ISSN: 0976-3031

Available Online at http://www.recentscientific.com

CODEN: IJRSFP (USA)

International Journal of Recent Scientific Research
Vol. 9, Issue, 2(L), pp. 24612-24617, February, 2018

International Journal of Recent Scientific Research

DOI: 10.24327/IJRSR

Research Article

A REPRESENTATIVE STANDARDIZED SAMPLE SET SELECTION FOR IMPROVING STUDENT'S PERFORMANCE PREDICTION

Sasi Regha, R[1]* and Uma Rani, R[2]

[1]Department of Computer Science, SSM College of Arts & Science, PinCode-638 183, Tamilnadu, India
[2]Department of Computer Science, Sri Sarada College for Women, PinCode-638 016, Tamilnadu, India

DOI: http://dx.doi.org/10.24327/ijrsr.2018.0902.1693

ARTICLE INFO

Article History:

Received 15th November, 2017
Received in revised form 25th December, 2017
Accepted 28th January, 2018
Published online 28th February, 2018

Key Words:

NMFC, modified Computer Aided Design of Experiments, Divergence, Modified Principal Component Analysis

ABSTRACT

A hybrid Artificial Fish Swarm-Cuckoo Search (AFSCS) and Non-negative Matrix Factorization Clustering (NMFC) was proposed for selecting optimal relevant features and removing redundant features in student academic dataset. However outliers and redundant data samples in the dataset are affecting the efficiency and effectiveness of classifiers. In this paper, initially, the most class specific representative samples are selected using Computer Aided Design of Experiments (CADEX) algorithm. The CADEX algorithm selects the sample that is closest to the mean and the next sample selection will be the one most distant from the already selected sample by using Euclidean distance. Principal Component Analysis (PCA) is used in the CADEX for selecting optimal components. Euclidean distance in CADEX only select right samples when all the attributes have the similar units. So, the Modified CADEX (MCADEX) is next proposed by using Kullback-Leibler divergence. In this approach, for each class mean value is calculated then divergence between mean and data samples are found with multiple reference data samples. The highly divergence data samples are selected for each class. In the MCADEX algorithm, MPCA is used instead of PCA because some attributes in dataset might be in orders of magnitude of others, this may lead to create highest variance while eigen values calculation. In MPCA, the eigenvectors of the covariance matrix is derived from various similarity measurements like Mutual information, angle information and hybrid Gaussian and polynomial kernel. The sample selected datasets are used for predicting student performance using Prism and J48 classifiers. The experimental results show that the proposed sample selection approaches are improving accuracy of classifiers.

INTRODUCTION

A representative sample is a group or set selected from a larger dataset. The representative samples have exact information of larger dataset and selected using some statistical and relational based calculations. It is a small amount of something which precisely reflects the larger entity. A selection process which selects samples more or less equally distributed over the calibration space will lead to a flat distribution. A distribution is more favorable from a regression point of view than the normal distribution of an equal number of samples, thus that the loss of predictive quality may be less than expected when looking only at the reduction of the number of samples (Hildrum, 1992). Several techniques are available for selecting representative samples for experimental design and regression validation (Ferré and Rius, 1996; Ferré and Rius, 1997).

A new hybrid optimization technique (Sasi regha and Uma rani, 2017) was proposed to select optimal relevant features for improving classification accuracy. AFSCS optimization was used for feature selection. The removal of irrelevant features leaves more relevant features in the dataset. The redundant feature within the relevant features was eliminated using NMFC. Prism and J48 classification was used to classify the student's performance. But, outliers and redundant samples available in the dataset are affecting the efficiency and effectiveness of classifiers. So in this paper, sample selection is proposed in efficient manner.

In this paper, by using CADEX algorithm, the most class specific representative samples are selected from the closet neighbours of mean value. The larger distance value from selected sample is selected by using Euclidean Distance. PCA is used in CADEX for finding principal components of selected samples. The Euclidean distance is not properly separate the samples when attributes in the sample are from different category. So, Kullback-Leibler divergence is used in CADEX known as modified CADEX(MCADEX) to improve the sample

*Corresponding author: **Sasi Regha, R**
Department of Computer Science, SSM College of Arts & Science, PinCode-638 183, Tamilnadu, India

selection .In MCADEX , MPCA is used with three similarity measurements (angle information and hybrid Gaussian kernel, mutual information and polynomial kernel similarity) for deriving covariance matrix. The three similarity measures and divergence used for sample selection only select most relevant samples for each class which reduces the execution time of classifier while maintaining high accuracy.

The remainder of the article is organized as follows: Section 2 explains about the existing sample selection techniques. Section 3 describes about the proposed methods. Section 4 demonstrates the overall performance valuation of the proposed techniques. Section 5 concludes the article work.

Related Work

The virtual sample selection using Gaussian distribution was proposed to predict heating energy consumption (Yuan *et al.* 2018). The similar days were found using grey correlation and entropy weight method. The selected sample set from similar days and virtual samples improved the prediction accuracies of back propagation neural network (BPNN) and multiple linear regression (MLR) models. However, the unavailability of practical dataset limits the effectiveness of sample selection.
The well known sample selection method Normal factor analysis was extended (Kim, 2018) to select samples from data set. This method was effectively select samples from dataset which contain more number of outliers. A Bayesian hierarchical model was used to estimate the samples to select. Markov chain Monte Carlo methods and bias corrections were also used to improve sample selection. The limitation of this scheme was selecting single sample for each Bayesian estimation.

The feature and sample selections were simultaneously proposed (Adeli *et al.* 2016) for Parkinson's disease (PD) classification. This approach removed irrelevant features and samples. This was a wrapper approach where classification and removing irrelevant feature and samples were iteratively processed until obtaining highest accuracy. The stopping criteria of iteration was not clearly discussed in this method.
A regression based multivariate sample selection method was proposed (Kim and Kim, 2016) by extending Heckman model. Heckman model was a univariate sample selection model. The equation of univariate selection was modified to select multivariate sample. The modified Monte Carlo approach was introduced to estimate the efficiency of multivariate sample selection. In this approach, the log-likelihood function was replaced by expectation/conditional maximization either function.

A novel approach (Lafférs and Nedela Jr, 2017) was proposed for a sensitivity analysis of the bounds of the average treatment effects (ATE) with sample selection. This approach was discovering assumptions for sharp bounds in the ATE. By using relaxation parameters, the departure from the exogeneity assumption was managed that was easy to interpret. The bounds were computed based on relaxed assumptions because an optimization issue.

A novel samples selection technique by using system identification (Li *et al.* 2018) was proposed. System identification was the selection of a model for a process based on a limited number of measurements of the input and outputs. This technique was selected the samples by using system identification. The selected training samples were contained the similar covariance matrices and the cluster model of the cell under test (CUT). The clutter model of the CUT was discovered by the neural network.

A joint two copula functions-sample selection technique (Sriboonchitta *et al.* 2017) was presented. The copula functions were utilized to model the dependence among the errors of the sample selection and the dependence of the error terms of the stochastic frontier equation. By using Akaike information criterion (AIC), the optimal model was selected.

A transfer learning scheme using sample selection (Duh and Fujino, 2012) was proposed for ranking. In this approach, small and large datasets were used for target domain of interest and source domain, respectively. While the functional relationship among features and labels, source domain training samples were selected. This approach was selected the labeled source domain samples which were related to the target domain. A conventional ranker on the joined data was trained. The relatedness measure was calculated by using Kullback–Liebler Importance Estimation Procedure algorithm for density ratio estimation. But, the proposed approach was needed linearity due to the similarity between rankers was computed by using a Gaussian kernel.

MATERIALS AND METHODS

In this section, a modified CADEX (MCADEX) algorithm and Modified PCA (MPCA) is described briefly. In MPCA, three similarity measurements are utilized for deriving the eigenvectors of the covariance matrix to support different unit measure data types in attributes.

Modified CADEX algorithm

KL divergence is measuring the difference among two probability distributions over the similar variable x. Here, divergence is measured by the multiple references among the mean value and each data sample. The mean value is calculated by using each class. This divergence is related to relative entropy, information divergence and information for discrimination, is a non-symmetric calculate of the difference among multiple references points $r_i(x), i = 1,2, \dots, n$. Initially, the mean value $M(x_c)$ computed as follows,

$$M(x_c) = \frac{1}{N}\sum_{j=1}^{L} p(x_j) \tag{1}$$

D_{KL} is computed as follows,

$$D_{KL} = \sum_{i=1}^{n} \frac{p(x)}{r_i(x)} - \sum \frac{r_i(x)}{M(x_c)} \tag{2}$$

Then it also computed as follows,

$$D_{KL} = \int_{-\infty}^{\infty} \sum_{i=1}^{n} \frac{p(x)}{r_i(x)} - \sum \frac{r_i(x)}{M(x_c)} dx \tag{3}$$

It is computes the distance among multiple reference points, it is not a distance measure.

Assume k is the optimal components selected in the modified PCA model, n referred as the number of samples and T symbolizes the score matrix of dimension $n\ x\ k$. In addition, the object that is nearest to the mean of the $T_{n\ x\ k}$ is assumed as the most representative of this input data set, it is integrated as the first point in the calibration set C of size m and symbolized as S_1. After that, between the remaining samples, the second object for the training set can be the one situated farthest away from S_1. The third object chosen is the one that is farthest away from both S_1 *and* S_2, etc.

Let $S_1, S_2, \dots, S_w (w < m)$ be w samples which have been assigned to C. The next object S_{w+1} added to this set is the object from the remaining $(m - w)$ samples which is farthest away from the samples already added to C by the equation as follows,

$$\Delta_{w+1}^2 = \max_{i\neq j}\{\Delta_i^2(w)\},\ for\ i = w+1,\dots,m \tag{4}$$

In equation (4),

$$\Delta_i^2(w) = \min_r\{D_{KL1}^2, D_{KL2}^2, \dots, D_{KLi}^2\}\ i \neq r \tag{5}$$

Modified PCA approach

This approach is presented three subspace models that correspond to three similarity measurements, in that order. In the figure 1, the matrices C and D have the similar nonzero eigenvalues (NEVs) among the matrices D, C, S, Ws. After that, from those of the matrix D, the eigenvectors (EVs) related to the NEVs of the C matrix are derived. A similarity matrix is known as S and a case of S is called as the D matrix. The relationship among the matrix S and Ws symbolizes the similar as which among the matrices C and D.

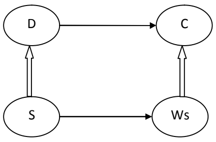

Figure 1 The matrices C, D, S and Ws

In Figure 1, this paper is to measure the EVs similar to the NEVs of the covariance C which is a huge scale once the samples are high dimensional, initially, the EVs is calculated related to the NEVs of the correlation D. After that, the EVs similar to the NEVs of the matrix C are recognized. The matrix C is viewed as a particular example of the matrices Ws which are achieved through the eigenvalues and EVs of the matrix S. Assume $X = [x_1, x_2, \dots, x_N], x_i \in R^M (M > N)$ indicates the centered training set. The NEVs and their EVs of the similarity matrix S of size $N \times N$ are λ_i *and* $\alpha_i (i = 1,2,\dots,r)$, respectively. Consider $\beta_i = X\alpha_i/\sqrt{\lambda_i}, (i = 1,2,\dots,r)$. These vectors are linearly independent, after that, there exists one or more matrices Ws of size $M \times M$ which is contain EVs β_i corresponding to $\lambda_i (i = 1,2,\dots,r)$ and the Ws are established using the data set X.

Given an arbitrary matrix W of size $M \times M$, if its NEVs are $\lambda_i\ (i = 1,2,\dots,r)$ and its EVs in proportion to these NEVs are $\beta_i (i = 1,2,\dots,r)$, next the following equation systems is established,

$$W\beta_i = \lambda_i \beta_i \qquad (i = 1,2,\dots,r) \tag{6}$$

$$W = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1M} \\ w_{21} & w_{22} & \cdots & w_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ w_{M1} & w_{M2} & \dots & w_{MM} \end{bmatrix}, \beta_i = \begin{pmatrix} b_{i1} \\ b_{i2} \\ \vdots \\ b_{iM} \end{pmatrix} \tag{7}$$

After that, M equation systems are obtained and the $jth\ (j = 1,2,\dots,M)$ is given below,

$$BW_j = Y_j, \quad (j = 1,2,\dots,M) \tag{8}$$

In equation (6), $Y_j = (\lambda_1 b_{1j}, \lambda_2 b_{2j}, \dots, \lambda_r b_{rj})^T$.

$$B = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_r \end{pmatrix} = \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1M} \\ b_{21} & b_{22} & \cdots & b_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ b_{r1} & b_{r2} & \dots & b_{rM} \end{pmatrix} \tag{9}$$

$$W_j = \begin{pmatrix} w_{j1} \\ w_{j2} \\ \vdots \\ w_{jM} \end{pmatrix} \tag{10}$$

In this assumption, $\beta_i \in R^M (i = 1,2,\dots,r)$ are linearly independent and $rank(B) = rank(BY_j) < N < M$. This assurances which the linear system (6) is infinitely numerous solutions. In equation (4) and (6), the Ws are determined based on λ_i *and* $\beta_i (i = 1,2,\dots,r)$ and $\beta_i = X\alpha_i/\sqrt{\lambda_i}$. Moreover, by using X, α_i is achieved. Therefore, Based on X, the Ws are determined.

Similarity Subspace Models in MPCA

Similarity subspace model_1 using mutual information

Mutual information and Entropy are essential concepts in information theory. Assume the discrete situation.

Entropy: A discrete random variable x, the entropy $H(x)$ is,

$$H(x) = -\sum_{x\in\chi} p(x) log p(x) \tag{11}$$

In equation (11), P(x) indicates the probability density function of a random variable x.

Joint Entropy: A pair of discrete random variables (x, y) with a joint distribution $P(x, y)$,

$$H(x, y) = -\sum_{x\in\chi}\sum_{y\in\gamma} p(x, y) \log p(x, y) \tag{12}$$

In equation (12), $H(x, y)$ denotes the joint entropy.

Mutual Information: Random variables (x, y) with a joint distribution p(x, y), their marginal possibility functions are $p(x)$ *and* $p(y)$.

$$I(x, y) = -\sum_{x\in\chi}\sum_{y\in\gamma} p(x, y) \log\left(\frac{p(x,y)}{p(x)p(y)}\right) \tag{13}$$

In equation (13), $I(x, y)$ indicates the mutual information.
The normalized mutual information is used that is calculated as follows,

$$NI(x, y) = \frac{I(x,y)}{min(H(x), H(y))} \tag{14}$$

The above equation is used to calculate the similarity among two training samples. So, the similarity matrix (SM) is measured as follows,

$$SM = \left(NI(x_i, x_j)\right), \qquad (i, j = 1,2,\dots,N) \tag{15}$$

The NEVs and their EVs of the SM are λ_j^m *and* $\alpha_j^m\ (j = 1,2,\dots,q)$, in that order. After that, the similarity subspace using mutual information is spanned by the below vectors,

$$\beta_j^m = \frac{X\alpha_j^m}{\sqrt{\lambda_j^m}}, \qquad (j = 1,2,\dots,q) \tag{16}$$

In the above equation, X denotes the vector based data set.

Similarity subspace model_2 based on angle information

We use the cosine distance to determine the similarity among the data points. The SM is described as follows,

$$Similarity\ Cosine\ (SC) = \left(cos(x_i, x_j)\right), \quad (i, j = 1,2, \dots, N) \qquad (17)$$

In the above equation, $cos(x_i, x_j)$ indicates the cosine distance among two data points $x_i\ and\ x_j$. That is, $cos(x_i, x_j) = x_i^T x_j / (\|x_i\|. \|x_j\|)$. We consider which the NEVs and their EVs of the similarity matrix SC are $\lambda_l^c\ and\ \alpha_l^c\ (l = 1,2, \dots, s)$, in that order. After that, the similarity subspace using cosine distance is spanned as follows,

$$\beta_l^c = \frac{X\alpha_l^c}{\sqrt{\lambda_l^c}}, \qquad (l = 1,2, \dots, s) \qquad (18)$$

In the above equation, X denotes the vector based data set.

Similarity subspace model_3 based on hybrid Gaussian and polynomial kernel

In this model, we utilize the Gaussian and polynomial kernel distance. That is, Gaussian kernel (GK), to calculate the similarity among the data points. The SM is described as follows,

$$SG\ and\ SP = \left(k(x_i, x_j)\right), \qquad (i, j = 1,2, \dots, N) \qquad (19)$$

In the above equation, $k(x_i, x_j)$ indicates the GK distance among two data points $x_i\ and\ x_j$. That is, $k(x_i, x_j) = exp\left(-\|x_i - x_j\|^2 / 2\sigma^2\right)$ for Similarity Gaussian (SG) and $k(x_i, x_j) = \left((x_i, x_j) + 1\right)^P$ for Similarity Polynomial (SP), in this equation, σ, P denotes the kernel parameter required to be specified. We consider which the NEVs and their EVs of the similarity matrix SG are $\lambda_h^k\ and\ \alpha_h^k\ (h = 1,2, \dots, t)$, in that order. After that, the similarity subspace using cosine distance is spanned as follows,

$$\beta_h^k = \frac{X\alpha_h^k}{\sqrt{\lambda_h^k}}, \qquad (h = 1,2, \dots, t) \qquad (20)$$

In the above equation, X denotes the vector based data set.

Algorithm

Input: dataset $D = \{d_1, \dots, d_n\}$
Output: optimal features and samples

```
1.  Assume $D = \{d_1, \dots, d_n\}, R = r_i, i = 1,2, \dots, n$
2.  Choose the relevant and non-redundant features $f_j\ (j = 1, \dots, m)$
3.  $for\ (i = 1; i \le n; i++)$
4.  $if\ (i \le n)$
5.  Calculate mean value using equation (1)
6.  Compute divergence value in equation (2) and (3)
7.  Compute the MPCA model_1 using (11) to (16)
8.  Compute the MPCA model_2 using (17) and (18)
9.  Compute the MPCA model_3 using (19) and (20)
10. Find the subset of samples $S_n\ (n = 1, \dots, m)$ using equation (4) and (5)
11. $else$
12. Go to step 3
```

EXPERIMENTAL RESULTS

In this section, the overall performance of the classification is compared with Optimized Hybrid Feature Selection (OHFS), Optimized Hybrid Feature Selection-CADEX-PCA-Sample Selection (OHFS-C-PCA-SS), Optimized Hybrid Feature Selection-Modified CADEX-MPCA-Sample Selection (OHFS-MC-MPCA-SS) in terms of True Positive rate, True Negative rate and classification accuracy.

Database Description

We are considered the students' dataset have 297 data example that is gathered from different colleges. In dataset 40 attributes are present that integrates students' name, course, age, gender and nature of college consists of medical/engineering, college type similar to government, self-financed, location feature, family belong to nuclear family or joint family, family factors such that occupation & educational qualification of family members, economic factors, college factors, social factors and spending time in television, mobile, computer, personal factors, academic factors etc.,. For example, location features described as the location in that students' home, school and college placed consists of rural area, urban area and semi-urban area. College features are one of the attributes that offers the information about whether student refer lecturer notes that is known by lecturer or books, techniques of teaching consists of lecturer technique/black board, number of students in class, whether college acceptable mobile phones or not, etc. Social features such as regulation of relatives for studies, number of friends and academic overall performance of friends.

In the data, the student's performance is evaluated consists of good /poor in the academy along with the features present. Data examples with these features are specified in the feature selection method then achieves chosen features. These chosen features are given to the classifiers for overall performance evaluation. In our experimentation, we are used Prism and J48 classifier. Prism and J48 are classification algorithms. Prism is used for inducing modular rules and J48 is used for building apruned or unpruned C4.5 decision tree. In our experimentation, the 150 data example is given as training data (with class label) to classifier for learning procedure and remaining data are assumed as test data (without class label) that is given to classifier with the intention of discovering the class label. At last, the output variable or attribute or class is to be determined in the dilemma is the academic status or student overall performance, which has two possible values: PASS (student who pass the course) or FAIL (a student who has to repeat the course).

True Positive rate

The TP rate represents the percentage of actual positives which are predicted to be positive. In this work, if the result class label from a prediction is PASS and the actual class label is also PASS, then it is called a TP rate. It is also known as sensitivity or recall. It is calculated as follows,

$$TP = \frac{TP}{TP + FN}$$

In the above equation, TP denotes the true positive and FN denotes the false negative.

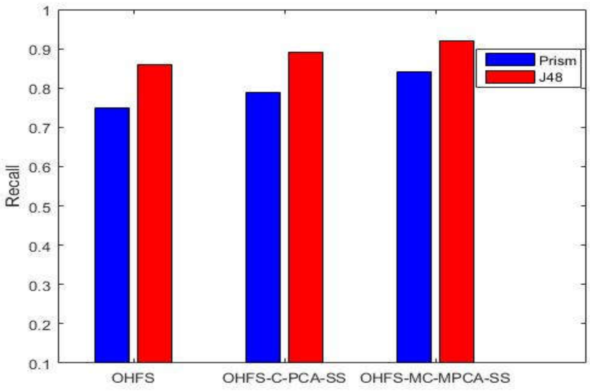

Figure 2 Comparison of True Positive Rate

From Figure 2, it is shown that the proposed OHFS-MC-MPCA-SS approach achieves high true positive rate compared to the other existing approaches. In this graph, X and Y axis are taken the classification schemes and true positive rate values, respectively.

True Negative Rate

TN rate describes the percentage of actual negatives which are predicted to be negative. In this work, if the result class label from a prediction is FAIL and the actual class label is also FAIL, then it is called a TN rate. It is also known as specificity. It is computed as follows,

$$TN = \frac{TN}{TN + FP}$$

In the above equation, FP denotes the False Positive.

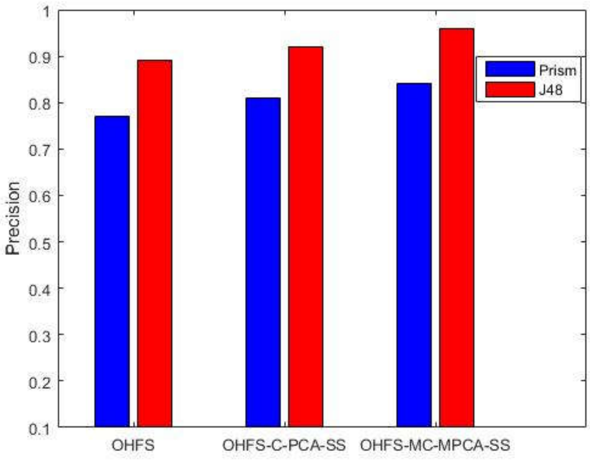

Figure 3 Comparison of True Negative rate

It is shown by Figure 3, comparison results of the proposed approach with existing approaches in terms of true negative rate. X-axis is taken as classification methods and Y-axis is taken as the true negative rate values. From the bar chart the proposed OHFS-MC-MPCA-SS approach provides high true negative rate.

Accuracy rate

It is described as the accuracy rate. False Positive (FP) rate represents as the percentage of actual negatives which are predicted to be positive. In this work, if the result class label from a prediction is PASS and the actual class label is FAIL, then it is called a FP rate. False Negative (FN) rate explains as the percentage of actual positives which are predicted to be Negative. In this work, if the result class label from a prediction is FAIL and the actual class label is PASS, then it is called a FN rate. Accuracy is computed as follows,

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

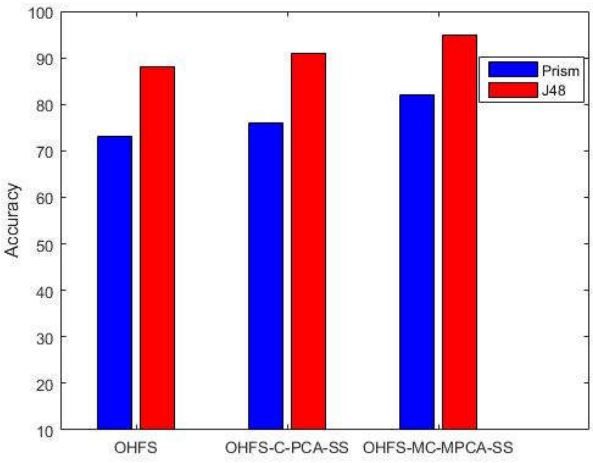

Figure 4 Comparison of Accuracy rate

The comparison of proposed and existing approaches for metric accuracy is shown in Figure 4. From the analysis, it is demonstrated that the proposed approach achieved very high than other existing approaches.

CONCLUSION

In this paper, a modified Computer Aided Design of Experiments (MCADEX) algorithm is proposed based on KL divergence for enhancing the accuracy of student's performance prediction. This approach is measured mean value for each class, after that divergence among mean and data samples are discovered by multiple references data samples. MPCA is used three similarity measurements for deriving the eigenvectors of the covariance matrix. In this method, the sample selected dataset is classified using Prism and J48 classifiers. The experimental outcomes illustrate which the proposed schemes are offering better outcomes in terms of True positive rate, Accuracy rate and True negative rate.

References

Hildrum, K. I. (1992). *Near infra-red spectroscopy: Bridging the gap between data analysis and NIR applications*. Ellis Horwood Ltd.

Ferré, J., & Rius, F. X. (1996). Selection of the best calibration sample subset for multivariate regression. *Analytical chemistry*, *68*(9), 1565-1571.

Ferré, J., & Rius, F. X. (1997). Constructing D-optimal designs from a list of candidate samples. *TrAC Trends in Analytical Chemistry*, *16*(2), 70-73.

Sasi regha, R. & Uma rani, R. (2017). An Efficient Clustering Based Feature Selection for Predicting Student Performance. iJET, 9(2), 524-531.

Yuan, T., Zhu, N., Shi, Y., Chang, C., Yang, K., & Ding, Y. (2018). Sample data selection method for improving the prediction accuracy of the heating energy consumption. *Energy and Buildings*, *158*, 234-243.

Kim, H. J. (2018). Bayesian hierarchical robust factor analysis models for partially observed sample-selection data. *Journal of Multivariate Analysis*, *164*, 65-82.

Adeli, E., Shi, F., An, L., Wee, C. Y., Wu, G., Wang, T., & Shen, D. (2016). Joint feature-sample selection and robust diagnosis of Parkinson's disease from MRI data. *NeuroImage*,*141*, 206-219.

Kim, H. J., & Kim, H. M. (2016). Elliptical regression models for multivariate sample-selection bias correction. *Journal of the Korean Statistical Society*, *45*(3), 422-438.

Lafférs, L., & Nedela Jr, R. (2017). Sensitivity of the bounds on the ATE in the presence of sample selection. *Economics Letters*, *158*, 84-87.

Li, H., Bao, W., Hu, J., Xie, J., & Liu, R. (2018). A training samples selection method based on system identification for STAP. *Signal Processing*, *142*, 119-124.

Sriboonchitta, S., Liu, J., Wiboonpongse, A., & Denoeux, T. (2017). A double-copula stochastic frontier model with dependent error components and correction for sample selection. *International Journal of Approximate Reasoning*, *80*, 174-184.

Duh, K., & Fujino, A. (2012). Flexible sample selection strategies for transfer learning in ranking. *Information Processing & Management*, *48*(3), 502-512.

How to cite this article:

Sasi Regha, R and Uma Rani, R.2018, A Representative Standardized Sample Set Selection for Improving Student's Performance Prediction. *Int J Recent Sci Res.* 9(2), pp. 24612-24617. DOI: http://dx.doi.org/10.24327/ijrsr.2018.0902.1693

****