www.ijecs.in
International Journal Of Engineering And Computer Science ISSN:2319-7242
Volume 4 Issue 3 March 2015, Page No. 10768-10771

Comparative Analysis Of Methods For Semi- Supervised Dimensionality Reduction

Archana H.Telgaonkar[1], Sachin Deshmukh[1]
[1] Department of computer science and IT, Dr. B.A.M. University,
Aurangabad-431004, India
archanashailu@email.com

Abstract: ***Data classification is one of the most challenging areas in the field of Machine Learning and Pattern Recognition application where data is represented as a point in high-dimensional space. The data can be classified using supervised learning if it is already labeled. Otherwise unsupervised learning is used. To get golden point between them, Semi supervised learning is introduced which uses both labeled and unlabeled data. Analyzing the high dimensional data is the biggest challenge that can be tackled with the help of dimensionality reduction techniques. When Dimensionality Reduction is embedded in Semi supervised learning, it gives superior performance. The purpose of dimensionality reduction is to reduce complexity of input data without losing important details.***

In this paper, Semi supervised learning is studied using four different approaches. Analysis and comparative study of these techniques is illustrated with the help of three datasets. Role of dimensionality reduction is also observed in the classification of data.

Keywords: Dimensionality reduction, Manifold regularization, Semi-supervised learning.

1. Introduction

Directly working on high dimensional data in many real application is not only time consuming but also computationally unreliable. Therefore dimensionality reduction plays an important role. By reducing data in to fewer dimensions it palliate the curse of dimensionality and provide efficient way for data processing. Many dimensionality reduction approaches have been proposed among that Principal Component Analysis (PCA) [1] and linear discriminant analysis (LDA) [2] are two popular methods which have been widely used in many classification applications. PCA is unsupervised while LDA is supervised and can achieve better classification results due to the utilization of label information. Labeled data are Limited, and to labeled large data would require expensive human labor in practice. While unlabeled data is relatively easier to obtain. To effectively utilize both labeled and unlabeled data simultaneously semi-supervised learning was proposed and used in many real applications [3], e.g., face recognition, cross-media retrieval, image segmentation, text categorization. In this paper we present a overview of semi-supervised dimensionality reduction methods namely SDA, Lap-Rs/L, LGC, GFHF along with performance evaluation on three dataset , CMU PIE [4], YALE-B [5] and COIL-20 [6].

1.1 Semi Supervised Learning

Semi supervised learning deal with both labeled and unlabeled data, large amounts of unlabeled data and a small quantity of labeled data. Where as in supervised learning the training dataset comprises of only labeled data. The process of finding a better classifier from labeled and unlabeled data is done by semi-supervised learning. The semi-supervised learning methodology can deliver high performance of classification by utilizing unlabeled data. Vapnik introduced a problem related to Semi Supervised Learning, already several decades ago: so called transductive learning, is to perform prediction only for test point not for new coming data point, in contrast to inductive learning which predicts training and unseen data. There are some semi-supervised learning models include co training, self training, graph-based methods, multi-view learning.

1.2 Dimensionality Reduction

Increase the dimensionality refers to the problems associated with multivariate data analysis. While working with such large data needs to reduce the dimensionality, the goal of dimensionality reduction is to represent the data in a lower-dimensional space, by keeping some of the original properties of the data. Equation.1 shows the reduction of high-dimensional data space data [K] in to lower-dimensional space [N].

$$X = \begin{bmatrix} a_1 \\ a_2 \\ a_n \end{bmatrix} - - > \text{ reduce dimensionality } - - > y = \begin{bmatrix} b_1 \\ b_2 \\ b_k \end{bmatrix} \quad (K << N) \qquad (1)$$

At the intersection of several disciplines, including statistics, databases, data mining, pattern recognition, text mining, artificial intelligence, visualization and optimization Dimensionality reduction is important. There are several supervised (LDA, SVM, HNN), unsupervised (PCA, SVD, ICA) and semi supervised algorithm for dimensionality reduction. In this paper we take overview of SDA, Lap-Rs/L, LGC, GFHF semi-supervised algorithm and make performance observation on dataset.

2. Methodologies

2.1 SDA (Semi-supervised Discriminant Analysis)

SDA is extension of supervised learning method LDA, unlike LDA semi supervised discriminant analysis method use both labeled and unlabeled samples. To maximize the separability between different classes the labeled data points are used, and to estimate the intrinsic geometric structure of the data the unlabeled data points are used. As mention previous manifold regularization, LDA extends to semi supervised learning method by adding a geometrically based regularization term. Nearby points will have similar representations in the lower-dimensional space is manifold smoothness assumption in SDA. Let $x_l=\{x_1, x2, xn\}$ as the data matrix of labeled data, and denote the number of the labeled samples in i-th class as n_i. Let us denote two graph similarity matrices $S^{\tilde{w}}$, $S^b \varepsilon R^{n\times n}$, where $S^{\tilde{w}}=\partial_{yi,yj}/n_{yi}$, $S^b_{ij}=\frac{1}{n}-S^{\tilde{w}}ij$ The corresponding Laplacian matrices of $S^{\tilde{w}}$, S^b are represented as $L_{\tilde{W}}$, $L_{\tilde{b}}$ respectively. The inter-class scatter S_b and intra-class scatter S_w of LDA can be rewritten as $S_b=\sum_{l=1}^{c} nc\,(\bar{x}_l-\bar{x})(\bar{x}_l-\bar{x})^T = x_l\tilde{L}_b x_l^T$ and sw= $\sum_{i=1}^{n}(x_i-\bar{x}_{yi})(x_{i-}\bar{x}_{yi})^T$, where $\bar{x}_l$ is the mean of the labeled samples in the l-th class and $\bar{x}$ is the mean of all the labeled samples. The objective function in SDA is

$$g_s(w)=\frac{|w^T x_l L_b\ x_l^T|}{|w^T(x_l(L_w+L_b)x_l^T+\alpha x l x^T+\beta I)w|} \quad (2)$$

where α and β are two parameters to balance three terms and $L\varepsilon R^{m\times m}$ is the graph Laplacian matrix for both labeled and unlabeled data[7].

2.2 LapRLS/L (Linear Laplacian Regularized Least Squares)

The Manifold regularization (MR) method [8] extends many existing methods such as least square and SVM to their semi-supervised learning methods by adding a manifold regularized term to preserve the geometrical structure. the linear Laplacian regularized Least Square method (referred as Lap-RLS/L) take as an example, Let $X=\{X_l,X_u\}=\{x_1,x_2,\ldots x_{l+u}\}\varepsilon R^{D\times(l+u)}$ be the data matrix where the first l columns are the labeled and the remaining u columns are unlabeled samples. The goal of Lap-RLS/L is to fix a linear model $y_j = V^T x_j+b^T$ by regressing X on Y and simultaneously to preserve the manifold smoothness embedded in both labeled and unlabeled set, where $V\varepsilon R^{D\times d}$ is the projection matrix and $b\varepsilon R^{1\times c}$ is the bias term. The objective function of Lap-RLS/L can be given as

$$J(v,b)=\min \sum_{j=1}^{l}\left\|v^T x_j+b^T-y_j\right\|_F^2+\alpha_t\|v\|_F^2+\alpha_m Tr(v^T xLx^T v) \quad (3)$$

Where L=D-W is the graph laplacian matrix associated with both labeled and unlabeled set [9], W is the weight matrix defined as: $w_{ij}=\exp\left(-\left\|x_i x_j\right\|^2 / 2\sigma^2\right)$, if xi is within the k nearest neighbor of x_j or x_j is within the k nearest neighbor of x_i ;$w_{ij}=0$, otherwise, D is a diagonal matrix satisfying $Dii=\sum_{j=1}^{l+u} w_{ij}$, αm and αt are the two parameters balance the tradeoff between manifold and Tikhonov regularized terms[10].

2.3 GFHF (Gaussian Fields and Harmonic Functions)

GFHF is graph based semi-supervised method where the nodes are labeled and unlabeled examples in the dataset, and edges it may be weighted reflect the similarity of examples. In this method estimating a function f on the graph that satisfies, fitness that is F should be close to the given labels on the labeled nodes. And the manifold smoothness that is F should be smooth on the whole graph. When this expressed in regularization framework first term is a loss function, and the second term is a regularizer. GFHF adopt Gaussian fields over a continuous state space rather than random fields over the discrete label set. Having a quadratic loss function with infinity weight can be viewed, so that the labeled data are fixed at given label values, and a regularizer based on the graph combinatorial LaplacianΔ: [11]

$$\sum_{i\varepsilon L}(f_i-y_i)^2+1/2\sum_{i,j}w_{ij}(f_i-f_j)^2 = \infty\sum_{i\varepsilon L}(f_i-y_i)^2+f^T\Delta f \quad (4)$$

GHFH estimate a prediction label matrix $F \in R^{m\times c}$ on the graph. Let us denote F_i and Y_i. As the i-th row of F and Y. The objective function

$$g_G(f)=\frac{1}{2}\sum_{i,j=1}^{m}\left\|f_i-f_j\right\|^2 s_{ij}+\lambda_\infty\sum_{i=1}^{n}\|f_i-y_i\|^2 \quad (5)$$

Where the coefficient λ balances the label fitness and the manifold smoothness, λ∞ is a very large number such that $\sum_{i=1}^{n}\|f_i-y_i\|^2=0$, or $f_i=y_i$ $\forall_i$=1, 2, and $\|A\|^2=trace(A^TA)$[7]. This method is used in image segmentation, colorization of gray scale images, sentiment analysis etc.

2.4 LGC (Local and Global Consistency)

As mentioned above graph based method, LGC estimate a prediction label matrix $F \in R^{m\times c}$ on the graph with respect to the label fitness and the manifold smoothness. The prior assumption of consistency is the key to semi-supervised learning problems. Often called the cluster assumption where, nearby points are likely to have the same label is Local consistency and Points on the same structure (typically referred to as a cluster or manifold) are likely to have the same label is Global consistency.

LGC minimize the objective function:

$$Q(F)=\frac{1}{2}\sum_{i,j=1}^{n}w_{ij}\left\|\frac{F_i}{\sqrt{D_{ii}}}-\frac{F_j}{\sqrt{D_{jj}}}\right\|^2+\mu\sum_{i=1}^{n}\|F_i-y_i\|^2 \quad (6)$$

Where μ > 0 is the regularization parameter. Classifying function is $F^*=\arg\min_{F\varepsilon f} Q(F)$. In the function $Q(F)$ first term of the right-hand side is the smoothness constraint, which shows that a good classifying function should not change too much between nearby points. And the second term is the fitting constraint, which revenue a good classifying function should not change too much from the initial label assignment. The positive parameter μ captured the trade-off between these two competing constraints. Both labeled and unlabeled data Contains in the fitting constraint [12].

3. Experimental Results

In the experiments we observe the performance of different semi-supervised algorithms applied on two face database, CMU PIE [4], YALE-B [5] and one object database COIL-20 [6].

3.1 Dataset

In COIL-20(Columbia Object Image Library) database consist of 20 objects varying angles at the interval of five degrees, resulting in 72 images of per object total samples are 1440. Each image is down-sample to the size of 32×32 and extracts a 1024 dimensional gray level feature for each image. CMU PIE is multi-view face database; consist of 575 images of 20

people. The size of each cropped image is 112×92 with 256 gray levels per pixel, resized each image to 28× 23 pixels. The CMU PIE database contains more than 40,000 facial images of 68 people [4]. The images were acquired under variable illumination conditions, over different poses, and with different facial expressions. The images from the frontal pose (C27) we choose here and each subject has around 49 images from varying illuminations and facial expressions. The images are cropped and then resized to 32 × 32 pixels.

YALE-B database consists of 38 subjects, with each person having around 64 near frontal Images under different illumination. The images are cropped and then resized to 32 × 32 pixels. Gray-level features are used for face recognition in this work.

3.2 Performance analysis

The performance (Mean Recognition Accuracy ± Standard Deviation %) of LapRLS/L[13], SDA[15], GFHF[14], LGC[12] semi-supervised algorithm over 20 random splits on three databases, For each dataset result shown in table Top-1 Recognition Performance, from A Flexible and Effective Linearization Method for Subspace Learning [7] and judged by t-test. The algorithms SDA, LapRLS/L nearest neighbor classifier is performed for classification after dimension reduction. For GFHF and LGC, we directly use the classification methods proposed in [14, 12] for classification. For LapRLS/L, SDA and GFHF, we need to determine the Laplacian matrix M (or L), in which the graph similarity matrix is set as $s_{ij} = \exp\left(-\|x_i - x_j\|^2/t\right)$, x_i (orx_j) is among k nearest neighbors of x_i (orx_j); s_{ij}=0, otherwise. For LGC, used the normalized graph Laplacian matrix$L' = I - D^{-\frac{1}{2}}SD^{-\frac{1}{2}}$, [12]. For LapRLS/L, GFHF and LGC the diagonal matrix U is determined [13, 14, and 12].

In all the experiments PCA is used as a preprocessing step to remove the null space of data covariance matrix and preserve 95% energy of the data. The final dimensions after dimension reduction are fixed as c for SDA, LapRLS/L. The regularization parameters (i.e. λI and λA in LapRLS/L, α and β in SDA) need to be set beforehand to balance different terms. For fair comparison, we set each parameter to $\{10^{-9}, 10^{-6}, 10^{-3}, 10^{0}, 10^{3}, 10^{6}, 10^{9}$ and then we report the top-1 recognition accuracy from the best parameter configuration.

We randomly selected 50% data as the training dataset and the remaining 50% data use as the test dataset. Among the training data, we arbitrarily label p samples per class and treat the other training samples as unlabeled data. Here p set as 1, 2 and 3 For CMU PIE, YALE-B and COIL-20 databases. The mean recognition accuracy and standard deviation over 20 random splits on the unlabeled and test respectively, is reported and result is judged by t-test with a 0.05significance level [7]. The recognition performance of GFHF on Coil-20 dataset, when considered 1, 2, 3 labeled sample respectively, shows unlabel (%): 78.6±2.1, 83.2±2.2, 85.6±2.0, same method apply on Yale-B dataset shows the results are 22.5±2.9, 35.9±3.3, 45.2±3.9. And for CMU PIE dataset results are 33.9±3.3, 478±2.6, 55.8±2.1. The same way LGC, SDA, LapRL/L methods are applied on these three dataset for unlabel and test data. By analyzing Top-1 Recognition Performance table, from A Flexible and Effective Linearization Method for Subspace Learning [7].

We observed that SDA is generally better on CMU PIE database as compare to other semi supervised algorithm, in terms of mean recognition accuracy. On the unlabeled dataset of COIL-20 the mean recognition accuracies of LGC and GFHF are generally better than SDA and LapRLS/L but worse on the unlabeled dataset of CMU PIE and Yale-B databases. LGC and GFHF cannot handle the unseen samples; the results for LGC and GFHF on the test dataset are not reported.

4. Conclusion

Transductive learning method such as GFHF, LGC, ISOMAP, LLE, LE, LSE predict labels for unlabeled data and obtain low dimensional coordinates but cannot deal with arbitrary new coming data which is known as out of sample problem, to handle this problem linear projection function is used for mapping new data. There are various methods which deals with such problem For example Locality Preserving Projections (LPP). The need of dimensionality reduction techniques presents new challenges, and novel methods are expected to be developed depending upon the application. In this paper we observe the performance of semi-supervised method GFHF, LGC, LapRLS/L, SDA on the three dataset, there is no consistent conqueror on all the databases and the unlabeled data can be used to improve the recognition performance.

References

[1] M. Turk, A. Pentland, "Face recognition using Eigenfaces", Computer Vision and Pattern Recognition, Proceedings CVPR'91, IEEE Computer Society Conference on (1991), pp. 586-59, 1991.

[2] P.N. Belhumeur, J.P. Hespanha, D.J. Kriegman, "Eigenfaces vs. fisherfaces: recognition using class specific linear projection", IEEE Trans. Pattern Anal. Mach. Intell. 19 (7), pp. 711–720, 1997.

[3] Feiping Nie, Dong Xu, Xuelong Li, Senior Member, IEEE, and Shiming Xiang: "Semisupervised Dimensionality Reduction and Classification through Virtual Label Regression", IEEE Transactions on Systems, Man, and Cybernetics—Part B: Cybernetics, 41(3), pp. 675-684, June 2011.

[4] T. Sim and S. Baker, "The cmu pose, illumination, and expression database", IEEE Transactions on Pattern Analysis and Machine Intelligence, 25(12), pp. 1615–1617, 2003.

[5] A. Georghiades, P. Belhumeur, and D. Kriegman, "From few to many: Illumination cone models for face recognition under variable lighting and pose", IEEE Transactions on Pattern Analysis and Machine Intelligence, 23(6), pp. 643–660, 2001.

[6] S. A. Nene, S. K. Nayar, and H. Murase, "Columbia object image library (coil-20)", Technical report, Columbia University, 1996.

[7] Feiping Nie, Dong Xu, Ivor W. Tsang, Changshui Zhang, "A Flexible and Effective Linearization Method for Subspace Learning, Graph Embedding for Pattern Analysis", pp.177-203, 2013.

[8] M. Belkin, P. Niyogi, V. Sindhwani, "Manifold regularization: a geometric framework for learning from labeled and unlabeled examples", J. Mach. Learn. Res. 7, 2399–2434, 2006.

[9] S. Yan, D. Xu, B. Zhang, H. Zhang, Q. Yang, S. Lin, "Graph embedding and extensions: a general framework

for dimensionality reduction", IEEE Trans. Pattern Anal. Mach. Intell. 29 (1), pp. 40–51, 2007.
[10] Mingbo Zhaoa, Bing Lib, n, Zhou Wu a, Choujun Zhanc, "Image classification via least square semi-supervised discriminant analysis with flexible kernel regression for out-of-sample extension", Neurocomputing 153, pp. 96–107, 2015.
[11] Xiaojin Zhu," Semi-Supervised Learning Literature Survey ",Computer Sciences TR 1530 University of Wisconsin – Madison Last modified on July 19, 2008.
[12] D. Zhou, O. Bousquet, T. N. Lal, J. Weston, and B. Sch¨olkopf," Learning with local and global consistency", In Advances in Neural Information Processing Systems, 2004.
[13] V. Sindhwani, P. Niyogi, and M. Belkin, " Linear manifold regularization for large scale semi supervised learning", In Workshop on Learning with Partially Classified Training Data, International Conference on Machine Learning, 2005.
[14] X. Zhu, Z. Ghahramani, and J. Lafferty," Semi-supervised learning using gaussian fields and harmonic functions ", In Proceedings of the International Conference on Machine Learning, 2003.
[15] D. Cai, X. He, and J. Han, "Semi-supervised discriminant analysis", In Proceedings of the IEEE International Conference on Computer Vision ,pp. 1-7, 2007.

Author Profile

Archana H. Telgaonkar received the B.Tech (CSE) degree from SGGS I E & T, Nanded in 2011. Now doing M.Tech (CSE) from Department of Computer Science & IT, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad.