Your Project Report Title Here  
CS7IS2 Project (2019-2020)

Student Name (Multiple names separated by comma)

email\_address\_1 (multiple email addresses separated by comma)

**Abstract.** This paper compares Principle Component Analysis (PCA), Independent Component Analysis (ICA) and Non-negative Matrix Factorization (NMF) based on their accuracy of predication on facial recognition and time it takes for the SVM classifier to train on them. This paper also tries to explain the underlying differences in each of the three-dimension reduction methods. In the comparison PCA outperforms ICA slightly and NMF performed the worst when it comes to prediction accuracy.

This document is a guideline for writing the final report for the CS7IS2 module *Artificial Intelligence*. You should follow its general structure as shown below.

You should not change its format (font, size, margin, space, etc.).

Report that not comply to the format or exceed the maximum length will be penalised (-5 marks).

Brevity is desirable in communication, however you should provide all those details necessary for the good understanding of the described methods and algorithms.

The report will be graded on the basis of:

* Originality;
* Technical soundness;
* Organisation;
* Clarity of presentation
* Adequacy of bibliography/Results (this last point strongly depends on the type of report)

**Your report should provide a survey and an experimental comparison of multiple solution approaches to a particular problem. This is a critical review of at least three papers that significantly contributed to advance the state-of-the-art for the problem you are analysing. It should not be a mere summary of the papers. You are expected to conduct an analytical review of the methods under analysis to try to find common aspect and differences, connections between methods, drawbacks and open problems. Unless the faced problem has emerged recently, students should choose their papers by diversifying the range of approaches used to solve the problem. A good guideline could be to choose a paper from a decade or two ago, and a couple of more recent papers. You need to experimentally evaluate approaches in a simulation of a problem, in a range of scenarios, and analyse the pros and cons of each approach.**

1 Introduction

In this section, you should introduce your work: what are the motivations behind this work? What is the relevant problem that you are investigating? Why is it relevant?

Briefly, introduce the background information required to understand the problem and the concepts that you will develop.

2 Related Work

Research in the field of face recognition started way back in 1960’s [13]. With the emergence of appearance- based most current face recognition techniques emerged in 1980’s and 1990’s. PCA was first applied to face images by Kirby and Sirovich. They showed that using it images can be compressed to their reconstructions with minimum mean squared error between them, i.e. it is an optimal compression scheme [26, 36]. PCA for face recognition was also used by Turk and Pentland [40]. In fact, they popularized its use. In their work compression to compressed subspace (eigenspace) was performed from a database of face images by computing a set of subspace basis vectors (eigenfaces) using PCA. The success of this method leads to the popularization of matching images in the compressed subspace. PCA produces spatially global feature vectors.

Researchers also put in efforts to create techniques that create spatially localized feature vectors, in the hopes that they would implement recognition by parts. ICA is the most general way to generate spatially localized features, which does it by producing statistically independent basis vectors [2]. One other method to generate localize feature vector is Non-negative matrix factorization (NMF) [27]. ICA can also be used to create feature vectors that uniformly distribute data samples in subspace [4, 5]. This is conceptually different from what ICA is believed to do, i.e. rather than producing features vectors which are spatially localized, it produces feature vectors which produce very fine distinction between images in order to spread the samples in subspace. To Keep up with the terminology, we refer to former as architecture I, and the latter as architecture II.

There are certain techniques which are formed by the combination of local linear subspaces. For instance, mix local PCA which is used to compress face data by Kambhatla and Leen [24], and a mixture of factor analyzers produced by Frey [19]. Although they have not yet been applied to face recognition, Tipping and Bishop, and Lee provide an algorithm for optimizing mixture models of PCA and ICA subspaces respectively [39, 28].

Altenative to these algorithms is a supervised learning algorithm knows as Fisher’s linear discriminant analysis (LDA, a.k.a. “fisherfaces”) [38]. Its goal is to produce N-1 basis vector in order to maximize and minimize the intra and inter class distances between N-clsses of N-class problem. Even though LDA differ PCA in terms that one is supervised learning algorithm and the other is unsupervised but when the data is labeled then either of the two can be used and in these cases LDA has been compared to PCA [7, 10, 31, 37]. One common characteristic of both PCA and LDA is that they produce spatially global feature vectors.

3 Problem Definition and Algorithm

This paper tries to implement numerous dimension reduction methods in conjunction with SVM classifier to aid facial recognition process. The dimension reduction methods used are PCA (Principle Component Analysis), ICA (Independent Component Analysis) and NMF (Non-negative Matrix Factoriztaion).

3.1 Principle Component Analysis (PCA)

This was introduced by Karl Pearson in 1901 and is mostly used for EDA and in predictive models. It is the eigenvector based multivariate analysis and is the simplest in its kind. It is basically reducing the dimensionality of the data in a way that it explains the variation (major features) in the best possible way to make it more understandable.

Say there are n\*n dimensional images i.e. n2 dimensions (pixels). When such a multivariate dataset is visualized as set of coordinates in high dimensional data space, then PCA can provide with the lower dimensional picture (eigen faces) of the object when viewed from its most informative viewpoint.

Mathematically by definition PCA is a procedure that converts M correlated into a set of K uncorrelated variables called principal components using orthogonal transformations (where K<=M).

Since PCA selects the principle components which show the direction of data and each proceeding component shows less direction and more noise. Hence first few principle components are enough to represent the original data.

Once we have found the ‘k’ principle components then image in the dataset or incoming new data can be represented as the linear combination (weighted sum) of k components, i.e. W\*K (W is a matrix of weight and k is a matrix of eigenfaces).

When we represent the data this way it reduces the number of values needed to recognize it. This makes the process faster and more error free because this discards the noise in the dataset. It is done by eigenvalue decomposition of a data covariance matrix. Results are usually discussed in form of principle components (how much of each of the k principle components) and loadings (weight).

3.2 Independent Component Analysis (ICA)

PCA is about finding correlation and the way that it does this is by maximizing variance, which in turn gives us the ability to reconstruct. ICA on the other hand tries to maximize independence. This when simply put means that it tries to find linear transformation of feature space into new feature space such that each of the individual new features are mutually independent. The new components produced are perpendicular to each other and there is no information loss between moving from observables to independent components.

There are certain hidden variables (independent components) which are random and mutually independent (value of one doesn’t tell us anything about the value of other). Then there are known observable variables which are given rise to by the linear combination of these hidden variables. The job of unobservable learner in this case is to find the hidden variables given the observables (under the assumption that hidden variables are independent of one another).

The classical example of this is the Blind Source Separation problem.

|  |  |  |
| --- | --- | --- |
|  | PCA | ICA |
| Mutually Orthogonal | Yes | No |
| Mutually Independent | No | Yes |
| Maximal Variance | Yes | No |
| Maximal Mutual Interaction | No | Yes |
| Ordered Features | Yes | No |
| Bag of Features | Yes | Yes |

*Table 1: Differences between PCA and ICA*

PCA is directional, and when it comes to face recognition problem it captures brightness, avg. (Eigen) face. Since PCA is based on global orthogonality hence it captures global features.

ICA on the other hand in face recognition problem captures nose, eye selectors, mouth selectors, hair selectors. Since ICA is based on local search hence it finds parts of face.

3.3 Non-negative Matrix Factorization (NMF)

Basic idea of NMF is to reduce a given matrix into two matrices which are both easier to work with and which when multiplied produce the original matrix. For a matrix V with n\*m dimensions the NMF will try and decompose it into two matrices W and H with n\*r and r\*m dimensions respectively, such that:

W\*H=V and Wij, Hij, Vij >=0

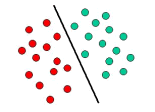
This problem cannot be solved analytically so it is generally solved numerically. NMF is relatively a new way of reducing dimensionality of our data into a linear combination of bases which in turn is necessary for machine learning algorithms for ease of computation. In the image example it;’s very difficult to consider all the pixels each time the image is handled so it is better to reduce the image into few representative pixels.

Due to fact that NMF has this non-negative constraint it can be used to can be used to depict data with non-negative features. It is quite similar to PCA but weights are only allowed to have positive values. This is compatible with the intuitive notion of combining parts to form the whole. Sparse bases and spares weightings are constructed by NMF assuming that there is an underlying structure in the data.

When it comes face recognition NMF forms bases that are parts of the face. The bases in this case are mostly empty and weighting matrix is also sparse, i.e. all parts are not used to form the image. On the contrary PCA has both its matrices densely populated and forms bases of positive and negative pixels and the weight blends them together.

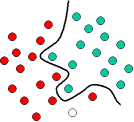
3.4 Support Vector Machine

Support Vector Machines work on the very idea of decision planes which states decision boundaries. Members belonging to different classes are separated by using decision planes. An example is shown below. The objects belong either to class red or they belong to the green one in this example. Both the classes are separated by the separating line such that all objects to the right are red and to the left are green. The new object falling to the right will be labelled red and to the left will be labelled green (or classified as RED should it fall to the left of the separating line).



Fig

The example shown above is of linear classifier i.e., it separates the set of objects into red and green classes with a line. Most of the classification tasks we come across are complex than this one and often more complex decision structures are required to classify them properly i.e., correctly classify new objects (test cases) based on the examples that are available (train cases). This is shown in the image below. AS compared to the last example the complete separation of red and green classes is not possible unless a curve is used. These type of classification tasks which use drawing separating lines are called hyperplanes. Support Vector Machines are particularly suited to handle such tasks.



Fig

Reasons for choosing these classifiers:

* Previously used.
* Gives the option to compare with previous models.
* These use different techniques and so it is reasonable to detect whether different defects are detected by each and whether the prediction consistency is different among the classifiers.

4 Experimental Results

4.1 Methodology

**Evaluation Criteria**: Results are being evaluated based on the accuracy of prediction on training as well as testing data using each of the dimension reduction techniques, namely PCA, ICA and NMF when used in conjunction with the SVM classifier.

**Data**: Labelled dataset named ‘labelled faces in the wild’ is used which consists of 5 users with more than 100 faces each. Test data is of 12 images.

**Modus Operandi:**

* The training data available at hand was already greyscale. So, it didn’t require any tweaking.
* Dimensionality is reduced using PCA, ICA and NMF. Twelve eigenfaces are created using each of the algorithms on same data.

A picture containing different

Description automatically generated A picture containing food

Description automatically generated

PCA ICA

A picture containing different, photo, room

Description automatically generated

NMF

*Fig. Eigenfaces of (a) NMF, (b) PCA and (c) ICA*

As expected, it is quite evident that the eigenfaces generated by PCA captures global features like brightness, avg. eigenfaces and that each next eigen face has less information and more noise. ICA in accordance with theory captured various facial elements in each of its eigenfaces captured the localized features like nose, eye selectors etc. NMF also performed as expected and produced sparsely information rich eigenfaces as per it inherent property due to sparsely positively populated decomposed matrices.

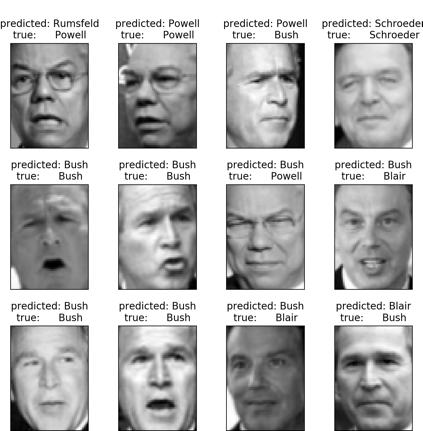
* SVM classifier is fit to this training data using each set of eigenfaces and then accuracy of SVM classifier is compared on the training data and test data based on the accuracy of prediction and time taken for a SVM classifier to be trained on it.

4.2 Results

Dataset of 500 images at hand after being reduced to 12 eigen vectors using PCA, ICA and NMF were compared for their accuracy of prediction on training as well as test data and for the time it took for the SVM classifier to run on them.

|  |  |
| --- | --- |
| Dimensionality Reduction | Accuracy |
| PCA | 0.90 |
| ICA | 0.87 |
| NMF | 0.57 |

*Table 2: Accuracy on training data*

 A picture containing photo, text, showing, newspaper

Description automatically generated

NMF PCA

A picture containing photo, text, showing, newspaper

Description automatically generated

ICA

*Fig. Prediction of test images by (a) NMF, (b) PCA and (c) ICA*

|  |  |
| --- | --- |
| Dimensionality Reduction | Accuracy |
| PCA | 0.83 |
| ICA | 0.87 |
| NMF | 0.50 |

*Table: Accuracy on testing data*

|  |  |
| --- | --- |
| Dimensionality Reduction | Time taken by SVM classifier |
| PCA | 17.068 seconds |
| ICA | 14.173 seconds |
| NMF | 25.999 seconds |

*Table: Time taken to fit SVM classifier to training data*

4.3 Discussion

The results show that performance of SVM classifier on training data when trained with dimensionally reduced data using PCA slightly edges ICA which are followed by NMF whereas when it comes to test data PCA and ICA has the same accuracy which is followed by NMF.

These results show that NMF is not a good enough algorithm to be used for facial recognition. PCA and ICA perform almost equally well and can be used to interchangeably according to results, but this second deduction is not in sync with the data in [3, 17]. Also, it took the most time for the SVM classifier to train on NMF generated eigenfaces.

5 Conclusions

Comparison between PCA, ICA and NMF is complex because of differences in underlying tasks, architectures, ICA algorithms and distance metrics must be considered. This paper explores the performance of PCA, ICA and NMF based on their accuracy of prediction on training and testing data when used in conjunction with SVM classifier and suggests that PCA and ICA performs equally well rather PCA edges ICA slightly which is in contrast with the literature which say that ICA edges PCA slightly when it comes to facial identity recognition.

This paper is limited in its scope because it compares the algorithms only on the basis of prediction accuracy and the time it took for the SVM classifier to be trained on the test data. This can be extended in future by finding the factors which influence the performance of algorithms and the time it takes for the classifiers to be trained on these eigenfaces. Numerous variations of these algorithms which are based on varied architectures (in ICA) and the distance matrices (in PCA) can be considered. These can further be explored and compared based on underlying pattern of the data distribution.

References

1. Baldonado, M., Chang, C.-C.K., Gravano, L., Paepcke, A.: The Stanford Digital Library Metadata Architecture. Int. J. Digit. Libr. 1 (1997) 108–121

2. Bruce, K.B., Cardelli, L., Pierce, B.C.: Comparing Object Encodings. In: Abadi, M., Ito, T. (eds.): Theoretical Aspects of Computer Software. Lecture Notes in Computer Science, Vol. 1281. Springer-Verlag, Berlin Heidelberg New York (1997) 415–438

3. van Leeuwen, J. (ed.): Computer Science Today. Recent Trends and Developments. Lecture Notes in Computer Science, Vol. 1000. Springer-Verlag, Berlin Heidelberg New York (1995)

4. Michalewicz, Z.: Genetic Algorithms + Data Structures = Evolution Programs. 3rd edn. Springer-Verlag, Berlin Heidelberg New York (1996)