International Journal Of Engineering Research And Development
e-ISSN: 2278-067X, p-ISSN: 2278-800X, www.ijerd.com
Volume 13, Issue 9 (September 2017), PP.43-51

Rough Set Based Opinion Mining in Tamil

*Sharmista A[1], Ramaswami M[2]
[1,2]Madurai Kamaraj University Tamil Nadu, India*
*Corresponding Author: *Sharmista A*

ABSTRACT: - As individuals impart on the Web about their sentiments on products and services they have used, it has become important to formulate methods of automatically classifying and judging them. The task of examining such data, collectively called client feedback data, is known as opinion mining. Opinion mining consists of several steps, and different techniques have been proposed at each stage in this process. This paper basically explains such techniques that have been used for the implementation of task of opinion mining in Tamil. On the basis of this analysis we provide an overall system design for the development of opinion mining approach. In this paper, the rough set theory is useful to extract the key sentences and its feature attribute after getting the opinion from the post will be evaluated. Before this operation the pre-processing steps will be discussed for finding the entity and its attribute, on the basis of the output the rough set theory is used for avoiding the ambiguities between the word sense sentences. Using Rough set we generated the result for three class (**நேர்மறை** (Positive), **யாரும்** (None), **எதிர்மறை** (Negative)) and five class (**மிகவும் சாதகமான** (More Positive), **நேர்மறை** (Positive), **யாரும்** (None), **மிகவும் எதிர்மறை** (More Negative), **எதிர்மறை** (Negative)) problems.

Keywords: Opinion Mining, Customer Feedback, Feature Extraction, Rough Set Theory, Sentiment Classification, Opinion Summarization.

I. INTRODUCTION

The text classification research in the field started with some problem related with sentiment and subjectivity classification. Sentiment classification classifies whether a feedback document (such as product reviews) or sentence expresses a positive or negative opinion. Subjectivity of classification determines whether a sentence is subjective or objective [2]. In day to day life user wants to know about the subject opinion. For example, from a product review, users want to know which product features consumers have praised and criticized. To explore this generic problem, let‘s use the following review segment on Camera as an example: (1) நான் சோனி ஐபோன் வாங்கினேன். (2) இது ஒரு நல்ல தொலைபேசி. (3) படம் தரம் மிகவும் குளிராக இருந்தது. (4) வீடியோ பதிவு தரம் கூட தெளிவாக இருந்தது. (5) இருப்பினும், நான் அதை வாங்கி முன் நான் சொல்லவில்லை என என் சகோதரன் என்னுடன் கோபமாக இருந்தார். (6) அவர் தொலைபேசி மிகவும் விலையுயர்ந்ததாக நினைத்தேன், அதை கடைக்குத் திருப்பிக் கொடுக்க விரு.

The question is, what do we want to extract from this review? The first thing that we might notice is that there are several opinions in this review. Sentences 2, 3, and 4 express three positive opinions, while sentences 5 and 6 express negative opinions or emotions.

We can also see that all the opinions are expressed about some targets or objects. For example, the opinion in sentence 2 is on the phone as a whole, and the opinions in sentences 3 and 4 are on the phone‘s picture and video recording quality, respectively. Importantly, the opinion in sentence 6 is on the phone‘s price, but the opinion/emotion in sentence 5 is about person, not the phone. In an application, the user might be interested in opinions on certain targets but not necessarily on user-specific information. Finally, we can also see the sources or holders of opinions. The source or holder of the opinions in sentences 2, 3, and 4 is the author of the review, but in sentences 5 and 6, it is about brotherhood.

In this example, we can extract several phrases such as நல்ல தொலைபேசி, உண்மையில், வீடியோ பதிவு தரம் தெளிவாக இருந்தது, அவரிடம் சொல்லவில்லை and தொலைபேசி மிகவும் விலை உயர்ந்தது which convey customer‘s opinion rather than facts. In particular, subjective words such as நல்ல, உண்மையில், தெளிவான, இல்லை and மிக விலை உயர்ந்த are used to express customer‘s positive/negative sentiment regarding the product features, which are referred by தொலைபேசி, படம், காணொளி. Although information gathered from multiple reviews are more reliable compared to information from only one review, manually sorting through large amounts of review one by one requires a lot of time and cost for both businesses and customers. Therefore it is more efficient to automatically process the various reviews and provide the necessary information in a summarized form.
Because of the importance of automatically extracting actionable knowledge from customer feedback data on the Web, opinion mining (OM) has become a significant subject of research in the field of data mining. The ultimate goal of OM is to extract customer opinions (feedback) on products and present the information in the most efficient way that serves the objectives chosen. This means that the necessary steps and techniques used for OM can be different depending on how the summarized information is presented [12].

For example, if we were to get the number of negative and positive reviews about a given product, classifying each review as positive or negative would be the most important task. On the other hand, if we want to show customer feedback on each of the different features of a product, it is necessary to extract product features and analyze the overall sentiment of each feature [1]. However, the focus of opinion mining is on the sentiment that the customer is expressing and this is where the methods are applied differently. As can be seen from the prior example, making the linguistic distinction between objective words that express facts and subjective words that express opinions is important. In this paper, we examine the two tasks that are specific to opinion mining: development of linguistic resources and sentiment classification in Tamil. And we have analysed the reviews for three class (Positive - நேர்மறை, None - யாரும், Negative - எதிர்மறை) and five class (More Positive - மிகவும் சாதகமான, Positive - நேர்மறை, None - யாரும், More Negative - மிகவும் எதிர்மறை, Negative - எதிர்மறை) problems.

The remainder of the paper is organized as follows. Section 2 introduces the proposed architecture for opinion mining and section 3 and 4 presents the methods to be used for OM in this architecture i.e. Rough Set Theory, dataset used and in section 5, we discuss about experimental results and discussions and in section 6 conclude the paper.

II. PROPOSED ARCHITECTURE FOR OPINION MINING

As mentioned in section 1, opinion mining can be roughly divided into three major tasks of development of linguistic resources, sentiment classification, and opinion summarization. We graphically present these tasks and the areas of research to which they are related in fig. 1. The overall architecture of the system is shown in fig. 1. There are four process phases organized as a pipeline in the system. The functions of these process phases are described in the following subsections. Basant Agarwal, Namita Mittal [1] are among the first to deal with opinion classification. They construct a rough set model so as to clarify whether two adjectives have the same orientation. The accuracy of this task is declared to be 82%. The techniques used for text classification and text summarization can also be applied to OM, along with linguistic resources [3]. Although sentiment classification and opinion summarization share several steps or techniques, sentiment classification focuses on classifying each review and opinion summarization is about how to effectively extract opinion expressions and summarize them from a large number of reviews of a given product.

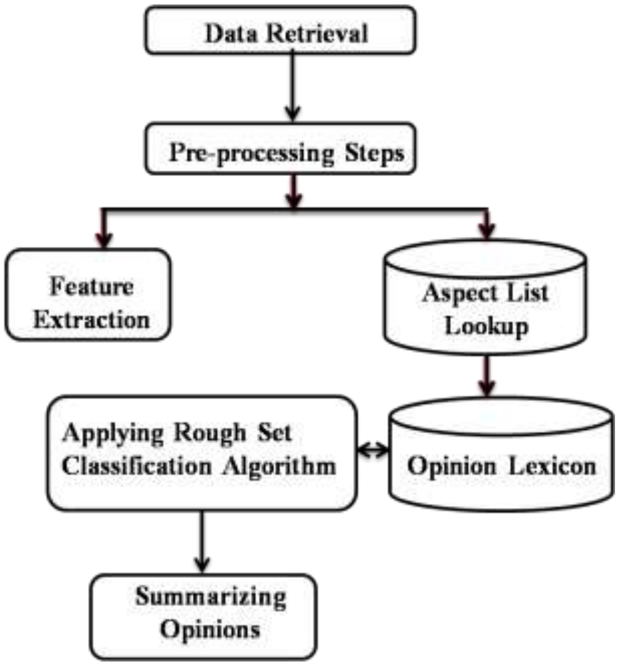

Fig.1: Flow for generating opinion from feedback

Sentiment related properties are well defined in appraisal theory [10], a framework of linguistic resources for describing how writers and speakers express inter-subjective and ideological positions. However, most researches for developing linguistic resources have focused on determining three properties: subjectivity, orientation, and strength of term attitude. For example, நல்ல, சிறந்த are positive terms while கெட்ட,தவறு and மோசமான are negative terms. செங்குத்து, மஞ்சள் and திரவ are objective terms. சிறந்த and மோசமான are more intense than நல்ல and மோசமான.

A. Document Pre-processing

Currently, input is of plain text format. By applying part of NLP [14] steps for obtaining the token of words is as follows:

Statement extraction: firstly, from each post, sentences are individuated, that are parts of text ending with a full stop, comma, question mark, exclamation mark or semicolon. Subsequently, conjunctions are analyzed for dividing sentences into statements, which are parts of text expressing only one meaning.

Anaphora resolution: Often, in informal text, subject and predicate could be understood; hence some statements could be incomprehensible. The goal of the anaphora resolution step is to restore a statement by adding understood parts.

Tokenization: In this step, each statement is divided into tokens, which are parts of text bounded by a separator (space, tab or end of line).

Stemming and Lemmatization: In order to reduce the number of different terms, each token is transformed reducing its inflectional forms to a common base form. The main difference among stemming and lemmatization‘s that the former extract ―brutally the root of a word (e.g. bio is the stem of biology, biocatalyst and biochemical); while the latter uses a vocabulary (often a lexical ontology) for returning the dictionary form of a word, that is the lemma (e.g. be is the lemma of is, are, was, and so on).

Tagging and Stop words elimination: Some word categories are too common to be useful to distinguish among statements. Hence, in this step articles, prepositions and conjunctions are first recognized and then removed. At the same time, we also remove proper nouns, which usually don‘t have an affective content.

B. Word Sense Disambiguation

In automatic text summarization, word sense disambiguation is important and many different approaches have been taken. In this process phase, the Lesk [13] approach is adopted and modified for word sense disambiguation. The Lesk approach assumes that words used in a sentence are collaborative in terms of topic and their dictionary definitions, thus must use some common words in their sense definitions. Based upon this assumption, the nouns and the verbs are first extracted from each sentence together with their senses given in WordNet [10, 13] as the input to the following process for sense disambiguation:

1) For a word to be disambiguated, the process first scores the semantic relatedness between any two senses, one for this word and the other for any other word in the same sentence.

2) Note each sense in WordNet is semantically related with a set of similar senses. The score computed in the previous step reflects the direct relatedness between any two senses. It is like a local link. To fully reflect the relatedness between two senses, their indirect relatedness should be taken into account. The indirect semantic relatedness of two senses is the sum of the pair wise relatedness scores of the two set of similar senses.

3) The final score of each sense for the word is the sum of the score given by Step 1.) And the half value of the score given by Step 2.).

4) Among all the senses of the word, the sense with the highest score is selected as the candidate sense of the word. After all words in a sentence are disambiguated, this phase builds and reports the sense representation for the sentence in terms of WordNet senses to indicate what concept the sentence may cover.

There are two main streams in developing linguistic resources for OM: the NLP techniques and the Rough Set Theory.

C. Natural Language Parsing Techniques

The main idea for entity discovery [6, 7] is to discover linguistic patterns and then use the patterns to extract entity names. However, traditional methods need a large number of manually labeled training examples, and labelling is very time consuming. For a different domain, the labelling process may need to be repeated. This section proposes an automated pattern discovery method for the task, which is thus unsupervised. The basic idea of the algorithm is that the user starts with a few seed entities. The system bootstraps from them to find more entities in a set of documents (or posts). Sequential pattern mining is employed at each iteration to find more entities based on already found entities [11]. The iterative process ends when no new entity names are found.

Pruning methods are also proposed to remove those unlikely entities. Given a set of seed entities E = {e1, e2 …en}, the algorithm consists of the following iterative steps:

Step 1 – data preparation for sequential pattern Mining

Step 2 – Sequential pattern mining

Step 3 – Pattern matching to extract candidate Entities

Step 4 –Candidate pruning

Step 5 –Pruning using brand and model relation and Syntactic patterns.

III. ROUGH SET THEORY

Rough set theory was developed by Grzymala-Busse JW [4] on the assumption that with each object of the universe of discourse we associate some information, and the objects can be - seen only through the accessible information. Hence, the object with the same information cannot be discerned and appear as the same. These results, that indiscernible object of the universe forms clusters of indistinguishable objects which are often called granules or atoms. These granules are called elementary sets or concepts, and can be considered as elementary building blocks of knowledge. Elementary concepts can be combined into compound concepts, i.e. concepts that are uniquely defined in terms of elementary concepts. Any union of elementary sets is called a crisp set, and any other sets are referred to as rough (vague, imprecise). Consequently, each rough set has boundary-line cases, i.e., objects which cannot be with certainty classified as members of the set or its complement. Obviously crisp sets have no boundary-line elements at all. This means that boundary-line cases cannot be properly classified by employing the available knowledge. The main goal of rough set theoretic analysis is to synthesis approximation (upper and lower) of concepts from the acquired data. Rough set theory assigns to each objects a grade of belongingness to represent an imprecise set, the focus of rough set theory is on the ambiguity caused by limited discernibility of objects in the domain of discourse. However, the rough set theory has been successfully applied to solve many real-life problems which involve decision making approaches. The main advantage of rough set theory is that it does not need any preliminary or additional information about data like probability in statistics and the grade of membership or the value of possibility in fuzzy set theory. It has been found by investigation that hybrid systems which consist of different soft computing tools combined into one system often improve the quality of constructed system [9].

Recently, rough sets have been integrated in soft computing framework, the aim being to develop a model of uncertainty stronger than either [5]. Therefore, Rough set systems have a significant potential.

A. Information Systems

An Information system can be viewed as a pair Ŝ=<U, A>, or a function f: U × A→V, where U is a non-empty finite set of objects called the Universe, A is a non-empty finite set of attributes, such that a: U→ V_a for every a ε A. The set V_a is called the value set of a. In many applications, there is an outcome of classification that is known. This is a posterior knowledge is expressed by one distinguished attribute called decision attribute, the process is known as supervised learning. Information systems of this kind are called decision systems. A decision system is any information system of the form Â= (U, A Ü{d}), where d ∉ A is the decision attribute. The elements of A are called conditional attributes or simply conditions. The sentence attribute may take several values though binary outcomes are rather frequent [6].

B. Indiscernibility and Set Approximation

A post expresses all the knowledge available about a product. This post may be unnecessarily large because it is redundant in at least two ways. The same or the indiscernible entity may be represented several times, or some of the attributes may be superfluous [7]. With every subset of attributes B⊆ A, one can easily associate an equivalence relation IB on U: IB= {(x, y) ε U: for every a εB, a(x) = a(y)}. IB is called B-Indiscernibility relation. If (x, y) ε IB, the entity x and y are indiscernible from each other by attributes B. The equivalence classes of the partition induced by the B-Indiscernibility relation are denoted by [x]B. These are also known as granules. The partition induced by the equivalence relation IB can be used to build new subsets of the universe. Subsets that are most often of interest have the same value of the outcome attribute. It is here that the notion of rough set emerges. Although we cannot delineate the concept crisply, it is possible to delineate the entity which definitely —belong to the concept and those which definitely - do not belong to the concept.

Let Â= (U, A) be an information system and let B⊆A and X⊆U. We can approximate X using only the information contained in B by constructing the B-lower and B-upper approximations of X, denoted as X and X respectively, where X=* |, - ⊆ + and * |, - +. The objects in X can be certainly classified as the members of X on the basis of knowledge in B and the objects in can only be classified as possible members of X on the basis of B. The set $B_N B(X)$ = B X. BX is called the B-boundary region of X and thus consists of those objects that we cannot decisively classify into X on the basis of knowledge of B. Thus, a set is said to be rough if the boundary region is non-empty otherwise crisp (boundary region is empty).

C. Reducts

Indiscernibility relation reduces the data by identifying equivalence feature, i.e. entity that is indiscernible, using the available attributes. Only one element of the equivalence feature is needed to represent the entire set. Reduction can also be done by keeping only those attributes that preserve the indiscernibility relation and, consequently, set approximation. So, one is, in effect, looking for minimal set of attributes taken from the initial set A, so that the minimal set induce the same partition on the domain of A. In other words, the essence of the information remains intact and the superfluous attributes are removed. The below sets of attributes are called reducts. Intersection of all reducts is called the core. Reducts have been clearly characterized in by discernibility matrices and discernibility functions [8].

Let us consider U={X_1,…,X_n}and A={a_1,…..,a_n} in the information system Ŝ=<U, A>.By the discernibility matrix M (Ŝ) of Ŝ is meant an n×n-matrix (symmetrical with empty diagonal) with entries Cij Ŝ as follows: Cij = {aεA: a (xi) ≠a (xj)}. A discernibility function fŜ is a function of m Boolean variables a 1,…, a m corresponding to the attributes a1,….,am respectively, and defined as follows: fŜ (a 1,…, a m) = { (Cij): 1≤i , j ≤n, j<i ,Cij ≠Ø} Where C_{ij} is the disjunction of all variables with aεC_{ij} . It is seen in [15] that {a_1,….,a_m} is a Reducts in Ŝ if and only a_{i1} ….. a_{ij} is a prime implicate (constituent of the disjunctive normal form) of fŜ.

IV. EXPERIMENTAL RESULTS AND DISCUSSIONS

R is a sophisticated statistical software package, which provides new approaches to data mining. We have analyzed an effect of product review dataset obtained from Opinion Lexicon. The rough set classification algorithm is executed to predict the best product by identifying the total number of positive opinions. The number of instances used for analysis of product data is 500.

Table I. shows results of rough set classification (3 class) with opinion lexicon for product reviews dataset run on R platform. It depicts the நேர்மறை (Positive), யாரும் (None), எதிர்மறை (Negative) reviews which are classified based on the number of instances 500. It is predicted that there are 207 நேர்மறை (Positive) reviews, 66 யாரும் (None) classified reviews with 61 error rate and 106 எதிர்மறை (Negative) reviews in case when matched with opinion lexicon.

Table I: Confusion Matrix of Rough Set Classification (3 Class) With Opinion Lexicon

	Predicted Class		
	நேர்மறை (Positive)	யாரும் (None)	எதிர்மறை (Negative)
Actual Class	**207**	0	0
	61	**66**	0
	60	0	**106**

Table II. Shows results of Rough set classification (5 Class) with opinion lexicon for product reviews dataset run on R platform. It depicts the மிகவும் சாதகமான (More Positive)**,** நேர்மறை (Positive), யாரும் (None), மிகவும் எதிர்மறை (More Negative), எதிர்மறை (Negative) reviews which are classified based on the number of instances 500. It is predicted that there are 17 மிகவும் சாதகமான (More Positive) reviews, 185 நேர்மறை (Positive) classified reviews , 65 யாரும் (None) reviews, 23 மிகவும் எதிர்மறை (More Negative) reviews and 89 எதிர்மறை (Negative) reviews in case when matched with opinion lexicon.

Table II: Confusion Matrix of Rough Set Classification (5 Class) With Opinion Lexicon

	Predicted Class				
	மிகவும் சாதகமான (More Positive)	நேர்மறை (Positive)	யாரும் (None)	மிகவும் எதிர்மறை (More Negative)	எதிர்மறை (Negative)
Actual Class	**17**	4	0	0	0
	0	**185**	0	0	0
	0	62	**65**	0	0
	0	4	0	**23**	0
	0	51	0	0	**89**

In general, the performance of sentiment classification is evaluated by using eight indexes. They are Specificity, Sensitivity, Precision, Recall, F1-measure, G-measure, Detection rate and Accuracy [12]. The common way for computing these indexes is based on the result shown below:

Table III: Overall Performance of Rough Set Classification (3 Class) with Opinion Lexicon

Machine Learning Approach			
Algorithm	Rough Set (3 Class)		
Class	நேர்மறை (Positive)	யாரும் (None)	எதிர்மறை (Negative)
Specificity	1	0.8594	0.8477
Sensitivity	0.6310	1	1
Precision	0.656	0.132	0.212
Recall	0.414	0.132	0.212
F1-measure	0.7738	0.6839	0.7794
G-measure	0.7944	0.7209	0.7991
Detection Rate	0.414	0.132	0.212
Accuracy	0.8155	0.9297	0.9239

Table IV: Overall Performance of Rough Set Classification (5 Class) with Opinion Lexicon

Machine Learning Approach					
Algorithm	Rough Set (5 Class)				
Class	மிகவும் சாதகமான (More Positive)	நேர்மறை (Positive)	யாரும் (None)	மிகவும் எதிர்மறை (More Negative)	எதிர்மறை (Negative)
Specificity	0.9917	1	0.8575	0.9916	0.8759
Sensitivity	1	0.6046	1	1	1
Precision	0.034	0.612	0.13	0.046	0.178
Recall	0.034	0.37	0.13	0.046	0.178
F1-measure	0.8947	0.7536	0.6771	0.92	0.7772
G-measure	0.8997	0.7775	0.7154	0.9230	0.7973
Detection Rate	0.034	0.37	0.13	0.046	0.178
Accuracy	0.9959	0.8023	0.9287	0.9958	0.9380

We calculate accuracy of classifier for both Rough set (3 Class) and Rough set (5 Class) classification, by training the algorithm on 500 sentences for each class of pre-classification dataset and applying it on the rest of the remaining datasets. Analysing the algorithm efficiency for the above dataset we achieved a 0.8155 value of correct classification of opinions for Rough set (3 Class). For Rough set (5 Class) most positive classification we have achieved 0.9959 and 0.8023 for correct positive classification. From the result we conclude that Rough set most positive (5 Class) classification has more accuracy than Rough set (3 Class) classification.

V. CONCLUSION

In this paper, we have presented a rough set aided extractive summarizer with better information coverage, redundancy reducing and techniques involved in developing an automated system for mining opinions that are found in customer feedback data on the Web. The rough set theory approach gives the exact result with the help of boundary extraction and also there are spaces for improvement if the tools and methods used for document pre-processing, word sense disambiguation and sequential pattern of sentences are perfected. In addition, the method implementation is in process and is not necessarily adequate. Furthermore, opinion mining has become important for all types of organizations, including for-profit corporations, government agencies, educational institutions, non-profit organizations, the military, etc. to gauge the opinions, likes and dislikes, and the intensity of the likes and dislikes, of the products, services, and policies they offer and plan to offer. As such, we believe that an understanding of the overall picture of the tasks and techniques involved in opinion mining is of significant importance. In future, more methods can be explored for making rough set based feature selection method computationally more efficient by incorporating evolutionary approaches in selecting feature subsets.

REFERENCES

[1]. Basant Agarwal, Namita Mittal, "Categorical Probability Proportion Difference (CPPD): A Feature Selection Method for Sentiment Classification", In Proceedings of the 2nd Workshop on Sentiment Analysis where AI meets Psychology (SAAIP), COLING 2012, pp 17–26, 2012.

[2]. Basant Agarwal, Namita Mittal, "Optimal Feature Selection Methods for Sentiment Analysis", In 14th International Conference on Intelligent Text Processing and Computational Linguistics (CICLing 2013), Vol-7817,pp:13-24, 2013.

[3]. Domenico Potena, Claudia Diamantini. Mining Opinions on the Basis of Their Affectivity in 2010 IEEE, pp.245-254.

[4]. Grzymala-Busse JW (2004c) Data with missing attribute values: Generalization of discernibility relation and rule induction. Transactions on Rough Sets, Lecture Notes in Computer Science Journal Subline, Springer-Verlag, 1: 78–95

[5]. Grzymala-Busse JW, Siddhaye S (2004). Rough set approaches to rule induction from incomplete data. Proceedings of the IPMU'2004, the 10th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems,Perugia, Italy, vol. 2, 923–930.

[6]. K. Khan, B. Baharudin, and A. Khan, "Identifying product features from customer reviews using hybrid patterns," Int. Arab J. Inf. Technol., vol. 11, no. 3, pp. 281–286, 2014. [Online].Available: http://www.ccis2k.org/iajit/index.php?option=com_content&task=blogcategory&id=92&Itemid=353

[7]. Lingyan Ji, Hanxiao Shi, Mengli Li, Mengxia Cai, Peiqi Feng. Opinion Mining of Product Reviews Based on Semantic Role Labeling in The 5th International Conference on Computer Science Education Hefei, China. August 24– 27, 2010,pp.1450-1453.

[8]. Q. Liu, B. Liu, Y. Zhang, D. S. Kim, and Z. Gao, "Improving opinion aspect extraction using semantic similarity and aspect associations," in Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, February 12-17, 2016, Phoenix, Arizona, USA., 2016, pp. 2986–2992.[Online]. Available: http://www.aaai.org/ocs/index.php/AAAI/AAAI16/paper/view/11973.

[9]. L. Polkowski, S. Tsumoto, and T. Y. Lin, Eds., Rough Set Methods and Applications – New Developments in Knowledge Discovery in Information Systems. Springer, 2000, to appear.

[10]. Rajendran, S., S. Arulmozi, B. Kumara Shanmugam, S. Baskaran, and S. Thiagarajan. 2002. "Tamil WordNet." Proceedings of the First International Global WordNet Conference. Mysore: CIIL, 271- 274.

[11]. Sandeep V. Sabnani. 2008. Computer Security: A Machine Learning Approach, Technical Report, MSc in Information Security at Royal Holloway, University of London.

[12]. Sumeet V.Vinchurkar, Smita M. Nirkhi. Feature Extraction of Product from Customer Feedback through Blog in International Journal of Emerging Technology and Advanced Engineering (IJETAE), Volume 2, Issue 1, January 2012 (ISSN 2250 – 2459) page No.314-323.

[13]. Satanjeev Banerjee and Ted Pedersen, ―An Adapted Lesk Algorithm for Word Sense Disambiguation Using WordNet‖ In Proceedings of the Third International Conference on Intelligent Text Processing and Computational Linguistics,Mexico City, pp. 136–145, 2002.

[14]. Thomas, L. a. (1993). A grammar of modern Tamil. Xiaowen Ding, Bing Liu, Lei Zhang. Entity Discovery and Assignment for Opinion Mining Applications KDD'09, June 28– July 1, 2009, Paris, France.

*Sharmista A. "Rough Set Based Opinion Mining In Tamil." International Journal Of Engineering Research And Development , vol. 13, no. 09, 2017, pp. 43–51.