2014 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)
July 6-11, 2014, Beijing, China

Clustering based Outlier Detection in Fuzzy SVM

Rahul K. Sevakula and Nishchal K. Verma
Department of Electrical Engineering, Indian Institute of Technology Kanpur, India
E-Mail : srahulk@iitk.ac.in, nishchal@iitk.ac.in

***Abstract*— Fuzzy Support Vector Machine (FSVM) has become a handy tool for many classification problems. FSVM provides flexibility of incorporating membership values to individual training samples. Performance of FSVM largely depends on how well these membership values are assigned to the training samples. Recently, a new approach for assigning membership values was proposed, where only possible outliers are allowed to have membership value lower than '1'. For doing the same, first DBSCAN clustering is performed to find the set of possible outliers and such possible outliers were then assigned membership values based on some heuristics. All other remaining samples were assigned a membership value of '1'. This paper extends the same approach by further analyzing the algorithm, introducing Fuzzy C-Means clustering based heuristic for assigning membership values and also comparing two methods of finding optimal parameters for FSVM model. Experiments have been performed over 4 real world datasets for comparing and analyzing the different methods.**

Keywords — fuzzy svm; clustering; outlier detection; dbscan; fuzzy c means

I. INTRODUCTION

Support Vector Machine (SVM) is a very popular classification algorithm. Its popularity lies in the fact that SVM gives consistently good performance over large number of datasets. It solves a convex objective function; hence given a set of parameters, the solution remains the same for every run. SVM targets at making a generalized classifier i.e. a classifier that would work well not only for training data, but also for most unseen data. Unlike other classification algorithms, it attempts to reduce structural error instead of training error [1][2].

The initial SVM, also termed as Maximal Margin Classifier suffered severely from outliers. Though the variations that were introduced later, like C-SVM and γ -SVM does allow misclassifications, they are surely not sufficient to solve the outlier problem of the classifier. Fuzzy SVM (FSVM), introduced by Lin et-al [3] allows one to incorporate fuzzy membership values to data samples. These membership values are to be given by the user. So they are either assigned based on prior knowledge or based on some heuristic membership function. As of now FSVM has been primarily used for reducing effect of outliers and for applying differential weights to different training samples. For e.g.: FSVM can be used in problems where samples are received over a period of time and recent samples are given higher weights as compared to the older samples.

Class Imbalance is another issue posed by SVM. Class Imbalance problem being very famous and common in other classification algorithms, many solutions have been proposed by researchers. Based on the paper by Verpoulos et-al [4], having differential costs in C-SVM is quite effective against class imbalance. Batuwita et-al incorporated this in FSVM by multiplying the already assigned membership values with differential costs [5]. In his paper, he also proposed many heuristic membership functions for FSVM which showed significant improvement in results of many datasets.

Recently a new manner of assigning membership values was proposed by a paper [6]. It basically tried to give the message that having a continuous heuristic membership function for assigning membership values may not be a good idea. This is because in such a case all training samples would be assigned a membership value lower than 1, which actually would defy the purpose of FSVM and SVM. The paper proposed that first DBSCAN clustering algorithm could be first used for finding the set of possible outliers and only these possible outliers be assigned a membership value that is lesser than 1. The paper also proposed some heuristic membership functions using Hausdorff distance.

This paper extends the same approach and further analyzing the algorithm. It proposes a new heuristic for assigning membership values based on Fuzzy C Means clustering. It also checks for *Gm* based optimal FSVM parameter search for imbalanced datasets. The flow of rest of the paper is as follows. Section II explains the relevant theory of all. Section III describes the methodology presented in [6] in detail. Section IV proposes new experiments for analyzing the algorithm's effectiveness and also proposes the new FCM based method. Section V describes the experimentation procedures. Section VI shows the results and tries to derive conclusions from the same. Section VII concludes the paper.

II. THEORY

A. Support Vector Machine

Vapnik first introduced SVM as Maximal Margin Classifier (MMC) [2]. MMC is a binary classifier, which in a linearly separable case finds the hyper-plane that separates the two classes with maximum possible margin and without misclassifying any of the training data. The mathematical background behind this is that maximal margin ensures lower VC dimension and hence low generalization error. To accommodate misclassifications of training data without altering the underlying principle, C-SVM was introduced. The objective function solved in C-SVM is given in Eq.1. In C-SVM, misclassifications are allowed with a penalty of cost 'C'. The ability to use kernel function is a major asset of SVM. Applying kernel trick allows SVM to perform nonlinear classification at very low computation.

$$\begin{aligned}&\min_{w,b}\ \frac{1}{2}w.w+C\sum_{i=1}^{l}\xi_i\\&s.t.\ \ y_i(w.\phi(x_i)+b)\geq 1-\xi_i\ \ ,i=1,2,.....l\\&\qquad \xi_i\geq 0\ \ ,\ i=1,2,.....l\end{aligned} \tag{1}$$

In Eq.1, 'C' is cost, ξ_i is margin error i.e. normal distance of the i^{th} sample from hyper-plane if misclassified, $\phi()$ is the mapping function used while performing non linear classification. This mapping function is taken care by the kernel trick in dual form of the objective function [2].

B. Fuzzy Support Vector Machine

FSVM as proposed by Lin et-al [3], gives the facility of incorporating fuzzy membership values in SVM. This is done by introducing a new variable 's_i' in the objective function, which denotes the membership value of i^{th} sample. It must be noted that the membership values will only come to effect for misclassified samples. This basically means that the penalty imposed due to a misclassified sample having low membership value will be low. The objective function of FSVM is given in Eq.2.

$$\begin{aligned}&\min_{w,b}\ \frac{1}{2}w.w+C\sum_{i=1}^{l}s_i\xi_i\\&s.t.\ \ y_i(w.\phi(x_i)+b)\geq 1-\xi_i\ \ ,i=1,2,.....l\\&\qquad \xi_i\geq 0,\ s_i\geq 0\ ,\ i=1,2,.....l\end{aligned} \tag{2}$$

C. Class Imbalance Learning

Class Imbalance is the case when one class is significantly larger than the other class. This affects SVM in the sense that the classifier may become biased in favor of the larger class. This is because it is highly probably that more misclassifications will be from the larger class; hence their effect in the objective function will also be more. Hence the classifier becomes biased. Haibo et-al [7] have given an elaborate review on various CIL techniques. Verpoulos et-al's Differential Error Cost (DEC) [4] is one such technique where different cost functions/values are given for the two classes. The objective function in such a case is given in Eq.3. In Eq.3 C^+,ξ_i^+,C^-,ξ_i^- are the cost values and margin errors of positive class and negative class respectively. Akbani et-al [8] found that optimal performance is shown when ratio of C^+/C^- is equal to the ratio of number of samples in negative class to that in positive class.

$$\begin{aligned}&\min_{w,b}\ \frac{1}{2}w.w+C^+\sum_{i=1}^{l}\xi_i^+ +C^-\sum_{i=1}^{l}\xi_i^-\\&s.t.\ \ y_i(w.\phi(x_i)+b)\geq 1-\xi_i\ \ ,i=1,2,.....l\\&\qquad \xi_i\geq 0\ \ ,\ i=1,2,.....l\end{aligned} \tag{3}$$

D. Fuzzy SVM with CIL

Batuwita et-al incorporated FSVM and DEC simultaneously by multiplying the DEC ratio with membership values, as shown in Eq.4. $f(x_i)$, r are the membership values and DEC ratio factor respectively. The DEC ratio is as per Akbani et-al's conclusion of optimal ratios. r lies in the range of [0,1] where r for minority class is equal to '1' and for majority class, it is equal to the ratio of minority to majority.

$$\left.\begin{aligned}s_i^+&=f(x_i^+)\,r^+\\ s_i^-&=f(x_i^-)\,r^-\end{aligned}\right\} \tag{4}$$

Batuwita et-al tested his work on 10 datasets and four membership functions shown in Eq.5-8.

$$f_{lin}^{cen}(x_i)=\ 1-\frac{d_i^{cen}}{\max(d_i^{cen})+\Delta} \tag{5}$$

$$f_{\exp}^{cen}(x_i)=\ \frac{2}{1+\exp\left(\beta d_i^{cen}\right)},\ \ \beta\in[0,1] \tag{6}$$

$$f_{lin}^{hyp}(x_i)=\ 1-\frac{d_i^{hyp}}{\max(d_i^{hyp})+\Delta} \tag{7}$$

$$f_{\exp}^{hyp}(x_i)=\ \frac{2}{1+\exp(\beta d_i^{hyp})},\ \beta\in[0,1] \tag{8}$$

In Eq.5-6, 'd_i^{cen}' refers to the Euclidean distance of sample x_i from the centre of the individual class. For Eq.7-8, a normal SVM is first run on the dataset, then absolute value of functional margin d_i^{hyp} is calculated from this initial hyper-plane, which is then later used in calculating the membership values. Δ and β are other user defined parameters. The purpose of Δ is to avoid having zero membership value. β on the other hand is generally given a value in the range [0,1].

III. FSVM WITH HAUSDORFF MEASURE

The paper on FSVM with Hausdorff Distance [6] argues with the traditional approach of assigning membership values as shown in Eq.5-8. It mentions that Eq.5,6 relies on the Euclidean distance from center of the class. Hence this would not suit cases where individual classes are made of irregular or broken structures. In addition to it, Eq.7-8 is run with the assumption that initial hyper-plane is quite good in separating the two classes. In addition to this, it can be computationally very expensive. The paper also argues that such general purpose continuous membership defy the principle with which SVM was actually built i.e. having zero training error and maximum margin. C-SVM and γ-SVM controls this strictness by allowing only some major outliers to be misclassified. It mentions that the best course of action would be to give membership values lower than '1' to only those samples that are possibly outliers and otherwise should be given a value of '1'.

The steps for the methodology begin here. DBSCAN clustering [9] is first used on individual classes to find the set of possible outliers. DBSCAN, a density based clustering method was selected because it is simple, it can recognize broken and irregular structures and also it is computationally very fast. Two parameters namely '*k*' and 'ε' need to be set here. To begin DBSCAN, a heuristic of $k = n+1$ is decided where n is the number of features of data samples, and ε is initialized to 0.1 times the largest diameter of the ellipsoid that can encompass all the training samples of the class called *diam_lar*. Those samples which do not belong to any of the formed clusters are considered as possible outliers and the ratio of the number of possible outliers to the total number of samples is termed as *out_ratio*. An estimate of *out_ratio* is to be given by the user based on prior knowledge. The aim of defining *out_ratio* is that number of possible outliers should not be more than a given estimate. So DBSCAN is first run and *out_ratio* is calculated. If the *out_ratio* found is greater than the estimate, then ε is incremented by *incr_fac* times. In the experiments, *incr_fac* was taken equal to 1.2. The same continues until *out_ratio* becomes just lesser than the estimate, after which DBSCAN runs are stopped. Two sets namely '*out_*class' which contains samples that are probably outliers and '*in_class*' which contain the remaining samples are formed. After this step, all the samples belonging to '*in_class*' set are assigned a membership value of '1'. For assigning membership values to '*out_class*' set samples, two heuristics using Hausdorff distance [10] were introduced in the paper.

IV. More Heuristics & Further analysis

A. Fuzzy C Means Clustering

Clustering methods can be broadly classified into two categories namely Hard Clustering and Soft Clustering. This categorization is done w.r.t. a sample's belongingness to a cluster. Fuzzy C Means (*FCM*) is a soft clustering method where a sample can belong to more than one cluster. A sample's degree of belongingness to a cluster is defined by its membership value for that cluster.

The popular version of the FCM algorithm was first introduced by J.C. Bezdek [11]. *FCM* is basically a fuzzy version of the *k-Means* algorithm. It tries to minimize the objective function given in Eq.9. The algorithm begins by initializing membership matrix $U = [\mu_{ij}]$ with random membership values, where μ_{ij} refers to i^{th} sample's membership value towards j^{th} cluster. Thus for N samples and C clusters, U forms an N x C matrix. Based on the current state of *U,* cluster centers are calculated as shown in Eq.10. The matrix U is then re-updated as per Eq.11. This process of consequent updating of U and cluster centers is repeated until the difference between consecutive U matrices is below a threshold value. In Eq.9-11, though m can take any integer value, generally m = 2 is used for most problems.

$$J = \sum_{i=1}^{N}\sum_{j=1}^{C}\mu_{ij}^{m}\left\|x_i - c_j\right\|^2 , \quad 1 < m < \infty \tag{9}$$

$$c_j = \frac{\sum_{i=1}^{N}\mu_{ij}^{m} . x_i}{\sum_{i=1}^{N}\mu_{ij}^{m}} \tag{10}$$

$$\mu_{ij} = \frac{1}{\sum_{k=1}^{C}\left(\frac{\left\|x_i - c_j\right\|}{\left\|x_i - c_k\right\|}\right)^{\frac{2}{m-1}}} \tag{11}$$

B. FCM Based Heuristic

Similar to *k-means* clustering, FCM also needs the value for number of clusters in prior. For finding this, we perform *k-means* clustering on the dataset while varying *k*, no. of clusters from 2 to $\sqrt{N}$. Global Silhouette Index (GSI) [13] is found for measuring the overall quality of all clusters. The value of k for which GSI is highest, is selected for performing FCM. After performing FCM, we get k membership values for each sample. For each sample i, their maximum membership value which is termed as max_mem_i is found. If a sample's *max_mem* is below a user defined threshold, it would naturally imply that the sample does not belong to any of the clusters prominently. Therefore the sample could possibly be an outlier. Following earlier methodology of assigning membership values, all non-outliers are assigned a membership value of '1' and the possible outliers are assigned membership value equal to their *max_mem.*This is attractive because unlike other heuristics, FCM has the capability of naturally finding fuzzy membership values for samples. This methodology for assigning membership values will be termed as f_{FCM}^{prop} for further reference.

C. Further Analysis of the Concept

In the earlier paper, much analysis was not presented for the introduced concept. This paper is aimed towards analyzing two parts.

1) Effect of assigning membership values below 1 to only the cluster defined outliers. For doing this, we will compare performance of FSVM with CIL while using Eq.5-8 with and without the introduced concept. The membership functions given in Eq.5-8 are termed as f_{lin}^{cen} , $f_{\exp}^{cen}$, f_{lin}^{hyp} , $f_{\exp}^{hyp}$.

Applying the introduced concept on them would mean we first find the set of possible outliers using DBSCAN clustering. Samples belonging to *in_class* set are assigned a membership value of '1' and samples belonging to *out_class* set are assigned membership values using Eq.5-8. Such assignment of membership values while using Eq.5-8 for possible outliers shall be termed respectively as $f_{cen\_lin}^{prop}, f_{cen\_\exp}^{prop}, f_{hyp\_lin}^{prop}, f_{hyp\_\exp}^{prop}$.

2) Accuracy is not considered to be a good measure of classifier performance on imbalanced datasets. This is because in highly imbalanced datasets, even if the classifier classifies all samples as majority class, still the accuracy for the classifier would be high, which obviously is undesirable. A popular performance measure for imbalanced datasets [5], [6] is *Gm*, gemoetric mean of Sensitivity (SE) and Specificity (SP). While finding optimal parameters for FSVM with RBF kernel, we generally check the 5 fold cross-validation Accuracy for each pair of C and γ. The pair giving highest Accuracy value is selected. This methodology is so common that many SVM packages like *LIBSVM* [13] give inbuilt functions for doing the same very easily. As the search for optimal parameters is being done on imbalanced datasets in this paper, using Accuracy for finding optimal parameters may not be the best course of action. In this paper, we wish to compare the option of using *Gm* instead of Accuracy for finding optimal FSVM parameters, i.e. find *Gm* for each pair of *C* and γ and then select that pair which gives highest *Gm* value.

V. Experimentation

As the work presented here is basically an extension of papers [5] and [6], similar training, testing and evaluation procedures have been followed. The performance of normal SVM with CIL and FSVM-CIL with membership functions f_{lin}^{cen}, $f_{\exp}^{cen}$, f_{lin}^{hyp}, $f_{\exp}^{hyp}$, $f_{cen\_lin}^{prop}$, $f_{cen\_\exp}^{prop}$, $f_{hyp\_lin}^{prop}$, $f_{hyp\_\exp}^{prop}$ and f_{FCM}^{prop} were compared on four benchmark datasets taken from UCI Repository. Details of the datasets are given in Table I. For comparing the options of finding FSVM optimal parameters with *Gm* and Accuracy, all experiments with both the options have been performed.

A variant of LIBSVM [12] that allows one to give variable weights to individual samples was used for implementing FSVM with CIL on MATLAB. An extensive fivefold cross validation was performed on each dataset for evaluating the performance. SE and SP values were found for the five folds and then averaged to find the final *Gm* value. RBF Kernel was used for all SVM and FSVM operations. For finding optimal parameters of C and γ for each fold, internal fivefold cross validation was performed for evaluating the pair. It must be noted that it is here where the two measures namely Accuracy and *Gm* were used for evaluating and comparing the

Table II: Details of Datasets Used

Dataset	Pos.	Neg.	Total	Imbalance Ratio	Total Classes	Positive Class
Breast Cancer	239	444	683	7:13	2	2
Ecoli	77	259	336	23:77	8	2
Haberman	81	225	306	26:74	2	2
Satimage	626	5809	6435	10:90	7	4

parameter pairs. Two levels of grid search were performed for finding optimal parameters. First level is the standard search where $log_2 C$ and $log_2 \gamma$ were varied across {1, 2, 3, ……..,. 15} and {-15, -14, -13, ……, -1} respectively. After finding optimal $\log_2$ parameters from current search say $\overline{C}$ and $\overline{\gamma}$, fine grid serach was performed by varying $log_2 C$ and $log_2 \gamma$ across {$\overline{C}$-0.75, $\overline{C}$-0.5, …., $\overline{C}$+0.75} and {$\overline{\gamma}$-0.75, $\overline{\gamma}$-0.5, …., $\overline{\gamma}$+0.75} respectively. The best pair among these was finally selected to build the FSVM.

For exponential based membership functions namely $f_{\exp}^{cen}$ and $f_{\exp}^{hyp}$, β was varied from 0.1 to 1.0 in steps of 0.1. The value of β which gave best results was used and noted down in round brackets in results section. The same optimal values of β were later used for $f_{cen\_\exp}^{prop}$ and $f_{hyp\_\exp}^{prop}$ as well. For $f_{cen\_lin}^{prop}$, $f_{cen\_\exp}^{prop}$, $f_{hyp\_lin}^{prop}$ and $f_{hyp\_\exp}^{prop}$, *out_ratio* parameter was varied from 0.1 to 1.0 in steps of 0.1, and best value for the same was noted down in round brackets in results section

VI. Results & Analysis

The results of different membership functions with Accuracy being used for finding optimal parameters are shown in Table II. Table III shows the same when *Gm* is instead used for finding optimal FSVM parameters.

By observing the results tables, we observe that the introduced heuristic f_{FCM}^{prop} has performed well like other membership function, but did not show spectacular or extraordinary results. It has surely given better results than Normal SVM for all four datasets in Table II. It also gave maximum *Gm* amongst all classifiers for Breast Cancer dataset.

For showing the two proposed analysis in a better fashion, Tables IV and V have been made from the results given in Table II and Table III. Table IV gives details regarding the number of datasets for which the performance improved by changing to the proposed membership function. It can be seen that in general the performance has improved, with maximum improvement occurring when membership functions changed from f_{lin}^{cen} to $f_{cen\_lin}^{prop}$ and from f_{lin}^{hyp} to $f_{hyp\_lin}^{prop}$. It was observed that whereas in some datasets where performance of proposed membership function is higher, the performance is actually significantly higher; in other cases when performance of proposed membership function is lower, the difference is

TABLE II: FSVM CLASSIFICATION RESULTS WHILE USING ACCURACY FOR FINDING OPTIMAL PARAMETERS

Dataset	Results	Normal SVM with CIL	FSVM f_{lin}^{cen} (%)	FSVM $f_{\exp}^{cen}$ (%)	FSVM f_{lin}^{hyp} (%)	FSVM $f_{\exp}^{hyp}$ (%)	FSVM $f_{cen\_lin}^{prop}$ (%)	FSVM $f_{cen\_\exp}^{prop}$ (%)	FSVM $f_{hyp\_lin}^{prop}$ (%)	FSVM $f_{hyp\_\exp}^{prop}$ (%)	FSVM f_{FCM}^{prop} (%)
Breast Cancer	*SE*	97.91	97.49	97.91	97.97	97.91	98.75	98.32	98.32	97.49	99.17
	SP	96.85	96.84	96.85	96.84	97.07	96.85	97.30	96.39	97.07	97.07
	Gm	97.38	97.17	97.38 (0.1)	97.37	97.49 (0.4)	97.79 (0.7)	97.81 (0.5)	97.35 (0.2)	97.28 (0.6)	98.11 (0.4)
Ecoli	*SE*	87.08	90.92	88.33	88.42	88.42	88.33	88.42	87.17	88.42	89.67
	SP	87.68	89.20	88.06	88.84	89.22	93.07	89.22	90.38	89.99	86.52
	Gm	87.38	90.05	88.20 (0.8)	88.63	88.82 (0.2)	90.67 (0.6)	88.20 (0.6)	88.76 (0.6)	89.20 (0.4)	88.08 (0.7)
Haberman	*SE*	33.38	35.81	38.38	49.63	47.94	42.13	37.13	42.13	40.88	43.38
	SP	88.89	85.33	87.56	66.67	81.78	84.00	88.44	84.00	85.33	79.56
	Gm	54.47	55.28	57.97 (0.1)	57.52	62.61 (0.7)	59.49 (0.1)	57.31 (0.1)	59.49 (0.1)	59.06 (0.1)	58.75 (0.9)
Sat-image	*SE*	69.17	68.69	69.49	69.01	91.85	68.69	69.17	69.49	69.49	69.81
	SP	95.64	95.54	95.68	95.71	88.14	95.75	95.64	95.59	95.59	95.47
	Gm	81.34	81.01	81.54 (0.1)	81.27	89.98 (1)	81.10 (0.1)	81.34 (0.1)	81.50 (0.2)	81.50 (0.2)	81.64 (0.5)

TABLE III: FSVM CLASSIFICATION RESULTS WHILE USING GM FOR FINDING OPTIMAL PARAMETERS

Dataset	Results	Normal SVM with CIL	FSVM f_{lin}^{cen} (%)	FSVM $f_{\exp}^{cen}$ (%)	FSVM f_{lin}^{hyp} (%)	FSVM $f_{\exp}^{hyp}$ (%)	FSVM $f_{cen\_lin}^{prop}$ (%)	FSVM $f_{cen\_\exp}^{prop}$ (%)	FSVM $f_{hyp\_lin}^{prop}$ (%)	FSVM $f_{hyp\_\exp}^{prop}$ (%)	FSVM f_{FCM}^{prop} (%)
Breast Cancer	SE	98.32	97.49	97.92	97.91	98.74	97.91	98.75	98.74	98.74	98.75
	SP	96.85	96.62	97.07	96.84	96.85	97.07	96.62	97.30	97.07	96.62
	Gm	97.85	97.05	97.49 (0.3)	97.37	97.79 (1)	97.49 (0.3)	97.68 (0.1)	98.02 (0.2)	97.90 (0.4)	97.68 (0.1)
Ecoli	SE	90.92	93.42	96.08	89.67	93.42	87.08	93.42	92.33	93.58	93.58
	SP	84.22	86.89	84.59	88.07	85.35	92.30	84.97	85.75	84.98	85.35
	Gm	87.50	91.10	90.15 (0.5)	88.86	89.29 (0.1)	89.65 (0.6)	89.09 (0.5)	88.98 (0.9)	89.18 (0.3)	89.37 (0.1)
Haberman	SE	49.34	50.51	55.44	56.91	55.44	54.26	54.19	50.59	53.16	50.44
	SP	79.11	80.44	78.22	76.89	78.67	77.78	80.00	80.00	78.67	79.56
	Gm	62.48	63.75	65.85 (0.9)	66.15	66.04 (0.1)	64.97 (0.2)	65.84 (0.7)	63.62 (0.6)	64.67 (0.9)	63.35 (0.8)
Sat-image	SE	91.05	88.50	91.85	88.34	90.25	89.77	91.21	91.53	90.09	90.26
	SP	89.10	90.81	88.14	89.58	88.93	90.15	88.69	88.23	89.71	89.14
	Gm	90.07	89.64	89.98 (1)	88.96	89.59 (0.6)	89.96 (0.4)	89.94 (0.6)	89.86 (0.1)	89.90 (0.9)	89.69 (0.3)

TABLE IV: FSVM RESULTS ANALYSIS: TRADITIONAL V/S PROPOSED MEMBERSHIP FUNCTIONS

Membership Function	Table II: No. of Datasets whose Results Improved	Table III : No. of Datasets whose Results Improved
f_{lin}^{cen} to $f_{cen\_lin}^{prop}$	4	3
$f_{\exp}^{cen}$ to $f_{cen\_\exp}^{prop}$	2	2
f_{lin}^{hyp} to $f_{hyp\_lin}^{prop}$	3	3
$f_{\exp}^{hyp}$ to $f_{hyp\_\exp}^{prop}$	1	2

TABLE V: FSVM RESULTS ANALYSIS: ACCURACY BASED OPTIMAL PARAMETERS V/S GM BASED OPTIMAL PARAMETERS

Membership Function	Table II Results are higher : No. of Datasets	Table III Results are higher : No. of Datasets	Table II Results are significantly higher : No. of Datasets	Table III Results are significantly higher : No. of Datasets
Normal SVM	0	4	0	2
f_{lin}^{cen}	1	3	0	2
$f_{\exp}^{cen}$	0	4	0	3
f_{lin}^{hyp}	0	4	0	2
$f_{\exp}^{hyp}$	1	3	0	1
$f_{cen\_lin}^{prop}$	2	2	0	2
$f_{cen\_\exp}^{prop}$	1	3	0	2
$f_{hyp\_lin}^{prop}$	0	4	0	2
$f_{hyp\_\exp}^{prop}$	1	3	0	2
f_{FCM}^{prop}	1	3	0	2

extremely small. Therefore going for the proposed membership functions could be a safer option. Table V on other hand gives performance details of different classifiers when Accuracy is used for finding optimal FSVM parameters as compared to when *Gm* is used for finding the same. The table has 4 columns apart from the membership function column. First & Second column gives details as to given a membership function, for how many datasets, Table II gave better results and for how many datasets, Table III gave better results respectively. It was observed that for some cases, the *Gm* difference between the two tables was negligible and in some cases it was quite significant. We would call the difference to be significant if the change in *Gm* is around 2% or more. For including this in our analysis, two more columns were introduced in Table V which showed the number of datasets where *Gm* results were significantly higher than the other table. As can be seen from the results tabulated in Table V, FSVM with *Gm* based optimal parameters are found to perform better in most cases. Though in some cases FSVM with Accuracy based optimal parameters gave better performance, but in no case was the performance difference significant. Whereas in the case of FSVM with *Gm* based optimal parameters, there were many instances where performance improvement was significant.

VII. CONCLUSIONS

This paper extended the work on clustering based outlier detection for assigning FSVM membership values. More analysis was performed to validate improvements by introducing the concept on 4 general purpose membership functions. It was found that though the results did not improve in all cases, in some cases the results improved dramatically and in no case did the results degrade significantly. A new heuristic f_{FCM}^{prop} was introduced, which worked well like other membership functions. Another analysis was done for checking *Gm* based FSVM optimal parameter search. The results have almost made it evident that for imbalanced datasets, *Gm* based FSVM optimal parameter search is a better and safer option than the Accuracy based FSVM optimal parameter search.

REFERENCES

[1] V.N. Vapnik, "An Overview of Statistical Learning Theory", *IEEE Trans. on Neural Networks,* vol. 10, no. 5, pp. 988-999, Sep. 1999.

[2] N. Cristianini and J. Shawe Taylor, An introduction to Support Vector Machines, Cambridge university press, 2000

[3] C.F. Lin and S.D. Wang, "Fuzzy Support Vector Machines", *IEEE Trans. on Neural Networks*, vol. 13, no .2, pp. 464-471, Mar. 2002.

[4] K. Veropoulos, C. Campbell and N. Christianini, "Controlling the sensitivity of Support Vector Machines", Proc. Int. Joint Conf. Artificial Intelligence, pp. 55-60, 1999.

[5] R. Batuwita and V. Palade, "FSVM-CIL: Fuzzy Support Vector Machines for Class Imbalance Learning", *IEEE Trans. on Fuzzy Systems*, vol. 18, no. 3, pp. 558-571, Jun. 2010.

[6] R.K. Sevakula and N.K. Verma, "Fuzzy Support Vector Machine using Hausdorff Distance", 2013 IEEE International Conference on Fuzzy Systems, pp.1-6, Jul. 2013

[7] H. Haibo and E. Garcia, "Learning from Imbalanced Data", *IEEE Trans. on Knowledge and Data Eng.,* vol. 21, no. 9, pp.1263-1284, Sep. 2009.

[8] R. Akbani, S. Kwek, and N. Japkowicz, "Applying support vector machines to imbalanced datasets", Proc. 15th Eur. Conf. Mach. Learning, pp. 39-50, 2004.

[9] M. Ester, H.P. Krigel, J. Sander and X. Xu, "A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise", Proc. Int. Conf. on Knowledge Discovery and Data Mining, pp. 226 – 231, 1996.

[10] J. Henrikson, "Completeness and Total Boundness of the Hausdorff Metric", MIT Maths UJM, vol. 1, pp. 69-79.

[11] James C Bezdek, "Pattern recognition with fuzzy objective function algorithms", Kluwer Academic Publishers, 1981

[12] C.C. Chang and C.J. Lin, " LIBSVM : a library for support vector machines", ACM Trans. on Intelligent Systems and Technology, 2:27:1--27:27, 2011

[13] N.K. Verma, A. Roy, "Self Optimal Clustering Technique using Optimized Threshold Function", IEEE Systems Journal, Early Access.