Computación y Sistemas. Vol. 7 No. 2 pp. 123 - 129
© 2003, CIC-IPN, ISSN 1405-5546 Impreso en México

Face Recognition Using Unlabeled Data

Reconocimiento de Rostros usando Datos No Etiquetados

Carmen Martínez and **Olac Fuentes**
Instituto Nacional de Astrofísica, Óptica y Electrónica
Luis Enrique Erro # 1
Santa Maria Tonanzintla, Puebla, 72840, México
carmen@ccc.inaoep.mx, fuentes@inaoep.mx

Abstract

Face recognition systems can normally attain good accuracy when they are provided with a large set of training examples. However, when a large training set is not available, their performance is commonly poor. In this work we describe a method for face recognition that achieves good results when only a very small training set is available (one image per person). The method is based on augmenting the original training set with previously unlabeled data (that is, face images for which the identity of the person is not known). Initially, we apply the well-known eigenfaces technique to reduce the dimensionality of the image space, then we perform an iterative process, classifying all the unlabeled data with an ensemble of classifiers built from the current training set, and appending to the training set the previously unlabeled examples that are believed to be correctly classified with a high confidence level, according to the ensemble.

We experimented with ensembles based on the k-nearest neighbors, feed forward artificial neural networks and locally weighted linear regression learning algorithms. Our experimental results show that using unlabeled data improves the accuracy in all cases. The best accuracy, 92.07%, was obtained with locally weighted linear regression using 30 eigenfaces and appending 3 examples of every class in each iteration. In contrast, using only labeled data, an accuracy of only 34.81% was obtained.

Resumen

Los sistemas de reconocimiento de rostros normalmente obtienen buenos resultados cuando tienen disponibles conjuntos de entrenamiento grandes. Sin embargo, cuando no hay un conjunto de entrenamiento grande disponible, su desempeño no es satisfactorio. En este trabajo presentamos un método para reconocimiento de rostros que obtiene buenos resultados cuando solo se tiene disponible un conjunto de entrenamiento pequeño (incluso una sola imagen por persona). El método se basa en expandir el conjunto de entrenamiento original usando datos no etiquetados previamente (esto es, imágenes de rostros con identidad desconocida). Inicialmente, aplicamos la técnica de eigenrostros para reducir la dimensionalidad del espacio de atributos, después realizamos un proceso iterativo, clasificando todos los datos no etiquetados con un ensamble de clasificadores construido a partir del conjunto de entrenamiento actual y agregando al conjunto de entrenamiento los ejemplos que han sido clasificados correctamente con un alto nivel de confianza, de acuerdo al ensamble.

Realizamos experimentos usando ensambles basados en el algoritmo de k vecinos más cercanos, redes neuronales artificiales, y regresión lineal localmente ponderada. Los resultados experimentales demuestran que el uso de datos no etiquetados mejora la clasificación en todos los casos. Los mejores resultados, con un porcentaje de clasificación correcta de 92.07, fueron obtenidos con regresión lineal localmente ponderada usando 30 eigenrostros y agregando 3 ejemplos de cada clase en cada iteración. Como comparación, usando únicamente los datos etiquetados, solo se clasificaron correctamente el 34.81% de los ejemplos.

1. Introduction

Face recognition has many important applications, including security, access control to buildings, identification of criminals and human-computer interfaces, thus, it has been a well-studied problem despite its many inherent difficulties, such as varying illumination, occlusion and pose. Another problem is the fact that faces are complex, multidimensional, and meaningful visual stimuli, thus developing a computational model of face recognition is difficult. The eigenfaces technique

[12] can help us to deal with multidimensionality because it reduces the dimension of the image space to a small set of characteristics called eigenfaces, making the calculations manageable and with minimal information loss.

The main idea of using unlabeled data is to improve classifier accuracy when only a small set of labeled examples is available. Several practical algorithms for using unlabeled data have been proposed. Most of them have been used for text classification; however, unlabeled data can be used in other domains.

In this work we describe a method that uses unlabeled data to improve the accuracy of face recognition. We apply the eigenfaces technique to reduce the dimensionality of the image space and ensemble methods to obtain the classification of unlabeled data. From these unlabeled data, we choose the 3 or 5 examples for each class that are most likely to belong to that class, according to the ensemble. These examples are appended to the training set in order to improve the accuracy, and the process is repeated until there are no more examples to classify. The experiments were performed using k-nearest-neighbor, artificial neural networks and locally weighted linear regression learning.

The paper is organized as follows: The next section presents the learning algorithms; section 3 presents the method to append unlabeled data; section 4 presents experimental results; finally, some conclusions and directions for future work are presented in section 5.

2. Learning Algorithms

In this section we describe the learning algorithms that we used in the experiments, ensemble methods, and the eigen faces technique.

2.1. K-Nearest-Neighbor

K-Nearest-Neighbor (K-NN) belongs to the family of instance-based learning algorithms. These methods simply store the training examples and when a new query instance is presented to be classified, its relationship to the previously stored examples is examined in order to assign a target function value.

This algorithm assumes all instances correspond to points in a *n*-dimensional space $<_n$. The nearest neighbors of an instance are normally defined in terms of the standard Euclidean distance.

One improvement to the k-Nearest-Neighbor algorithm is to weight the contribution of each neighbor according to its distance to the query point, giving larger weight to closer neighbors. A more detailed description of this algorithm can be found in [8]. In this work we use distance-weighted K-NN.

2.2. Artificial Neural Networks

An artificial neural network (ANN) is an information processing system that has certain performance characteristics in common with a biological neural network. Neural networks are composed by a large number of elements called *neurons* and provide practical methods for learning real valued, discrete-valued, and vector-valued target functions. There are several network architectures; the most commonly used are: feed forward networks and recurrent networks.

Given network architecture the next step is the training of the ANN. One learning algorithm very commonly used is backpropagation. This algorithm learns the weights for a multilayer network, given a network with a fixed set of units and interconnections. It uses gradient descent to minimize the squared error between the network output values and the target values. More detailed information can be found in the books [1] and [4]. In this work we apply a feed forward network and the brack propagation algorithm.

2.3. Locally Weighted Linear Regression

Like k-nearest neighbor, locally weighted linear regression belongs to the family of instance-based learning algorithms. This algorithm uses distance-weighted training examples to approximate the target function *f* over a local region around a query point x_q.

In this work we use a linear function around a query point to construct an approximation to the target function. Given a query point x_q, to predict its output parameters we assign to each example in the training set a weight given by the inverse of the distance from the training point to the query point

$$w_i = \frac{1}{|x_q - x_i|} \qquad (1)$$

Let *X* be a matrix compound with the input parameters of the examples in the training set, with addition of a 1 in the last column. Let *Y* be a matrix compound with the output parameters of the examples in the training set. Then the weighted training data are given by

$$Z = WX \qquad (2)$$

and the weighted target function is

$$V = WY \qquad (3)$$

where W is a diagonal matrix with entries w1, ...,wn. Finally, we use the estimator for the target function [10]

$$y_q = x_q^T (Z^T Z)^{-1} Z^T V \qquad (4)$$

2.4. Ensemble Methods

An ensemble of classifiers is a set of classifiers whose individual decisions are combined in some way, normally by voting, to classify new examples. There are two types of ensembles: homogeneous and heterogeneous. In a homogeneous ensemble, the same learning algorithm is implemented by each member of the ensemble, and they are forced to produce non-correlated results using different training sets, however, in heterogeneous ensemble combines different learning algorithms.

Several methods have been proposed for constructing ensembles, such as bagging, boosting, error-correcting output-coding and manipulation of input features [3]. In this work homogeneous ensembles with manipulation of input features are used.

2.5. The Eigenfaces Technique

Principal component analysis (PCA) finds the vectors which best account for the distribution of face images within the entire image space. Each vector is of length N^2, describes an N-by-N image, and is a linear combination of the original face images.

Let the training set of face images be $\Gamma_1, \Gamma_2, \ldots \Gamma_M$

The average of the training set is $\Psi = \frac{1}{M}\sum_{i=1}^{M} \Gamma_i$. Each example of the training set differs from the average by $\Phi_i = \Gamma_i - \Psi$. The set $[\Phi_1 \Phi_2 \ldots \Phi_M]$ is then subject to PCA, which finds a set of M orthonormal vectors U_k and their associated eigenvalues λ_k, which best describe the distribution of the data. The vectors U_k and scalars λ_k are the eigenvectors and eigenvalues, respectively, of the covariance matrix

$$C = \frac{1}{M}\sum_{n=1}^{M} \Phi_n \Phi_n^T \qquad (5)$$

$$C = AA^T \qquad (6)$$

where matrix $A = [\Phi_1 \Phi_2 \ldots \Phi_M]$.

The eigenvectors U_k correspond to the original face images, and these U_k are face-like in appearance, so in [12] they are called "eigenfaces". The matrix C is N^2-by-N^2, and determining the N^2 eigenvectors and eigenvalues, it is an intractable task for typical image sizes. Turk and Pentland use a trick to get the eigenvectors of AA^T from the eigenvectors of A^TA. They solve a much smaller M-by-M matrix problem, and taking linear combinations of the resulting vectors. The eigenvectors of C are obtained in this way. First, the eigenvectors of:

$$D = A^T A \qquad (7)$$

are calculated. The eigenvectors of D are represented by V , and V has dimension $N^2 \cdot M$. Since A has dimension $N^2 \cdot M$ then:

$$E = V^T \cdot A \qquad (8)$$

Thus, E has dimension $M \cdot M$.

With this technique the calculations are greatly reduced from the order of the number of pixels in the images N^2 to the order of the number of images in the training set M, and the calculations become quite manageable.

3. Using Unlabeled Data

Several practical algorithms for using unlabeled data have been proposed. Blum *et al*. [2] present the co-training algorithm that is targeted to learning tasks where each instance can be described using two independent sets of attributes. This algorithm was used to classify web pages.

Nigam *at el*. [9] proposed an algorithm based on the combination of the naive Bayes classifier and the Expectation Maximization (EM) algorithm for text classification. Their experimental results showed that using unlabeled data improves the accuracy of traditional naive Bayes.

McCallum and Nigam [7] presented a method that combines active learning and EM on a pool of unlabeled data. Query-by-Committee [5] is used to actively select examples for labeling, then EM with a naive Bayes model further improves classification accuracy by concurrently estimating probabilistic labels for the remaining unlabeled examples. Their work focussed on the text classification problem. Also, their experimental results show that this method requires only half as many labeled training examples to achieve the same accuracy as either EM or active learning alone.

Solorio and Fuentes [11] proposed an algorithm using three well known learning algorithms, artificial neural networks (ANNs), naive Bayes and C4.5 rule induction to classify several datasets from the UCI repository. Their experimental results show that for the vast majority of the cases, using unlabeled data improves the quality of the predictions made by the algorithms. Solorio [10] describes a method that uses a discriminative approach to select the unlabeled examples that are incorporated in the learning process. The selection criterion is designed to diminish the variance of the predictions made by an ensemble of classifiers. The algorithm is called Ordered Classification (OC) algorithm and can be used in combination with any supervised learning algorithm. In the work the algorithm was combined with Locally Weighted Linear Regression for prediction of stellar atmospheric parameters. Her experimental results show that this method outperforms standard approaches by effectively taking advantage of large unlabeled data sets.

3.1. The Method for Appending Unlabeled Data

We want to use unlabeled data for improving the accuracy of face recognition. The description of the method proposed to append unlabeled data to the training set is the following:

1. One example is chosen randomly from each class with its respective classification from the data set. These examples are the training set and the remaining are considered as the original test set or unlabeled examples.

2. Then an ensemble of five classifiers is used to classify the test set. This ensemble assigns the classification to each example of the test set by voting.

3. After that, for each class, the examples that received the most votes to belong to it are chosen from the test set, always extracting the same number of examples for each class. These examples are appended to the examples of the training set. Now we have a new training set and the remaining examples from the test set are the new test set.

4. Steps two and three are repeated until the examples for each class in the test set are fewer than the number of examples that we are extracting in each step.

5. Finally, the accuracy of the original test set is obtained. The accuracy is obtained comparing the real classification of each example with the classification assigned by the ensemble.

4. Experimental Results

We used the UMIST Face database from the University of Manchester Institute of Science and Technology [6]. From this database we used images of 15 people with 20 images per person. Figure 1 shows one example of one class that exists in the UMIST Face database. The feedforward network was trained with the back propagation learning algorithm during 500 epochs. Three nearest neighbors were used in the Distance-Weighted Nearest Neighbor algorithm and each ensemble was composed with five classifiers.

We performed two different experiments: In the first experiment, one example is chosen randomly from each class with its respective classification from the data set. These examples are the training set and the remaining are the test set. Then an ensemble is used to classify the test set. Table 1 shows the accuracy rates and the best results are obtained with KNN.

In the second experiment, we apply the algorithm described in the previous section, using the same training and test sets as before. From the test set a fixed number of examples for each class are chosen according to the highest voting obtained by the ensemble. Then, these examples are appended to the training set. Now we have a new training set and the remaining examples from the test set are the new test set. This process is repeated until the examples for each class in the test set are less than the number of examples that we are extracting. Table 2 shows the accuracy rates; the best results are obtained with LWR.

The experiments were performed with 30 eigenvectors, containing about 80% of the information in the original data set, or 60 eigenvectors, containing about 90% of the information, and appending 5 and 3 examples of each class to the training set in each iteration. The accuracy rates shown in the tables are the average of five runs. The best classification was obtained using the locally weighted linear regression algorithm with 30 eigenvectors and appending 3 examples for each class.

5. Conclusions and Future Work

We have presented a method for face recognition using unlabeled data. Our experimental results show that using unlabeled data improves accuracy in all cases. The best results were obtained when a small set was aprended to the training set. This is because, by choosing fewer examples in each iteration, we are increasing the probability that each of them is correctly classified by the ensemble. The experiments were performed using three different learning algorithms: k-nearest neighbor, artificial neural networks and locally weighted linear regression. Locally weighted linear regression gives the best results, with an accuracy of 92.07% with 30 eigenvectors and appending 3 examples for each class in each iteration, using a single example of each class as the original training set. In contrast, using only labeled data, the accuracy is 34.81%. Future work includes extending the experiments to other databases and using other learning algorithms.

Figure 1: One example of one class that exist in the UMIST Face database.

KNN		ANN		LWR	
30	60	30	60	30	60
57.43	56.70	18.24	16.56	34.67	45.46

Table 1: Comparison of the accuracy rates of face recognition using only labeled data, with 30 or 60 eigenvectors.

KNN		ANN		LWR	
30	60	30	60	30	60
64.28	70.18	36.14	41.12	86.46	85.12
74.67	74.25	44.56	42.81	**92.07**	90.18

Table 2: Comparison of the accuracy rates of face recognition using unlabeled data. The experiments were performed appending 5 examples (first row) and 3 examples (second row) for each class and 30 or 60 eigenvectors respectively.

6. References

[1] **C. M. Bishop.** Neural Networks for Pattern Recognition. Oxford Universiy Press, Oxford, England, 1996.

[2] **A. Blum and T. Mitchell.** Learning to classify text from labeled and unlabeled documents. Conference on Computational Learning Theory, 1998.

[3] **T. G. Dietterich.** Machine learning research: Four current directions. The AI Magazine, 1997.

[4] **L. Fausett**. Fundamentals of Neural Networks: Architectures, Algorithms and Aplications. Prentice- Hall, 1994.

[5] **Y. Freund, H. Seung, E. Shamir, and N. Tishby.** Selective sampling using the query by committee algorithm. Machine Learning, 1997.

[6] **D. Graham and N. Allinson.** Character-rising Virtual Eigensignatures for Face Recognition. In Face Recognition: From Theory to Applications. Springer-Verlag, 1998.

[7] **A. McCallum and K. Nigam**. Employing EM in pool-based active learning for text classification. In Proceeding of the 15th International Conference on Machine Learning, 1998.

[8] **T. M. Mitchell.** Machine Learning. McGraw-Hill, 1997.

[9] **K. Nigam, A. McCallum, S. Thrun, and T. Mitchell**. Learning to classify text from labeled and unlabeled documents. Machine Learning, 1999.

[10] **T. Solorio.** Using unlabeled data to improve classifier accuracy. Master's thesis, Computer Science Department, Instituto Nacional de Astrofísica, Óptica y Electrónica, 2002.

[11] **T. Solorio and O. Fuentes.** Improving classifier accuracy using unlabeled data. International Conference on Artificial Intelligence and Applications, 2001.

[12] **M. A. Turk and A. P. Pentland.** Face recognition using eigenfaces. Proc. of IEEE Conf. on Computer Vision and Pattern Recognition, pages 586–591, 1991.

Olac Fuentes. *Recibió el grado Doctor en Ciencias de la Computación en la Universidad de Rochester en 1997. Sus áreas de interés incluyen el aprendizaje automático y sus aplicaciones, particularmente en astronomía, visión por computadora y robótica. Ha publicado más de 50 artículos de investigación en las áreas antes mencionadas. Actualmente es Investigador Titular "B" en la Coordinación de Ciencias Computacionales del Instituto Nacional de Astrofísica, Óptica y Electrónica.*

Carmen Martínez.. *Obtuvo el grado de Licenciada en Ciencias de la Computación en la Benemérita Universidad Autónoma de Puebla en 2000 y el de Maestra en Ciencias en Ciencias Computacionales en el Instituto Nacional de Astrofísica Óptica y Electrónica en 2004. Sus intereses de investigación incluyen visión por computadora y aprendizaje automático.*