Behaviour-Based Clustering of Neural Networks Applied to Document Enhancement*

F. Zamora-Martínez[1], S. España-Boquera[2], and M.J. Castro-Bleda[2]

[1] Departamento de Lenguajes y Sistemas Informáticos
Universitat Jaume I, Castellón (Spain)
fzamora@guest.uji.es
[2] Departamento de Sistemas Informáticos y Computación
Universidad Politécnica de Valencia, Valencia (Spain)
{sespana,mcastro}@dsic.upv.es

Abstract. This work proposes an agglomerative hierarchical clustering algorithm where the items to be clustered are supervised-learning classifiers. The measure of similarity to compare classifiers is based on their behaviour. This clustering algorithm has been applied to document enhancement: A set of neural filters is trained with multilayer perceptrons for different types of noise and then clustered into groups to obtain a reduced set of neural clustered filters. In order to automatically determine which clustered filter is the most suitable to clean and enhance a real noisy image, an image classifier is also trained using multilayer perceptrons.

1 Motivation

The field of off-line optical character recognition (OCR) has been a topic of intensive research for many years [1,2,3,4]. One of the first steps in the classical architecture of a text recognizer is preprocessing, where noise reduction and normalization take place. Many systems do not require a binarization step, so the images are maintained in gray-level quality. Document enhancement not only influences the overall performance of OCR systems, but it can also significantly improve document readability for human readers.

In many cases, the noise of document images is heterogeneous, and a technique fitted for one type of noise may not be valid for the overall set of documents. One possible solution to this problem is to use several filters or techniques and to provide a classifier to select the appropriate one.

Neural networks have been used for document enhancement (see [5] for a review of image processing with neural networks). One advantage of neural network filters for image enhancement and denoising is that a different neural filter can be trained for each type of noise.

This work proposes the clustering of neural network filters to avoid having to label training data and to reduce the number of filters needed by the enhancement system. An

* This work has been partially supported by the Spanish Government under contract TIN2006-12767, by the Generalitat Valenciana under contract GVA06/302, and the Universidad Politécnica de Valencia under contract 20070448.

F. Sandoval et al. (Eds.): IWANN 2007, LNCS 4507, pp. 144–151, 2007.
© Springer-Verlag Berlin Heidelberg 2007

agglomerative hierarchical clustering algorithm of supervised classifiers is proposed to do this. The technique has been applied to filter out the background noise from an office (coffee stains and footprints on documents, folded sheets with degraded printed text, etc.).

2 Behaviour-Based Clustering of Supervised Classifiers

2.1 Agglomerative Hierarchical Clustering

Agglomerative hierarchical clustering is considered to be a more convenient approach than other clustering algorithms, mainly because it makes very few assumptions about the data [6,7]. Instead of looking for a single partition (based on finding a local minimum), this clustering constructs a hierarchical structure by iteratively merging clusters according to certain dissimilarity measure, starting from singletons until no further merging is possible (one general cluster). The hierarchical clustering process can be illustrated with a tree that is called dendogram, which shows how the samples are merged and the degree of dissimilarity of each union. The dendogram can be easily broken at a given level to obtain clusters of the desired cardinality or with a specific dissimilarity measure. A general hierarchical clustering algorithm can be informally described as follows:

1. Initialization: M singletons as M clusters.
2. Compute the dissimilarity distances between every pair of clusters.
3. Iterative process:
   (a) Determine the closest pair of clusters i and j.
   (b) Merge the two closest clusters selected in (a) into a new cluster $i + j$.
   (c) Update the dissimilarity distances from the new cluster $i + j$ to all the other clusters.
   (d) If more than one cluster remains, go to step (a).
4. Select the number N of clusters for a given criterion.

2.2 Behaviour-Based Clustering

When the points of the set to be clustered are supervised classifiers, both a dissimilarity distance and the way to merge two classifiers must be defined (see Figure 1):

- The dissimilarity distance between two clusters can be based on the behaviour of the classifiers with respect to a validation dataset. This dissimilarity is defined as the combination of distances between the output of two classifiers for a given validation dataset.
- To merge the closest pair of clusters, a new classifier is trained with the associated training data of the two merged clusters. Another possibility, which has not been explored in this work, might be to make an ensemble of the two classifiers.

We have used Multilayer Perceptrons (MLP) with the same input-output scheme as supervised classifiers. When two clusters are merged, a new MLP is trained with the associated training data of the two merged clusters.

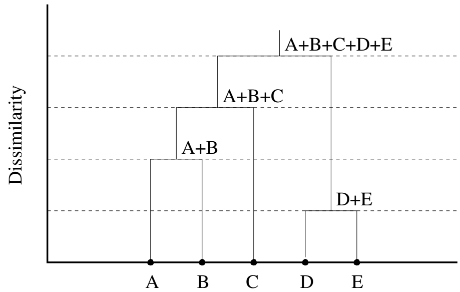


Fig. 1. Behaviour-based clustering. The dissimilarity distance is based on the performance of the filters on a validation dataset. For example, cluster D+E is trained with data used to train the classifiers D and E.

3 An Application of Behaviour-Based Clustering of MLPs to Document Enhancement

Neural networks have been used in previous works for image restoration: the input to the MLP are the pixels in a moving window, and the output is the restored value of the current pixel [8,5,9]. We have also used neural network filters to estimate the gray level of one pixel at a time [10]: the input to the MLP consisted of a square of pixels that was centered at the pixel to be cleaned, and there were four output units to gain resolution (see Figure 2). Given a set of noisy images and their corresponding clean counterparts, a neural network was trained. With the trained network, the entire image was cleaned by scanning all the pixels with the MLP. The MLP, therefore, functions like a nonlinear convolution kernel. The universal approximation property of a MLP guarantees the capability of the neural network to approximate any continuous mapping [11].

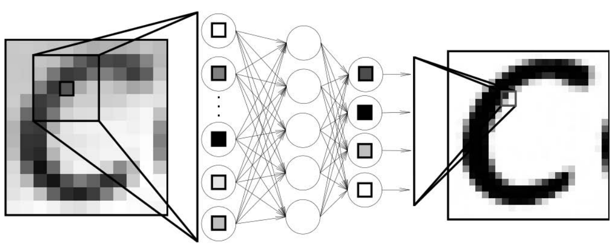


Fig. 2. Architecture of the artificial neural network to enhance and clean images. The entire image is cleaned by scanning it with the neural network.

This approach clearly outperforms other classic spatial filters for reducing or eliminating noise from images (the mean filter, the median filter, and the closing/opening filter [12]) when applied to enhance and clean a homogeneous background noise [10]. However, if the images are degraded by heterogeneous types of noise, two different techniques can be used: 1) train a single neural filter that is capable of cleaning all types of noise (generic filter), or 2) train specific neural filters for each kind of noise. Specific neural filters are expected to be easier to train and to perform better than general filters.

Nevertheless, specific neural filters have some drawbacks: the set of noisy images must be labeled with the kind of noise in order to train each specific filter; the number of different neural filters can be too high; and whenever a new noisy image must be enhanced, it first has to be classified in order to select the filter to apply. To avoid having to label noisy images and to reduce the number of neural filters, a hierarchical clustering of filters has been applied, initializing the process with a neural filter for each image. This method is based on the performance of the filtering process and not on any intrinsic quality of the noise. More precisely, the overall process can be summarized as follows:

1. *Obtain N neural clustered filters and N associated groups of images by hierarchical clustering.*
   (a) Given a set of M unclassified pairs of noisy and clean images, a specific neural filter is trained for every image.
   (b) The iterative agglomerative hierarchical clustering algorithm presented in Section 2 is used to merge the two filters that produce the most similar cleaned images. The merging process consists of training a new filter using the training data that was used to train the two filters.
   (c) The result of this clustering process is a dendogram where, given a desired number of clusters N, $1 \leq N \leq M$, a set of N filters and N associated groups of images is obtained. Note that although there are $\sum_{i=1}^{M} i$ filters in all the clusters, only $2M - 1$ are different.
2. *Obtain a classifier for the N types of neural clustered filters.* Once the number of neural filters is selected, a filter classifier is needed to select the appropriate filter to clean and enhance a new image. To this end, the set of noisy images that is associated to each filter can be used to train a classifier. The proposed solution consists of training a MLP that receives a fixed dimension square of pixels from the image. The output layer has a neuron for each class and uses the softmax activation function to estimate the posterior probability of the cluster class, given a portion of the input image. Since a small portion of the image does not provide very information, the classifier is applied to the entire image to be cleaned one portion at a time. The estimates of all these portions are then averaged, and the most probable class is chosen.
3. *Denoise and enhance a real noisy image.* Finally, when a real noisy image is to be cleaned, a clustered filter must be selected with the filter classifier and then applied to the image.

4 Experimentation

4.1 "Noisy Office": Simulated Noisy Image Dataset

A database of printed documents with typical noises from an office was built for the experiments. We scanned at 300 dpi noisy printed documents obtained by crossing the following parameters: type of noise (folded sheets, wrinkled sheets, coffee stains and footprints), font type (typewriter, serif, roman), yes/no emphasized font, and font size (footnote size, normal, large), obtaining 72 types of noisy documents.

The filtering process is based on MLPs that require a corpus of training pairs "(clean image, noisy image)". It is much easier to obtain a simulated noisy image from a clean one than to clean noisy images or estimate a document degradation model [13]. A dataset of simulated noisy images was obtained by combining noisy-background scanned images with clean text following the scheme shown in Figure 3.

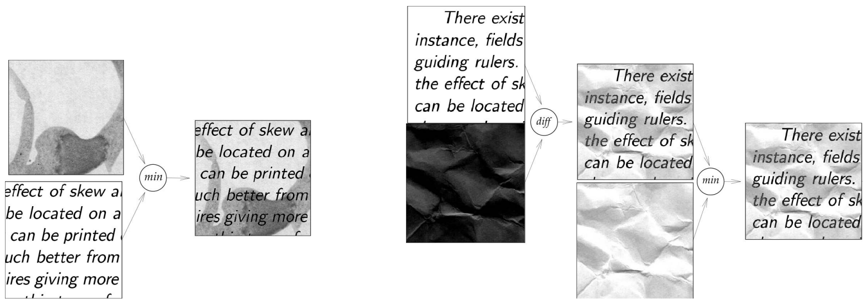


Fig. 3. Simulated noisy process for "coffee-noise" (*left*) and "wrinkle-noise" (*right*). Pixels are codified as gray-levels in the interval [0,1], where 0 means "black" and 1 means "white".

4.2 Agglomerative Hierarchical Clustering of MLPs

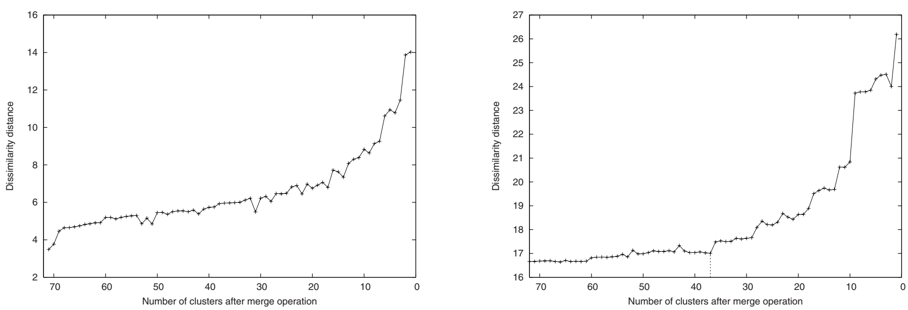
We instantiated the Agglomerative Hierarchical Clustering algorithm presented in Section 2 as follows:

1. *Initialization.* Each initial *singleton* is a trained MLP specific filter for each type of noise (we started with $M = 72$ types of filters).

   The input to the network consisted of a $(2p + 1)$-sided square of pixels that was centered at the pixel to be cleaned (see Figure 2), where the value p is defined as the neighborhood. The activation function of the units of the hidden layer(s) was the sigmoid function, while the activation function of the output unit was the identity function or the logistic function. Better results were obtained with the logistic function. The trained neural networks differed in the number of neighbor pixels (from 4 to 6), the number of hidden layers (one or two hidden layers), and the number of hidden neurons in each layer (from 16 to 64 hidden units). In every case, the online version of the backpropagation learning algorithm with momentum was used [11]. For the same topology, several trainings were performed varying the learning rate and the momentum term. The stopping criterion was the mean squared error of the validation set. The robustness of the methodology has been proved since many of the best results were achieved with many MLPs. We chose an MLP topology with four neighbors and two hidden layers of 64 units for the clustering process.

2. The *dissimilarity distance* is the distance between the images cleaned by two filters. A weighted euclidean distance, where edge pixels have a weight of 1.0 and other pixels have a weight of 0.25, was used.
3. To *merge* the closest pair of filters, a new MLP was trained with the associated training data of the two merged clusters.
4. To select the *number of clusters*, the dissimilarity distance between the closest pair of clusters at each iteration of the clustering algorithm was plotted in order to find the point where an abrupt growth in the dissimilarity distance occurred [6,7] (see Figure 4a). Also, in order to measure the final behaviour of the system, a set of simulated noisy images was cleaned with the true-class neural clustered filters at each iteration. The average weighted euclidean distance between the clean and the cleaned images was plotted (see Figure 4b). As can be observed, the second criterion shows a more abrupt point (at the point with 37 clusters) and, therefore, was used in our experiments.

Each step of the clustering algorithm used a different subset of the simulated noisy dataset.



(a) (b)

Fig. 4. (a) Dissimilarity distance between the closest pair of clusters throughout the clustering process. (b) Dissimilarity distance between a validation image dataset cleaned with the specific filter and the same images cleaned with the true-class neural clustered filter.

4.3 Training a Classifier for the N Types of Neural Clustered Filters

A classifier is needed to select the neural clustered filter that is the most suitable to enhance a given noisy image. A MLP that estimates the posterior probability of the cluster class given a fixed dimension square of pixels (from real noisy images) was trained. The input was fixed to 21×21 pixels. The output layer was composed of M units corresponding to the M neural clustered filters. The classifier was applied one portion at a time to the entire noisy image, and the estimates of all these portions were averaged in order to choose the most probable neural clustered filter. The MLP was trained for 29 cycles, achieving a classification rate of 68.05%. A subset of the real noisy images dataset was used to train and validate this MLP classifier. The rest of the dataset was used to perform the evaluation of the enhancement system.

4.4 Evaluation of the Enhancement System

The proposed approach was objectively evaluated by using the real noisy images. To this end, each real image was cleaned with its specific neural filter (trained with the corresponding type of noise). These "reference" cleaned images were compared with the output of the proposed enhancement system: the euclidean distance between the reference cleaned image and the same image cleaned with the classified-class neural clustered filter was computed. The average distance of this experiment was 37.88.

In order to determine how much the proposed system improves a general neural filter, we trained one MLP with all types of noise, and we calculated the distance from the reference cleaned images to the same images cleaned with this general neural filter. The obtained distance was 62.46, which is much higher than the value obtained by our system.

Finally, to see the effect of the error caused by the filter classifier on the overall system, we simulated the enhancement system with an error-free filter classifier. This simulation reduced the average distance from 37.88 to 28.92.

In order to perform a subjective evaluation of the cleaned test database, we visually inspected a subset of the cleaned images. An example of the performance of the proposed neural method is shown in Figure 5. As can be observed from the examples, the result clearly improved the image quality.

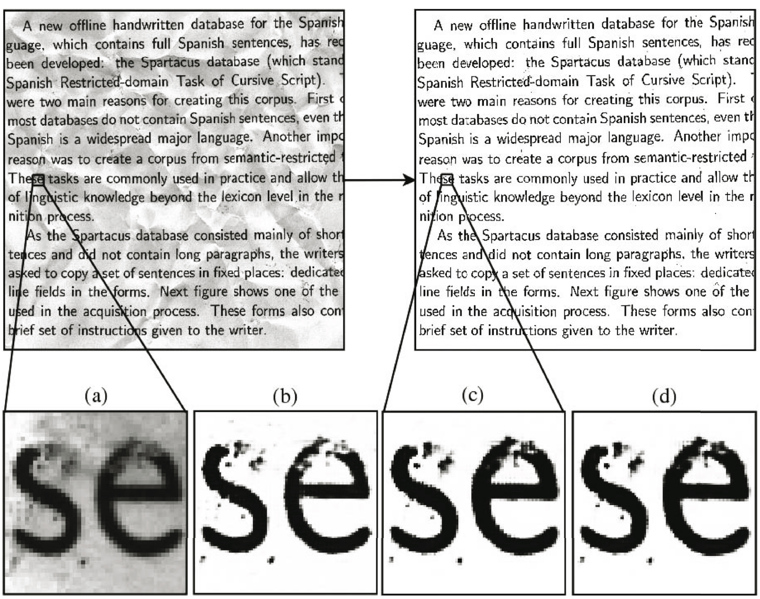


Fig. 5. An example of the enhancement and cleaning process. (a) Original real noisy image. (b) Result of applying a neural filter trained with all types of noise. (c) Result of applying the proposed neural clustered filter. (d) Result of applying the neural filter trained with only that type of noise.

5 Summary and Conclusions

An agglomerative hierarchical clustering of supervised-learning classifiers that uses a measure of similarity among classifiers based on their behaviour on a validation dataset has been proposed. As an application of this clustering procedure, we have designed an enhancement system for document images using neural network filters. Both objective and subjective evaluations of the cleaning method show excellent results in cleaning noisy documents. This method could also be used to clean and restore other types of images, such as noisy backgrounds in scanned documents, folded documents, stained paper of historical documents, vehicle license recognition, etc.

As inmediate future work, we plan to perform a systematical evaluation by studying the differences in the OCR performance for real and enhanced images using our proposed system and other enhancement filters. In order to show that this cleaning process is independent of the features or methods that are used in the recognizers, both a standard HMM-based recognition system developed in our research group and more sophisticated commercial products will be used.

References

1. Plamondon, R., Srihari, S.N.: On-line and off-line handwriting recognition: A comprehensive survey. IEEE Trans. on PAMI 22(1), 63–84 (2000)
2. Bozinovic, R.M., Srihari, S.N.: Off-Line Cursive Script Word Recognition. IEEE Trans. on PAMI 11(1), 68–83 (1989)
3. Bunke, H.: Recognition of Cursive Roman Handwriting – Past, Present and Future. In: Proc. ICDAR, pp. 448–461(2003)
4. Toselli, A.H., Juan, A., González, J., Salvador, I., Vidal, E., Casacuberta, F., Keysers, D., Ney, H.: Integrated Handwriting Recognition and Interpretation using Finite-State Models. IJPRAI 18(4), 519–539 (2004)
5. Egmont-Petersen, M., de Ridder, D., Handels, H.: Image processing with neural networks – a review. Pattern Recognition 35(10), 2279–2301 (2002)
6. Jain, A.K., Murty, M.N., Flynn, P.J.: Data clustering: a review. ACM Comput. Surv. 31(3), 264–323 (1999)
7. Mollineda, R.A., Vidal, E.: A relative approach to hierarchical clustering. In: Pattern Recognition and Applications, vol. 56, pp. 19–28. IOS Press, Amsterdam (2000)
8. Stubberud, P., Kanai, J., Kalluri, V.: Adaptive Image Restoration of Text Images that Contain Touching or Broken Characters. In: Proc. ICDAR. Vol 2, pp. 778–781 (1995)
9. Suzuki, K., Horiba, I., Sugie, N.: Neural Edge Enhancer for Supervised Edge Enhancement from Noisy Images. IEEE Trans. on PAMI 25(12), 1582–1596 (2003)
10. Hidalgo, J.L., España, S., Castro, M.J., Pérez, J.A.: Enhancement and cleaning of handwritten data by using neural networks. In: IbPRIA 2005. LNCS, vol. 3522, pp. 376–383. Springer, Heidelberg (2005)
11. Bishop, C.M.: Neural Networks for Pattern Recognition. Oxford University Press, Oxford (1996)
12. Gonzalez, R., Woods, R.: Digital Image Processing. Addison-Wesley Pub. Co, London (1993)
13. Kanungo, T., Zheng, Q.: Estimating Degradation Model Parameters Using Neighborhood Pattern Distributions: An Optimization Approach. IEEE Trans. on PAMI 26(4), 520–524 (2004)