ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 2, February 2015

Sensor Drift Compensation in Time Series Prediction through Regularized Ensemble of Classifiers

Shruti Asmita[1], K.K. Shukla[2]

M.Tech. Student, Department of Computer Science, Banasthali University, Jaipur, India[1]

Professor and Head of Department, Department of Computer Science and Engineering, Indian Institute of Technology, Banaras Hindu University, Varanasi, India [2]

Abstract: Concept drift is a major problem that the chemical sensor community is facing in their research and development. This arises when the predictive characteristic feature of the target variable in the sensor setup, changes due to some chemical or physical interaction of the environment elements with the surface of the sensor and other factors such as aging and poisoning of the surface. This is not a new problem but is having a wide scope of further developments. Recent research suggests an upcoming solution for this drift compensation as ensemble of classifiers with uniform weightage to all the participating classifiers or some non-uniform weightage according to performance of individual classifier. There arise some ill posed problems with this ensemble of classifiers. This paper introduces a new machine learning approach to solve this problem through regularized weighted ensemble of classifiers for overcoming the time dependent drift occurrence. The weights are chosen by regularized majority voting which are then associated with the individual classifiers to form the ensembles. This regularized drift compensation algorithm is applied to solve the gas discrimination problem for classification of 6 different volatile organic compounds over time series data set consisting of data recorded using an array of 16 metal oxides sensors for 3 years. Our experiment tries various different regularization approaches and finds best improvement in case of double regularization of Single Value Decomposition and L2 regularization. Results clearly indicate improved classification accuracy in presence of drift and better results than recent reporting. To the best of our knowledge, such a machine learning approach has not yet been applied for drift correction in sensor community.

Keywords: concept drift, support vector machine, ensemble learning, regularization, singular value decomposition

I. INTRODUCTION

Concept refers to the whole distribution of the problem in a certain point of time. Concept drift represents a change in the distribution of the problem, possibly being a feature change (change only in unconditional probability distribution function) or conditional change (change only in posterior probabilities) or a dual change (change in both the unconditional probability distribution function and posterior probabilities) [1]. In general terms, one can say that concept drift is the difference in the reading taken from the sensor to what the actual reading should be. This is a problem of increasing importance in machine learning because the data sets under analysis are no more only the static data set but also the data streams in which the concept and distribution may vary due to various internal or external factors. Presently this is seriously affecting the electronic nose (EN) or the odour sensing research community. EN refers to a device of reproducing human sense of smell based on sensor arrays of smell and pattern recognition methods [2]. They widely uses chemical gas sensors which shows drift in time series computations and starts to give improper readings containing certain amounts of drift. There are several commercial applications of EN emerging such as food industry [3], environment monitor and control [4], public safety [5], fire detection [5], space applications [6], etc. In the human olfactory system the main components are mucous membrane (which dissolves the molecule of the breath in air) epithelium (recepting layer of human nose which produces electrical signals on each chemical reaction), bulb (receives the electrical signal sent by the epithelium after reception), cortex (receiving end of the brain and further it redirects the signals to other regions of brain accordingly) and higher brain (which finally recognizes the odour). In the EN the corresponding function units are odour molecule delivery system, sensor array, data processing system and data recording system. The sensor array is hence directly exposed to the environment and the analyte due to which there occurs concept drift in data processing phase. Few sensors that are applied to EN are conducting polymer [7], quartz crystal microbalance [8], surface acoustic wave sensor [7] and semiconductor metal oxide gas sensors [8]. All these sensors experience concept drift through some or the other factors.

Concept drift is usually an outcome of non-stationary learning problems over time since the data distribution differs with time [9]. The categorization of concept drift is shown in Fig 1. Such algorithms may be active or passive. Active algorithms are trigger based algorithms which sets a flag as soon as the change in the model is required. However the passive algorithms are evolving one. They assume the presence of drift according to the general

ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 2, February 2015

associated facts. The passive concept drift is the one, which we can compensate through any machine learning technique specially the ensemble of classifiers where the adaptivity is achieved by assigning weights to individual models output at each instance of time [10]. The passive concept drift in isolation can be a sudden (abrupt) or gradual drift. Sudden drift in the reading arises due to certain failure in apparatus or the power damage cases. Gradual drift can be first order drift or second order drift. First order drift or the real drift arise due to the chemical or physical interaction that continuously goes on between the analyte present in the environment and the exposed sensor surface. This results in aging i.e. reorganization of the sensor composition over time and poisoning i.e. irreversible binding due to chemical contamination. Second order drift arise due to external and uncontrollable environmental conditions such as humidity, thermal fluctuations, delivery systems, noise, and memory affect etc. On the other side passive concept drift in sequences can be discriminated on the basis of predictability (predictable or non-predictable), frequency (periodic or non-periodic), recurrence (cyclic recurrent, unordered recurrent or non-recurrent).In the case of EN or the chemical gas sensor, there occurs passive gradual drift in isolation both real and second order may be individually or in combination. When in sequence, chemical sensor drift is periodic (the time of start of deterioration is predictable), non-recurrent (once drift has occurred, the sensor cannot go back to the previous concept since the change in chemical composition of the sensor occurs) and predictable (the sensitivity and reaction rate of the sensor material with analyte in contact is known beforehand).

In this paper we deal with compensation of concept drift that arise in EN, using regularized weighted ensemble of classifiers. This is a supervised machine learning approach that is used for time series predictions. Here the weights are learned by regularized majority voting technique. These weights are associated with the individual models which are then ensemble to provide a group opinion in any classification. The obtained classification accuracy by this approach is giving better results than a recent reporting in this field [11]. To the best of our knowledge, such technique which deals with the problem of overfitting of the classifiers and their ensembles in the supervised machine learning has yet not been applied to the sensor' s drift correction. This technique reduces overfitting by reducing the curvature of each depression in the fitted curve and hence promote generalization which is an essential factor in classification. In the stretch of this paper, we describe the data set, the feature extraction, basic ensemble learning, regularization and other essential concepts. Further we describe the regularized drift compensation algorithm and framework followed by detailed description of our experimental findings. Experiments are all done on the dataset that records the readings through metal oxide gas sensors. Finally we present the concluding comments drawn from the results.

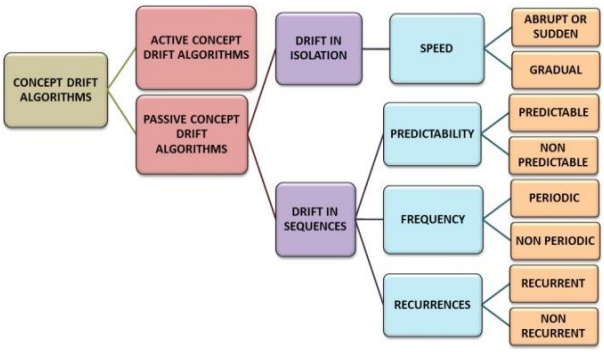

Fig. 1. Hierarchy categorization of concept drift

II. BACKGROUND.

A. General Approach

The problem of detecting concept drift is not a new one but is a challenging domain for finding improved solutions to for it. Existing techniques to detect concept drift is classified as follows (Fig 2). Baseline manipulation is a data pre-processing method which includes differential, relative or fractional transformation of individual signals based on the initial value of transient response. But this method can work only in some special cases of drift detection. Frequency Domain Filtering methods focus on removing those components of signals which are producing drift. Discrete wavelet transform is a powerful tool of filtering without creating distortion in the original data. Under Periodic Calibration category, Multiplicative drift correction method is improved approach of univariate calibration, in which the temporal variation of the system with the multiplicative drift correction factor is taken as calibration measure. Next on the same line, the component correction is a multivariate drift correction approach. It includes two correction methods i.e. Principal Component Analysis (PCA) and Partial Least Square (PLS) method. Component Deflation is another multivariate drift correction method which correspond drift to the variance produced in analysis. However Component correction methods suffer limitations in handling non linarites under their respective restrictions. Further the Attuning methods perform component correction without resorting to the use of calibration samples, but trying to deduce drift components directly from the training data. For dealing with sensor drift we can also take into account the disturbances derived from the measuring environment. In methods like PCA, the computed principal components are mutually uncorrelated. But non correlation does not guarantee statistical independence. Hence Independent Component Analysis (ICA) was introduced as a technique to separate data matrix into series of components each independent of the others. In this case independence implies that the information carried by each component cannot be inferred from the others. Orthogonal Signal Correction (OSC) is another attuning method based on a signal processing technique. Although attuning methods proved to be most promising in this stream but in case of chemicals the choice of celebrants is application specific [12]. This leads to loss of generalization and

ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 2, February 2015

standardization. Adaptive methods are another category of algorithms for drift detection and correction which works on pattern recognition and correction model. In this direction, neural network models are quite successful in detecting drifts. They are data driven, self-adaptive and nonlinear methods which give flexible modelling of real world complex relationships. But this had several drawbacks such as the selection of optimal value for learning rate, requirement of large number of training samples, slow rate of convergence etc. Evolutionary algorithms are more robust than neural network models to the discontinuous data but couldn't completely overcome the limitations of neural network model. All the above mentioned methods assume that data are linear in the feature space. Kernalized version of component analysis such as kernalized PCA can handle nonlinear data well but this techniques has not been investigated much [12]. Recently a research on chemical gas sensor drift compensation using classifier ensembles has been done [11]. This study discusses a supervised learning algorithm of drift compensation for non-linear data using weighted ensembles of classifiers on a time series data set.

Moving a step further, this paper discusses a new approach which solves the mentioned problem by regularizing the weights in weighted ensemble of classifiers in supervised learning models over time series data set. Hence the overfitting that arise in the weighted ensemble of classifiers models, is dealt with and generalization is promoted. Learning of weights is done by regularized majority voting. Overfitting arises when the generalization is decreased. The effect of overfitting can be handled either by reducing the number of features under consideration or by regularization. Reducing number of features is not a good option as far as accuracy of performance is considered. Hence regularization is applied which handles this issue by keeping all the features but reducing value of each feature parameter accordingly. Here the weights assigned to a particular classifier in the ensemble classification, acts as feature which is required to be regularized. By the term time series data set, what meant is that the data is collected over a large time and is batched in small time intervals so that the chances of presence of concept drift is higher in later batches than in earlier batches. Individual classifiers in the ensemble learning are the model of classifiers that evaluate their own opinion about the class of a particular vector. These opinions are ensemble to give the final decision. There are many possible classifiers which can take part in the ensemble decision such as decision tree classifier, k nearest neighbour classifier, bayes classifier, frequent pattern classifier, rule based classifier, support vector machine (SVM) classifier etc. Out of all these SVM is highly accurate and lies exceptional in its ability to model complex non-linear decision boundaries by mapping non-linear data to higher dimensions. Hence it can classify both the linear as well as non-linear data. Because of the support vectors, the feature vector lying on the decision boundary, the classification is much more compact than the other methods. The chances of overfitting lies least in SVM hence handling overfitting and improving generalization are most promising in this particular technique of classification. This is also the best for the time series predictions and online learning. Hence the proposed work in this paper uses the SVM model as individual classifier model which is weighted ensemble to deduce a common decision.

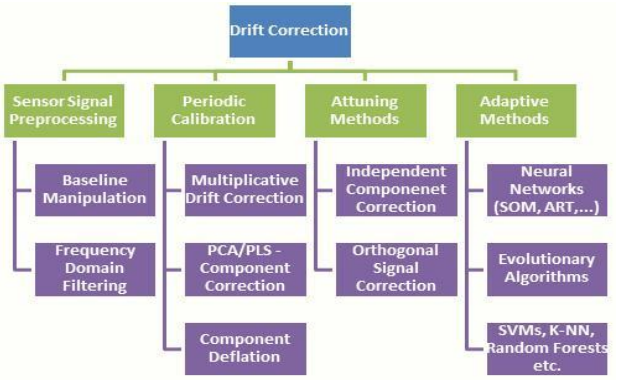

Fig. 2. Existing sensor drift correction methods

B. Dataset

Gas sensor array drift at different concentration dataset [11,13] is used for the whole experiment and analysis. It's a multivariate time series dataset which contains no missing values. This archive contains 13910 measurements from 16 chemical sensors exposed to 6 different gases at various concentration levels. The dataset was gathered during the period of January 2008 to February 2011 (36 months) in a gas delivery platform facility situated at the Chemo Signals Laboratory in the BioCircuits Institute, University of California San Diego. The measurement system platform provides versatility for obtaining the desired concentrations of the chemical substances of interest with high accuracy and in a highly reproducible manner, minimizing thereby the common mistakes caused by human intervention and making it possible to exclusively concentrate on the chemical sensors[11] The resulting dataset comprises recordings from six distinct pure gaseous substances, namely Ammonia, Acetaldehyde, Acetone, Ethylene, Ethanol, and Toluene, dosed at a wide variety of concentration levels in the intervals (50,1000), (5,500), (12,1000), (10,300), (10,600), and (10,100) ppmv, respectively. The responses of the said sensors are read in the form of the resistance across the active layer of each sensor; hence, each measurement produced a 16-channel time series, each represented by an aggregate of features reflecting the dynamic processes occurring at the sensor surface in reaction to the chemical substance being evaluated In particular, two distinct types of features were considered in the creation of this dataset: (i) the steady-state feature , And (ii), an aggregate of features reflecting the sensor dynamics of the increasing/decaying transient portion of the sensor response during the entire measurement. Each feature vector contains the 8 features extracted from each particular sensor, resulting in a 128-dimensional feature vector (8 features x 16 sensors). Our goal is to

DOI 10.17148/IJARCCE.2015.4215

ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 2, February 2015

discriminate the six different analytes regardless of their concentration. For processing purposes, the dataset is organized into ten batches, each containing the number ofmeasurements per class and month indicated in the tables below. This reorganization of data was done to ensure having a sufficient and as uniformly distributed as possible number of experiments in each batch.

TABLE I.
DISTRIBUTION OF MONTHLY DATA INTO BATCHES AND NUMBER OF SAMPLES IN EACH GAS OF CORRESPONDING BATCH

Batch ID	Month ID	Ethanol	Ethylene	Ammonia	Acetaldehyde	Acetone	Toluene
Batch 1	1,2	83	30	70	98	90	74
Batch 2	3,4,8, 9,10	100	109	532	334	164	5
Batch 3	11,12, 13	216	240	275	490	365	0
Batch 4	14,15	12	30	12	43	64	0
Batch 5	16	20	46	63	40	28	0
Batch 6	17,18, 19,20	110	29	606	574	514	467
Batch 7	21	360	744	630	662	649	568
Batch 8	22,23	40	33	143	30	30	18
Batch 9	24,30	100	75	78	55	61	101
Batch 10	36	600	600	600	600	600	600

C. Feature Extraction

The gas sensor array drift data set used in whole of the experiments discussed in this paper use metal oxide gas sensors. They response slow but best in controlled environment conditions i.e. constant air flow and fixed operating temperature. The identification of the analyte is done by observing the smooth change in conductance/resistance across sensing layer due to adsorption/desorption reaction of chemical analyte at the micro porous surface of the sensor [11]. Hence throughout the process, the static factors in determining speed and amount of reactions are analyte identity, sensor surface type and surface temperature and the only factors which actually determine identification is analyte concentration [14]. Feature extraction is defined as transformation mapping of sensor response to a lower dimension space such that the relevant information from the sensor signal is not lost. Two types of features are considered in tabulating the dataset. First is the adsorption, desorption and steady state responses of sensor elements. Second is the normalized value of difference of maximal resistance change and the baseline (equation 1).

$$|| \Delta R || = (\max_k r[k] - \min_k r[k]) / \min_k r[k] \quad (1)$$

Where r[k] is the sensor resistance, k is discrete time indexing the recording interval [0,T] when chemical vapour is present in the gas chamber. The aggregate of features reflecting rising/decaying sensor response is evaluated by exponential moving average ema_α. Ema_α is determined by calculating maximum/minimum y[k] (equation 2) for respective rising/decaying evaluation.

$$y[k]= (1-\alpha)\, y[k-1] + \alpha(r[k]-r[k-1]) \quad (2)$$

for y=1,2,3.. T where y[0]=0 is the initial condition, $\alpha=\{0,1\}$ is the smoothing parameter. Three variations of α are used i.e. 0.1, 0.01 and 0.001 for determining three values in rising observation and three values for the decaying observation. This paper also uses the same feature extraction technique on data set as above in all the reported experimental findings.

D. Ensemble Learning

The group of people can often make better decisions than individuals especially when group members come in with their own biases [15]. In the process of making a machine learn something in presence of a supervisor, we train them on several feature inputs providing them with corresponding label. Later we consider the learning good if the classification accuracy of the machine in generating the output as label, taking a new feature vector as input, is appreciable. In this we have a single decision maker and it is the single learned entity. Ensemble Learning is the process of training multiple learning machines and combines their outputs, treating them as a “committee" of decision makers. The principle is that the committee decision, with individual predictions combined appropriately, should have better overall accuracy, on average, than any individual committee member [16]. Numerous empirical and theoretical studies have demonstrated that ensemble models very often attain higher accuracy than single models. This ensemble model is also known as the Multi Classifier Systems. In the ensemble learning, the individual classifiers are aggregated together through any of the voting techniques such as majority voting, behaviour knowledge based aggregation, borda count aggregation, dynamic classifier selection etc. [17]. Out of these, majority voting has a wide successful use in time series experimentations. The three versions of majority voting are unanimous voting in which chosen class label is the one for which all classifiers agree (most difficult to achieve), simple voting in which chosen class label must be predicted by at least one more than half the number of classifiers and plurality voting in which the chosen class label must have received the highest number of votes (may or may not exceeds 50%). Plurality voting is the most optimal form of majority voting. Mathematically, if $f_k(k=1,2,\ldots..K)$ be a decision function of the k^{th} classifier in the ensemble of K number of classifiers and Cj (j=1,2,…..C) denote a label of the j^{th} class. Then equation 3 represents the number of classifiers whose decisions are known to the j^{th} class. The final decision of the ensemble of classifiers $f_{mv}(x)$ for a given test vector x due to the majority voting is determined by equation 4[18].

$$N_j = \#\{k \mid f_k(x)=C_j\} \quad (3)$$

$$f_{mv}(x) = \text{argmax} N_j \quad (4)$$

DOI 10.17148/IJARCCE.2015.4215

ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 2, February 2015

In the proposed statement of this paper the weighted majority voting is achieved with the fact that if certain experts are more qualified than others, weighting their decisions more heavily may further improve the overall performance than that can be obtained by the plurality voting.

E. Regularization

The concept of regularization was introduced in 1990's. The goal of learning is prediction. The supervised learning has a strong perspective of statistical learning theory. After learning a function for classification on the basis of training data, the function is validated on test data, data that did not appear in training set. To know how predictive the learned function is, classification function uses percentage of input that was correctly classified at the time of training. This is known as loss function. Hence statistical theory states that learning is generalization/inference problem from usually small sets of high dimensional, noisy data. But in this theory, the probabilities of acceptable predictivity are unknown. Also the constraints needed to achieve generalization are not defined. In supervised machine learning problems, the demand is not to find a function that most closely fits the data but to find one that will most accurately predict output from future input. Hence generalization of function has to be improved. Also this can be said as the overfitting of function has to be decreased. In fig 3 the blue curve is a 2 degree curve, red curve is a 4 degree curve and the green curve is the 8 degree curve which is the maximum out of the three. The green curve fits the data points the most, but the test accuracy decreases. However the blue curve shows minimum training accuracy but chances of betterment in test accuracy is the maximum in this case. Green curve shows the overfitting. Hence over fitting occurs when generalization is decreased. To prevent this overfitting, either number of features under consideration is reduced or secondly all the features are kept but value of each feature parameter is reduced. This second solution is known as Regularization of the loss function. This provides problem stability. Hence regularization can be accomplished by restricting the hypothesis space to a linear function or a polynomial of a particular degree according to the scenarios. In general, regularization deals with inverse problems (determination of the forces from the knowledge of trajectories) converting then to direct problems (computation of the trajectories of bodies from the knowledge of forces) by model selection, complexity control or by incorporating prior knowledge to the solution. Direct problems satisfy uniqueness, existence and stability and are termed as well posed problems whereas inverse problems are termed as ill posed problems. In terms of vector space, Banach space is a complete vector space endowed with method of calculating size of the vector (norm operation). A Hilbert space 'H' is a banach space further endowed with a dot product operation. Reproducing kernel hilbert space (RKHS) are built on 'H' and requires that all Dirac evaluation function in 'H' are bounded and continuous. If 'x' is some real vector and 'f' is a function from this vector space to RKHS, then dirac function in one that maps 'f' to the value 'f' has at 'x' [19]. Regularizing solutions are derived by differential linear operator (applied either in spatial domain or fourier domain) and its green function. Given a positive integral operator 'T', its inverse operator 'D' corresponds to inner product of RKHS, then the kernel function associated with 'D' is known as its Green's function. Hence smoothness to the function is provided by putting the function in RKHS. A regularization parameter 'λ' associated with the regularization term of optimization function controls the trade-off between stability and accuracy.

In case of the ensemble learning, to increase the test classification accuracy, if overfitting of the model function has to be checked, the regularization can be applied to the optimization of the loss function. This reduces the degree of the best fit polynomial so that the training classification accuracy is reduced but the test classification accuracy is improved. On the other side, if the degree of the best fit function is kept constant, overfitting can also be checked by regularization of the weightage associated to each individual classifiers participating in the ensemble learning. This reduces the curvature of each depression in the curve without reducing the degree of whole curve. Hence the loss function is modified to provide the curve fitting over the input feature vectors. Another statistical technique is bootstrap resampling in which we draw out from the dataset DS, a new set dataset DS' by random sampling with replacement. Applying several such DS' to ensemble of classifier gives a technique known as Bagging. For a large DS, the number of individual samples that are not present in any of the bootstrapped dataset is large. The probability that first training sample is not selected once is (1- 1/N) and not selected at all is $(1-1/N)^N$ [15]. Since N -> ∞, 1/e = 0.36 .Hence only about 63% of original training samples are represented in any bootstrapped set. Since bagging reduces variance, it provides an alternative approach to regularization [15] because even if each classifier is individually overfit, they are likely to be overfit to different things.

III. PROPOSED WORK

In our work, regularized ensemble of classifier has been used to cope with sensor drift. Considering a classification problem we have set of features x as inputs and class label y as output. At every time step we have a batch of data of size mt containing (feature vector, label) pairs i.e. $S_t=\{(x_1,y_1), (x_2,y_2)\ldots.. (x_{mt},y_{mt})\}$ [11]. For training and optimization of our problem, we have used a popular library libSVM [20,21]. At any time step, for current batch data we first create classifiers for all the previous batches of data. Then we perform weighted ensemble of all those classifiers using regularized majority voting technique. This regularizes the sum of weightage given to each individual classifiers participating in the ensemble learning. Fig 4 describes this whole framework.

For SVM, the loss function optimized is the hinge loss $L(f(x),y)=\max(0,1-y.f(x))$. The regularization factors that

DOI 10.17148/IJARCCE.2015.4215

ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 2, February 2015

generates the best accuracy in our case, is double regularization with combination of singular value decomposition (SVD) of weight matrix with regularization parameter λ_1 and square of norm 2 of weight matrix with regularization parameter λ_2. Other regularization factors are L1, L2 and tikhonov regularization. The objective function is:

$$\text{argmin}_{\beta 1 \ldots . \beta t} \sum_{i=1:mt} \sum_{j=1:t} \max(0, 1-\beta_i . y_i . f(x_i)) \quad (5)$$

Here β_i is achieved through regularized majority voting for the double regularization with combination of SVD and norm 2 regularization (equation 6), L1 (equation 7), L2 (equation 8) and tikhonov regularization (equation 9) as follows:

$$h_{t+1}(x) = \text{argmax}_{y=\{1..L\}} \Sigma t: f_t(x)=y.\beta_t + \lambda_1 . \text{SVD}(\beta) + \lambda_2 . (||\beta||_2) \quad (6)$$

$$h_{t+1}(x) = \text{argmax}_{y=\{1..L\}} \Sigma t: f_t(x)=y\ \beta_t + \lambda_1 . (||\beta||_1) \quad (7)$$

$$h_{t+1}(x) = \text{argmax}_{y=\{1..L\}} \Sigma t: f_t(x)=y\ \beta_t + \lambda_1 . (||\beta||_2) \quad (8)$$

$$h_{t+1}(x) = \text{argmax}_{y=\{1..L\}} \Sigma t: f_t(x)=y\ \beta_t + (\lambda_1)^2 . (||\beta||_2)^2 \quad (9)$$

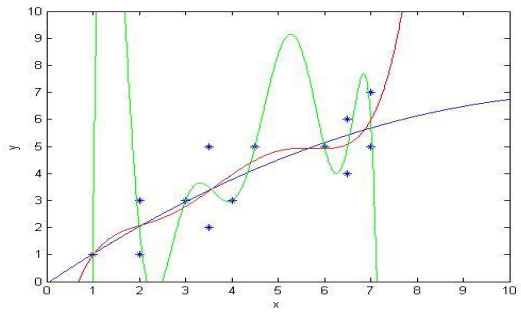

Fig. 3. Fitting of classifiers onto an example set of points

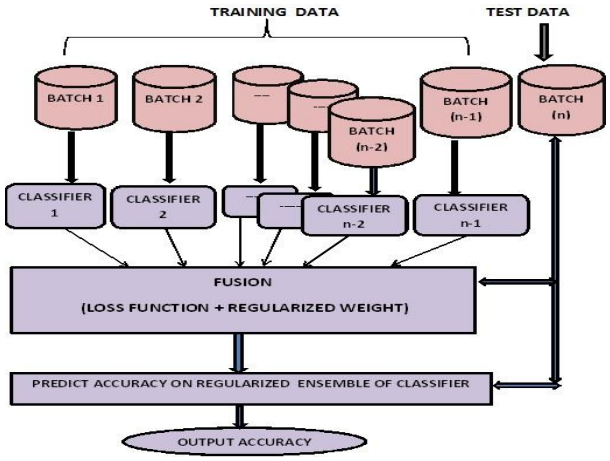

Fig. 4. Framework of proposed work

A. Algorithm

1: for each batch N data
2: load S_t={ ($data_1$,$label_1$)…….($data_{N-1}$, $label_{N-1}$)}
3: Train a SVM classifier on S_t
4. end for
5: Estimate the weights {$\beta_{1.....}$ β_{N-1}} through regularized majority voting technique
6: Evaluate ensemble of classifier model
8: Receive S_{test}= ($data_N$,$label_N$)
9: Report classification accuracy on S_{test}

The λ in the regularization controls trade-off between stability and accuracy. There are many regularization techniques in existence and this is also a topic under further research. L1 Regularization is norm 1 regularization factor which penalizes all the factors equally. It can be viewed as the selection of only the relevant factors. Defined as $\lambda_1.||\beta||_1$, this regularizer has best usage in signal processing, compressed sensing, wavelet thresholding, geophysics problem, decoding linear codes etc. However, this regularizer is slow for large scale problems. L2 regularization defined as $\lambda_1.||\beta||_2$ |restricts large value components and can use iterative methods such as conjugate gradient method for its computation. It adds less complexity to the desired output in comparison to L1 norm regularization. L2 attempts to minimize curvature at all the points in the curve by applying penalty that scales square of curvature whereas L1 penalty is linear which tends to produce many points with zero curvature. Tikhonov regularizer is a special case of L2 Regularization represented by term $(\lambda_1)^2.(||\beta||_2)^2$. Double regularization is the addition of two regularization factors to objective function. SVD is a factorization of a real or complex matrix, with many useful applications in signal processing and statistics. The minimum singular value of a matrix not only specifies the rank of the matrix, it also gives a measure of distance of the matrix from the set of matrices having a rank less than its rank. This distance is used as a measure to compare the ability of inputs to control a mode. This SVD in combination with norm 2 regularization is represented as λ_1. SVD(β) + λ_2. ($||\beta||_2$). Regularization path varies with the experimental conditions.

IV. EXPERIMENTAL RESULTS

Kernel methods are a class of algorithm for pattern analysis used when it is hard to classify data in the lower dimensions. The boundary line between two sets of data points becomes increasingly complex as the number of classes and data points are increased. Hence transformation of data to higher dimensions is required so that simple hyperplanes could be constructed. Transforming data into higher dimensions can be done only for those algorithms which use dot products of vectors in the higher dimension to come up with the boundary. SVM is one such classifier. The choice of a kernel function depends on the model to plot. A polynomial kernel allows to model feature conjunctions up to the order of the polynomial. Radial basis functions (RBF) allows circular boundaries (or hyper spheres in higher dimensions). Linear kernel allows putting linear boundaries (or hyper planes in higher dimensions). Multiclass classification is best achieved through RBF. If γ is the kernel bandwidth parameter and (x_i , x_j) is vector to be transformed to higher dimensions, equations 10 shows RBF kernel equation.

$$K(x_i , x_j) = \exp(-\gamma ||x_i - x_j||^2) , \gamma > 0 \quad (10)$$

Grid search is a parameter estimation algorithm. In v-fold cross-validation, the training set is divided into v subsets of equal size. Sequentially one subset is tested using

ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 2, February 2015

classifier trained on v-1 subsets. Hence each instance is predicted once and so the cross validation accuracy is the percentage of data which are correctly classified. The kernel parameters (C, γ) are estimated using cross-validation. Various combination of (C, γ)are tried and one with best cross validation accuracy is picked. In the experiments of our proposed work, time series multi class SVMs with RBF kernel is trained using libSVM library [20,21]. The features in the training and test datasets were scaled between -1 and +1. The kernel bandwidth parameter γ and SVM C parameter were chosen using 10 fold cross validation by performing grid search in the range $[2^{-10}, 2^{-9}, \ldots..2^{5}]$ and $[2^{-5}, 2^{-4}, \ldots..2^{10}]$, respectively. For regularization parameter, the condition $0 < \lambda_1 < 0.5$, $0 < \lambda_2 < 0.5$ is followed in choosing values of λ_1 and λ_2. The chemical gas sensors drifts and this drift causes degradation in performance of the classifier ensembles. Our goal is to cope with the sensor drift and give maximum possible classification accuracy even in the presence of drift. For the analysis, we are considering following setting.

Setting 1: Simple ensemble of classifier
Setting 2: Weighted ensemble of classifier
Setting 3: Bagging ensemble of classifier
Setting 4: Regularized weighted ensemble of classifier
  Setting 4.1: Double regularization (SVD + L2)
Setting 4.2: L1 regularization
Setting 4.3: L2 regularization
Setting 4.4: Tikhonov regularization

Setting 1, 2, 3 does not take into account any regularization technique. Setting 1 is assumes equal weightage to the individual classifiers in the ensemble. Setting 2 evaluates the weight associated with each individual classifier by majority voting. Setting 4 is same as setting 2 but weights are evaluated using regularized majority voting. Setting 3 describes regularization through bagging technique. Setting 4 has four subcases in which different regularizers are used in regularized majority voting for learning the weights to associate with each classifier. This provides smoothness to curve so that prediction accuracy could be improved. Double regularization (SVD +L2) i.e. Setting 4.1 reports best average classification accuracy than all other setting of experiment. Fig 5 shows the comparative experiment results of L1, L2, tikhonov, bagged and double regularization(SVD+L2) respectively.

The setting 4.1.results are better than the most recent reporting [11] for the same goal i.e. compensation of drift arising in the chemical gas sensor setup. Also according to another reporting [15] bagging ensemble of classifier is also an approach of regularization of objective function. Setting 4.1 result are even better than this bagging ensemble of classifier approach. Fig 6 shows the comparison between our best result (setting 4) and the corresponding recent reported result. Fig 8 reports the total execution time in ensemble learning for each batch, training all the previous batches. The execution time of batch 7 and batch 10 are very high because the testing set of these 2 batches is relatively much higher than other batches and this test batch is used twice in the algorithm, once in majority voting and once in finding prediction accuracy at the time of fusion.

Analysis of the regularizers applied in setting 4 can be done on the basis of worst case time complexity. In L1 regularization, there are total of (t-1) sum operations computed at run of algorithm. Time Complexity O(t) is reported. In L2 regularization, there are total of (t-1) sum operations, t operations to square all the elements, and 1 square root operation is computed. Time complexity O(3t) is reported. One degree regularization parameter is applied. In the tikhonov regularization, time complexity O(3t) is same as L2 regularization but here 2 degree regularization parameter is applied. In double regularization (SVD+L2), there are two expressions involved. $O(t^2)$ for SVD computation summed with O(3t) for norm 2 computation. Hence time complexity $O(t^2)$ is reported

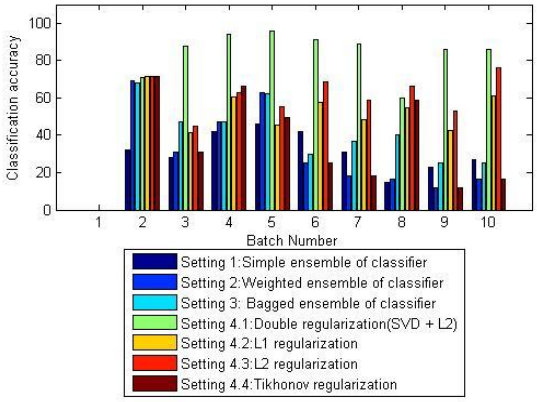

Fig. 5. Comparative results of Setting 1, 2, 3, 4.1, 4.2, 4.3 amd 4.4

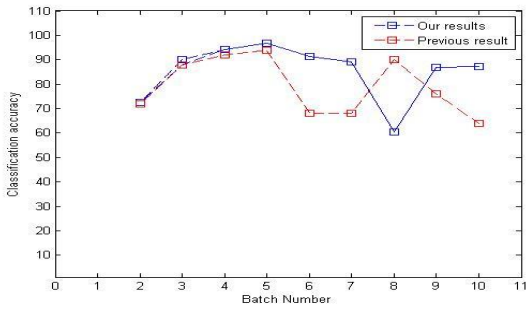

Fig. 6. Comparison of best result of proposed work (double regularization Setting 4.1.) with the most recent reported result of same aim.

The final observations from all the experiments suggests that the complexity of double regularization is more than thecomplexity of single regularization but the prediction accuracy in double classifier is far more than that in single classifier. Further the penalty in L1 regularization is not differentiable at zero so it is able to delete many noise features by estimating their coefficients to zero. Whereas the penalty in L2 regularization is differentiable everywhere and so it uses all the input features in

ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 2, February 2015

classification. Hence L2 regularization achieves higher order smoothness for curve estimation. Next, since the bagging model shows the inclusion of only about 63% of the original training samples in any bootstrapped set (as discussed in section 2.5), the regularization provided by this technique is not as smooth as the double regularization.

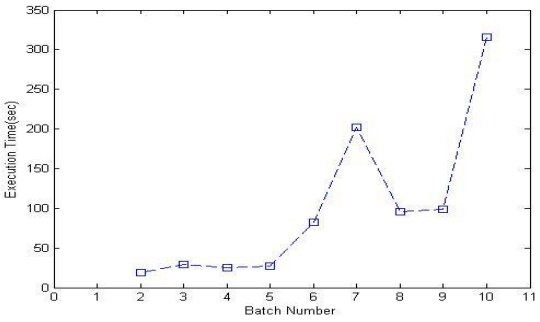

Fig. 7. Execution time in setting 4.1. i.e. double regularization of SVD+ L_2

V. CONCLUSION

A supervised machine learning approach has been discussed for compensation of the drift produced in the sensors due to physical and chemical interaction of the analyte with the exposed sensor surface. The various categories of drift that ever arise and the drift of which the sensor community feels it to be a victim, both are briefly described. We have described the approaches experimented till now to deal with this problem of drift and how are approach is better over the recent reporting. The time complexity issues and various variations possible are all well discussed. For further extension of this idea, one can suggest improved regularizers to deal with this concept drift.

REFERENCES

[1] Leandro L. Minku, Allan P. White, Xin Yao, Impact of diversity on online ensemble learning in the presence of concept drift, IEEE Transaction of knowledge and data engineering,Vol 22, No.5, 2010
[2] Michifumi Yoshioka, Toru Fujinaka, SigeruOmatu, Intelligent Electronic Nose Systems with MetalOxide Gas Sensors for Fire Detection, IJAIS, Vol 2 No 1 2009
[3] Headspace-Analysis and its Use for Rapid Determination of volatile Organic Compounds in Food Quality Monitoring", *Sensors and ActuatorB*, Vol. 114, 2006.
[4] M. Kusuke, A.C. Romain, and J. Nicolas, "Microbial Volatile Organic Compounds as Indicators of Fungi. Can an Electronic Nose Defect Fungi in Odoor Environments? ", *International Journal of Building Scienceand its Applications*, Vol. 40, 1995.
[5] A. Norman, F. Stam, A. Morrissey, M. Hirschfelder, and D. Enderlein, "Packaging Effects of a Novel Explosion-Proof Gas Sensor", *Sensorsand Actuator B*, Vol. 95, 2003.
[6] R. C. Young, W. J. Buttner, B. R. Linnel, R. Ramesham, "Electronic Nose for Space Program Applications", *Sensors and Actuator B*, Vol. 93, 2003.
[7] N. Baric, M. Bucking, and M. Rapp, " A Novel Electronic Nose based on Minimized SAW sensor arrays coupled with SPME Enchanced Headspace-Analysis and its Use for Rapid Determination of volatile Organic Compounds in Food Quality Monitoring", *Sensors and Actuator B*, Vol. 114, pp. 482–488, 2006
[8] B. Charumporn, M. Yoshioka, T. Fujinaka, and S. Omatu, "An Electronic Nose System Using Back Propagation Neural Networks with a Centroid Training Data Set", *Proc. Eighth International Symposium on Artificial Life and Robotics, Japan*, 605–608, 2003.
[9] I. Zliobaite, "Learning under concept drift: an overview," *CoRR* , 2010.
[10] FabricioBreve, Liang Zhao , "Particle Competition and Cooperation in Networks for Semi-Supervised Learning with Concept Drift" , WCCI- IEEE , 2012
[11] A.Vergara, Shankar Vembu, Tuba Ayhanb, Margaret A. Ryanc, Margie L. Homerc, Ramon Huertaa "Chemical gas sensor drift compensation using classifier ensembles" . Sensors and Actuators B, 166-167 2012
[12] S. Di Carlo and M. Falasconi, "Drift correction method for Gas Chemical Sensors in artificial Olfaction Systems : Techniques and Challenges",
[13] Irene Rodriguez-Lujan, JordiFonollosa, Alexander Vergara, Margie Homer, Ramon Huerta. On the calibration of sensor arrays for pattern recognition using the minimal number of experiments.Chemometrics and Intelligent Laboratory Systems (2013) in press.
[14] A. Vergara, M.K. Muezzinoglu, N,Rulkov, R. Huerta, Information theoretic optimization of chemical sensors, Sensor and Actuators B , 148(1),2010
[15] Hal Daume III, A course in Machine Learning, Chapter Ensemble learning
[16] Gavin Brown, Encyclopaedia of Machine Learning 1, 2010
[17] RobiPolaker, Ensemble based systems in decision making
[18] Hyun-Chul Kim, Shaoning Pang, Hong-Mo Je, Daijin Kim, and Sung-Yang Bang, Support Vector Machine Ensemble with Bagging, LNCS 2388, 2002
[19] Hal Daume III, From zero to reproducing kernel hilbert spaces in twelve pages or less, 2004
[20] C.-C. Chang, C.-J. Lin, LIBSVM: A Library for Support Vector Machines,Software. Available at http://www.csie.ntu.edu.tw/cjlin/libsvm, 2001.
[21] Chih-Wei Hsu, Chih-Chung Chang, and Chih-Jen Lin, Department of Computer Science, National Taiwan University,Taipei 106, Taiwan, 2003, Practical Guide to Support Vector Classification.

BIOGRAPHIES

Ms.ShrutiAsmita (B.Tech., 2013 – KEC Ghaziabad, Uttar Pradesh Technical University, Lucknow) is a M.Tech. student at Department of Computer Science, Banasthali University, Jaipur and research intern at Department of Computer Science and Engineering, Indian Institute of Technology, Banaras Hindu University, Varanasi. Her research interests include machine learning, sensor networks, image processing and data mining.

Dr. K.K. Shukla (Ph. D., 1993 - Institute of Technology (BHU), Varanasi) is a professor at department of CSE, IIT (BHU), Varanasi. He completed his B.Tech in 1980 from APSU, Rewa, M.Tech. in 1982 from IT (BHU) and PhD in 1993 from IT (BHU). He is working as a professor at department of CSE-IIT, BHU He has 30 years of research and teaching experience, more than 120 research papers in reputed journals and conferences and more than 90 citations. Presently he has research collaboration with Space Applications Center, ISRO, Tata Consultancy Services, Institut National de Recherche en Informatique et en Automatique (INRIA), France and École de TechnologieSupérieure (ÉTS), Canada. He has written 4 books onNeuro-computers, RTS Scheduling, Fuzzy modelling, Image Compression. Current fields of Interest include Pattern Recognition, Graph Problems, Wireless Sensor Networks, Machine Learning.