RESEARCH ARTICLE OPEN ACCESS

Survey on Opinion Mining Algorithms and Applications

Anjanabhargavi A. Kulkarni[1],Vidyashree A. Hundekar[2],S. S. Sannakki[3],Vijay S. Rajpurohit[4]

(Department of Computer Science and Engineering, Gogte Institute of Technology Belagavi.)

----------**********----------

Abstract:

The Opinion mining is an ongoing field of research and development in web text mining domain. It is the computational treatment of opinions, sentiments and subjectivity of text. This survey paper focus on a comprehensive overview of the Opinion mining algorithms and the different classification with their field of applications.

Keywords **–Opinion Mining, web text mining, sentiment analysis**

----------**********----------

I. INTRODUCTION

The Social media is one of the open area to express opinions. The term sentiment analysis is the procedure by which information is extracted from the opinions, appraisals and emotions of people in regards to entities, events and their attributes.

Opinion Mining is to analyze and classify the user generated data like reviews, blogs, comments, articles etc. The main objective of Opinion mining is Sentiment Classification i.e. to classify the opinion into positive or negative classes. The proposed work is able to collect information from various sites and perform a Opinion *analysis and collaboration of algorithms with application.*

II. PURPOSE OF OPINION MINING

Today a day to day proliferation of the current digital based economy a large amount of information is available in the form of textual data which can often be used more easily if it is categorized or classified into some predefined classes.

In any business or industrial environment corporate information may be available in multiple different formats, about 80% of which is in text documents. This information exists in the form of descriptive data formats which include service reports repair information, manufacturing quality documentation, customer help desk notes and product reviews and opinions.[8]

In order to enhance customer satisfaction and their shopping experiences, it has become a common practice for E-Commerce.

III.OPINION MINING TERMINOLOGIES

In this section we define the basic terminologies currently used in the area of opinion mining.

- Fact: A fact is something that has really occurred or is actually the case.
- Opinion: An opinion is a belief about matters commonly considered to be subjective, and is the result of emotion or interpretation of facts.
- Subjective/opinionated text**:** A text is subjective or opinionated if it expresses personal feelings or beliefs, e.g. opinions.
- Objective text: An objective text expresses some factual information about the world.
- Item: An item is a concrete or abstract object such as product, service, person, event, organization. An item can be represented as a hierarchy of components, sub-components, etc.
- Review: A review is a subjective text containing a sequence of words describing opinions of reviewer regarding a specific item. Review text may contain complete sentences, short comments, or both.

- Short comments or pros/cons: The reviewer can describe pros and cons of the item.

IV.MATHEMATICAL MODEL

To aid the extraction of opinions from text, recent work has tackled the issue of determining the orientation of subjective terms contained in text, i.e. deciding whether a term that carries opinionated content has a positive or a negative connotation.
It is mathematically represented as, an opinion is a quintuple

$$O_j, f_{jk}, SO_{ijkl}, h_i, t_l$$

Where,

- O_j is a target object.
- f_{jk} is a feature of the object O_j
- SO_{ijkl} is the sentiment value of the opinion of the opinion holder h_i on feature f_{jk} of object O_j at time t_l. SO_{ijkl} is +ve, -ve, or neutral, or a more granular rating.
- h_i is an opinion holder.
- t_l is the time when the opinion is expressed.

V.COMPONENTS OF OPINION MINING

- Opinion Holder: Is the holder of a particular opinion, it may be a person or an organization that holds the opinion.
- Opinion Object**:** Is an object on which the opinion holder is expressing the opinion.
- Opinion Orientation**:** It determines whether the opinion of an opinion holder about an object is positive, negative or neutral.

SCENARIO: James bought an iPhone a few days ago. He expresses his opinion *"It was such a nice phone. The touch screen was really cool. The voice quality was clear too".*

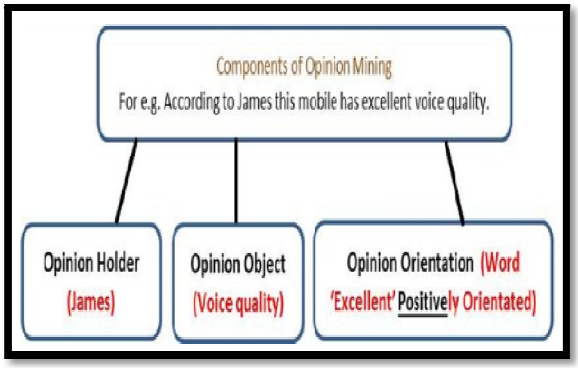

Figure 1. Components of Opinion Mining

VI. TECHNIQUES IN OPINION MINING

The data mining algorithms can be classified into different types of approaches as Supervised, Unsupervised or Semi supervised algorithms. Supervised approaches works with set of examples with known labels. In unsupervised approaches aims to obtain the similarity of the attribute values without knowing the labels of the example in the dataset. Semi supervised approaches are being[1] used when the examples in the dataset is the combination of both the labelled and unlabelled examples.

A. Classification

Classification is the Supervised technique in which every instances belongs to the specific class, it is being indicated by the value of class attribute or any special goal attribute. The categorical values are taken by the goal attribute in which each attribute belongs to the corresponding class. Two different parts that exist in each example are set of predictor attribute values and goal attribute value[1]. In classification technique the mining function can be classified into set of tasks such as the training and test set. In the training phase the model that is to be used for the effective classification will be formed up from the training set and in testing phase the model will be evaluated on the test set. Main goal of the classification algorithm is to improve the predictive accuracy in training the model.

The algorithms being discussed includes the following

1. K-Nearest Neighbor
2. Support Vector Machines

3. Decision Tree
4. Naive Bayes
5. The Apriori Algorithm
6. Multi-Layer Perceptron (MLP)

1) K-Nearest Neighbor:

K-Nearest Neighbor algorithm that is being widely used for classification and regression and also it is a non-parametric method. Every training set that is being present in the multidimensional feature space are the vectors with the specific class labels specified. In n-dimensional space each attribute will be pointing to the training samples with n dimensional numeric attributes. The training phase of the algorithm it acts by storing the feature vectors and class labels. When an unknown sample is given to k-nearest neighbor algorithm it searches for the pattern space for the k training samples that are closer to the unknown samples. Euclidean distance determines the property of the "closeness" measures. When KNN approach is to be applied value should be appropriate and the effectiveness of this approach mainly depends upon this value.

Advantages of KNN Algorithm

- It can be widely adopted for multi-class model classes and also for the objects with multiple class labels.
- KNN is an efficient classification algorithm that is easy for understanding.
- Building of the model is also inexpensive with extremely flexible classification.
- It is robust even in the case of large dataset with noisy training data being used.

Application of KNN Algorithm

It is widely used in the application areas of legal, medical, agriculture, news and banking for problem solving, functional learning and for user training purposes(Teaching and aiding).

Finance: KNN plays an important core role in identifying the stock market forecasting such as analysing the market trends, planning for investment strategies and identifying the time period for obtaining the stocks and for credit ratings. Used in the banks for loan management, bank customer profiling.

Medicine: To predict patient's health based on the clinical record persisting to each patient. For example, in estimating the amount of glucose level in the diabetic patients from the existing health record. When coupled with genetic algorithms, KNN is being used for analyzing the micro-array gene expression data.

Agriculture: Used in the climate forecasting and to estimate the soil water parameters.

In simulation and precipitation of weather variables the KNN algorithm is widely used.

2) Support Vector Machines:

SVM was introduced by Boser, Guyon and Vapnik and widely being used for classification, regression and pattern recognition.SVM has capability to classify indeed of the dimensions or size of the input space. It acquires the major advantage because of its high generalization performance with indeed of the much prior knowledge. The goal of the SVM lies in finding the best classification function and also it aims to distinguish between members of the two classes in training data. The major idea behind the SVM is construction of the optimal hyper plane that is widely used for the problems of classification and for patterns identification .From the set of hyper planes the hyper plane that is of optimal is needed to be selected for pattern classification and thus to improve the margin of the hyper plane. SVM needs to classify the given patterns correctly so that it can maximize the margin that determines the efficiency of the SVM algorithm. The accuracy in classifying pattern will improve based on the size of the margin i.e. greater the margin size more exactly it classifies the patterns. The equation for the hyper plane is given below. Hyper plane, the above pattern can be mapped into high dimensional space.SVM tries in finding the hyper plane accurately that separates the two different samples with the set of independent training samples being specified.

Advantages of SVM algorithm

- It provides the greater benefits on the text classification when the high-dimensional spaces are being used.
- Accuracy in the prediction is comparatively high with other classification algorithms. Fast evaluation of the learned target function.
- Used widely in various real time applications with the high scope in evaluating the good outcome.
- Without the dependence of the dimensionality of feature space it has the good ability in learning.
- It interprets the inherent characteristics of the data better when comparing to artificial neural networks.

Application of SVM algorithm

- It is been widely used for many real world problems such as
- Text categorization: To categorize the text documents i.e. natural text, based on their content, for example in email filtering, web searching, sorting the documents to specific topic.In assigning documents to more than one category such that for series of binary classification problem.
- Image classification: Used in validating and testing the bacterial image, pathogens and for the classification SVM is used widely.
- Medicine: It is used in detecting the micro calcifications in mammograms which is an indicator for the breast cancer.

3) Decision Tree

A Decision Tree Classifier consists of a decision tree generated on the basis of instances. A decision tree is a classifier expressed as a recursive partition of the instance space. The decision tree [7] consists of nodes that form a rooted tree, meaning it is a directed tree with a node called "root" that has no incoming edges. All other nodes have exactly one incoming edge.

A node with outgoing edges is called an internal or test node. All other nodes are called leaves. In a decision tree, each internal node splits the instance space into two or more sub-spaces a certain discrete function of the input attributes.

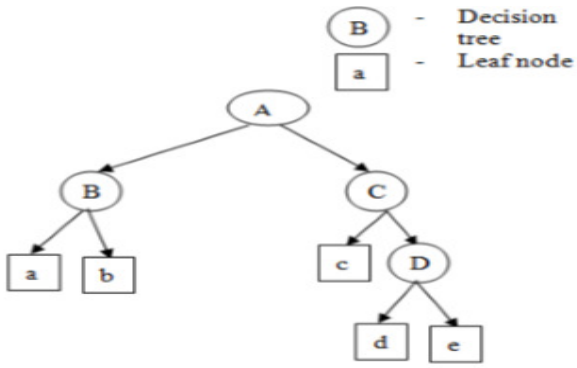

Figure 2 : Decision tree model.

The root and the internal nodes are associated with attributes, leaf nodes are associated with classes. Basically, each non-leaf node has an outgoing branch for each possible value of the attribute associated with the node. To determine the class for a new instance using a decision tree, beginning with the root, successive internal nodes are visited until a leaf node is reached. At the root node and at each internal node, a test is applied. The outcome of the test determines the branch traversed, and the next node visited. The class for the instance is the class of the final leaf node.

The estimation criterion [7] in the decision tree algorithm is the selection of an attribute to test at each decision node in the tree. The goal is to select the attribute that is most useful for classifying examples. A good quantitative measure of the worth of an attribute is a statistical property called information gain that measures how well a given attribute separates the training examples according to their target classification. This measure is used to select among the candidate attributes at each step while growing the tree.

Advantages of Decision Trees:

- Easy to interpret and explain
- Non-parametric, so you don't have to worry about outliers or whether the data is linearly separable (e.g., decision trees easily take care of cases where you have class A at the

low end of some feature x, class B in the mid-range of feature x, and A again at the high end).

- Their main disadvantage is that they easily overfit, but that's where ensemble methods like random forests (or boosted trees) come in.
- Plus, random forests are often the winner for lots of problems in classification (usually slightly ahead of SVMs, I believe), they're fast and scalable, and you don't have to worry about tuning a bunch of parameters like you do with SVMs, so they seem to be quite popular these.

4) NAIVE BAYES

Given a set of objects, each of which belongs to a known class, and each of which has a known vector of variables, our aim is to construct a rule which will allow us to assign future objects to a class, given only the vectors of variables describing the future objects. Problems of this kind, called problems of supervised classification, are ubiquitous, and many methods for constructing such rules have been developed[5]. One very important one is the naive Bayes method—also called idiot's Bayes, simple Bayes, and independence Bayes. This method is important for several reasons. It is very easy to construct, not needing any complicated iterative parameter estimation schemes. This means it may be readily applied to huge data sets. It is easy to interpret, so users unskilled in classifier technology can understand why it is making the classification it makes. And finally, it often does surprisingly well: it may not Probabilistic approaches to classification typically involve modeling the conditional probability distribution P(C|D), where C ranges over classes and D over descriptions, in some language, of objects to be classified. Given a description d of a particular object, we assign the class argmax_c P(C = c|D = d). A Bayesian approach splits this posterior distribution into a prior distribution P(C) and a likelihood P(D|C):

$$\text{argmax}_c\, P(C = c|D = d) = \text{argmax}_c \frac{P(D = d|C=c)\, P(C=c)}{P(D=d)} \qquad (1)$$

The denominator P(D = d) is a normalising factor that can be ignored when determining the maximum a posteriori class, as it does not depend on the class. The key term in Equation (1) is P(D = d|C = c), the likelihood of the given description given the class (often abbreviated to P(d|c)). A Bayesian classifier estimates these likelihoods from training data, but this typically requires some additional simplifying assumptions. For instance, in an attribute-value representation (also called propositional or single-table representation), the individual is described by a vector of values a1, . . . , an for a fixed set of attributes A1, . . . ,An. Determining P(D = d|C = c) here requires an estimate of the joint probability P(A1 = a1, . . . ,An = a_n|C = c), abbreviated to P(a1, . . . ,a_n|c). This joint probability distribution is problematic for two reasons: (1) its size is exponential in the number of attributes n, and (2) it requires a complete training set, with several examples for each possible description. These problems vanish if we can assume that all attributes are independent given the class:

$$P(A1 = a1, \ldots, An = an|C = c) = \prod_{i=1}^{n} P(A_i = a_i|C = c) \qquad (2)$$

This assumption is usually called the naive Bayes assumption, and a Bayesian classifier using this assumption is called the naive Bayesian classifier, often abbreviated to 'naive Bayes'. Effectively, it means that we are ignoring interactions between attributes within individuals of the same class.

Advantages of Naive Bayes:

- Super simple, you're just doing a bunch of counts.
- If the NB conditional independence assumption actually holds, a Naive Bayes classifier will converge quicker than discriminative models like logistic regression, so you need less training data.

- And even if the NB assumption doesn't hold, a NB classifier still often performs surprisingly well in practice.

5) THE APRIORI ALGORITHM

One of the most popular data mining approaches is to find frequent item sets from a transaction dataset and derive association rules. Finding frequent item sets is not trivial because of its combinatorial explosion. Once frequent item sets are obtained, it is straightforward to generate association rules with confidence larger than or equal to a user specified minimum confidence. Apriori is a seminal algorithm for finding frequent item sets using candidate generation [5]. It is characterized as a level-wise complete search algorithm using anti-monotonicity of item sets, "if an item set is not frequent, any of its superset is never frequent". By convention, Apriori assumes that items within a transaction or itemset are sorted in lexicographic order. Let the set of frequent item sets of size k be Fk and their candidates be Ck . Apriori first scans the database and searches for frequent item sets of size 1 by accumulating the count for each item and collecting those that satisfy the minimum support requirement. It then iterates on the following three steps and extracts all the frequent item sets.

1. Generate C_{k+1}, candidates of frequent item sets of size k +1, from the frequent item sets of size k.
2. Scan the database and calculate the support of each candidate of frequent item sets.
3. Add those item sets that satisfies the minimum support requirement to F_{k+1}.

F_1=(Frequent itemsets of cardinality 1);

for($k = 1; F_k \neq \phi; k++$) do begin

C_{k+1} = apriori-gen(F_k); //New candidates

for all transactions $t \in$ Database do begin

C'_t = subset(C_{k+1}, t); //Candidates contained in t

for all candidate $c \in C'_t$ do

$c.count++;$

end

$F_{k+1} = \{C \in C_{k+1} \mid c.count \geq$ minimum support $\}$

end

end

Answer $\cup_k F_k$;

Figure 3. Apriori algorithm is given in the below

1. Function apriori-gen in line 3 generates C_{k+1} from F_k in the following two step process: 1. Join step: Generate RK+1, the initial candidates of frequent item sets of size k + 1 by taking the union of the two frequent item sets of size k, P_k and Q_k that have the first kí1 elements in common.

$R_{k+1}=P_k \cup Q_k = \{item_1, item_2,.....,item_k, item_{k'}\}$

$P_k = \{item_1, item_2,.....,item_k, item_k\}$

$Q_k = \{item_1, item_2,....., item_{k'}\}$

Where $item_1<item_2<\ldots\ldots\ldots<item_k<item_{k'}$

2. Prune step: Check if all the item sets of size k in R_{k+1} are frequent and generate Ck+1 by removing those that do not pass this requirement from R_{k+1}. This is because any subset of size k of C_{k+1} that is not frequent cannot be a subset of a frequent item set of size k +1.

Function subset in line 5 finds all the candidates of the frequent item sets included in transaction t. Apriori, then, calculates frequency only for those candidates generated this way by scanning the database. It is evident that Apriori scans the database at most k_{max+1} times when the maximum size of frequent item sets is set at k_{max}.

6) Multi-Layer Perceptron (MLP):

An MLP also known as Artificial Neural Network (ANN) can be considered as network of neurons called perceptrons[2]. The perceptron computes a single output from multiple inputs.MLP is also known as feed forward networks and can have one or more hidden layers between input and output layer. The MLP networks can be used for both supervised and unsupervised learning process.

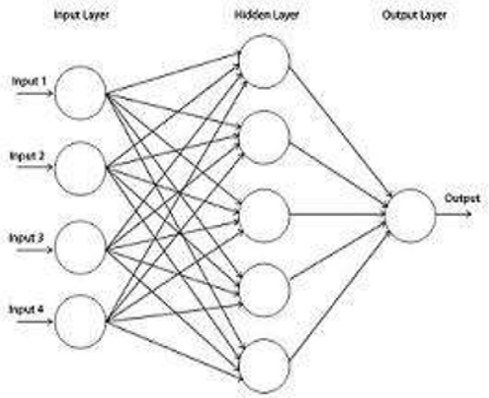

Figure 4:MLP

The above architecture has the following properties: 1.There is no connections within a layer, 2.There is no direct connections from input to output layers, 3.The layers are fully connected, 4.Generally there are more than 3 layers, 5.It not necessary that the no. of input units are equal to the no. of output units, 6.No.of hidden units in each layer can be more or less than input or output units. The MLP network should have minimum three hidden layers for any valid representation and such a network takes much time for its training process.MLP is the most used type of neural network algorithm and having huge number of applications. It is capable of modelling complex functions. It is very good at ignoring irrelevant inputs and noise and it can be used even if a few knowledge available about the relationship of the function to be modelled.

B. Clustering

Clustering is the unsupervised technique that performs natural grouping of instances[1]. It is the method of dividing the data into different groups with the similar objects. Every group, called cluster, consisting of several objects that are similar within the particular cluster and dissimilar to the objects of the other clusters. Clustering algorithms are also used for data compression too rather than the categorizing and organizing the data An effective clustering algorithms aims in obtaining the effective clusters irrespective of their shapes and size of data. Most Commonly used algorithms in the clustering falls into any of the following categories as Hierarchical, Partitioning, Grid based, Density based, Model Based and Constraint based algorithms.

The algorithms that are being described are

1. K-Means Clustering
2. SOM(self organized map)
3. Hierarchical Clustering
4. DBSCAN Clustering
5. Optics
6. Sting

1) K-means clustering algorithm:

K-means algorithm is most common and popular clustering tool that is widely used in many applications and it falls under the partitioning algorithms that aims in constructing the various patterns and evaluates them by using some criterion. With the given collection of n data, k different clusters are formed with each cluster having an unique centroid (mean) and thus the partitioning is made. The letter k describes the number of clusters needed to be made. When number of n objects is to be grouped into k clusters, K cluster centre is to be initialized. Every object will be given to the closest cluster centers and .the centre of cluster is updated every time until state of no change occurs in the each cluster. The elements in each cluster will be in close contact with centroid of that particular cluster and will be different to the elements belonging to other clusters.

$E = \sum_{i=1}^{k} \sum_{p \in Ci} |p - m_i|^2$

The sum of the discrepancies between the point and the centroid expressed by specific distance is used as the objective function. Total intra-cluster variance describes the sum of the squares of the error between the point and respective centroids.

Advantages of K-means clustering algorithm

- The k-means algorithm produces an relatively scalable results when handing the large data sets.
- It is used for undirected knowledge discovery and very simple to use.
- Used widely in various number of applications ranging from unsupervised learning of neural networks, image processing, pattern recognition and may other applications.
- It can provide the best results when the datasets are distinct and well separated from each other.

Applications of K-means clustering algorithm

- Used in acoustic data for converting the waveforms into some category for speech understanding
- Colour based image segmentation is possible by the use of K-means clustering technique.
- It can be used in machine learning and also for the data mining.

2) Self organized Map (SOM) algorithm:

SOM is a type of the artificial neural network (ANN) that is unsupervised learning methodology introduced by the professor Kohenon in 1980 s and though it is also called as Kohonen's Self Organizing Map. It is widely used in vector quantization and it belongs to the category of competitive learning networks[1]. The SOM can be used to detect the features that is inherent to the problem and thus it can also been known as SOFM, i.e. Self- Organizing Feature Map. It is majorly used in representing the data in low dimensional discretized representation from the high dimensional representation. SOM consists of numerous components called as nodes or neurons. Each neuron will be assigned with the specific weight in the output space. Based on the weights it will reflect on the cluster content. It provides the topology preserving map from the high dimensional space to map units that are used is preserving the relative distance between the points. In the SOM the points that are near to each other in the input space are mapped to nearest map units-Matrix representation is used in Self organized maps in order to specify or identify the distance between the existing other neurons. With the U-matrix representation the cluster boundaries are easily recognized.

Advantages of SOM algorithm

- Since it is the unsupervised learning method it does not need any human intervention but only needs some knowledge about the input data.
- Ability of the network to generalize and characterize the inputs that are not being encountered before.
- It can be applied to compare the various maps of different sizes and can be used effectively in vector quantization.

Applications of SOM algorithm

- Used in many real world problem such as Speech recognition and analysis: Used in creating the representation of the spectra for the different speech samples to different parts of the map.
- Visualization property of the SOM is used in the voice analysis applications.
- Interpretation of ECG data: With the two-dimensional display it is used to monitor the ECG signal as a trajectory. By the use of clustering the data decisive features in the sleep ECG is being identified with the SOM.

3) Hierarchical Clustering

A hierarchical method creates a hierarchical decomposition of the given set of data objects. Here tree of clusters called as dendro grams is built[6]. Every cluster node contains child clusters, sibling clusters partition the points covered by their common parent. In hierarchical clustering we assign each item to a cluster such that if we have N items then we have N clusters. Find closest pair of clusters and merge them into single cluster. Compute distance between new cluster and each of

old clusters. We have to repeat these steps until all items are clustered into K no.of clusters.
It is of two types:

a) Agglomerative (bottom up)-

Agglomerative hierarchical clustering is a bottom-up clustering method where clusters have sub-clusters, which in turn have sub-clusters, etc. It starts by letting each object form its own cluster and iteratively merges cluster into larger and larger clusters, until all the objects are in a single cluster or certain termination condition is satisfied. The single cluster becomes the hierarchy's root. For the merging step, it finds the two clusters that are closest to each other, and combines the two to form one cluster [5].

b) Divisive (top down)-

A top-down clustering method and is less commonly used. It works in a similar way to agglomerative clustering but in the opposite direction. This method starts with a single cluster containing all objects, and then successively splits resulting clusters until only clusters of individual objects remain [4].

4) DBSCAN Clustering

DBSCAN (Density Based Spatial Clustering of Application with Noise).It grows clusters according to the density of neighborhood objects. It is based on the concept of "density reachibility" and "density connectability", both of which depends upon input parameter- size of epsilon neighborhood e and minimum terms of local distribution of nearest neighbors. Here e parameter controls size of neighborhood and size of clusters. It starts with an arbitrary starting point that has not been visited [6]. The points e-neighbourhood is retreived, and if it contains sufficiently many points, a cluster is started. Otherwise the point is labelled as noise. The number of point parameter impacts detection of outliers. DBSCAN targeting low-dimensional spatial data used DENCLUE algorithm.

5) OPTICS

Ordering Points To Identify Clustering Structure) is a density based method that generates an augmented ordering of the data's clustering structure. It is a generalization of DBSCAN to multiple ranges, effectively replacing the e parameter with a maximum search radius that mostly affects performance. MinPts then essentially becomes the minimum cluster size to find. It is an algorithm for finding density based clusters in spatial data which addresses one of DBSCAN"S major weaknesses i.e. of detecting meaningful clusters in data of varying density. It outputs cluster ordering which is a linear list of all objects under analysis and represents the density-based clustering structure of the data. Here parameter epsilon is not necessary and set to maximum value. OPTICS abstracts from DBSCAN by removing this each point is assigned as "core distance", which describes distance to its MinPts point. Both the core-distance and the reachability-distance are undefined if no sufficiently dense cluster w.r.t epsilon parameter is available [6].

6) STING

STING (STastical INformation Grid) is a grid-based multi resolution clustering technique in which the embedded spatial area of input object is divided into rectangular cells[6]. Statistical information regarding the attributes in each grid cell, such as the mean, maximum, and minimum values are stored as statistical parameters in these rectangular cells. The quality of STING clustering depends on the granularity of the lowest level of grid structure as it uses a multi resolution approach to cluster analysis. Moreover, STING does not consider the spatial relationship between the children and their neighbouring cells for construction of a parent cell. As a result, the shapes of the resulting clusters are is thetic, that is, all the cluster boundaries are either horizontal or vertical, and np diagonal boundary is detected. It approaches clustering result of DBSCAN if granularity approaches 0. Using count and cell size information, dense clusters can be identified approximately using STING.

REFERENCES

[1] "Algorithms for Opinion Mining and Sentiment Analysis: An Overview" G. Sneka, CT. Vidhya International Journal of Advanced Research in Computer Science and Software Engineering Research Paper Volume 6, Issue 2, February 2016.

[2] "A Survey on Sentiment Classification and Analysis Using Data Mining" Ashish Shukla, Shweta Shukla, International Journal of All Research Education and Scientific Methods.

[3] "Survey on Classification Algorithms for Data Mining:(Comparison and Evaluation)" Delveen Luqman Abd AL-Nabi, Shereen Shukri Ahmed, Computer Engineering and Intelligent Systems ISSN 2222-1719 (Paper) ISSN 2222-2863 (Online) Vol.4, No.8, 2013.

[4] "A Survey of Classification Methods and Applications for Sentiment Analysis" M.Govindarajan ,Romina M. The International Journal Of Engineering And Science (IJES) ISSN(e): 2319 – 1813 ISSN(p): 2319 – 180.

[5] "Classification Algorithms for Data Mining: A Survey" Raj Kumar, Dr. Rajesh Verma International Journal of Innovations in Engineering and Technology (IJIET).

[6]"A Review: Comparative Study of Various Clustering Techniques in Data Mining" Aastha Joshi, Rajneet Kaur International Journal of Advanced Research in Computer Science and Software Engineering ISSN:2277128X Volume3, Issue 3, March 2013 Research Paper.

[7] "Classification algorithm in Data mining: An Overview" S.Neelamegam Dr. E.Ramaraj International Journal of P2P Network Trends and Technology (IJPTT) – Volume 4 Issue 8- Sep 2013.

[8] Opinion Mining From Text Reviews Using Machine Learning Algorithm *Poobana S, Sashi Rekha k*