



Application of Regularization to Ensemble of Classifiers for Drift Compensation in Electrical and Chemical Equipments

Shruti Asmita¹, K.K. Shukla²

¹Department of Computer Science, Banasthali University,
C Scheme, Sarojini Marg, Jaipur 302001, Rajasthan, INDIA

²Department of Computer Science and Engineering,
Indian Institute of Technology,
Banaras Hindu University, Varanasi 221005, Uttar Pradesh, INDIA

Abstract: Any electrical or chemical device undergoes a problem of concept drift with time due to various chemical, physical interactions of the environment elements with the exposed surface of the device and other factors such as aging and poisoning of the surface. This has been a major problem faced by such equipments when they are used as an experimental tool during any research and developments. In this paper, we introduce a hard machine learning approach to solve this problem by the application of regularization to the ensemble of classifiers for overcoming the time dependent drift occurrence. We have applied our regularized drift compensation algorithm on two real time data and one synthetic data. Our experiment finds best improvement on the application of Single Value Decomposition and Norm 2 combination regularization to the ensemble of classifiers. To the best of our knowledge, regularization has not yet been applied for drift correction in such electrical and chemical equipments.

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

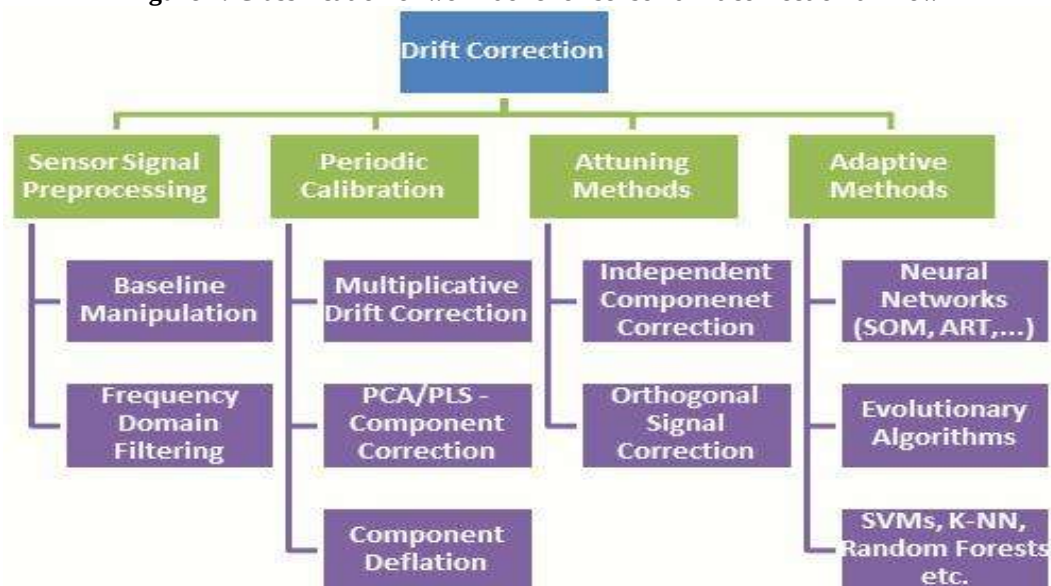
I. Introduction

Electrical and Chemical equipments are the frequently used devices in various experimental setups for research and development. Such devices produce output in form of electrical or optical signals. This output may be taken as input to other experiments or may be taken as final output depending on the scenario. There are various problems faced in such measurement like noise, drift, hysteresis, digitization errors, sampling errors, aliasing errors, calibration and precision errors etc. Such devices use, an electrical or chemical sensor for generating the reading. Two main steps are involved in any sensor detection is recognition and transduction. The experimental setting parameters produce a pattern which a sensor recognize and then generate signals accordingly to produce the detections.

Any electrical or chemical device's sensitivity indicates how much the device output changes when the input quantity being measured changes. When the output of signal changes slowly with time, independent of measured property, then this is the reported drift occurred in the equipments. Any long term drift indicates slow degradation of the equipments. Concept drift is a problem which can be dealt by learning the patterns of the data recorded from the start of equipment's life and compensating the degradation factor. Drift has become a great problem for research community, decreasing the performance of the recordings of data. There are two primary sources of drift. First order drift, also known as real drift arise due to interaction between detection agent and exposed equipment's surface. It produces effects like aging and poisoning. Second order drift occurs due to external and uncontrollable changes in operating environment conditions such as thermal effects, temperature and humidity variations etc. The problem of detecting concept drift is not a new one but is a challenging domain for finding improved solutions to for it [1]. Techniques to detect concept drift is classified as follows (Fig 1). 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 [1]. 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[1]. 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 [1]. 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 linearities 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 means that the information carried by each component cannot be inferred from the others [1]. Orthogonal Signal Correction (OSC) is another attuning method based on a signal processing technique [1,2]. Although attuning methods proved to be most promising in this stream but in case of chemicals the choice of celebrants is application specific. This leads to loss of generalization and standardization.

Figure 1: Classification of work done for sensor drift correction till now



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 [1]. They are data driven, self-adaptive and nonlinear methods which give flexible modeling 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. Kernelized version of component analysis such as kernelized PCA can handle nonlinear data well. These techniques however have not been investigated much.

In this paper, we introduce a new approach which solves the mentioned problem by regularizing the weights in weighted ensemble of classifiers in supervised learning models. Hence learning of model by regularized weight is done. To the best of our knowledge, such technique which deals with the problem of over fitting of the classifiers and their ensembles in the supervised hard machine learning has yet not been applied to the electrical and chemical equipment's drift correction. Over fitting arise once the generalization is decreased. The effect of over fitting 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. In the stretch of this paper, we first describe the data set used and the background terms. Further we describe the regularized drift compensation algorithm, followed by detailed description of our experimental findings. And finally we present the concluding comments drawn from the results.

II. The Dataset

We have used 3 different datasets for our experiments. First data set has been recorded at Italy Electricity Company by the power grids, which is an electrochemical device. Second data set has been recorded at Australian Electricity Company by the power measurement device, which is an electrical equipment planted at different regions of location under experiment. Third is a Streamline Ensemble data having an artificial

introduced drift. First two data are the real time data while the third is the artificial data. The detailed description of the data set is as follows.

A. Power Supply Stream data set

It contains hourly power supply of an Italy electricity company which records the power from two sources: power supply from main grid and power transformed from other grids [3]. This stream contains three year power supply records from 1995 to 1998, and our learning task is to predict which hour (1 out of 24 hours) the current power supply belongs to. The concept drifting in this stream is mainly driven by the issues such as the season, weather, hours of a day (e.g., morning and evening), and the differences between working days and weekend. It is characterized by 2 attributes and 24 class labels. There are a total of 29928 instances. We have divided the whole collection into 10 equal Batches for experimental convenience and demand. Each batch of data contains recorded value of approximately 3.5 months in time series sequence. Being time dependent, the data for later batches show more drift than initial batches.

B. Electricity Pricing data set

This is a streaming classification data [20,21,22]. It holds information for the Australian New South Wales (NSW) Electricity Market, containing 27552 records dated from May 1996 to December 1998, each referring to a period of 30 minutes subsampled as the completely observed portion of nearly 45k total records with missing values. The records have been divided into 13 batches each containing nearly 20k- 30k records in sequence of time. These batches are made for experimental convenience. The reformulated data contains five numeric attributes capturing aspects of electricity demand and a class label. Electricity prices are affected by the demand and supply ratio depending on the season change, wind pattern etc attributes. Price directly depends on the consumption. The electrical device used for reporting the consumption starts getting affected with time due to concept drift.

C. SEA data set

Streaming ensemble algorithm (SEA) data set is an artificial data set contains abrupt concept drift, first introduced in [23]. It is generated using three attributes, where only the two first attributes are relevant. All three attributes have values between 0 and 10. The points of the dataset are divided into 4 blocks with different concepts. We have further divided these four block into 12 batches in sequence of time. In each block, the classification is done using $f1+f2 \leq \theta$, where $f1$ and $f2$ represent the first two attributes and θ is a threshold value. The most frequent values are 8, 9, 7 and 9.5 for the data blocks. 10% of class noise was then introduced into each block of data by randomly changing the class value of 10% of instances. The variation of threshold gives the same effect of concept drift as it arises with time in equipments.

III. BACKGROUND TERMS

Teaching machines that work in ones and zeros to reach their own conclusions about the world. i.e. machine Learning is about learning to do better in future based on what was experienced in the past. Tom Mitchell definition of Machine learning [6] "For task T, a computer program learns from experience E and performs P, if its performance P improves with experience E." For instance if we want to make a machine learn to play chess game, T is game playing, E is playing 100 or 1000 times and P is win against a new opponent. One kind of learning is the ability to classify, to differentiate between two or more groups of available instances from environment. We have hard supervised classification technique for dealing with large number of features i.e. Support Vector Machines (SVM). An Ensemble model for Support Vector Machine gives improved classification. Moreover if an efficient Regularizer is applied to the ensemble of classifiers, it promotes the generalization and further increases classification accuracy. Followed is a brief discussion of these concepts.

A. Ensemble of Classifiers

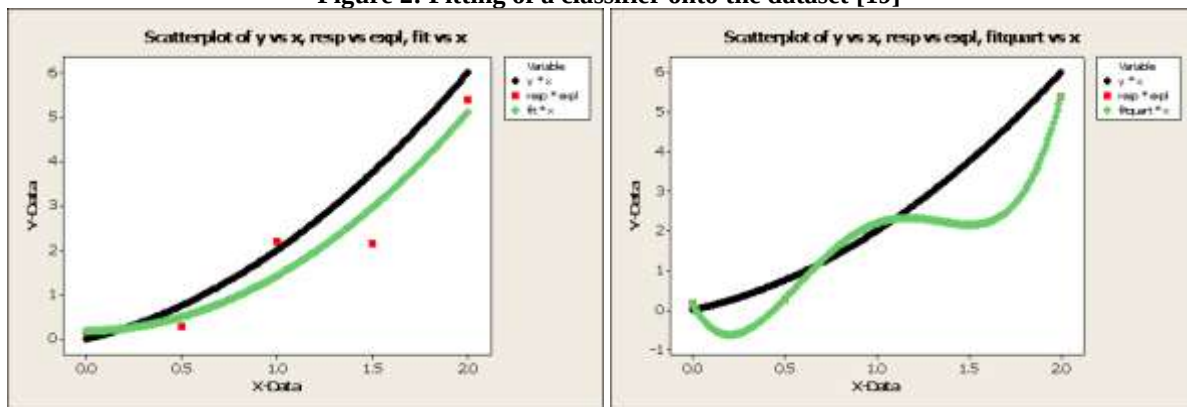
Kernel methods are a class of algorithm for pattern analysis. It is hard to classify data in the lower dimensions [7]. The boundary line between two sets of data points becomes increasingly complex as we increase the number of classes and data points. Hence we need a way to transform this data into higher dimensions where we can draw simple hyper planes to separate multiple classes. If the boundary between two sets of points is too crooked, the algorithm will take a lot of time to converge. Even when it converges, it might not be the most optimal boundary. Kernel Functions applies a distinct approach to this problem. Instead of taking the hard route of classifying the data in lower dimension by putting a really curvy line, Kernel Functions map the data into higher dimensional spaces in the hope that the data is more easily separated there. Support vector machine (SVM) is an application of kernel functions. Each instance in the training set contains one target value (i.e. the class labels) and several attributes (i.e. the features or observed variables). The goal of Support Vector Machine (SVM) is to produce a model (based on the training data) which predicts the target values of the test data given only the test data attributes.

SVM was initially proposed to provide good generalization performance, but the classification results of practically implemented SVM is often far from theoretical expected level. SVM primarily has two drawbacks which are also the reasons of this difference stated above. First, since it is originally a model for the binary-class classification, we should use a combination of SVMs for the multi-class classification. (This is not giving parallel performance to binary SVM). Second, since learning of the SVM is a very time consuming for a large scale of data, we should use some approximate algorithms. Using the approximate algorithms can reduce the computation time, but degrade the classification performance. SVM ensembles improve upon these drawbacks.

B. Regularization

Regularization is a parameter added to objective function which is required to be optimized to prevent problems like over-fitting and thereby improving Generalization. If we are given 'n' distinct 'x' values and corresponding 'y' values for each, it is possible to find a curve going exactly through all n resulting points (x,y). This can be done by setting up a system of equations and solving simultaneously. We can also draw Conditional Mean line which considers average of only those data points that fall into our scenario chosen or taken. However drawing Regression under responses, they are not expected to go through all the data points. It aims at establishing a relationship among all the data points and gives best fit analysis.

Figure 2: Fitting of a classifier onto the dataset [19]



In Fig 2 (left) Black solid Line shows Conditional mean curve of all Data Points. $y=x^2$ and Green Solid Line shows Regression Line (also a second degree regression curve: $y= \alpha + \beta x + \gamma x^2$) [19]. Both lines are not same but nearly close to each other. We cannot have conditional mean for every Real Scenario. But we try achieving Conditional mean through Regression curve. Suppose taken fourth degree Regression Curve: $y= \alpha + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \beta_4 x^4$. Regression curve fits all the data (Fig 3). But it is not a good regression curve because we didn't achieved conditional mean curve. This is over-fitting. This occurs even in third degree regression curve because our conditional mean is a second degree curve. Hence over fitting occurs when generalization is decreased. To prevent this over-fitting we can either reduce number of features under consideration or secondly we can keep all the features but reduce value of each feature parameter. This second solution is dealt under Regularization.

C. Concept drift

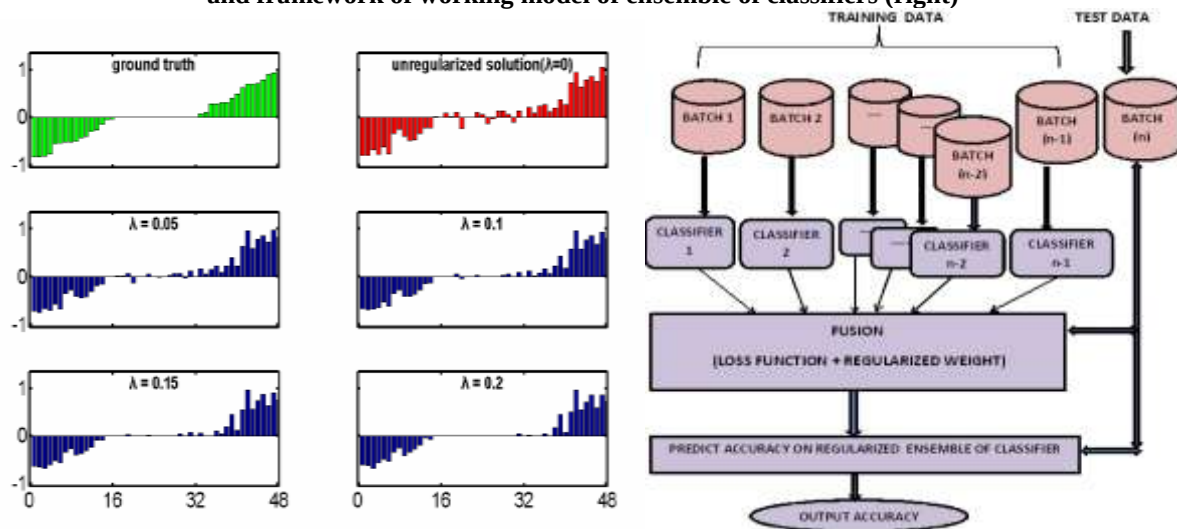
In predictive analytics and machine learning, the concept drift means that the statistical properties of the target variable, which the model is trying to predict, change over time in unforeseen ways. This causes problems because the predictions become less accurate as time passes. [15,16,17]. The term concept refers to the quantity to be predicted. To prevent deterioration in prediction accuracy because of concept drift, both active and passive solutions can be adopted. Active solutions rely on triggering mechanisms, e.g., change-detection tests to explicitly detect concept drift as a change in the statistics of the data-generating process. In stationary conditions, any fresh information made available can be integrated to improve the model. Differently, when concept drift is detected, the current model is no more up-to-date and must be substituted with a new one to maintain the prediction accuracy. On the contrary, in passive solutions the model is continuously updated, e.g., by retraining the model on the most recently observed samples, or enforcing an ensemble of classifiers.

IV. REGULARIZERS UNDER STUDY

Regularization is a measure of handling the problem of over fitting. Fig 3 shows the effect of regularization. λ used in the graph is the regularization parameter. The first top left graph shows the ground truth value. If over fitting is not dealt with in the classification, graph generated is unregularized (top right). Moving to other graphs, we can see that as we move from $\lambda=0.05$ to $\lambda=0.2$, the smoothness of the output has been improved.

This shows the effect of regularization on optimization of loss function. There are many regularization techniques in existence and this is also a topic under further research. L_1 Regularization is norm 1 regularization factor which penalizes all the factors equally. It can be viewed as the selection of only the relevant factors and is defined as $\lambda \|w\|_1$. This has best usage in signal processing, compressed sensing, wavelet thresholding, geophysics problem, decoding linear codes etc. This is slow for large scale problems. Regularization path is $(0, \infty)$. L_2 Regularization is norm 2 regularization factor defined as $\lambda (\|w\|_2)^2$. It 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 L_1 norm regularization. Tikhonov regularization is a special case of L_2 Regularization represented by term $\lambda^2 (\|w\|_2)^2$. Finally, In linear algebra, the singular value decomposition (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 [14]. This SVD in combination with norm 2 regularization is represented as $\lambda_1 \cdot \text{SVD}(w) + \lambda_2 \cdot (\|w\|_2)^2$. Regularization path varies with the experimental conditions.

Figure 3: Effect of application of regularization to loss function in optimization of a problem [18] (left) and framework of working model of ensemble of classifiers (right)



V. REGULARIZED DRIFT COMPENSATION ALGORITHM

Regularized Ensemble of Classifier has been used [8,10,11] to cope with concept 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. $S_t = \{(x_1, y_1), (x_2, y_2), \dots, (x_{mt}, y_{mt})\}$. For training and optimization of our problem, we have used a popular library libSVM [12,13]. 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 majority voting technique. There are three versions of majority voting. Unanimous voting i.e. on which all classifiers agree, simple voting i.e. predicted by at least one more than half the number of classifiers and plurality voting i.e. the highest number of votes, whether or not the sum of those votes exceeds 50%. This is most optimal form of majority voting. If we have evidence that certain experts are more qualified than others, weighting the decisions of those qualified experts more heavily may further improve the overall performance than that can be obtained by the plurality voting. In our experiments we have used majority voting fusion technique as also shown in framework (Fig 3)(right). At this stage, the weight matrix obtained acts as feature vector for the ensemble of classifiers and hence this is to be regularized to overcome problem of over fitting. The loss function which has to be optimized is the hinge loss $L(f(x), y) = \max(0, 1 - y \cdot f(x))$. The regularization factors, application of which to the loss function, generates the best accuracy in our case, is 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 . Hence the resulting objective function is:

$$\text{argmin}_{\beta_1, \dots, \beta_t} \sum_{i=1:mt} \sum_{j=1:t} \max(0, 1 - \beta_i \cdot y_i \cdot f(x_i)) \quad (1)$$

$$\text{where } h_{t+1}(x) = \text{argmax}_{y \in \{1, \dots, L\}} \sum_{t: f_t(x) = y} \beta_t + \lambda_1 \cdot \text{SVD}(\beta) + \lambda_2 \cdot (\|\beta\|_2)^2 \quad (2)$$

In the algorithm below 'T' is the total number of batches in data set. 'k' batches of data forms training data. 't' batch of data forms the testing batch. 'C' is the Support vector machine parameter and 'y' is the kernel bandwidth parameter. Model[] is an struct array for storing each batch classifier.

1. for $t = 2 \dots T$ do
2. for $k = 1 \dots (t-1)$
3. Load ($data_k, label_k$)
4. Scale $data_k$ matrix in range $[-1, 1]$
5. $['C', 'y'] =$ output generated by grid search on ($data_k, label_k$)
6. $Model_t = \text{Train}(data_k, label_k)$ for respective ($'C', 'y'$) to form classifier
7. end for
8. Perform weighted ensemble of the $Model[t-1]$
9. Apply regularization to the weighted in ensemble of classifier obtained in step 7 (ref equation 3).
10. Load ($data_t, label_t$)
11. Scale of $data_t$ matrix in range $[-1, 1]$
12. Predict and aggregate classification accuracy of ensemble model with ($data_t, label_t$)
13. end for
14. Output aggregated prediction accuracy generated at step 11.

VI. EXPERIMENTAL RESULTS

We trained multi class SVMs (one Vs one strategy) with RBF kernel using LibSVM software [12,13]. 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 $[2^{-10}, 2^{-9}, \dots, 2^5]$ and $[2^{-5}, 2^{-4}, \dots, 2^{10}]$, respectively. The regularization path for the regularization parameters is $[0, 1]$. The condition $0 < \lambda_1 < 1$, $0 < \lambda_2 < 1$ has to be true while choosing values of λ_1 and λ_2 . The electrical and chemical equipments used in research, medical and other development setups shows drift in reading taken with time. This causes degradation in performance of the classifiers to assign correct class to output. Our goal is to cope with the concept drift and give maximum possible classification accuracy even in the presence of drift. For processing purposes, all the data sets is organized into certain number of batches [Batch 1, Batch 2, Batch 3..... Batch T]. For the analysis, we are considering following 4 cases in comparison to the base case i.e. simple ensemble of classifier case:

Case 1 : SVD and Norm 2 combination regularization is applied on Ensemble of classifier.

Case 2 : Norm 1 regularization is applied on Ensemble of classifier

Case 3 : Norm 2 regularization is applied on Ensemble of classifier

Case 4: Tikhonov regularization is applied on Ensemble of classifier

The Simple Ensemble of classifier does not take into account any regularization technique. It is a simple weighted ensemble model. SVD and norm 2 combination regularization is the one, our proposed algorithm takes into account. We provide experiments with norm 1 regularization, norm 2 regularization, tikhonov regularization along with SVD and norm 2 combination regularization, to give a comparative analysis. Training and optimization of loss is done by libraries of libSVM[12,13].

A. Results on Power Supply data set

Graphs are placed in order of the cases described above. Even the improved classification lies below 30% accuracy because the data collected is just over 3 years and hence has not shown much larger drift has occur.

Figure 4 Comparison of SVD and norm 2 combination regularization with the simple ensemble of classifiers on power supply stream data set (left) and Comparison of norm 1 regularization with the simple ensemble of classifiers on power supply stream data set (right).

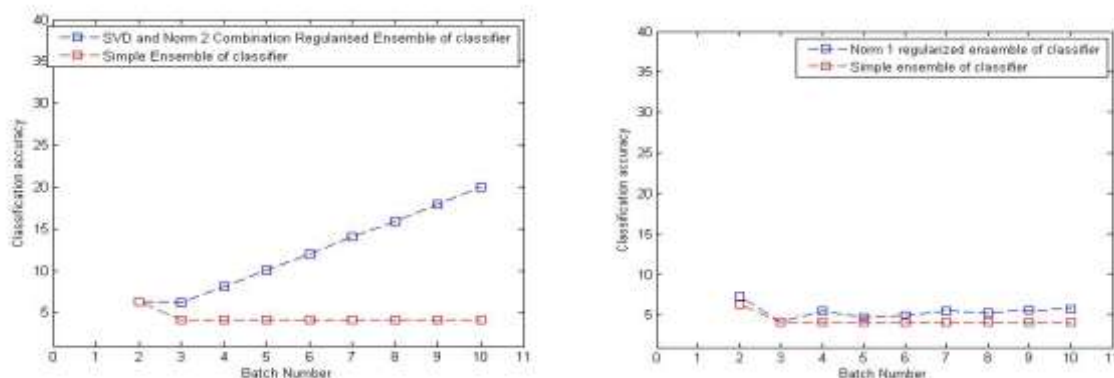


Figure 5 Comparison of norm 2 regularization with the simple ensemble of classifiers on power supply stream data set(left) and Comparison of tikhonov regularization with the simple ensemble of classifiers on power supply stream data set(right).

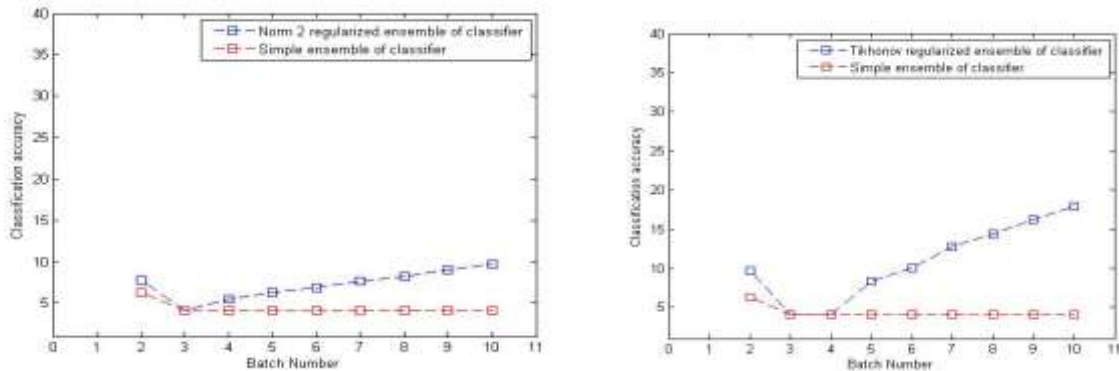


Table I Percentage increase in classification accuracy on application of regularization to weighted ensemble of classifiers in power supply data set

S. No.	Regularization Technique Applied with loss function of ensemble of classifier	Maximum Percentage increase in classification accuracy (%) after regularization
1.	Norm 1 Regularization	3
2.	Norm 2 Regularization	5.56
3.	Tikhonov Regularization	13.74
4.	SVD and norm 2 combination regularization	15.80

B. Results on Electricity Pricing data set

Case 1 regularization shows maximum classification accuracy in range [75-98]% The concept drift arise with time in the recordings. Hence the improvement in classification accuracy by drift compensation is also seen better in later batches of data than in starting batch of data.

Figure 6 Comparison of SVD and norm 2 combination regularization with the simple ensemble of classifiers on electricity pricing data set(left) and Comparison of norm 1 regularization with the simple ensemble of classifiers on electricity pricing data set(right).

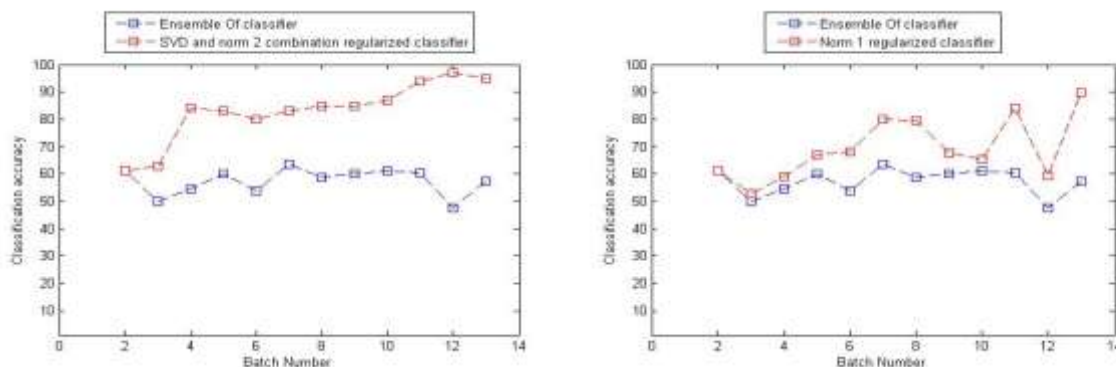


Figure 7 Comparison of norm 2 regularization with the simple ensemble of classifiers on electricity pricing data set(left) and Comparison of tikhonov regularization with the simple ensemble of classifiers on electricity pricing data set (right).

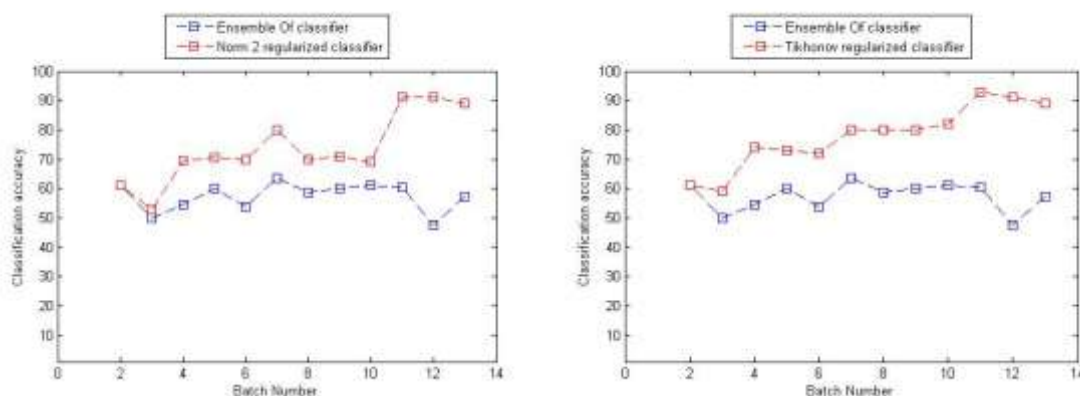


Table II Percentage increase in classification accuracy on application of regularization to weighted ensemble of classifiers in electricity pricing data set

S. No.	Regularization Technique Applied with loss function of ensemble of classifier	Maximum Percentage increase in classification accuracy (%) after regularization
1.	Norm 1 Regularization	30.43
2.	Norm 2 Regularization	31.47
3.	Tikhonov Regularization	44.01
4.	SVD and norm 2 combination regularization	46.98

C. Results on SEA data set

The graphical results have been placed in the order of the cases mentioned above. The case 1 regularization of ensemble of classifier gives the maximum classification accuracy in range [85-99] %.

Figure 8. Comparison of SVD and norm 2 combination regularization with the simple ensemble of classifiers on SEA data set(left) and Comparison of Norm 1 regularization with the simple ensemble of classifiers on SEA data set (right).

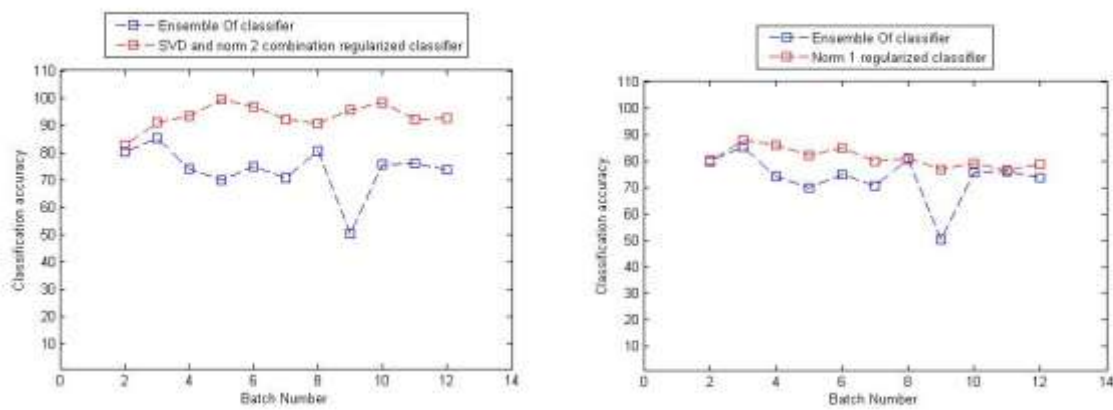


Figure 9 Comparison of Norm 2 regularization with the simple ensemble of classifiers on SEA data set (left) and Comparison of tikhonov regularization with the simple ensemble of classifiers on SEA data set (right).

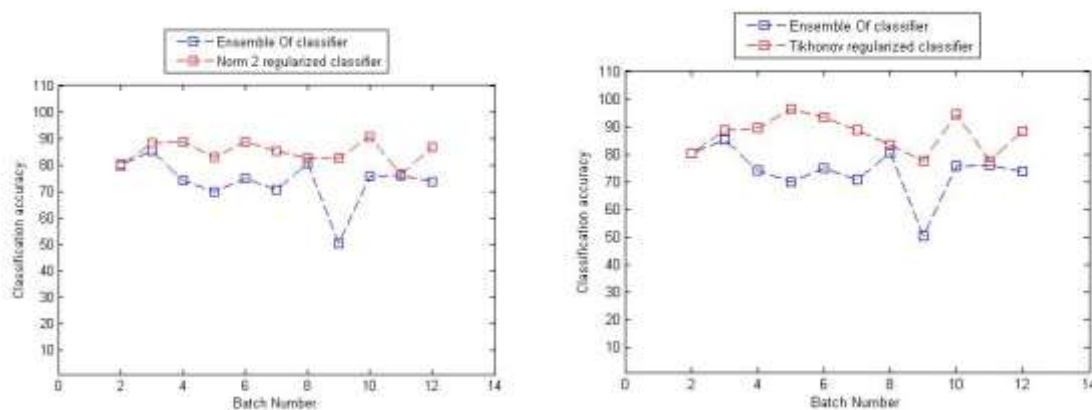


Table III Percentage increase in classification accuracy on application of regularization to weighted ensemble of classifiers in SEA data set

S. No.	Regularization Technique Applied with loss function of ensemble of classifier	Maximum Percentage increase in classification accuracy (%) after regularization
1.	Norm 1 Regularization	12.54
2.	Norm 2 Regularization	14.52
3.	Tikhonov Regularization	18.68
4.	SVD and norm 2 combination regularization	30.42

Table III Overall Time Complexity Analysis of Regularizers under study

S.No.	Regularization Techniques	Complexity Analysis
1.	Norm 1 Regularization	There are total of $(t-1)$ sum operations computed at run of algorithm. Time Complexity $O(t)$ is reported..
2.	Norm 2 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.
3.	Tikhonov Regularization	There are total of $(t-1)$ sum operations, t number of multiplication operations to perform square, and 1 square root operation is computed. Time complexity $O(3t)$ is reported. 2 degree regularization parameter is applied.
4.	SVD and norm 2 combination regularization	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.

VII. APPLICATIONS

The concept drift arise in sensors with time due to physical and chemical reactions of sensor surface with the environment and sensing agent. This drift is a major problem arising, wherever there is continuous sensing real or experimental setup having sensors involved in infrastructure. Here in this paper we have applied drift correction for compensating drift in power supply stream sensor. We can also use the same algorithm for dealing with drift in chemical sensors such as electrode sensors, thin film sensors etc which have degrading on their surface due to reaction with chemical analyte and other environment conditions. Another place where such sensors are in use is the pharmacy. Sensors there are used for continuous monitoring of drug reagents and their reactions on beings etc. Other places are weather data recordings, satellite data recording and fetching, forensic science labs etc.

VIII. CONCLUSION

Our study has mainly focused on the improvement of the classification accuracy by application of regularization techniques to weights in weighted ensemble of classifiers so as to reduce the complexity of fusion of the classifiers. Over the existing regularization techniques, SVD and norm 2 combination regularization gives the best result. In future we can further apply some new regularization techniques to improve present results. In our experiments, the fusion of data is done at the prediction level. However we can also apply other fusion techniques for aggregation of ensemble of classifiers if it gives any better results than current.

IX. REFERENCES

- [1] S. Di Carlo and M. Falasconi, Drift correction method for Gas Chemical Sensors in artificial Olfaction Systems : Techniques and Challenges
- [2] Ali Azizi, Amir Azizi, Orthogonal Signal correction partial least square method method for simultaneous spectrometric determination of Nickle, Cobolt and Zinc.
- [3] X. Zhu, Stream Data Mining Repository, . Power supply stream data set. <http://www.cse.fau.edu/~xqzhu/stream.html>, 2010
- [4] Alexander Vergara and Shankar Vembu and Tuba Ayhan and Margaret A. Ryan and Margie L. Homer and Ramon Huerta, Chemical gas sensor drift compensation using classifier ensembles.
- [5] Irene Rodriguez-Lujan, Jordi Fonollosa, 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
- [6] Jaime G Carbonell, Ryszard S Michalski, Tom M. Mitchell , An overview of Machine Learning
- [7] Ethem Alpaydin, second edition, MIT press, Introduction to machine learning
- [8] H. Wang, W. Fan, P.S. Yu, J. Han, Mining concept-drifting data streams using ensemble classifiers.
- [9] J.Z. Kolter, M.A. Maloof, Dynamic weighted majority: A new ensemble method for tracking concept drift.
- [10] J.Z. Kolter, M.A. Maloof, Using additive expert ensembles to cope with concept drift.
- [11] J.Z. Kolter, M.A. Maloof, Dynamic weighted majority: An ensemble method for drifting concepts.
- [12] C.-C. Chang, C.-J. Lin, LIBSVM: A Library for Support Vector Machines, Software. Available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>,
- [13] Chih-Wei Hsu, Chih-Chung Chang, and Chih-Jen Lin, Department of Computer Science, National Taiwan University. Practical Guide to Support Vector Classification.
- [14] A.M.A. Hamdan , An investigation of the significance of singular value decomposition in power system dynamics.
- [15] Zliobaite, I., Learning under Concept Drift: an Overview. Faculty of Mathematics and Informatics, Vilnius University: Vilnius, Lithuania.
- [16] Jiang, J., A Literature Survey on Domain Adaptation of Statistical Classifiers.
- [17] Kuncheva L.I. Classifier ensembles for detecting concept change in streaming data: Overview and perspectives.
- [18] Y Albert Park CSE 291 4. L1 Norm regularization.
- [19] <http://www.ma.utexas.edu/users/mks/statmistakes/overfitting.html>
- [20] M. Harries. Splice-2 Comparative Evaluation: Electricity Pricing”. University of New South Wales, School of Computer Science and Engineering
- [21] M. Baena-Garcia, J. del Campo-Avila, R., Fidalgo, A. Bifet, R. Gavalda and R. Morales-Bueno. Early drift detection method
- [22] L.I. Kuncheva C.O. and Plumpton, Adaptive Learning Rate for Online Linear Discriminant Classifiers.
- [23] Albert Bieft and Ricard Gavalda, Learning from time changing data with adaptive windowing,”