concludes the paper and focus on future work.

II. LITERATURE SURVEY

Many articles have been presented till now in the area of sentiment analysis. The number of research work has been increased till now. Some related research has been discussed in the survey of this paper. In [6], a linguistic pattern based method is proposed to analyze the text content and perform sentiment analysis. The researcher focused on extra sentimental words like adjectives covering additional sentiment. To discover such additional words having sentimental expression, they applied a set of linguistic rules of connectives words like "but", "or", "and", "either", "neither" etc. An updated method is proposed by Jia *et al.,* [7], which includes negation identification in the sentence. Negation identification is very important to decide the actual polarity of the sentence. Sentence polarity using linguistic rules with identification of negation improves the performance of the system. It required a deep analysis to identify implicit negation of any sentence [8]. Explicit negative is easy to identify. In the proposed work, an implicit negation detection technique has three-layer each consist of eight features. In a sentiment analysis method is proposed to improve the analysis of textual data based on exploiting discourse similarities [9]. Here conjunctions, models, conditionals and connection words may have a significant effect on the polarity of tweets and textual posts. In [10], some rules applicable to opinion mining are proposed and these rules are combined in a concept that categorizes the sentence as positive and negative. This work analyzes the sentence having two or more words holding different polarities. These types of words together in a group produce a reverse meaning like "terrible good" is producing a negative feel. This is declared a sentiment conflicts. In analysis of conditional sentences is explained using the study of their linguistic structure [11]. They applied machine learning techniques to analyze such content. They perform dynamic analysis using supervised learning. In a framework of sentiment analysis is proposed which is based on text classification using a bag of word models [12]. The result of the model as compared to the basic bag of word model is improved for text classification. In a new concept based method proposed in which text is represented as a vector of concept [13]. It is known as a bag of concept [BOC] model. This vector container represents the main concept related to the particular text. For Example, a ball pen, cello pen, gel pen, etc are considered a pen only. This model enhances the accuracy of SVM for classification. In this model concept, vector and BOW are synonyms set to each other. The enhancement of the BOC model is proposed by [14]. The work shows that the words of a similar class would also have a similar representation. It represents a set of entities as a concept in a particular domain. In SVM and syntax inspired feature selection based sentimental analysis approach is proposed for different applications like customers review analysis to get good products in online shopping [15]. In predictions of several classifiers are combined for better sentimental results [16].

III. SEMANTIC RULES OF WORD CONCEPT

These are well designed semantic rules for analysis of sentences to evaluate its polarity. The main idea of these rules is to utilize them in sentiment analysis to function the system more accurately. Here we define some rules on different English words with the example; those apply to a highly significant word category like conjunction [17].

The proposed rule-based sentiment analysis algorithm is defined as follows: The basic concept of this algorithm is to find the polarity score of a sentence by the use of SenticNet [18]. SenticNet is a knowledge base that contains all words with its polarity and multiword polarity. The SenticNet version 3.0 is referring here. This knowledge-base has more than 25 thousand concepts of common sense.

These concepts contain situation oriented word sequence applicable on occasion_based_celebration, fun_supporting_situation, feeling_on_events, and favorable_activity, etc. These all word sequences represent some significant identity of sentiment. The computational sentiment analysis system needs to find the polarity of the sentence to declare it as positive or negative. The base approach to finding sentence polarity is just counting the word polarity with their positive and negative lexicon based on polarity magnitude.

Example 1: The subject is very tuff but it is rather not horrible.

Some words of this sentence are neutral and some are negative. The most significant words of the sentence (like tuff and horrible) are negative. The noun is neither negative nor positive. So the total polarity of this sentence must be negative as most of the words have negative polarities but it a positive sentence. In this example, the main role played by the words is “but” and “rather”. These two words change the overall polarity of the sentence from negative to positive. The flowchart for this situation is given using the dependency tree as follows.

Example 2: The class teacher made me pissed off.

Some words of this sentence are neutral and some are negative. The most significant words of the sentence (like tuff and horrible) are negative. The noun is neither negative nor positive. So the total polarity of this sentence must be negative. The sentence is finally declared as negative.

Example 3: Professor of this university hurt the cute girl.

Some words of this sentence are neutral, some are positive and some are negative. The most significant words of the sentence (like hurt and cute) are both positive and negative. The noun is neither negative nor positive. So the overall polarity of this sentence is not easy to evaluate using the above procedure. So some more rules need to define here.

According to the word type, a particular rule will be triggered and appropriate action will be performed [19]. The algorithm uses these rules in appropriate manners. The main word type in the English language is subject, object, verb, adjective, adverb, conjunction, proposition, etc.

Rule 1: The main logic of this rule is based on the relation between noun and verb. The logic of sentiment with noun and verb is defined here as follows. If both noun and verb found together in SenticNet, then polarity is directly assigned using the knowledge base of SenticNet, otherwise, the system needs to apply some further steps.

(a) If both noun and verb have positive polarity then would also be positive.
(b) If the noun has positive polarity and noun is the first person and verb have negative polarity then the sentence is positive.
(c) If the noun has negative polarity and verb have positive polarity then the sentence is considered as positive.
(d) If the sentence is not an active voice but passive and noun and verb have negative polarity then the sentence would be positive.

Examples

1. Student's issues were resolved by principle.

In this example, the verb 'resolve' has a relation with the subject 'student's issues'. Negative sentimental polarity has shown by word 'issue' while another word 'resolve' has positive sentimental polarity. In this situation, according to the rules overall sentimental polarity of the sentence is referred from the word 'resolve'. So the positive polarity is applicable to this whole sentence.

2. Their all achievements pissed off them.

In this example, 'achievement' is a positive sentimental word, which has a relation with the verb 'pissed off' which is a negative sentimental word. So the overall sentiment of the sentence declared as negative.

3. Internship offer was disgusting.

In this example, the subject 'internship offer' have a relation with adjective 'disgusting'. Positive sentimental polarity is showing by 'offer' and negative sentimental polarity shown by 'disgusting'. The sentence will be evaluated according to rules as negative.

Rule 2: When a sentence contains some conjunctive words like 'that', 'those', 'whether' etc. then sentence have two parts in the sentence. In this situation, we need to find the sentiment of both parts separately and then apply the following rules to generate final sentimental polarity of the sentence.
(a) If the first part is showing any type of sentiments, then this one will be the overall sentiment of both parts. The second part has no role in sentiment evaluation.
(b) If the first part is not showing any sentiment then sentiment shown by second part is final for the overall sentence.
(c) If there is negation in the first part and it also not has any sentiment then resultant sentiment, the sentiment of the second part will be inverted.

Examples

1. The student love that the tuition fees did not increase in their university.

The sentence has positive sentimental polarity in overall because according to rules, the first part of the sentence is a positive and second part is not considered.

2. Professor does not check whether students are cheating during examination.

The sentence has negative sentimental polarity in overall because according to rules, the first part of the sentence is a negative and second part is not considered.

3. The student doesn't know whether their faculty is gentle and highly qualified.

In this sentence, there is no sentiment shown by the first part of the sentence. But it has one negation that gives alertness to the second part of the sentence. So by the rules, the sentiment of the second part is inverted and the overall sentence is finally declared as negative.

Rule 3: When a sentence contains 'while' words then this rule is applicable. 'while' works the same as 'but' words. The main sentence breaks down into two parts using the subject and comma presented in the sentence has two parts in the sentence. In this situation, we need to find the sentiment of both part separately and then apply the following rules to generate final sentimental polarity of the sentence is evaluated by the second part of the sentence.

Examples

1. While I am sure the preparation is fine, the result of the exam is very different.
2. While walking, his brother completed today's exercise.
3. While students shouting, the teacher is explaining confidently.

Rule 4: When the sentence contains preposition 'against' then this rule is applicable. This word shoes negative polarity so the overall sentence polarity is inverted.

Examples

1. The professor is against all cheating during examination.
2. Our director is against our internship.
3. Students are against hostel warden.

IV. COMPUTATIONAL MODEL

The working of the computational model to evaluate sentiment polarity is explained here. Here we describe the approach which is responsible for the utilization of all semantic rules in the sentence analysis. The first task is to find the type of word present in the sentence and find the relation between them. The words that are linked together are mainly responsible for sentiment of overall sentence. This applies syntax analysis to validate and indent its type of each word of the sentence and prepare a dependency tree as shown below.

The dependency tree is modified with level to each node, leaf and root word, which provides a flow of computation to evaluate the sentiment of the overall sentence. Leaf word levels are associated with the relation of each other and flow of execution finally reached to root word to generate final sentiment. Once the dependency tree is generated with the proper level, the parsing needs to start from leftmost leaf and parallelly the linguistic rules are applied to evaluate sentiment polarity of the sentence. This procedure is the same as semantic rules are applied.

For every sentence, we find a relation between the words. Relation (w1, w2) represents the relation between the words of the sentence. Every relation is processed by SenticNet and linguistic rules to generate the polarity as discussed in the previous section. According to the category of w1 and w2, the particular rule is to be activated, and computation performed. If no rule is suitable for relation (w1, w2), the control is transferred to the next relation of dependency tree towards the right side. Words involved in previous relations are also checked with another word combination and the above procedure is repeated. The algorithm gives the polarity of relation (w1, w2) that was the last relation of the parse tree. It holds the sentiment of all sentences because the final relation relies on the previous relation of left to right scanning of the dependency tree. Sentiment polarity mainly depends on the efficiency and accuracy of SenticNet and proposed sentence is the main key technique with SenticNet knowledge base. We have also combined the general classification methods based on machine learning to classify and generate sentiment polarity for enhancing the system performance [20].

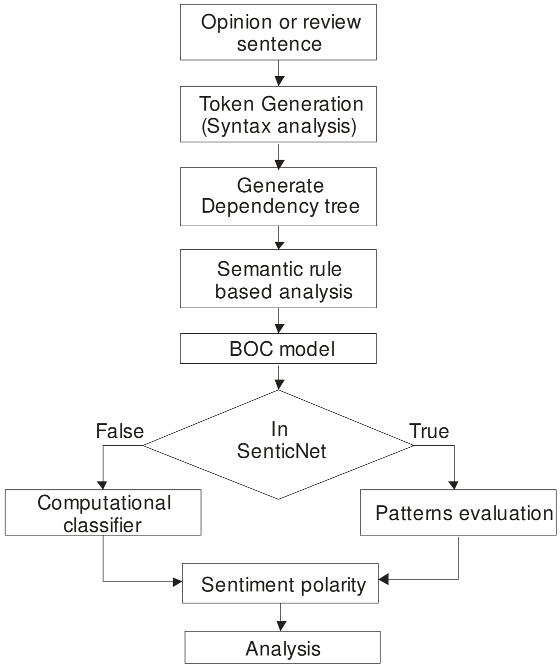

Fig. 2. Computational model of sentiment analysis.

The classifier has been trained using 60% data of predefined dataset. The control is transferred to classifier for generating sentiment polarity of the sentence if the above algorithm does not find any rule suitable for that sentence. All sentences those are processed by SenticNet rules do not require classifier.

V. RESULT AND DISCUSSION

For sentiment analysis, SenticNet computational model outperforms on a different dataset. Some of the results are very impressive. The proposed system is robust and stable across the dynamic dataset. It improves the functionality of the general classification technique

The proposed algorithm is very helpful in all kind of dataset because general classification techniques have also included covering all kind of sentences. Semantic analysis enhanced the quality of the system. SenticNet and linguistic rules help to improve the performance of the system.

The experiments show the performance of the proposed model with a comparison of existing techniques. The main parameter used in the experiment is Accuracy based on some standard like precision and recall initially proposed by Baeza Yates [1999]. The proposed method has been applied to the data used in SemEval–2014 [21]. The dataset has more than 2500 sentences for training and approx 750 sentences for testing. Each sentence contains one or more sentimental words.

In the dataset, the sentences having at most two sentimental words are 80%, remaining is having more than 2 sentimental words. This is the main reason why rules. Concept of the relation between two words of the the semantic linguistic rule is effectively applied to the dataset.

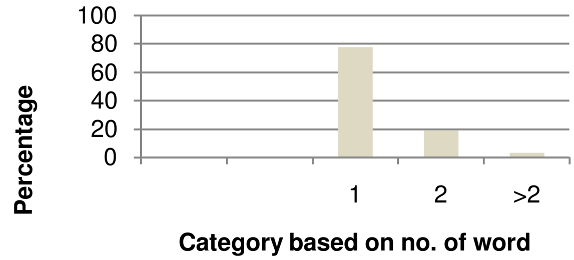

Fig. 3. Distribution of the number of word categories per sentence.

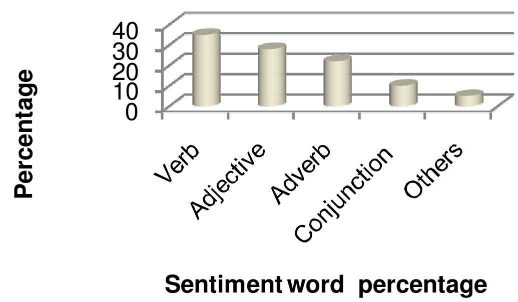

Fig. 4. Relative frequency of the sentimental word categories.

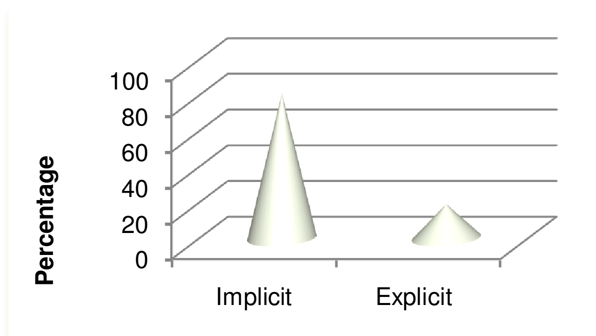

Fig. 5. Ratio of implicit and explicitly sentiment categories.

Percentage of the sentimental word shown by Fig. 4, represent that adjective and verb are more than 65%, the rest of 35% words are others.

Verb and adjective come in the big category found in the dataset. It is also found that either sentence has an implicit or explicit sentimental presence. It is shown by Fig. 5, indicating that most of the sentence is in the implicit category, few are in the explicit category.

Precision indicates the capability of a system to retrieve only relevant sentences. It is the fraction of retrieved sentences that are relevant to each product, event or system.

Recall means the capability of a system to retrieve only relevant content useful in sentiment analysis. It is the fraction of total relevant content that has been retrieved from the system [22] and was evaluated using the confusion matrix terms.

	Positive	Negative	
System Positive	True Positive	False Positive	Precision= $\frac{TP}{TP+FP}$
System Negative	False Negative	True Negative	
	Recall= $\frac{TP}{TP+FN}$		Accuracy=$\frac{TP+TN}{TP+FP+TN+FN}$

Fig. 6. Formulation of precision recall and accuracy.

The true positive (TP) is the number of the positive sentence in the system. It is the relevant sentences of the dataset that are correctly identified by the system [23]. False positive (FP) is the number of sentences that are not relevant in the system but all have been incorrectly identified by the system as relevant sentences. False Negative (FN) is the number of relevant sentences that are not identified by the system just because some miss happening occurs [24].

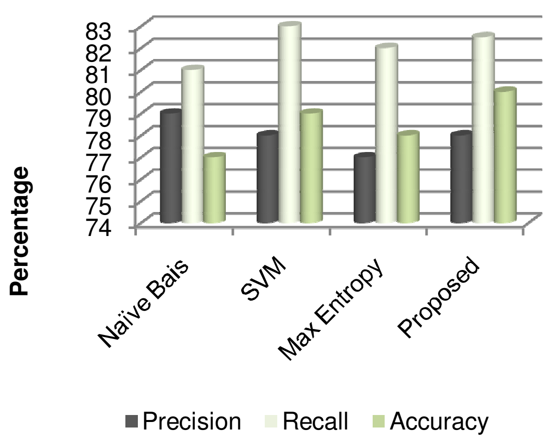

Fig. 7. Result of different technique with proposed.

All classifiers are implemented using Python code. Naive Bays has better precision compared to the other three classifiers, but slightly lower accuracy and recall. SVM, Maximum Entropy Classifier and proposed system have similar precision and recall but the proposed system has better accuracy.

VI. CONCLUSION

The model shows, how linguistics pattern can be blended in order to understand sentiment associated with the text. The presented approach will combine the use of various semantic rule based on the syntactic structure of the sentences. There are many approaches that analyze sentiments but hardly any work accomplished on word category based analysis. Data gather by user reviews are enormous, noisy, unstructured, and dynamic in nature, and thus novel challenges arise. Overall, the proposed approach, relying on the linguistic sentiment flow algorithm, has outperformed the majority of the main existing approaches on the benchmark datasets, showing outstanding effectiveness.

ACKNOWLEDGEMENT

We are very thankful to all the faculty members, research scholars and other supporting staffs of MANIT, Bhopal, India, for the continued encourages and support to complete this research work.

CONFLICT OF INTEREST

There is no conflict of research in this research article. The manuscript has not been submitted to, nor is under review in another journal or other publishing venue.

REFERENCES

[1]. Cambria, E., Schuller, B., Liu, B., Wang, H. & Havasi, C. (2013). Knowledge-Based Approaches to Concept-Level Sentiment Analysis. *IEEE Intelligent Systems,* Vol. *28*(2): 12-14.

[2]. Rizwana, K. H., & Kalpana, B. (2018). An ontology based sentiment analysis using protege software. *International Journal of Theoretical & Applied Sciences,* Vol. *10*(1): 41-46.

[3]. Pan, S. J., Ni, X., Sun, J. T., Yang, Q., & Chen, Z. (2010). Cross domain sentiment classification via spectral feature alignment. *19th International Conference on World Wide Web,* 751-760.

[4]. Bollegala, D., Weir, D., & Carroll, J. (2013). Cross domain sentiment classification using a sentiment sensitive thesaurus. *IEEE Transaction Knowledge,* Vol. *25*(8): 1719-1731.

[5]. Poria, S., & Cambria, E. (2015). Sentiment Data Flow Analysis by means of Dynamic Linguistic Pattern. *IEEE Computational Intelligence Magazine,* Vol. *10*: 26-36.

[6]. Hatzivassiloglou, V., & McKeown, K. R. (1997). Predicting the semantic orientation of adjectives. *35th Annual Meeting Association Computational Linguistics 8th Conference European Chapter Association Computational Linguistics*, 98/EACL: 174-181.

[7]. Jia, L., Yu, C., & Meng, W. (2009). The effect of negation on sentiment analysis and retrieval effectiveness. *International Conference on Information and Knowledge Management, Proceedings*, 1827-1830.

[8]. Reyes, A., & Paolo, R. (2013). On the difficulty of automatically detecting irony: Beyond a simple case of negation. *Knowledge and Information Systems*, *40*.

[9]. Chawla, K., Ramteke, A., & Bhattacharyya, P. (2013). IITB Sentiment Analysts: Participation in Sentiment Analysis in Twitter SemEval Task. *2nd Joint Conference Lexical Computational Semantics*, SEM-2: 495-500.

[10]. Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies*. Vol. *5*: 1-167.

[11]. Narayanan, R., Liu, B., & Choudhary, A. (2009). Sentiment Analysis of Conditional Sentences. *EMNLP '09 Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, Vol. *1*: 180-189.

[12]. Kim, Y. H., Hahn, S. Y., & Zhang, B. T. (2000). Text filtering by boosting naive bayes classifiers.

Proceedings of the 23rd annual international ACM SIGIR conference on Research and development in information retrieval, 168-175.
[13]. Sahlgren, M., & Cöster, R. (2004). Using bag-of-concepts to improve the performance of support vector machines in text categorization. *Proceedings of the 20th international conference on Computational Linguistics*, 487.
[14]. Wang, F., Wang, Z., & Wen, J. R. (2014). Concept-based short text classification and ranking. *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*, 1069-1078.
[15]. Venkatesan, D., Vithiya Ruba, K., & Sekar, K. R. (2019). Investigation of various techniques and classification methods on cognitive sentimental learning. *International Journal on Emerging Technologies*, Vol. *10*(2): 15-18.
[16]. Arya, P., Bhagat, A., & Nair, R. (2019). Improved performance of machine learning algorithms via ensemble learning methods of sentiment analysis. *International Journal on Emerging Technologies*, Vol. *10*(2): 110-116.
[17]. Poria, S., Cambria, E., Winterstein, G., & Huang, G. B. (2014). Sentic Patterns: Dependency based rules for concept-level sentiment analysis. *Knowledge based system*. Vol. *69*: 45-63.
[18]. Cambria, E., Olsher, D., & Rajagopal, D. (2014). SenticNet 3: A common and common sense knowledge base for cognition driven sentiment analysis. *Proceedings of the Twenty Eighth AAAI Conference on Artificial Intelligence*, 1515-1521.
[19]. Poria, S., Cambria, E., Erik, G., Alexander, B., Federica, H., & Hussain, A. (2015). sentiment data flow analysis by means of dynamic linguistic patterns. *IEEE Computational Intelligence Magazine.* Vol. *10*: 26-36.
[20]. Schouten, K., Weijde, O., & Frasincar, F. (2017). Supervised and unsupervised aspect category detection for sentiment analysis with co-occurrence data. *IEEE Transactions on cybernetics*, 1-13.
[21]. Pontiki, M. Galanis, D., Pavlopoulos, J., Papageorgiou, H., Androutsopoulos, I., & Manandhar, S. (2014). SemEval-2014 Task 4: aspect based sentiment analysis. *Proceedings of the 8th international workshop on semantic evaluation*, 27-35.
[22]. Vaghela, V. B., & Jadav, B. M. (2016). Analysis of various sentiment classification techniques. *International Journal of Computer Applications, Foundation of Computer Science (FCS), NY, USA,* Vol. *140*(3): 22-27.
[23]. Basha, S. M., Bagyalakshmi, K., Ramesh, C., Rahim, R., Manikandan, R., & Kumar, A. (2019). Comparative study on performance of document classification using supervised machine learning algorithms: KNIME. *International Journal on Emerging Technologies*, Vol. 10(1): 148-153.
[24]. Kechaou, Z., Ammar, M. B., & Alimi, A. M. (2011). Improving E-learning with sentiment analysis of users' opinions. *Global Engineering Education Conference (EDUCON), 2011 IEEE*, 1032-1038.

How to cite this article: Sahu, Y., Thakur, G.S. and Dhyani, S. (2019). Semantic Rules based linguistic Computational Model of Sentiment Analysis. *International Journal on Emerging Technologies*, **10**(3): 238–243.