IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

IMPROVED HARMONIC LINEAR DISCRIMINANT ANALYSIS PAIRWISE BASED DATA DIMENSIONALITY REDUCTION

FouziyaBanu.H[*1], Thenmozhi.N[2]

[*1]Department of Information Technology, Government Arts College, Coimbatore,Tamilnadu,India.
fouziyahabeeb@gmail.com[1]
[2]Department of Information Technology,Government Arts College,Coimbatore,Tamilnadu, India.
nthenmozhi300@gmail.com[2]

ABSTRACT

Data mining and machine learning techniques extract valuable information from database to exploit problem-solving and decision-making. Dimensionality reduction plays a key role in decision-making. Dimensionality reduction methods effectively reduce data dimensionality for efficient decision-making process. Linear Discriminant Analysis (LDA) is the most popular dimensionality reduction algorithm that found a subspace which minimize within-class distance and maximize between-class distance. LDA used arithmetic mean to calculate between-class distance. However, it has some limitations which were overcome by Harmonic Linear Discriminant Analysis (HLDA) and HLDA pairwise (HLDAp). HLDA and HLDAp used the harmonic mean to compute the between-class distance. However, the time complexity is high for the initialization of HLDAp. In this paper, Improved HLDAp (IHLDAp) is introduced to reduce the time complexity of HLDAp. An initialization schema is used in IHLDAp which initializes the composition matrix in a consistent and robust way. In the initialization schema, a datum is generated based on the identity of data, uniqueness of the data and noise. Then, the error of the initialization process is reduced by using Alternating Direction Method of Multipliers (ADMM). By minimizing the error, a proper initialization of composition matrix is achieved which helps to reduce the dimensionality of data. Finally, the reduced data is given as input to Support Vector Machine (SVM). The propose IHLDAp can be applied for both single label classification and multi-label classification problem. The experiments are conducted to prove the effectiveness of IHLDAp in terms of accuracy, precision and recall.

Keywords: Dimensionality reduction, Linear Discriminant Analysis, Harmonic Linear Discriminant Analysis, Harmonic Linear Discriminant Analysis pairwise, Improved Harmonic Linear Discriminant Analysis pairwise.

1. INTRODUCTION

In order to learn potential information by mapping input data to target outputs machine learning algorithms are used. But, when the input data consist of noise or redundant features, the generalization capability of machine learning methods will be negatively affected, and thus dimensionality reduction [1, 2] techniques are often required to pre-process training data to achieve better generalization performance. Dimensionality reduction is a task of reducing the number of variables which enhance the performance of the classification. Processing of high dimensional data leads the increase of complexity both in memory usage and execution time. There are different techniques have been proposed to reduce the dimensions of the dataset.

Nowadays, Linear Discriminant Analysis (LDA) [3] is the most powerful and fundamental tools of dimensionality reduction for extracting effective features on high-dimensional vectors in input data. The LDA used arithmetic mean of between-class distances and it doesn't consider pairwise between-class distance. So some classes may overlap with each other in subspace. This is the major limitation of LDA based dimensionality reduction method. To overcome the limitations in LDA, HLDA and HLDAp [4] were proposed for dimensionality reduction.

The HLDA objective using the harmonic mean-based pairwise between-class distance for dimensionality reduction which overcomes the disadvantage of LDA. HLDA was proposed to weight more heavily the close distance pairs of classes in the optimization, the difficult part of the discriminant function. In most of the datasets, various classes has within-class covariance, the global average of within-class would differ significantly from each class. But usually, the average of two classes is likely to be close to each of the two classes. Based on this idea, HLDAp was proposed which believed that using the global average of within-class distances of all classes is a less accurate representation as compared to use average of two class covariance. However, the time consumption is high for initialization in HLDAp.

In order to reduce the time complexity problem in HLDAp, IHLDAp is introduced in this paper for dimensionality reduction. In IHLDAp, an initialization schema is introduced which initializes the composition matrix in a consistent and robust way. Also, ADMM is used to reduce the error of initialization process. A proper initialization is achieved by minimizing the error. The reduced data is processed by SVM for classification process. By the proper initialization of HLDAp, an effective reduced data with less time is obtained which enhance the classification accuracy.

2. LITERATURE SURVEY

A feature ranking method is proposed [5] for intrusion detection system. This method was based on the gain and correlation of information. Through integrating the ranks obtained from both correlation and information gain the novel approach reduced the features. Finally, the reduced features were given for training and testing of KDD99 dataset as input to the feed forward neural network. Prior to practice, the KDD99 dataset is pre-processed by normalizing each class's number of instances. Then, the system behaved smartly to classify the test data into classes of attack and non-attack. Pre-processing is done manually in this method and it is possible to minimize the number of features in the reduced feature set which are the key drawbacks of this approach.

A Hybrid Genetic Algorithm with Wrapper- Embedded feature approach (HGAWE) [6] was proposed for feature selection in big data analysis. Through combining the genetic algorithm with the embedded regularization approach, this approach merged global search and local search. In addition, a new representation of chromosomes has been suggested in HGAWE for local and global optimization procedures. The regularization method has been chosen for the relevant features in big data, as per the representation of the chromosome. A training framework was developed simultaneously. The genetic operations were used to maximize the control parameters in non-convex regularization. However, the genetic algorithm is sensitive to the initial population.

A novel framework [7] was presented which combined distributed feature selection approach and econometric models for efficient economic big data analysis. To recognize the important attributes in economic data, a subtractive clustering-based feature selection algorithm was developed. Subtractive clustering is a clustering algorithm based on density that examined the correlation of data samples. The identification of the representative attributes was then merged with attribute coordination. To capture the hidden patterns for economic development, these feature selection processes combined with econometric model construction. However, this framework has high computational complexity problem.

A holistic approach [8] approach was presented to distributed dimensionality reduction of big data. A chunk tensor was presented in this approach to develop a unified model by merge the unstructured, structured and semi-structured data. In the unified model, all characteristics of the heterogeneous data were approximately ordered along the tensor orders. The dimensionality of the unified model was reduced by a proposed Lanczos based High Order Singular Value Decomposition algorithm. A Transparent Computing paradigm and linear predictive model were employed to construct the distributed computing platform and to partition the data blocks respectively. It executed the dimensionality reduction task effectively. However, this approach can provide a low rank approximation for the initial tensor which is not the best approximation of the initial data.

A Genetic Programming approach [9] was presented for high efficient feature selection technique that an efficient selection of the significant features was offered. Here, two main challenges such as curse of dimensionality and skewed data classification were considered for Automatic Document Classification (ADC). The proposed solution used the space of possible combinations of features selected via basic metrics to establish an unbiased estimator of the features ' discriminative power. Numerous feature space projections were combined with the proposed approach, optimizing classification accuracy and capturing the strongest feature-class relationships. In this method, due to data skewness, the problem of weighting and combining numerical values ranging from different scales to poor feature choice was avoided.

A hybrid approach called Ant Colony Optimization- Artificial Neural Network (ACO-ANN) [10] was proposed for feature selection in big data environment. The selection process was analyzed using the ACO algorithm and the ANN was used as the ACO-ANN method classification. By updating the position and velocity of each ant in the population, ACO minimized the dimensionality of original data by choosing optimal features. The selected features were used in ANN, which listed the best subset of features from all subsets and categorized the text. However, it has low accuracy problem.

3. METHODOLOGIES

In this section, dimensionality reduction based on LDA, HLDA, HLDAp and IHLDAp is described in detail.

3.1 Linear Discriminant Analysis

LDA is a widely used supervised dimensionality reduction algorithm. Assume, $X \in \mathbb{R}^{p\times n}$ be the data matrix, where $X = (x_1, x_2, \dots x_n)$ and p is the data dimension and n is the number of data points. Here, k represent the class index, c represent the desired sub space dimension, K represents the total number of classes. A transformation matrix to a c-dimensional subspace is represented as $H \in \mathbb{R}^{p\times c}$. The between-class scatter matrix M_{bc},within-class scatter matrix M_{wc} and total scatter matrix M_{total} is given as follows:

$$M_{bc} = \sum_{k=1}^{K} NP_k (s_k - s)(s_k - s)^T \qquad (1)$$

$$W_k = \frac{1}{NP_k} \sum_{x_i \in k} (x_i - s_k)(x_i - s_k)^T \qquad (2)$$

$$M_{wc} = \sum_{k=1}^{K} NP_k W_k \qquad (3)$$

$$M_{total} = M_{bc} + M_{wc} \qquad (4)$$

where, K is the total class number, NP_k is the number of points in class k,s_k is the mean of class k and s is the mean of entire dataset

s_k and s are calculated as,

$$s_k = \frac{\sum_{x_i \in k} x_i}{NP_k} \qquad (5)$$

$$s = \frac{\sum_{x_i \in k} x_i}{N} \qquad (6)$$

M_{bc}, M_{wc} and M_{total} are semi-positive definite matrices. Traditional LDS determines a transformation matrix H by solving the issue:

$$\max_H Tr \frac{H^T M_{bc} H}{H^T M_{wc} H} \qquad (7)$$

The concentrate of LDA is to maximize the between-class distance during minimizing within-class distance. The problem specified in Eq. (8) can be maximized to maximize between-class distance in the subspace of H.

$$\max_H \sum_{k=1}^{K} \|H^T(s_k - s)\|^2 = Tr(H^T M_{bc} H) \qquad (8)$$

The problem specified in Eq. (3.9) can be minimized to minimize the sum of within-class distance in the sub-space of H.

$$\min_H \sum_{k=1}^{K} \sum_{x_i \in k} \|H^T(x_i - s_k)\|^2 = Tr(H^T M_{wc} H) \qquad (9)$$

Another similar LDA objective function is formed by combining the above two tasks together which is given as follows,

$$\max_H \frac{Tr(H^T M_{bc} H)}{Tr(H^T M_{wc} H)}$$

$$\text{Subject to:} \quad H^T H = 1 \qquad (10)$$

Based on the objective function in Eq. (10), the dimensionality of the data is reduced.

3.2 Harmonic Linear Discriminant Analysis

The HDLA objective using the harmonic mean-based pairwise between-class distance for dimensionality reduction that overcomes the disadvantages of LDA. The arithmetic mean formulation is equivalent to the Eq. (3.10). The objective function of HLDA is given as follows:

$$\min_H I_1(H) = \sum_{k=l}^{K} n_k n_l \frac{Tr(H^T M_{wc} H)}{Tr(H^T B_{kl} H)}$$

$$\text{Subject to: } H^T H = 1 \qquad (11)$$

where, B_{kl}- pairwise between-class scatter matrix for class k and l.

HLDA was proposed to weight more heavily the close distance pairs of classes in the optimization, the difficult part of the discriminant function; whereas the conventional LDA weights more of large distance pairs, the less important part of the discriminant function. Hence, the HLDA is more robust.

An efficient algorithm was introduced to minimize the HLDA objective. The gradient of Eq. (11) is given as follows,

$$\nabla I_1 \triangleq \frac{\partial I_1}{\partial H} = 2\sum_{k<l} n_k n_l \frac{M_{wc}H}{Tr(H^T B_{kl} H)} - 2\sum_{k<l} n_k n_l B_{kl} H \frac{Tr(H^T M_{wc} H)}{(H^T B_{kl} H)^2} \quad (12)$$

Limitation $H^T H = I$ enforces H on the Stiefel manifold. Variations of H on this manifold is parallel transport, which gives some constraint to the gradient. The gradient that reserves the manifold structure is

$$\nabla I_1 - H|\nabla I_1|^T H \quad (13)$$

Hence the algorithm computes the new H is given as follows:

$$H \leftarrow H - \varphi(\nabla I_1 - H[\nabla I_1]^T H) \quad (14)$$

In Eq. (14), φ- step size. The manifold preserving the dimension which is typically small and it is very fast.

Stiefel gradient descent algorithm for HLDA

```
Input: $X \in \mathbb{R}^{p\times n}$, class indicator matrix $Y \in \mathbb{R}^{n\times k}$, $K$, $c$
Output: Projection matrix $H \in \mathbb{R}^{p\times c}$
1. Initialize $H$
2. Calculate $M_{wc}$ and $B_{kl}$
3. while objective in Eq. (3.11) not converge do
4. Calculate Stiefel manifold gradient
5. Update $H$
6. end while
```

3.3 Harmonic Linear Discriminant Analysis pairwise (HLDAp)

In most of the datasets, various classes has within-class covariance, the global average of within-class M_{wc} would differ significantly from each class. But usually, the average of two classes is likely to be close to each of the two classes. Based on this idea, it is believed that using the global average of within-class distances of all classes is a less accurate representation as compared to use average of two class covariance. Providentially, this pairwise average can be accommodated into Eq. (11). For this purpose, the pairwise within-class covariance of class k and l is introduced in HLDAp which is expressed as follows:

$$W_{kl} = \frac{1}{n_k+n_l}(n_k W_k + n_l W_l) \qquad (15)$$

In Eq. (12), W_k and W_l is calculated as in Eq. (2). The objective function of HLDAp is given as follows,

$$\min_H I_2(H) = \sum_{k=l}^{K} n_k n_l \frac{Tr(H^T W_{kl} H)}{Tr(H^T B_{kl} H)}$$

Subject to: $H^T H = 1$ (16)

A Stiefel gradient descend method is used to solve the minimization problem which is given as follows,

$$\nabla I_2 \triangleq \frac{\partial I_1}{\partial H} = \sum_{k<l} 2 n_k n_l \left[\frac{W_{kl} H}{Tr(H^T B_{kl} H)} - B_{kl} H \frac{Tr(H^T W_{kl} H)}{Tr(H^T B_{kl} H)^2}\right] \qquad (17)$$

3.4 Multi-label HLDA and HLDAp

In multi-class classification problem, each object is classified into more than one class. The objective of multi-label HLDA is formed by considering multi-label between-class scatter matrix $\widetilde{M_{bc}}$ and within-class scatter matrix $\widetilde{M_{wc}}$:

$$\min_H \sum_{k<l}^{K} n_k n_l \frac{Tr\left(H^T \widetilde{M_{wc}} H\right)}{Tr\left(H^T \widetilde{B_{kl}} H\right)}$$

Subject to: $H^T H = 1$ (18)

This is solved using similar approach as Eq. (11).

The multi-label pairwise within-class scatter matrix $\widetilde{W_{kl}}$ is defined as,

$$\widetilde{W_{kl}} = \frac{1}{n_k+n_l}\left(n_k \widetilde{W_k} + n_l \widetilde{W_l}\right) \qquad (19)$$

The objective of multi-label HLDAp is given as follows,

$$\min_H \sum_{k<l}^{K} n_k n_l \frac{Tr\left(H^T \widetilde{W_{kl}} H\right)}{Tr\left(H^T \widetilde{B_{kl}} H\right)}$$

Subject to: $H^T H = 1$ (20)

Eq. (20) is solved using Eq. (16).

3.3 IMPROVED HARMONIC LINEAR DISCRIMINANT ANALYSIS PAIRWISE

In IHLDAp, the initialization problem is formulated as,

$$x_{i,j} = Yh_i + Gw_{i,j} + r_{i,j} \qquad (21)$$

In Eq. (21), $x_{i,j}$ is the data point, Y is the subspace that provides different objects that exist in the training set, h_i is the vector that defines the identity of the particular data point, G is the projection matrix that defines the uniqueness for each data, $w_{i,j}$ is the vector that provides the uniqueness for the particular data and $r_{i,j}$ is the missing value or noise in the image.

Without any loss of generality, suppose that exist a total of P data and a total of o identities. By heaping all data in a column-wise manner, the model in (21) is reframed as,

$$X = Y[h_1 1^T \dots h_I 1^T] + G\widetilde{W_{kl}}\widetilde{B_{kl}} + R \qquad (22)$$

In Eq. (22),

$$X = \left[x_{1,1} \dots x_{1,J} \dots x_{I,1} \dots x_{I,J}\right] \qquad (23)$$

$$1^T = \underbrace{[1 \dots 1]}_{\text{J-times}} \qquad (24)$$

$$R = \left[r_{1,1} \dots r_{1,J} \dots r_{I,1} \dots r_{I,J}\right] \qquad (25)$$

The error R is minimized to provide subspaces Y and G as informative as possible. In order to achieve this, the following Eq. (3.26) needs to be solved.

$$\min_R \frac{1}{2}\|R\|_Y^2$$

$$\text{Subject to: } X = Y[h_1 1^T \dots h_I 1^T] + G\widetilde{W_{kl}}\widetilde{B_{kl}} + R \qquad (26)$$

where, $Y^T Y = I$, $G^T G = I$, $Y^T G = 0$

In Eq. (26), $\|X\|_Y = \sqrt{tr(X^T X)}$ is the Forbenius norm and $tr(.)$ is the trace of a square matrix. In order to ensure the subspaces are not correlated, a orthonormality constraints is introduced in the minimization function and hence the solution is identifiable. Eq. (26) is solved by using the Alternating Direction Method of Multipliers (ADMM) where an augmented Lagragian is minimized. The augmented Lagrangian can be written as follows:

$$\mathcal{L}(\theta) = \min_\theta \left\{\frac{1}{2}\|R\|_Y^2 + tr\left[\Lambda^T\left(X - YH - G\widetilde{W_{kl}}\widetilde{B_{kl}} - R\right)\right] + \frac{\mu}{2}\left\|X - YH - G\widetilde{W_{kl}}\widetilde{B_{kl}} - R\right\|_Y^2\right\}$$

$$\text{Subject to: } Y^T Y_t = I, Y_t^T G_t = 0 \qquad (27)$$

For deriving subspace G, the following Eq. (28) need to solve

$$G_{t+1} = \underset{G_t}{\operatorname{argmin}} \left\| X - R_t + \frac{\Lambda_t}{\mu_t} G \widetilde{W_{kl\_t}} \widetilde{B_{kl\_t}} - Y_{t+1} H_t \right\|_Y^2$$

$$\text{Subject to: } G_t^T G_t = I, G_t^T Y_{t+1} = I \qquad (28)$$

Eq. (28) is solved using Eq. (16). Thus the dimensionality of the data is reduced effectively. The reduced dimensionality data is given as input to the Support Vector Machine (SVM) to classify the data.

4. RESULT AND DISCUSSION

In this section, the performance of LDA-SVM, HLDA-SVM, HLDAp-SVM and IHLDAp-SVM are tested in terms of accuracy, precision and recall. Single label datasets YaleB and ATT and multi-label datasets MediaMill and Barcelona datasets are used for experimental purpose. Table 1 shows the dataset description.

Table.1 Dataset Description

Data	Dataset Type	Dimension	Sample Number	Class Number
YaleB	Single-label dataset	504	1984	31
ATT	Single-label dataset	644	400	40
MediaMill	Multi-label dataset	120	6601	74
Barcelona	Multi-label dataset	48	139	4

4.1 Accuracy

Accuracy is the measure of correctly classify the data based on the reduced dimensionality in all instances. It can be calculated by

$$Accuracy = \frac{(True\ Positive\ (TP) + True\ Negative (TN))}{(TP + TN + False\ Positive\ (FP) + False\ Negative\ (FN))}$$

where, TP is actual positive data which are exactly classified as positives, TN is the actual negative data which are classified exactly as negatives, FP is known negative data which are wrongly classified as positives and FN is known positive data which are wrongly classified as negatives.

Figure.1 shows accuracy of LDA-SVM, HLDA-SVM, HLDAp-SVM and IHLDAp-SVM on YaleB and ATT datasets. On YaleB dataset, the accuracy of IHLDAp-SVM is 46.03% greater than LDA-SVM, 29.58% greater than HLDA-SVM and 8.24% greater than HLDAp-SVM. On ATT dataset, the accuracy of IHLDAp-SVM is 15.38% greater than LDA-SVM, 9.76% greater than HLDA-SVM and 3.45% greater than HLDAp-SVM. From this analysis, it is proved that the proposed IHLDAp-SVM has high accuracy than other methods for single-label datasets.

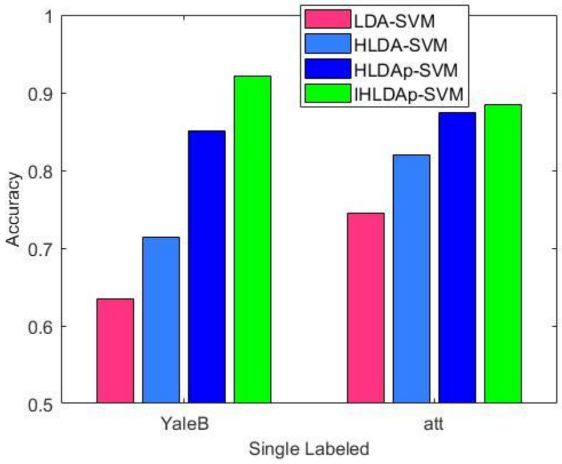

Figure.1 Evaluation of Accuracy on Single-label Dataset

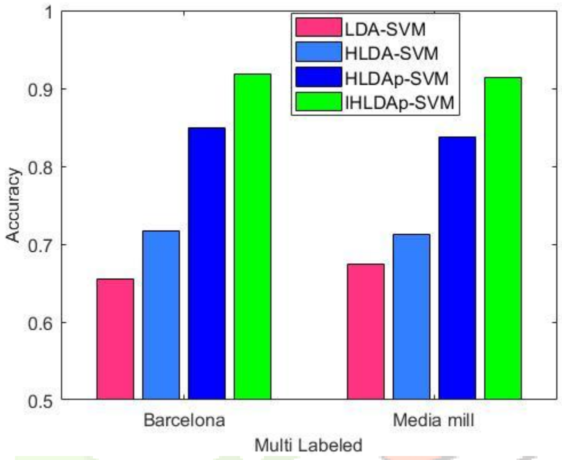

Figure.2 Evaluation of Accuracy on Multi-label Dataset

Figure.2 shows accuracy of LDA-SVM, HLDA-SVM, HLDAp-SVM and IHLDAp-SVM on Barcelona and MediaMill datasets. On Barcelona dataset, the accuracy of IHLDAp-SVM is 36.36% greater than LDA-SVM, 26.76% greater than HLDA-SVM and 5.88% greater than HLDAp-SVM. On MediaMill dataset, the accuracy of IHLDAp-SVM is 34.63% greater than LDA-SVM, 28.86% greater than HLDA-SVM and 5.99% greater than HLDAp-SVM. From this analysis, it is proved that the proposed IHLDAp-SVM has high accuracy than other methods for multi-label datasets.

4.2 Precision

Precision value is evaluated according to the classification at true positive prediction and false positive prediction.

$$Precision = \frac{TP}{(TP + FP)}$$

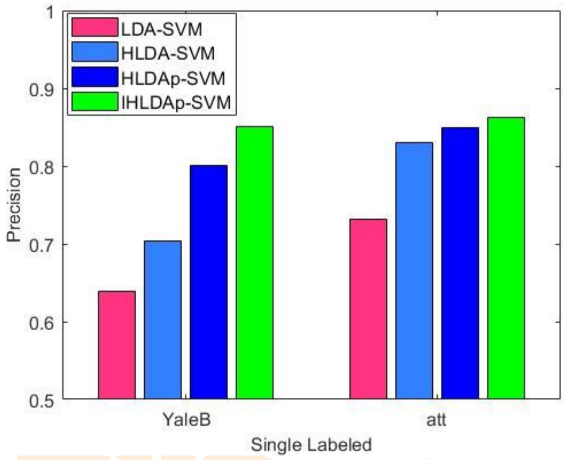

Figure.3 Evaluation of Precision on Single-label Dataset

Figure.3 shows precision of LDA-SVM, HLDA-SVM, HLDAp-SVM and IHLDAp-SVM on YaleB and ATT datasets. On YaleB dataset, the precision of IHLDAp-SVM is 32.81% greater than LDA-SVM, 21.43% greater than HLDA-SVM and 6.25% greater than HLDAp-SVM. On ATT dataset, the precision of IHLDAp-SVM is 15.79% greater than LDA-SVM, 6.02% greater than HLDA-SVM and 3.53% greater than HLDAp-SVM. From this analysis, it is proved that the proposed IHLDAp-SVM has high precision than other methods for single-label datasets.

Figure.4 shows precision of LDA-SVM, HLDA-SVM, HLDAp-SVM and IHLDAp-SVM on Barcelona and MediaMill datasets. On Barcelona dataset, the precision of IHLDAp-SVM is 27.69% greater than LDA-SVM, 18.57% greater than HLDA-SVM and 3.75% greater than HLDAp-SVM. On MediaMill dataset, the precision of IHLDAp-SVM is 25% greater than LDA-SVM, 19.57% greater than HLDA-SVM and 4.43% greater than HLDAp-SVM. From this analysis, it is proved that the proposed IHLDAp-SVM has high precision than other methods for multi-label datasets.

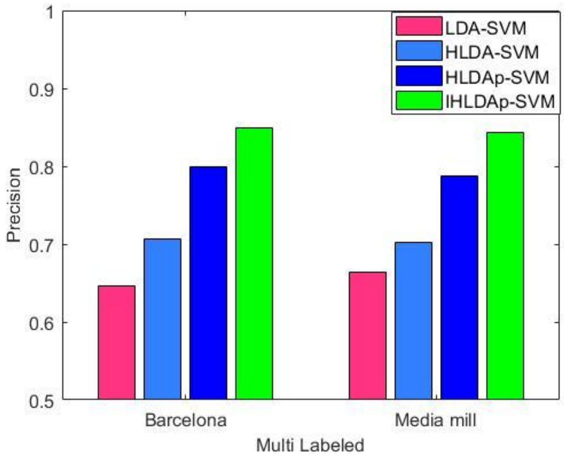

Figure.4 Evaluation of Precision on Multi-label Dataset

5.2.3 Recall

Recall is evaluated according to classification at true positive and false negative predictions.

$$Recall = \frac{TP}{(TP + FN)}$$

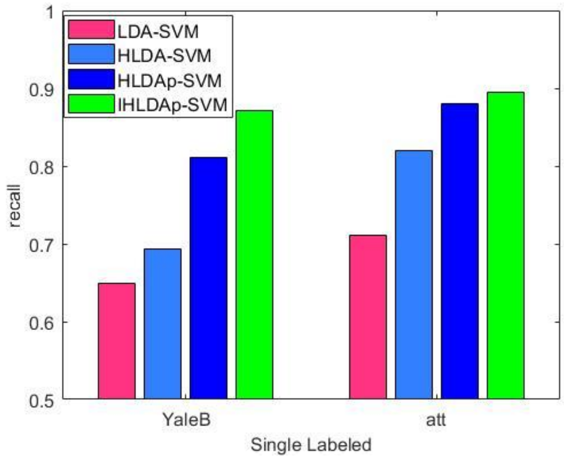

Figure.5 Evaluation of Recall on Single-label Dataset

Figure.5 shows recall of LDA-SVM, HLDA-SVM, HLDAp-SVM and IHLDAp-SVM on YaleB and ATT datasets. On YaleB dataset, the recall of IHLDAp-SVM is 33.85% greater than LDA-SVM, 24.29% greater than HLDA-SVM and 7.41% greater than HLDAp-SVM. On ATT dataset, the recall of IHLDAp-SVM is 22.97% greater than LDA-SVM, 10.98% greater than HLDA-SVM and 3.41% greater than HLDAp-SVM. From this analysis, it is proved that the proposed IHLDAp-SVM has high recall than other methods for single-label datasets.

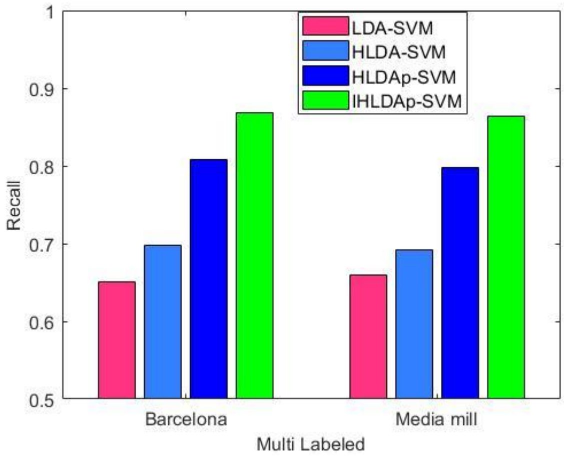

Figure.6 Evaluation of Recall on Multi-label Dataset

Figure.6 shows recall of LDA-SVM, HLDA-SVM, HLDAp-SVM and IHLDAp-SVM on Barcelona and MediaMill datasets. On Barcelona dataset, the recall of IHLDAp-SVM is 30.77% greater than LDA-SVM, 21.43% greater than HLDA-SVM and 6.25% greater than HLDAp-SVM. On MediaMill dataset, the recall of IHLDAp-SVM is 29.09% greater than LDA-SVM, 25.29% greater than HLDA-SVM and 5.19% greater than HLDAp-SVM. From this analysis, it is proved that the proposed IHLDAp-SVM has high recall than other methods for multi-label datasets.

5. CONCLUSION

In this paper, IHLDAp is proposed for dimensionality reduction which solves the time complexity problem in HLDAp. An initialization schema is introduced in IHLDAp which eliminate the local maxima problem and initializes the composition matrix in more efficient way. In IHLDAp, a datum is created according to the identity of data, uniqueness of the data and noise. Also, the error is reduced by using ADMM. Thus, proper initialization of composition matrix is achieved which enhance the performance of dimensionality reduction and data classification. The experimental results prove that the proposed IHLDAp-SVM has high accuracy, precision and recall for both single label and multi-label datasets.

REFERENCES

[1] Kasun, L. L. C., Yang, Y., Huang, G. B., & Zhang, Z. (2016). Dimension reduction with extreme learning machine. *IEEE Transactions on Image Processing*, *25*(8), 3906-3918.

[2] Tian, H., Lan, L., Zhang, X., & Luo, Z. (2019). Neighbors-Based Graph Construction for Dimensionality Reduction. *IEEE Access*, *7*, 138963-138971.

[3] Ye, J., & Ji, S. (2010). Discriminant analysis for dimensionality reduction: An overview of recent developments. *Biometrics: Theory, Methods, and Applications. Wiley-IEEE Press, New York*.

[4] Zheng, S., Ding, C. H., Nie, F., & Huang, H. (2018). Harmonic Mean Linear Discriminant Analysis. *IEEE Transactions on Knowledge and Data Engineering*.

[5] Manzoor, I., & Kumar, N. (2017). A feature reduced intrusion detection system using ANN classifier. *Expert Systems with Applications*, *88*, 249-257.

[6] Liu, X. Y., Liang, Y., Wang, S., Yang, Z. Y., & Ye, H. S. (2018). A hybrid genetic algorithm with wrapper-embedded approaches for feature selection. *IEEE Access*, *6*, 22863-22874.

[7] Zhao, L., Chen, Z., Hu, Y., Min, G., & Jiang, Z. (2018). Distributed feature selection for efficient economic big data analysis. *IEEE Transactions on Big Data*, (2), 164-176.

[8] Kuang, L., Yang, L. T., Chen, J., Hao, F., & Luo, C. (2018). A Holistic Approach for Distributed Dimensionality Reduction of Big Data. *IEEE Transactions on Cloud Computing*, (2), 506-518.

[9] Viegas, F., Rocha, L., Gonçalves, M., Mourão, F., Sá, G., Salles, T., ... & Sandin, I. (2018). A Genetic Programming approach for feature selection in highly dimensional skewed data. *Neurocomputing*, *273*, 554-569.

[10] Manoj, R. J., Praveena, M. A., & Vijayakumar, K. (2019). An ACO–ANN based feature selection algorithm for big data. *Cluster Computing*, 1-8.