

# A Holistic Classification System for Check Amounts Based on Neural Networks with Rejection\*

M.J. Castro<sup>1</sup>, W. Díaz<sup>2</sup>, F.J. Ferri<sup>2</sup>, J. Ruiz-Pinales<sup>3</sup>, R. Jaime-Rivas<sup>3</sup>,  
F. Blat<sup>1</sup>, S. España<sup>1</sup>, P. Aibar<sup>4</sup>, S. Grau<sup>1</sup>, and D. Griol<sup>1</sup>

<sup>1</sup> Dep. Sistemas Informáticos y Computación, Univ. Politécnica de Valencia, Spain

<sup>2</sup> Dep. Informática, Univ. de València, Burjassot, Valencia, Spain

<sup>3</sup> FIMEE - Univ. de Guanajuato, Salamanca, Guanajuato, Mexico

<sup>4</sup> Dep. Lenguajes y Sistemas Informáticos, Univ. Jaume I, Castellón, Spain

mcastro@dsic.upv.es

**Abstract.** A holistic classification system for off-line recognition of legal amounts in checks is described in this paper. The binary images obtained from the cursive words are processed following the human visual system, employing a Hough transform method to extract perceptual features. Images are finally coded into a bidimensional feature map representation. Multilayer perceptrons are used to classify these feature maps into one of the 32 classes belonging to the CENPARMI database. To select a final classification system, ROC graphs are used to fix the best threshold values of the classifiers to obtain the best tradeoff between accuracy and misclassification.

## 1 Introduction

The processing of legal amounts in checks involves the recognition of handwritten cursive words, which remains a challenging problem due to the great variability in writing styles and devices. In general, there are two main approaches in handwriting word recognition: analytic and holistic [1]. Analytical approaches assume that words are constructed from characters and characters are produced by strokes, so an explicit or implicit segmentation is needed to perform classification. However, correct segmentation requires that individual characters be first identified, which leads to a paradox besides that it is not always possible to recognize every character. Thus, errors made by most analytic systems are mainly due to segmentation errors of the connected characters. On the other hand, holistic approaches, which avoid the segmentation problem and treat the word as a single, have been limited to small size lexicons. However, there still exists a big difference between the performances of human readers and automatic recognition systems even for small-sized vocabulary tasks.

---

\* Thanks to the Spanish CICYT under contract TIC2002-04103-C03-03 and the Generalitat Valenciana under contract 20040479 for partial funding.

In the rest of the paper, we will describe the handwritten CENPARMI database and its preprocessing, the methods used to classify, present some experimental results, and discuss the obtained system.

## 2 The CENPARMI Database

The off-line handwritten English legal word database of CENPARMI [2] has been created from 2500 handwritten checks written in English. The number of writers is estimated to be close to 800. All checks images were scanned at 300 dpi, transformed into binary images, and segmented into tagged word images. The database is composed of the 32 different legal word classes (from “one” to “twenty”, “thirty”, “forty” and so on, plus the words “and”, “dollars” and “only”). In the recognition experiments, 3069 images of cursive words were used for training and 341 images for testing.

A check may contain different backgrounds consisting of pictures printed in various colors and intensities. In order to isolate the handwritten information, it is necessary to eliminate or reduce the background. The next step is to locate the handwritten items of information (in this case, the legal amount written in words) to be processed, which can greatly vary in different check forms. The items are found through locating the baselines under the items. This task of baseline detection of cursive words is very important, because the detection of ascenders and descenders and other important features (such as the surface of closed loops above, within and below the central zone) requires a reliable detection of the baseline, something that is very difficult to obtain. In a previous work [3], we proposed that the position of these features can be encoded into several feature maps in order to avoid detecting them explicitly and a new method to extract these features in gray-level images is accomplished.

Global features are extracted from each word using a feature extraction approach to the problem of representing shape (for a survey see [1]). The binary images obtained from the cursive words are processed following the human visual system, employing a Hough transform method to extract perceptual features and they are coded into a bidimensional feature map representation. Images are represented in two different encodings as 7 feature maps of size  $12 \times 3$  and  $24 \times 6$ . Details of the preprocessing can be found in [3].

## 3 A Holistic Classification System Based on NNs

The holistic approach to word recognition treats the word as a single entity and attempts to recognize it as whole [1]. This approach avoids the difficult problem of segmenting the word into characters, graphemes or other smaller units. However, since each word is encoded as a different class, it is usually restricted to applications involving small lexicons, as the case for the CENPARMI database is.

**Classification using neural networks.** Multilayer perceptrons (MLPs) are commonly used in handwritten character or word recognition tasks. For this

purpose, the number of output units is defined as the number of classes,  $|\mathcal{C}|$ , and the input layer must hold the input samples. If a softmax activation function [4] is used, the activation level of an output unit is an approximation of the a posteriori probability that the input sample belongs to the corresponding class [5]. Therefore, given an input sample  $\mathbf{x}$ , the trained MLP computes  $g_k(\mathbf{x}, \omega)$  (the  $k$ -th output of the MLP with parameters  $\omega$  given the input sample  $\mathbf{x}$ ) which is an approximation of the a posteriori probability  $\Pr(k|\mathbf{x})$ . Thus, for MLP classifiers we can use the Bayes decision rule:

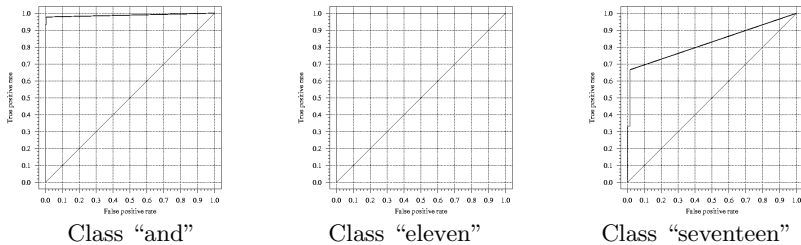
$$k^*(\mathbf{x}) = \underset{k \in \mathcal{C}}{\operatorname{argmax}} \Pr(k|\mathbf{x}) \approx \underset{k \in \mathcal{C}}{\operatorname{argmax}} g_k(\mathbf{x}, \omega). \tag{1}$$

The set of classes  $\mathcal{C}$  are the 32 words of the task and classification follows the “winner-takes-all” (WTA) strategy. Nevertheless, in real applications, such as recognition of checks amounts, misclassifications are considered much more costly than rejections, and it is preferred to introduce some rejection criteria [6]. This can be done through the concept of confidence threshold (below which the classifier abstains) and equation (1) is converted to:

$$k^*(\mathbf{x}) = \begin{cases} \underset{k \in \mathcal{C}}{\operatorname{argmax}} g_k(\mathbf{x}, \omega), & \text{if } g_k(\mathbf{x}, \omega) \geq \mathcal{T}_k, \\ 0, & \text{in other case.} \end{cases} \tag{2}$$

With this rule, the classification is always given by the class with greatest a posteriori probability, but the rejection will depend on thresholds which are different for each class. We will refer to classification rules from equations (1) and (2) as “WTA Rule” and “WTA +  $\mathcal{T}$  Rule”, respectively.

**Training the multilayer perceptrons.** The corpus was divided into a training set, a validation set and a test set. The trained neural networks differed in the number of hidden layers (one or two hidden layers) and the number of hidden neurons in each layer (from 10 to 200 hidden units). In every case, the online version of the backpropagation learning algorithm with momentum was used [7]. For the same topology, several trainings were performed varying the learning rate, the momentum term and using different initializations of the weights. The stopping criteria was the classification error in the validation set.



**Fig. 1.** ROC curves on the threshold values for classification for three classes

**Fixing threshold parameters with ROC graphs.** Receiver Operating Characteristics (ROC) graphs [8] are a useful technique for visualizing, organizing and selecting classifiers based on their performance. In this work, we use ROC graphs in order to select the threshold values for “WTA +  $\mathcal{T}$  Rule” classifiers to obtain the best tradeoff between accuracy and misclassification: we performed classification with the optimal configuration of MLPs on the patterns of the validation set and analyzed the ROC curves. Finally, we measure network performance on the test set for the best configuration of MLP and the best threshold for each class. Figure 1 shows some examples of the ROC curves for classification with feature maps of size  $24 \times 6$  for three classes.

## 4 Experiments

In this section, the best result obtained with each approach is given. Performances with MLPs for both tested parametrizations with and without using thresholds for rejection are shown in Table 1. The recognition rate of each legal word class for the best classifier (MLP with two hidden layers of 150 and 100 units for the  $24 \times 6$ -sized feature maps with rejection) is 86.26% (with no rejection) and 69.88% with rejection. It is worth noting that there are very few samples of each word in the test set (for instance, only 3 and 5 instances of the words “seventeen” and “thirteen”, respectively).

**Table 1.** Classification measures (W=Wrong, R=Right, U=Unknown)

Experiment		Classification	Test		
Classifier	Feature maps	rule	%W	%R	%U
MLP 80-60	$12 \times 3$	WTA	18.77	81.23	0.00
		WTA + $\mathcal{T}$	0.88	51.32	47.80
MLP 150-100	$24 \times 6$	WTA	13.74	86.26	0.00
		WTA + $\mathcal{T}$	0.00	69.88	30.12

## 5 Summary and Conclusions

A holistic neural network classification system with rejection has been proposed for handwriting images recognition. Each image is represented as a whole with feature maps of different sizes. As we are interested in cautious classifiers, ROC graphs are employed to obtain the best tradeoff between accuracy and misclassification. The best performance, close to 70% of classification with rejection of 30% and no errors, is achieved with an MLP for a representation of seven feature maps of size  $24 \times 6$ . This result shows the effectiveness of the proposed approach to avoid misclassifications and still achieving a competitive performance.

The task of legal amount recognition from checks is a limited domain application, so a model of language can greatly help in recognition. For future work we are also considering to use multiple classifier systems, that is, combining

the expertise of multiple classifier methods (through voting, linear confidence accumulation or weighted schemes).

## References

1. Madhvanath, S., Govindaraju, V.: The role of holistic paradigms in handwritten word recognition. *IEEE Trans. on PAMI* **23** (2001) 149–164
2. Suen, C.Y., et al.: Computer recognition of unconstrained handwritten numerals. Special Issue of *Proc IEEE* **7** (1992) 1162–1180
3. Ruiz-Pinales, J., Lecolinet, E.: Cursive handwriting recognition using the Hough transform and a neural network. In: *ICPR*. (2002) 231–234
4. Bridle, J.S.: Training stochastic model recognition algorithms as networks can lead to maximum mutual information estimation of parameters. In: *Advances in Neural Information Processing Systems 3*, Morgan Kaufmann (1990)
5. Duda, R.O., Hart, P.E., Stork, D.G.: *Pattern Classification*. Second edn. John Wiley and Sons, New York, NY, USA (2001)
6. Ferri, C., Hernández-Orallo, J.: Cautious classifiers. In: *ROCAI*. (2004) 27–36
7. Rumelhart, D.E., Hinton, G.E., Williams, R.J.: Learning internal representations by error propagation. *MIT Press* (1986) 319–362
8. Fawcett, T.: *ROC Graphs: Notes and Practical Considerations for Researchers*. HP Technical Report HPL-2003-4, HP Labs (2003) Revised March, 2004.