Experiments with the n-tuple Method of Pattern Recognition

J. R. ULLMANN

***Abstract*—The n-tuple method of pattern recognition has been simulated on a somewhat larger and more comprehensive scale than previously reported. The nonweighted version has been found to work better than the maximum likelihood weighted version, and to achieve about 93 percent successful recognition of unconstrained hand-printed numerals, but at the cost of about 42 million bits of storage.**

***Index Terms*—Computer simulation, handprinted numerals, nonlinear decision-making, n-tuple method, pattern recognition, statistical approximation.**

Introduction

The n-tuple method of pattern recognition was introduced by Bledsoe and Browning [3], and has been described more formally by Steck [6]. Simulation of the maximum likelihood version has been reported by Bledsoe and Bisson [2], and Chow [4] and Roy and Sherman [7] have pointed out that this is a method of approximating a higher order distribution by a product of lower order distributions. So one might expect that the larger the sample size (the value of n), the better the approximation, and the lower the error rate of the system. A purpose of the present work was to investigate this suggestion, which seemed sufficiently fundamental to deserve experimental attention.

Experimental Data

Six-hundred and fifty different subjects each wrote a set of ten numerals in ink or ball point in preprinted boxes on standard paper. They were told: "Your best writing is not required but use the same standard that you would use if you were writing the post town part of an address."

The characters were binarized into a 22-column by 30-row array by means of a flying spot scanner which did not have automatic centering or size-normalizing facilities. So the characters were centered and roughly size-normalized by a human operator. But there was no normalization of height-to-width ratio or orientation, and the characters were used directly, without line width standardization, noise reduction, or any other preprocessing.

Recognition Experiments

A. Performance Versus n

n is the number of pattern element locations constituting an n-tuple. The purpose of this experiment was to find how performance at optimal n varied with the size of the "training" or "design" set.

Using a training set of ten characters per class, Bledsoe and Browning's original nonweighted method [3] was applied to the ten numerals, using in turn various values of n. The results form the bottom curve in

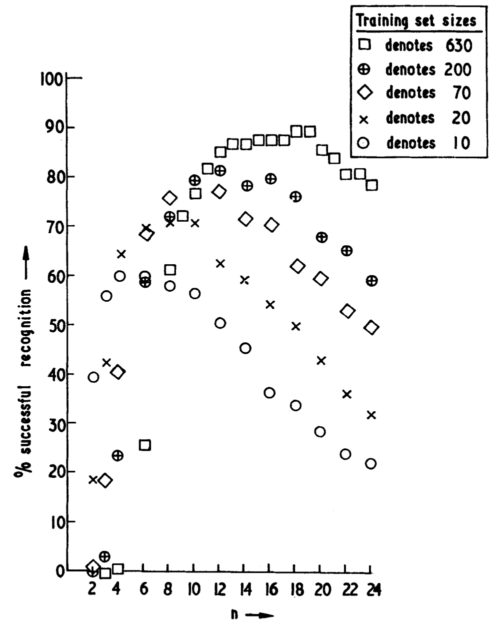

Fig. 1. Recognition performance versus n for nonweighted method.

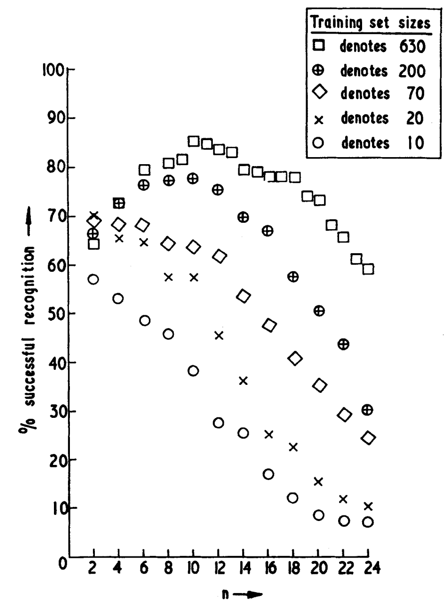

Fig. 2. Recognition performance versus n for maximum likelihood weighted method.

Fig. 1. The vertical axes of Figs. 1, 2, and 3 are the percentages of characters correctly recognized, that is, neither substituted nor rejected. The other curves in Fig. 1 are the results of this experiment repeated, respectively, with 20, 70, 200, and 630 patterns per class in the training set instead of 10.

Fig. 2 gives results of simulation of the maximum likelihood weighted n-tuple method, again with 10, 20, 70, 200, and 630 patterns per class in the training set, and various values of n. The actual n-tuples, training set, and test set used in any run with the maximum likelihood method were the same as those used in the corresponding run with the primitive nonweighted method. For example, with a training set of 200 per class and $n=12$, the same training set, test set, and n-tuples were used with the maximum likelihood weighted and

Manuscript received May 8, 1969; revised July 2, 1969.
The author is with the Division of Computer Science, National Physical Laboratory, Teddington, Middlesex, England.

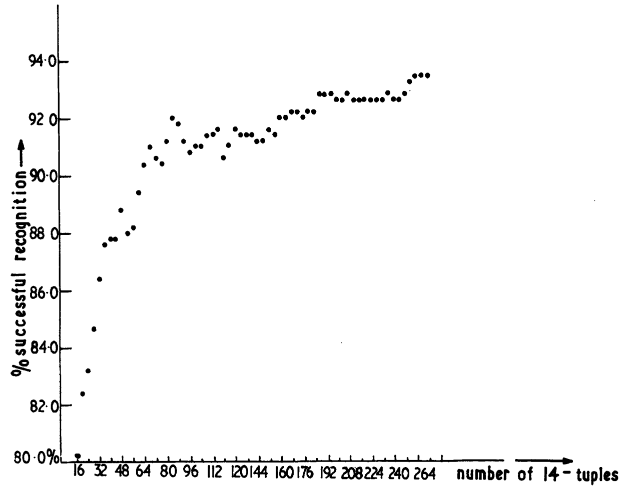

Fig. 3. Recognition performance versus number of 14-tuples for nonweighted method.

nonweighted methods. (Eight bits per weight were used in the simulation of the maximum likelihood method.)

All these experiments used 40 randomly chosen n-tuples. The test set always consisted of 20 patterns per class, the patterns used for testing being different to those used for training. The test set size was sufficient to reveal the flattening out trend of the locus of the maxima, in Figs. 1 and 2, as the size of the training set was increased. For Bayesian pattern classifiers, an analogous effect has been predicted by Hughes [5], and it is interesting that the family of curves in Fig. 1 fit his quantitative predictions better than those in Fig. 2.

B. Performance of Nonweighted Method Versus Number of n-Tuples

Fig. 3 is a plot of performance versus number of random n-tuples, with $n=14$ and training and testing sets of 600 and 50 per class, respectively. This experiment, like all the previous ones, was carried out with the ten numerals, and again the patterns used for testing were different from those used for training. Fig. 3 was obtained with Bledsoe and Browning's original nonweighted n-tuple method. (We could not afford to run the maximum likelihood version on this scale.)

Some Experimental Statistics of n-Tuple States

On a given n-tuple, different states generally occur different numbers of times in the training set of a given class. For each (n-tuple, class) pair, we determined these numbers of occurrences, and obtained from them a histogram showing the numbers of states occurring once, twice, and so on. The numbers of states occurring once in all the (n-tuple, class) histograms were added together to yield the leftmost point in Fig. 4. The numbers of states occurring twice were added together to yield the second-from-left point in Fig. 4, and so on.

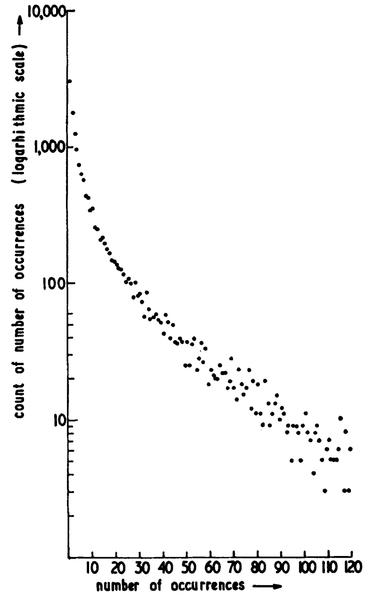

Fig. 4. Histogram of counts of numbers of occurrences of 6-tuple states.

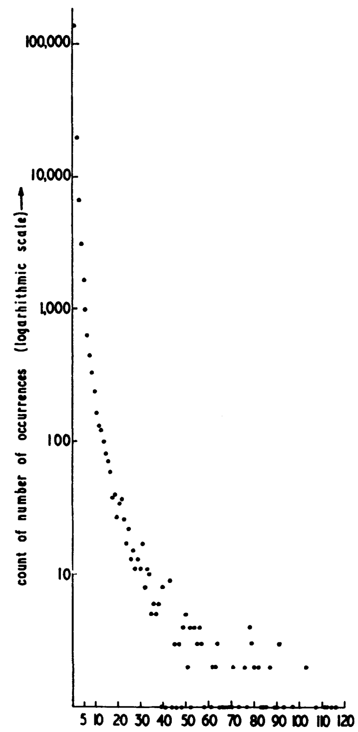

Fig. 5. Histogram of counts of numbers of occurrences of 18-tuple states.

Fig. 4 was obtained from the ten numerals with 40 6-tuples, using a training set of 650 per class. The vertical scale is logarithmic, to fit it on the page, and the horizontal scale is truncated at 120 occurrences. For instance, the leftmost point in Fig. 4 signifies that the number of 6-tuple states which occur only once is 3002 (out of a possible 25 600). Fig. 5 differs from Fig. 4

only in that it was obtained with $n=18$ instead of $n=6$.

For each n-tuple we obtained the number of different states occurring in the training set of each class at least once. This number was averaged over all (n-tuple, class) pairs, working with 40 random n-tuples, 630 patterns per class in the training set, and various values of n. Fig. 6 is a plot of the average number of different states per n-tuple versus n, which may be of interest from the point of view of associative memory, as well as giving together with Figs. 4 and 5 a quantitative indication of the diversity and quality of our experimental data.

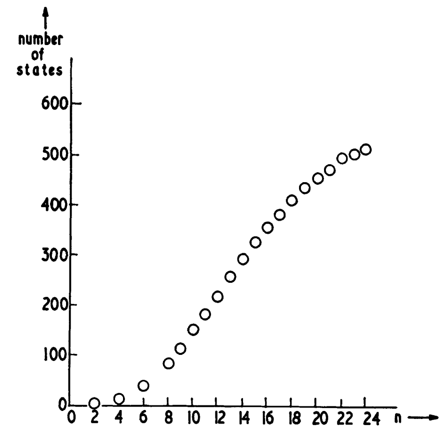

Fig. 6. Average number of difference states per n-tuple per class versus n.

Concluding Remarks

A. Weakness of Maximum Likelihood Method

In a run with 40 n-tuples and a training set of 650 patterns per class, altogether $40\times10\times650=260\ 000$ n-tuple states occurred during training. Of these more than half (actually 134 369, c.f. Fig. 5) only occurred once per n-tuple per class, when $n=18$. One occurrence is of course insufficient to allow an adequate estimation of the class membership conditional probability of an n-tuple state, and this indicates and exemplifies the cause of weakness of the maximum likelihood n-tuple method.

B. Bits per Weight when $n=1$

Extrapolating from Fig. 2 to the case where $n=1$, we see that beyond a certain size, any further increase in the size of the training set will cause a deterioration of performance. It may be of interest to relate this to a previous finding, reported elsewhere in an elementary paper [8], that with $n=1$ and a fixed sized training set, beyond a certain number, a further increase in the number of bits per weight caused a deterioration of performance.

C. Pattern Probability Distributions

Figs. 4 and 5 suggest a trend such that, for larger n, the occurrence more than once of any state of an n-tuple in any class would become rare. This confirms the impression that a machine recognizing unconstrained hand-printed characters must continually deal with new characters which have never occurred before. Instead of having to recognize a limited repertoire of patterns, each of them occurring quite frequently, the machine has to recognize a virtually unlimited repertoire of patterns, most of which occur very infrequently. This is why, as Nagy [1] has said, " . . . In practice the training set is always too small." The training set is never sufficiently representative of the virtually unlimited repertoire of patterns to be recognized. This underlies the diminishing returns trend of the locus of the maxima in Fig. 1, which suggests that however big the training set, there will always in practice be a pool of rarely occurring patterns which the system will fail to recognize correctly.

This is of course a drastic weakness of the n-tuple method. But one should also consider to what extent the same weakness is manifest in all other known recognition systems, even those which are intuitively designed.

Acknowledgment

The experimental data was made available by the authorities of the Post Office Research Station, Dollis Hill, London, N.W.2. The work described above was carried out at the National Physical Laboratory.

References

[1] D. H. Andrews, A. J. Atrubin, and K. C. Hu, "The IBM 1975 optical page reader. Part III: Recognition logic development," *IBM J. Res. and Develop.*, vol. 12, pp. 364–371, September 1968.
[2] W. W. Bledsoe and C. L. Bisson, "Improved memory matrices for the n-tuple pattern recognition method," *IRE Trans. Electronic Computers* (Correspondence), vol. EC-11, pp. 414–415, June 1962.
[3] W. W. Bledsoe and I. Browning, "Pattern recognition and reading by machine," *1959 Proc. Eastern Joint Computer Conf.*, Boston, Mass., December 1–3, 1959, pp. 225–232.
[4] C. K. Chow, "A class of nonlinear recognition procedures," *IEEE Trans. Systems Science and Cybernetics*, vol. SSC-2, pp. 101–109, December 1966.
[5] G. F. Hughes, "On the mean accuracy of statistical pattern recognizers," *IEEE Trans. Information Theory*, vol. IT-14, pp. 55–63, January 1968.
[6] G. P. Steck, "Stochastic model for the Browning–Bledsoe pattern recognition scheme," *IRE Trans. Electronic Computers*, vol. EC-11, pp. 274–282, April 1962.
[7] R. J. Roy and J. Sherman, "Two viewpoints on the k-tuple pattern recognition," *IEEE Trans. Systems Science and Cybernetics*, vol. SSC-3, pp. 117–120, November 1967.
[8] J. R. Ullmann and P. A. Kidd, "Recognition experiments with typewritten numerals from envelopes in the mail," *Pattern Recognition*, vol. 1, pp. 273–289, July 1969.