The Influence of Spatial Pixel Relations on the Image Recognition Performance

Robert P.W. Duin, Pattern Recognition Group
Faculty of Applied Physics, Delft University of Technology
duin@ph.tn.tudelft.nl

Keywords: neighborhood analysis, characters, nearest neighbors, small sample size classification

Abstract

Pixels are spatially related. In the direct use of pixels as features for a traditional classifier this information is lost. It is therefor to be expected that classifiers fed with such features perform worse (learn slower) than classifiers that can make use of the spatial pixel relations in some way or another. It is attempted to quantify this fact for two classifiers in a character recognition problem.

Introduction

It is the intuition of many researchers in the field of image analysis that the use of the spatial relationship between pixels is crucial for an optimal result. For that reason it is considered to be a bad idea to use pixels directly as features for image recognition. Classifiers that use such features are expected to perform worse than classifiers that do use the spatial connectivity of the pixels. In this research a possible setup is investigated by which the influence of this relationship is quantified in terms of the learning power of a classifier, i.e. the speed by which its error decreases with an increasing size of the training set.

As an example we will consider the recognition of characters. Usually this is based on images varying from 16x16 pixels to as large as 128x128 pixels. The possible set of characters has thereby a finite size. In principle a table of all possible characters would solve the problem. Even for binary (0/1) 16x16 images this would need a table of 2^{256} entries which is entirely unrealistic. This theoretical possibility, however, proves that any recognition system that asymptotically converges to such a table thereby also converges to a zero-error. The 1-nearest neighbor classifier (1-NNC) that uses the pixels as features may be considered as such a system. An infinite training set covers all possibilities and has a zero error if no two characters are exactly identical with in the image resolution. We demand that any system that makes use of the spatial pixel relations should preserve the asymptotic zero error, otherwise it is outperformed at the end. So it should be better by faster learning. In the next section the '*learning power*' of a classifier is introduced that defines the speed of learning with respect to increasing training sizes for asymptotic zero error problems.

One simple way to relate spatially neighboring pixels in binary images is to blur these images [1], replacing the 0/1 features by continuous values. A feature value is herewith related to the spatial distance to the neighboring pixels that have originally the value '1'. If a linear filter is used for blurring then this operation corresponds with a linear transformation of the feature space (we will use a 'wrap around' image boundary condition). As this transformation is reversible it does not influence the class overlap. Separated classes remain separated. The above demand is in this way fulfilled. In the below described experiments it is investigated whether the use of blurred characters improves the learning power of the 1-NNC.

The blurring operation increases the correlation between neighboring pixels. As a result feature reduction might improve the learning power as well, due to lower dimensional feature spaces. This is investigated in the experiments too. Beside the 1-NN we used as a reference the Small Sample Size Classifier (SSSC) as this classifier may yield good results for small training sets in high dimensional feature spaces [2].

The Learning Power

The error ε of a classifier can be approximated as a function of the size of the training set *m* by:

$$\varepsilon = am^{-V} \tag{1}$$

in which *a* is some arbitrary constant and *V* is the classification power. V can be estimated from a set of observed errors for different values of *m* by:

$$V = -\frac{\log(\varepsilon) - \log(a)}{\log(m)} \tag{2}$$

using a linear approximation in the double logarithmic domain. See fig. 1 for an example of the relation between ε and *m*. The resulting values of V for several classifiers are shown in fig. 3 and fig. 4. Note that each point in these figures corresponds with the estimated angle of a line in fig. 1.

Experiments

We used a simple example for this initial research based on the handprinted characters '3' and '8' from the NIST-3 database [3]. This set has been converted as described in [4] to 16x16 binary images, see fig. 2 for some examples. In this procedure the character were normalized on size and line width.

It is shown in [4] and [5] that the 1-NNC on this dataset can perform over 98% using 1000 objects per class. In the present study we used a more restricted set of 200 examples for each class. Randomly chosen subsets of *m* samples per class were used for training, $1 <= m <= 100$. Each time the remaining samples were used for testing. This procedure was repeated 50 times and the resulting errors were averaged, see fig. 1 for the 1-NNC results on the unblurred characters. In this double logaritmic plot the error curves show almost straight lines. This is in agreement with the previous section.

Feature reduction was obtained by equidistant pixel sampling. Thereby feature sizes have been used of 256, 128, 64 and 32 features. Note that smaller feature sizes yield worse results for all sample sizes.

As was explained in the introduction, the experiments were repeated for uniformly blurred versions using filter sizes of 3x3, 5x5, up to 15x15. In this way pixel values represent the distance to the character boundary thereby introducing a spatial pixel relation in the recognition system. The resulting values for the learning power of the 1-NNC are presented in fig. 3. Finally all experiments were repeated for the SSSC, see fig. 4. This classifier finds that linear discriminant plane between the classes that maximizes the margin with the two class boundaries. Where the NNC can be characterized as a locally sensitive method, the SSSC has a somewhat more globally sensitive character. For both classifiers no assumptions are made on the class densities. The NNC, however, is heavily dependent on the metric. If the metric does not agree with the class properties, this classifier is able to overcome this using a large training set.

The resulting classification powers for the two classifiers (fig. 3 and 4) show an almost complementary character. The SSSC preserves its power for large filter sizes while the NNC deteriorates. This can be explained from the high correlations that are caused by large filters. There are many samples needed for the NNC to restore the proper metric for this space.

Discussion and conclusions

The main result of our experiments is that the NNC is most powerful for small filtersizes. So it is good for the NNC to have some pixel neighborhood information. If this is pushed too far, the bad metric of the highly correlated features deteriorates the power of the NNC. The SSSC appears to be almost independent of the feature transformations. It has been observed in other studies that the NNC is good for images, Michie [6], and for character recognition, De Ridder [4].

From fig. 1 it can be observed that in this problem and for the studied feature sizes, the NNC does not suffer from the 'curse of dimensionality': the peaking phenomenon. More features yield better results even for very small sample sizes. It can be concluded from this that the classification problem under study is a very simple one. This might be caused by the normalizing preprocessing of the data.

In conclusion it can be stated that the phenomenon we intended to study, the influence of the spatial pixel relation on the recognition performance, has been shown, but only weakly. In order to see this better we need a classifier that is not disturbed by highly correlated features. Besides, this is only of interest if this classifier preforms better than the SSSC which seems to be to some extend insensitive for transformations that emphasize spatial pixel relations. Finally, a more complicated classification problem might be needed as the problem at hand already shows good performances without using the pixel relations.

Acknowledgment

Arnold Smeulders is cordially acknowledged for raising the author's interest into the problem and for drawing the attention to reference [1]. The conversion of the NIST data to the 16x16 binary format was done by Dick de Ridder [4].

References

[1] L. Guyon.V. Vapnik, B. Boser, L. Bottou, and S.A. Solla, *Capacity Control in Linear Classifiers for Pattern Recognition*, , Proceedings 11th IAPR International Conference on Pattern Recognition, Volume II, Conference B: Pattern Recognition Methodology and Systems (ICPR11, The Hague, The Netherlands, August 30 - September 3, 1992), IEEE Computer Society Press, Los Alamitos, California, 1992, 385-388.

[2] R.P.W. Duin, *Small sample size generalization*, SCIA'95, Proc. 9th Scandinavian Conference on Image Analysis, Volume 2 (Uppsala, Sweden, June 6-9, 1995), 1995, pp. 957-964.

[3] C.L. Wilson, M.D. Marris, *Handprinted character database 2*, april 1990. National Institute of Standards and Technology; Advanced Systems division.

[4] D. de Ridder, *Shared weights neural networks in image analysis*, Master thesis, Delft University of Technology, February 1996.

[5] D. de Ridder, A. Hoekstra, R.P.W. Duin, *Shared weights neural networks in image analysis*, ASCI '95, Proc. Second Annual Conference of the Advanced School for Computing and Imaging, June 1996.

[6] D. Michie, D.J. Spiegelhalter, C.C. Taylor, eds., *Machine learning, neural and statistical classification*, Ellis Horwood, New York, 1994.

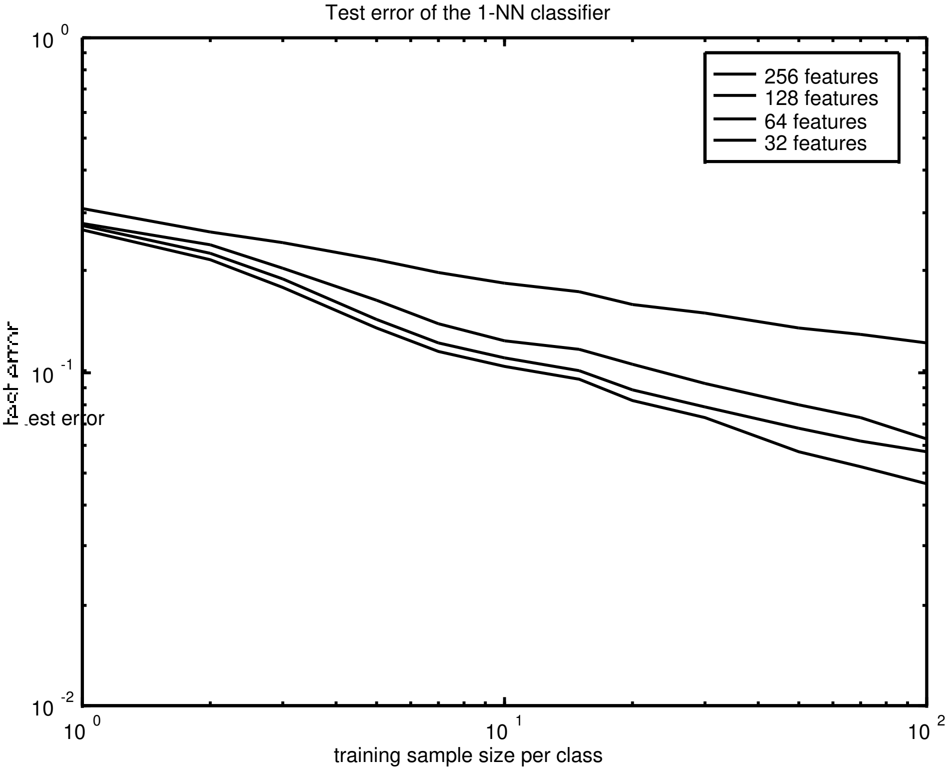

Fig. 1. Test error as function of the training size for the 1- Nearest Neighbor Classifier.

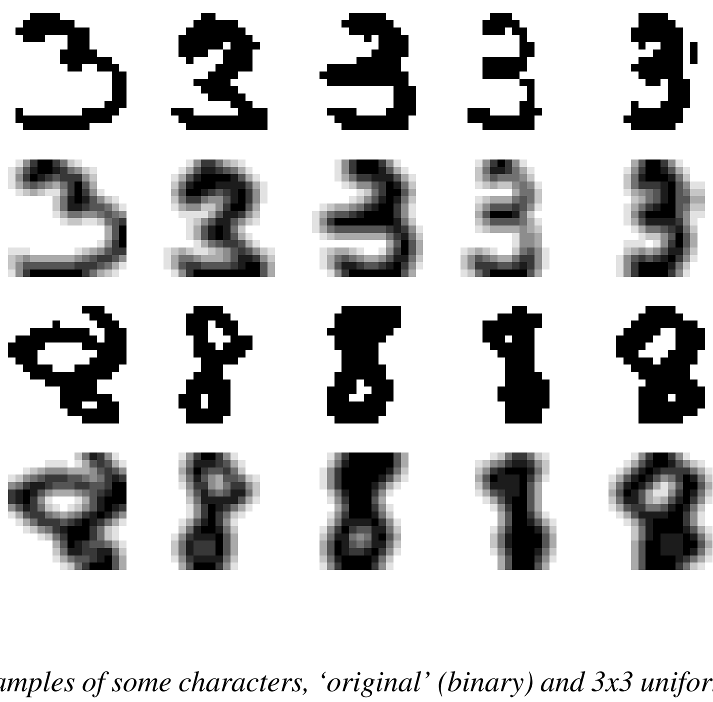

Fig. 2. Examples of some characters, 'original' (binary) and 3x3 uniform blurred. The 'original' characters are normalized on shape and line width before.

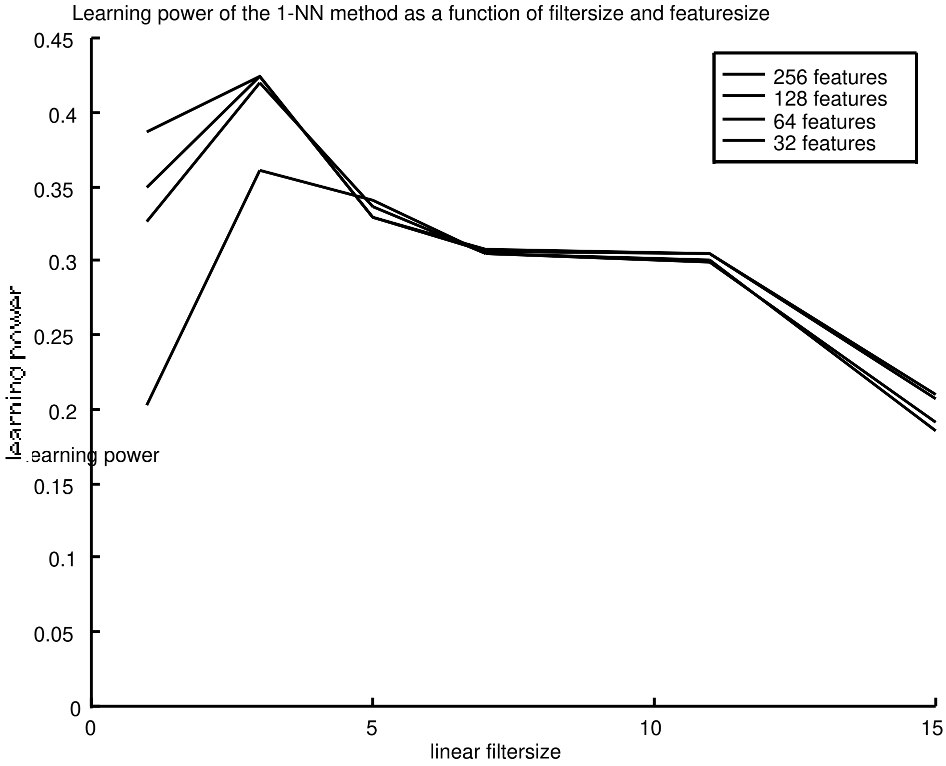

Fig. 3. Learning power V of the 1-Nearest Neighbor Classifier on the given data as a function of the blurring filtersize, repeated for some feature subsets.

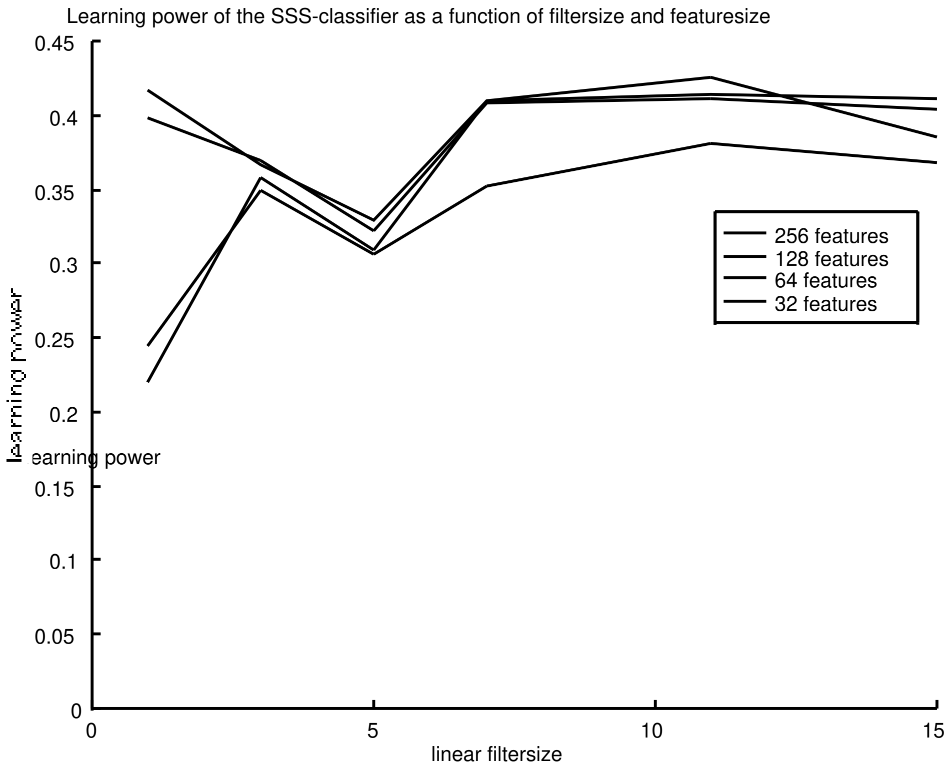

Fig. 4. Learning power V of the Small Sample Size Classifier on the given data as a function of the blurring filtersize, repeated for some feature subsets.