International Journal of Scientific Research in Science, Engineering and Technology (www.ijsrset.com)
© 2020 IJSRSET | Volume 7 | Issue 1 | Print ISSN: 2395-1990 | Online ISSN : 2394-4099
DOI : https://doi.org/10.32628/IJSRSET207135

Sentimental Analysis and Deep Learning : A Survey

Princy Baby[1], Krishnapriya B[2]

[1]M. Tech Scholar, M Tech Scholar, Department of Computer Science and Engineering, GEC Idukki, Kerala, India

[2]Assistant Professor, Department of Computer Science and Engineering, GEC Idukki, Kerala, India

ABSTRACT

Sentiment Analysis is an ongoing field of research in text mining. Sentiment Analysis is the computational treatment of opinions, Sentiments, and subjectivity of text. Many recently proposed algorithms enhancements and various Sentiment Analysis applications are investigated and presented briefly in this survey. The related fields to Sentiment Analysis that attracted researchers recently are discussed. The main target of this survey is to give nearly full image of Sentiment Analysis techniques and the related fields with brief details. In recent years machine learning has received greater attention with the success of deep learning. Deep learning can create deep models of complex multivariate structures in structured data. Though deep learning can be characterized in several different ways, the most important is that deep learning can learn higher-order interactions among features using a cascade of many layers. Deep learning has been applied to neural networks and across many fields, with significant successes in many applications. Convolution neural networks, deep belief networks, and many other approaches have been proposed to enhance the abilities of deep structure networks

Keywords : Classification, deep learning, Sentiment Analysis

I. INTRODUCTION

Sentiment Analysis or Opinion Mining is the computational study of people's opinions, attitudes, and emotions toward an entity. The entity can represent individuals, events or topics. These topics are most likely to be covered by reviews. The two expressions of sentiment analysis are interchangeable. They express a mutual meaning. However, some researchers stated that OM and Sentiment Analysis have slightly different notions Opinion Mining extracts and analyzes people's opinions about an entity while Sentiment Analysis identifies the Sentiment expressed in a text then analyzes it. Therefore, the target of Sentiment Analysis is to find opinions, identify the Sentiments they express, and then classify their polarity as shown in Fig. 1. Sentiment Analysis can be considered a classification process as illustrated in Fig. 1. There are three main classification levels in document-level, sentence-Sentiment Analysis level, and aspect-level. Document-level Sentiment Analysis aims to classify an opinion document as expressing a positive or negative opinion or Sentiment. It considers the whole document a basic information unit. Sentence-level Sentiment Analysis aims to classify Sentiment expressed in each sentence. The first step is to identify whether the sentence is subjective or objective. If the sentence is subjective, Sentence-level Sentiment Analysis will determine whether the sentence expresses positive or negative opinions. Sentiment expressions are not Sentiment Analysis subjective in nature. However, there is no fundamental difference between document and sentence level classifications because sentences are just short documents. Classifying text at the document level or the sentence level does not provide the Sentiment Analysis detail needed opinions on all aspects of the entity which is needed in many applications, to obtain these details; we need to go to the aspect level. Aspect-level Sentiment Analysis aims to classify the Sentiment concerning the specific aspects of entities. The first step is to identify the entities and their aspects.

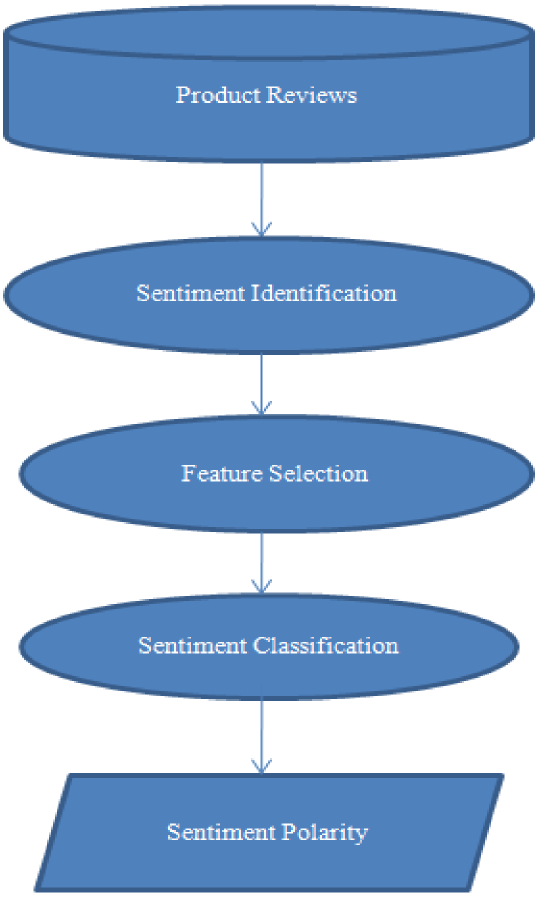

Fig 1 : Sentimental Analysis Process on Product Review

II. FEATURE SELECTION IN SENTIMENT CLASSIFICATION

A. Feature Selection Methods

Feature Selection methods can be divided into lexicon-based methods that need human annotation, and statistical methods which are automatic methods that are more frequently used. Lexicon-based approaches usually begin with a small set of 'seed' words. Then they bootstrap this set through synonym detection or on-line resources to obtain a larger lexicon.

The feature selection techniques treat the documents either as a group of words (Bag of Words (BOWs)) or as a string that retains the sequence of words in the document. BOW is used more often because of its simplicity for the classification process. The most common feature selection step is the removal of stop-words and stemming (returning the word to its stem or root i.e. flies fi fly).

B. Sentiment Classification Techniques

Sentiment Classification techniques can be roughly divided into machine learning approach, lexicon-based approach and hybrid approach. The Machine Learning Approach (ML) applies the famous ML algorithms and uses linguistic features. The Lexicon-based Approach relies on a Sentiment lexicon, a collection of known and precompiled Sentiment terms. It is divided into a dictionary-based approach and a corpus-based approach that uses statistical or semantic methods to find Sentiment polarity. The hybrid Approach combines both approaches and is very common with Sentiment lexicons playing a key role in the majority of methods.

The supervised methods make use of a large number of labeled training documents. The unsupervised methods are used when it is difficult to find these labeled training documents. The lexicon-based approach depends on finding the opinion lexicon which is used to analyze the text. There are two methods in this approach. The dictionary-based approach which depends on finding opinion seed words, and then searches the dictionary of their synonyms and antonyms. The corpus-based approach begins with a seed list of opinion words and then finds other opinion words in a large corpus to help in finding opinion words with context- specific orientations. This could be done by using statistical or semantic methods.

III. RELATED WORKS AND SURVEY

In this section, Give a brief introduction to previous work on Sentiment Analysis and user rating prediction.

A. Sentiment Analysis

Sentiment Analysis learns the positivity or negativity of products or services from user comments. Sentiment Analysis is based on an opinion lexicon. An opinion lexicon is a dictionary containing words, which express the polarity of words through positive or negative Sentiments, such as happy, good, bad, or disgusting. These opinion words are used in Sentiment Analysis as the key indicator to calculate the opinion of the user.

Y. Choi and C. Cardie [1] Introduced a set of heuristic- based methods for determining the polarity of a Sentiment bearing expression. Each assesses the polarity of the words or constituents using a polarity lexicon that indicates whether a word has positive or negative polarity, and finds negators in the given expression using a negator lexicon. The methods then infer the expression-level polarity using voting-based heuristics or heuristics that incorporate compositional semantics. The lexicons are described Compose first checks whether the first argument is a negator, and if so, flips the polarity of the second argument. Otherwise, Compose resolves the polarities of its two arguments. Note that if the second argument is a negator, we do not flip the polarity of the first argument, because the first argument, in general, is not in the semantic scope of the negation. Instead, we treat the second argument as a

constituent with negative polarity. Experiment with two variations of the Compose function depending on how conflicting polarities are resolved: COMPOMC uses a Compose function that defaults to the Majority Class of the polarity of the data, while COMPOPR uses a Compose function that selects the polarity of the argument that has higher semantic Priority. For brevity, we refer to COMPOPR and COMPOMC collectively as COMPO.

B. Phrase Dependency Parsing

Y. Wu, Q. Zhang, X. Huang, and L [2] Introduced a Dependency grammar is a kind of syntactic theories presented. Independency grammar, structure is determined by the relation between a head and its dependents. In general, the dependent is a modifier or complement; the head plays a more important role in determining the behaviors of the pair. Therefore, criteria of how to establish dependency relations and how to distinguish the head and dependent in such relations are a central problem for dependency grammar. There are seven pairs of dependency relationships, depicted by seven arcs from heads to dependents. the mainstream of dependency parsing is conducted on lexical elements: relations are built between single words. A major information loss of this word-level dependency tree compared with the constituent tree is that it doesn't explicitly provide local structures and syntactic categories of phrases. On the other hand, the dependency tree provides connections between distant words, which are useful in extracting long-distance relations. Therefore, compromising between the two, we extend the dependency tree node with phrases. That implies a noun phrase "Canon SD500 PowerShot" can be a dependent that modifies a verb phrase head “really enjoy using” with relation type “dobj”. The feasibility behind is that a phrase is a syntactic unit regardless of the length or syntactic category .and it is acceptable to substitute a single word by a phrase with the Sentiment Analysis syntactic category in a sentence define the dependency parsing with phrase nodes as phrase dependency parsing. A dependency relationship which is an asymmetric binary relationship holds between two phrases. One is called the head, which is the central phrase in the relation. The other phrase is called dependent, which modifies the head. A label representing the relation type is assigned to each dependency relationship, such as subj (subject), obj (object), and so on.

C. CRF-based Approach for Opinion Target Extraction Token

N. Jakob and I. Gurevych [3]Introduced this feature represents the string of the current token as a feature. Even though this feature is rather obvious, it can have a considerable impact on the target extraction performance. If the vocabulary of targets is rather compact for a certain domain (corresponding to a low target type/target ratio), the training data is likely to contain the majority of the target types, which should hence be a good indicator. It will refer to this feature as tk in our result tables.

POS

This feature represents the part-of-speech tag of the current token as identified by the Stanford POS Tagger. It can provide some means of lexical Sentiment Analysis e.g. indicate that the token “sounds” is a noun and not a verb in a certain context. At the Sentiment Analysis time, the CRF algorithm is provided with additional information to extract opinion targets which are multiword expressions, i.e. noun combinations.It will refer to this feature as pos in our result tables.

Short Dependency Path

The dependency parse tree to link opinion expressions and the corresponding targets. Both works identify direct dependency relations such as “amod” and “nsubj” as the most frequent and at the Sentiment Analysis time highly accurate connections between a target and an opinion expression. Hence label all tokens which have a direct dependency relation to an opinion expressed in a sentence. The Stanford Parser is employed for the constituent and dependency parsing.

Word Distance

It can infer that opinion expressions and their targets are not always connected via short paths in the dependency parse tree. Since we cannot capture such paths with the above- mentioned feature we introduce another feature that acts as a heuristic for identifying the target to a given opinion expression. Have shown that (base) noun phrases are good candidates for opinion targets in the datasets of product reviews. Therefore label the tokens in the closest noun phrase regarding word distance to each opinion expressed in a sentence. It will refer to this feature as wrdDist in our result tables.

Opinion Sentence

With this feature, we simply label all tokens occurring Our goal is to extract individual instances of opinion targets from sentences that contain an opinion expression. This can be modeled as a sequence segmentation and labeling task. The CRF algorithm receives a sequence of tokens t1...tn for which it has to predict a sequence of labels l1...ln. It represents the possible labels following the IOB scheme: B- Target, identifying the beginning of an opinion target, I- Target identifying the continuation of a target, and O for other (non-target) tokens. We model the sentences as a linear chain CRF, which is based on an undirected graph. In the graph, each node corresponds to a token in the sentence and edges connect the adjacent tokens as they appear in the sentence. In our experiments, we use the CRF implementation from the Mallet toolkit.

W. Medhat, A [4] Proposed a method the training objective of the Skip-gram model is to find word representations that are useful for predicting the surrounding words in a sentence or a document. More formally, given a sequence of training words w1,w2,w3, . . . , wT, the objective of the Skip-gram model is to maximize the average log probability.

A computationally efficient approximation of the full softmax is the hierarchical softmax. In the context of neural network language models. The main advantage is that instead of evaluating W output nodes in the neural network to obtain the probability distribution, it is needed to evaluate only about log2(W) nodes. The hierarchical softmax uses a binary tree representation of the output layer with the W words as its leaves and, for each node, explicitly represents the relative probabilities of its child nodes. These define a random walk that assigns probabilities to words.

An alternative to the hierarchical softmax is Noise Contrastive Estimation (NCE). NCE posits that a good model should be able to differentiate data from noise using logistic regression. This is similar to the hinge loss used. Trained the models by ranking the data above noise. While NCE can be shown to approximately maximize the log probability of the softmax, the Skip-gram model is only concerned with learning high-quality vector representations, so we are free to simplify NCE as long as the vector representations retain their quality. We define Negative Sentiment Analysismpling(NEG) by the objective.

T. Mikolov, I. Sutskever, K. Chen [5]Proposed a method Sentiment Analysis or Opinion Mining is the computational study of people's opinions, attitudes, and emotions toward an entity. The entity can represent individuals, events or topics. These topics are most likely to be covered by reviews. The two expressions are interchangeable. They express a mutual meaning. However, some researchers stated that and Sentiment Analysis have slightly different notions. Opinion Mining extracts and analyzes people's opinions about an entity while Sentiment Analysis identifies the Sentiment expressed in a text then analyzes it. Therefore, the target of Sentiment Analysis is to find opinions, identify the Sentiments they express, and then classify their polarity. There are three main classification levels I Sentiment Analysis n: document-level, sentence- level, and aspect-level. Sentiment Analysis Document-level Sentiment Analysis aims to classify an opinion document as expressing a positive or negative opinion or Sentiment. It considers the whole document a basic information unit Sentence-level Sentiment Analysis aims to classify Sentiment expressed in each sentence. The first step is to identify whether the sentence is subjective or objective. If the sentence is subjective, Sentence-level Sentiment Analysis will determine whether the sentence expresses positive or negative opinions.

G. Zhao, X. Qian, and X. Xie [6] Introduced predict user- service ratings, It focuses on users' rating behaviors. It fuses four factors, personal interest, interpersonal interest similarity, interpersonal rating behavior similarity, and interpersonal rating behavior diffusion, into matrix factorization. Among these factors, interpersonal rating behavior similarity and interpersonal rating behavior diffusion are the main contributions of our approach. Hereinafter we turn to the details of the approach.

1) It focuses on exploring user rating behaviors. A concept of the rating schedule is proposed to represent user daily rating behavior. The factor of interpersonal

rating behavior diffusion is proposed to deep understand users' rating behaviors. We consider these two factors to explore users' rating behaviors.

2) It fuses three factors, interpersonal interest similarity, interpersonal rating behavior similarity, and interpersonal rating behavior diffusion, together to directly constrain users' latent features, which can reduce the time complexity.

Z. Hai, G. Cong, K. Chang [7] Proposed SJASM model: The generation for aspect-specific Sentiments depends on the aspects. This means that we first generate latent aspects, on which we subsequently generate corresponding Sentiment orientations. The generation for aspect terms depends on the aspects, while the generation for opinion words relies on the Sentiment orientations and semantic aspects. The formulation is intuitive, for example, to generate an opinion word “beautiful”, we need to know its Sentiment orientation positive and related semantic aspect appearance. The generation for overall ratings of reviews depends on the semantic aspect-level Sentiments in the reviews. Based on the model assumptions, to generate a review document and its overall rating, we first draw hidden semantic aspects conditioned on document-specific aspect distribution; We then draw the Sentiment orientations on the aspects conditioned on the per-document aspect-specific Sentiment distribution; Next, we draw each opinion pair, which contains an aspect term and corresponding opinion word, conditioned on aspect and Sentiment specific word distributions; We lastly draw the overall rating response based on the generated aspect and Sentiment assignments in the review document.

Y. Fang, H. Tan, and J. Zhang [8]Inroduced SVM (Support vector machine) is a machine learning classification technique that is highly effective at traditional text categorization. It seeks a decision hyperplane and divides the data sets into two types. In Sentiment Analysis, SVM can tell the Sentiment polarities of articles according to a feature vector that is composed of Sentiment features. The feature vector is described as, a set of features extracted from the training corpus. It's common to use Sentiment phrases within the training corpus as a Sentiment feature in the feature vector. Though there are other ways to represent Sentiment features, most researchers tend to use Sentiment phrases as features, because they are basic symbols of Sentiment information. For each Sentiment phrase, we need to measure it with a certain weight value. There are different methods of weighting feature phrases, for example, Document Frequency (DF), Feature Presence (FP) and Term Frequency (TF) – Inverse Document Frequency (IDF). We believe that phrase with big frequency difference in two kinds Balanced Corpus makes many contributions to Sentiment classification. So, these phrases should be used as features, and when a phrase appears equal in two kinds of corpus, it has poor classification ability and should be abandoned. We call our method as Relative Document Frequency (RDF).

The Naïve Bayes (NB) algorithm is widely used in document categorization as a classification. Given feature labels, it computes the posterior probabilities of a document

corresponding to different classes and then assigns it to the highest probability class. In Sentiment Analysis, Naïve Bayes first deals with labeled training corpus where the Sentiment polarities of each document are known. It tokenizes each article in training corpus and extracts Sentiment words. Then it computes the posterior probability of each Sentiment word and records them in a probability table. When we need to tell the Sentiment polarities of a document, we tokenize the document and extract Sentiment phrases. Then we get their posterior probabilities from the probability table and count the total posterior probabilities of Sentiment phrases. Finally, assign the document to the Sentiment polarity with a higher total posterior probability.

Y. Ma, Z. Xiang, Q. Du [9] Introduced numerically embed each word into the model and represent the holistic meaning of each review, recurrent neural networks (RNN) was used for sequence encoding. LSTM, an algorithm that has demonstrated state-of-the-art performance on various sequence modeling tasks, was deployed. It is well known that words in sentences or paragraphs, short or long, have dependencies between each other, which is the default behavior of LSTMs. To extract such dependencies between words, LSTM employs two vectors that encode meanings as a result of long-term and short-term dependencies. Both long-term memory and short-term memory as shown in Fig. 2 earlier can be understood as semantic meanings and will be slightly updated as the recurrent neural network moves along different words of the sequence. It used the embedding hf at the last time step, i.e., of the last word token in the review, as our feature representation for the textual content. This can be thought of as the representation of the holistic meaning of the text in a specific review.

Compared to text embedding, image processing is more challenging as what the computer "sees" are just millions of pixels. Motivated by limitations of previous machine learning techniques and inspired by how mammalian vision system works convolution and pooling-based blocks in neural networks achieved great success. Like the architecture depicted, a convolutional neural network (CNN) aims to detect features layer by layer through multiple simple-to-complex feature extraction steps. Based on simpler features like edges or corners detected by lower layers, CNN can identify more complicated ones such as rectangles or circles in intermediate layers, which can then be stacked to build sophisticated objects like animals or buildings. To represent the whole image, CNN oftentimes adds a fully connected layer at last, which is guided by the Sentiment Analysis approximation theorem. In this study, the 152-layer deep residual network (ResNet) was adopted for our final image feature representation. In the end, each image in the sequence was wrapped and we obtained an image feature sequence C.

B. Purkaystha, T. Datta [10] Proposed a method that emphasizes on factorizing each user and each item individually to overcome the limitation of the statistical assumption that the similar users will have similar opinions on an item. The main contribution of our work lies in how we model each user profile and each item characteristic without using any demographic information. This way of factorization is particularly helpful in scenarios where only a huge amount of rating data exists without any explicit information about the users or the items. This has also contributed to eliminating the limitation of the traditional linear factorization models (i.e., the number of user factors and the number of item factors be the Sentiment Analysis) as our deep model allows any number of factor units at the bottommost layer. We constructed the user behaviors and the item characteristics in the factor layer and modeled the relationship in the upper layers. shown that our model produces better predictions than the other factorization models and the models which use contextual information or content information.

Rung-Ching Chen and Hendry [11] Introduced learning user ratings by incorporating the vector representation of the word, deep learning, and tradeoff of emotions in user comments. The system begins by extracting user comments from the data set. The system applies word stemming to stem all the words from the sentences in the user comments. We used a Sentiment Analysis word dictionary of positive and negative opinion words and the bag of words algorithm to count the words.

The 50-D input feature vector is created from the WordNet Sentiment Analysis. The system implements a three-step noise reduction to remove noise from the data set. The noise is categorized by short user comments, phrases without opinion words, and Sentiment Analysis sarcasm (user rating is expressed by opposite impression). Although it is interesting to deal with Sentiment Analysisrcasm, remove the Sentiment Analysis sarcasm and label it as noise.

The deep learning model uses the 50-D input feature vector as input nodes. RBM is implemented as a pretraining to perform features extraction from the input feature vector. As the input feature vector is usually sparse, RBM is used to learn, which represented words that have significant features. RBM then produces a new input feature vector, which is designed to give better features to the DBN learning model, which then applies a backpropagation network to learn user ratings from the input feature vector, which is produced from the RBM pretraining model. The user rating prediction is obtained from the output layer of the DBN.

IV. CONCLUSION

Sentiment Analysis is an ongoing field of research in text mining field. Sentiment Analysis is the computational treatment of opinions, Sentiments, and subjectivity of text. This survey paper tackles a comprehensive overview of the

last update in this field. Many recently proposed algorithms' enhancements and various Sentiment Analysis applications are investigated and presented briefly in this survey. These articles are categorized according to their contributions to the various Sentiment Analysis techniques. The related fields to Sentiment Analysis that attracted researchers recently are discussed. The main target of this survey is to give nearly full image of Sentiment Analysis techniques and the related fields with brief details. The main contributions of this paper include the sophisticated categorizations of a large number of recent articles and the illustration of the recent trend of research in the Sentiment Analysis and its related areas.

V. REFERENCES

[1]. Y. Choi and C. Cardie, "Learning with compositional semantics as structural inference for subsentential Sentiment Analysis," in Proc. Conf. Empirical Methods Natural Lang. Process., Stroudsburg, PA, SENTIMENT ANALYSIS, Oct. 2008, pp. 793–801.

[2]. Y. Wu, Q. Zhang, X. Huang, and L. Wu, "Phrase dependency parsing for opinion mining," in Proc. Conf. Empirical Methods Natural Lang.Process. (EMNLP), vol. 3, Aug. 2009, pp. 1533–1541.

[3]. N. Jakob and I. Gurevych, "Extracting opinion targets in a single-and cross-domain setting with conditional random fields," in Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP), Oct. 2010, pp. 1035–1045

[4]. W. Medhat, A. HasSentiment Analysis, and H. Korashy, "Sentiment Analysis algorithms and applications: A survey," Ain Shams Eng. J., vol. 5, no. 4, pp. 1093–1113, Dec. 2014.

[5]. T. Mikolov, I. Sutskever, K. Chen, G. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," in Proc. Adv. Neural Inf. Process. Syst., Oct. 2013, pp. 3111–3119.

[6]. G. Zhao, X. Qian, and X. Xie, "User-service rating prediction by exploring social users' rating behaviors, "rating behaviors," IEEE Trans.Multimedia, vol. 18, no. 3, pp. 496–506, Mar. 2016.

[7]. Z. Hai, G. Cong, K. Chang, P. Cheng, and C. Miao, "Analyzing Sentiments in one go: A supervised joint topic modeling approach," IEEE Trans. Knowl. Data Eng., vol. 29, no. 6, pp. 1172–1185, Jun. 2017

[8]. Y. Fang, H. Tan, and J. Zhang, "Multi-strategy Sentiment Analysis of consumer reviews based on semantic fuzziness," IEEE Access, vol. 6,p. 20625–20631, 2018.

[9]. Y. Ma, Z. Xiang, Q. Du, and W. Fan, "Effects of user-provided photos on hotel review

helpfulness: An analytical approach with deep learning," Int. J. Hosp itality Manage., vol. 71, pp. 120–131, Aug. 2018.

[10]. B. Purkaystha, T. Datta, M. S. Islam, and M.-E-Jannat, "Rating prediction for recommendation: Constructing user-profiles and item characteristics using backpropagation," Appl. Soft Comput., vol. 75, pp. 310–322, Feb. 2019

[11]. Rung-Ching Chen and Hendry "User Rating Classification via Deep Belief Network Learning and Sentiment Analysis" IEEE TranSentiment Analysisction On Computational social System,2019

Cite this article as :

Princy Baby, Krishnapriya B, "Sentimental Analysis and Deep Learning : A Survey", International Journal of Scientific Research in Science, Engineering and Technology (IJSRSET), Online ISSN : 2394-4099, Print ISSN : 2395-1990, Volume 7 Issue 1, pp. 212-220, January-February 2020. Available at doi : https://doi.org/10.32628/IJSRSET207135
Journal URL : http://ijsrset.com/IJSRSET207135