tion techniques as and when introduced may thus be incorporated in the system and tested.

ACKNOWLEDGMENT

We are grateful to Prof. R. Narasimhan, Director, NCSDCT for providing us with the opportunity to experiment with computer animation. We are thankful to R. K. Joshi, a professional calligrapher, for the design of the character font and V. Rammohan, a professional animator, for the animation idea of the example presented. We thank P. Kumaran for his meticulous typing of this work.

REFERENCES

[1] W. M. Newman and R. F. Sproull, "An approach to graphics system design," in *Proc. IEEE*, vol. 62, pp. 471–483, Apr. 1974.
[2] J. Halas, Ed., *Computer Animation*. London and New York: Focal, 1974.
[3] D. D. Weiner and S. E. Anderson, "A language for educational pictures," in *Computer Animation*, J. Halas, Ed. London and New York: Focal, 1974, pp. 129–131.
[4] S. P. Mudur and J. H. Singh, "Computer animation: Some experiments," Tech. rep., NCSDCT, TIFR, Bombay, May 1977, to be published.
[5] G. Bracchi and D. Ferrari, "A graphic language for describing and manipulating two-dimensional patterns," in *Advanced Computer Graphics*, R. Parslow and E. Green, Eds. London and New York: Plenum, 1971, pp. 263–278.
[6] R. M. Baecker, "A conversational extensible system for the animation of shaded images," *Computer Graphics* (quarterly report of SIGGRAPH-ACM), vol. 10, pp. 32–39, Summer 1976.
[7] ——, "Interactive computer GENESYS mediated animation," in *Computer Animation*, J. Halas, Ed. London and New York: Focal, 1974, pp. 97–115.
[8] G. Courtieux *et al.*, "An interpreter for the interactive generation and animation of two dimensional pictures," in *Graphic Languages*, F. Nake and A. Rosenfeld, Eds. Amsterdam: North-Holland, 1972.
[9] N. Burtnyk and M. Wein, "Computer generated key-frame animation," in *Computer Animation*, J. Halas, Ed. London and New York: Focal, 1974, pp. 39–44.
[10] N. Burtnyk and M. Wein, "Interactive skeleton techniques for enhancing motion dynamics in key-frame animation," *Comm. Ass. Comput. Mach.*, vol. 19, pp. 564–569, Oct. 1976.

A Note on Distance-Weighted k-Nearest Neighbor Rules

T. BAILEY AND A. K. JAIN

***Abstract*—A distance-weighted k-nearest neighbor rule is not necessarily better than the majority rule for small sample size if ties among classes are broken in a judicious manner. The behavior of several tie-breaking procedures is demonstrated using the bivariate distributions for three classes used by Dudani. In the infinite sample case, the majority rule is the best among all distance-weighted rules.**

I. INTRODUCTION

In a recent paper, Dudani [1] introduced the concept of a distance-weighted k-nearest neighbor rule. This rule differs from a majority k-nearest neighbor rule in that it assigns a weight to each of the k nearest neighbors. A pattern with unknown classification is then assigned to that class for which the weights assigned to its training samples sum to the greatest value. Dudani proposed that when the number of training samples is small or of moderate size, then the distance-weighted rule will yield a smaller probability of error than the majority rule. He showed the admissibility of the distance-weighted k-nearest neighbor rule by demonstrating that the probability of error, for a particular data set, obtained by the distance-weighted rule is strictly lower than that for the majority rule.

The main reason for the superior performance of the distance-weighted rule over the majority rule seems to be the absence of ties in the former. While Dudani's implementation of the majority rule counts all ties as errors, there are several intuitive methods to resolve ties. In this correspondence, we have duplicated the experiments reported by Dudani and demonstrated that if ties occurring in the majority rule are resolved in a judicious manner, then the admissibility of the distance-weighted rule is questionable. We have also proved that, in the infinite sample case, the majority rule has the lowest probability of error among all distance-weighted rules.

II. RESOLUTION OF TIES FOR SMALL SAMPLE CASE

The following strategies have been implemented in our experiments to resolve ties which occur when the majority k-nearest neighbor rule is used.

1) Ties are resolved by randomly selecting one of the tied pattern classes. This method is frequently employed [2, p. 223].

2) Ties are resolved by considering fewer than k nearest neighbors. That is, if a tie occurs among the k nearest neighbors, then we consider $(k - l)$, $l = 1, 2, \cdots, k - 1$, nearest neighbors successively until the tie is resolved. Patrick [2, p. 224] mentions a similar approach.

3) Ties are resolved by considering more than k nearest neighbors. As in method 2), if a tie occurs among the k nearest neighbors, then we use $(k + l)$, $l = 1, 2, \cdots$ nearest neighbors successively until the tie is resolved.

Computationally, method 3) is least attractive. However, implementing any one of the above methods to resolve ties takes no more effort than computing the weights in [1]. Each training set used in our experiments consists of 50 samples from each of the three bivariate distributions described in [1]. The test set contains 1000 samples per class for a total of 3000 samples. An average performance was obtained by considering six independent training sets thus duplicating Dudani's experiments. The results of our experiments are shown in Fig. 1. Curves (△) and (▲) in Fig. 1 are identical to the two curves in Fig. 1 of [1]. The first two strategies for resolving ties as discussed above give almost identical results, and their performance is shown by curve (□). These results show that if any one of the methods for resolving ties is used, the majority rule gives better performance than the distance-weighted rule for moderate values of $k (k \leq 26)$. For large values of $k (k > 26)$ the error rate of the majority rule starts to increase, while the error curve of the distance-weighted rule remains essentially flat.

One advantage claimed in [1] for the distance-weighted rule is that its performance does not peak. While this is true for $k \leq 50$, as far as the average performance on this data set is concerned, for specific training sets there are instances where the probability of error of the distance-weighted rule attains a minimum and starts to increase as k increases. The deterioration in the performance of the majority rule, even with a provision to resolve ties, for large values of k is to be expected, and it is a consequence of the "curse of finite sample size." The same phenomenon occurs for the distance-weighted rule. Thus, when only a finite number of training samples is available, the value of k chosen should be a small fraction of the number of training samples [3, p. 105].

III. ASYMPTOTIC PERFORMANCE OF A WEIGHTED RULE

We now define the problem in a more formal way. Let x be a test pattern, and let θ be the class of x. Let X_n be a set of n training

Manuscript received May 13, 1977; revised October 7, 1977. This work was supported by the National Science Foundation under Grant ENG 76-11936.

The authors are with the Department of Computer Science, Michigan State University, East Lansing, MI 48824.

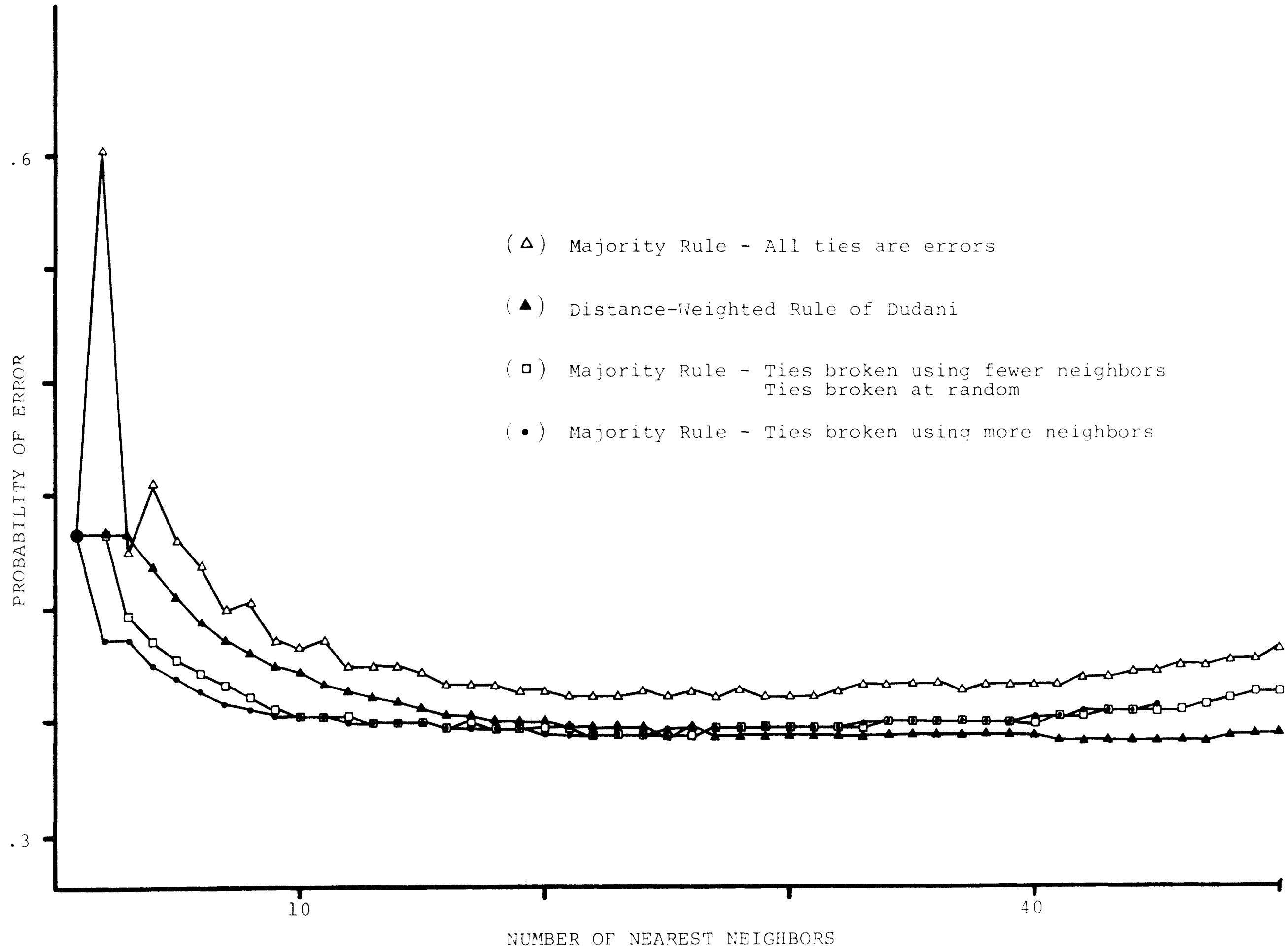

Fig. 1. Error rates for several nearest neighbor rules.

patterns whose classes are known. Let $X_k = \langle x_i, i = 1, k\rangle$ be the sequence of k patterns in X_n which are closest to x. Let $\theta_k = \langle \theta_i, i = 1, k\rangle$ be the sequence of classes of the k patterns in X_k. Suppose there are c classes $\{w_j, j = 1, c\}$. A weighted k-nearest neighbor rule T classifies x using a weight function $t_i = t_i(x,X_k)$, $i = 1, k$ which assigns a nonnegative weight to each element of X_k. The weight assigned to class w_j is

$$\tau_j = \sum_{i=1}^{k} t_i \delta_{\theta_i, w_j}$$

where δ is the Kroneker delta. The rule T is

$$\text{classify } x \text{ as } w_t \text{ if } \tau_t = \max_{j=1,c} \tau_j.$$

If the weights of the k closest patterns are all the same, $t_i = t_1$, $i = 1, k$, then the weighted rule T reduces to the usual majority rule M. We assume ties are broken by randomly choosing from among the tied classes.

Theorem: In the infinite sample case ($n \to \infty$) the probability of error of the majority k-nearest neighbor rule is minimum among all weighted k-nearest neighbor rules.

Proof: The probability of error of a weighted rule T may be written

$$P_n(e \mid T) = \iint P_n(e \mid x,X_k,T)\, dF(X_k \mid x)\, dF(x). \tag{1}$$

Let θ' be the class assigned to x by rule T. Then the probability of error for a given sample x and a given sequence of closest neighbors X_k is

$$P_n(e \mid x,X_k,T) = 1 - \sum_{j=1}^{c} P_n(\theta = w_j, \theta' = w_j \mid x, X_k, T)$$
$$= 1 - \sum_{j=1}^{c} P(w_j \mid x) P_n(w_j \mid X_k,T) \tag{2}$$

since θ is determined independently of θ'. In (2), $P_n(w_j \mid X_k,T)$ is an average over all c^k possible values for the sequence of classes θ_k:

$$P_n(w_j \mid X_k,T) = \sum_{l=1}^{c^k} P_n(w_j \mid A_l,T) P_n(A_l \mid X_k) \tag{3}$$

where $A_l = \langle a_i, i = 1, k\rangle$ is one of the possible values of θ_k, $a_i \in \{w_j, j = 1, c\}$.

For a particular x and X_k, the difference in the probability of error between the weighted rule T and the majority rule M may be written as

$$\Delta_n = P_n(e \mid x,X_k,T) - P_n(e \mid x,X_k,M). \tag{4}$$

Using (2) and (3) in (4),

$$\Delta_n = \sum_{l=1}^{c^k} P_n(A_l \mid X_k) \sum_{j=1}^{c} P(w_j \mid x) \cdot \{P_n(w_j \mid A_l,M) - P_n(w_j \mid A_l,T)\}. \tag{5}$$

For a particular A_l, let w_t be the classification of x using rule T

$$(\tau_t = \max_{j=1,c} \tau_j),$$

and let w_m be the classification of x using rule M

$$(k_m = \max_{j=1,c} k_j),$$

where k_j is the number of times class w_j appears in A_l.

Then

$$P_n(w_j | A_l, M) = \delta_{w_j, w_m},$$

$$P_n(w_j | A_l, T) = \delta_{w_j, w_t},$$

and (5) becomes

$$\Delta_n = \sum_{l=1}^{c^k} P_n(A_l | X_k)\{P(w_m | x) - P(w_t | x)\}. \qquad (6)$$

Since the k samples in X_k are chosen independently, the probability of a particular class sequence A_l is

$$P_n(A_l | X_k) = \prod_{i=1}^{k} P_n(a_i | x_i).$$

Now,

$$\lim_{n \to \infty} P_n(A_l | X_k) \to \prod_{i=1}^{k} P(a_i | x). \qquad (7)$$

Since class w_j appears k_j times in A_l,

$$\prod_{i=1}^{k} P(a_i | x) = \prod_{j=1}^{c} [P(w_j | x)]^{k_j}. \qquad (8)$$

Using (6)–(8), the difference in the asymptotic probabilities of error for the weighted and majority rules is

$$\Delta = \lim_{n \to \infty} \Delta_n = \sum_{l=1}^{c^k} \{P(w_m | x) - P(w_t | x)\} \cdot \prod_{j=1}^{c} [P(w_j | x)]^{k_j}.$$

The sum over all c^k possible values of θ_k can be written in another order. For each A_l there is an unique A_l' such that $(a_i = w_m) \Rightarrow (a_i' = w_t)$, $(a_i = w_t) \Rightarrow (a_i' = w_m)$, and $(a_i = w_j) \Rightarrow (a_i' = w_j)$ for all $j \neq m, t$. Thus, for A_l', rule T classifies x as w_m, and rule M classifies x as w_t. Since $\{A_l\}$ and $\{A_l'\}$ are identical, Δ may be written as

$$\Delta = \tfrac{1}{2}(\Delta_{A_l} + \Delta_{A_l'})$$

$$= \frac{1}{2} \sum_{l=1}^{c^k} \{P(w_m | x) - P(w_t | x)\} \prod_{j=1}^{c} [P(w_j | x)]^{k_j}$$

$$+ \frac{1}{2} \sum_{l=1}^{c^k} \{P(w_m' | x) - P(w_t' | x)\} \prod_{j=1}^{c} [P(w_j' | x)]^{k_j}$$

where $w_m' = w_t$, $w_t' = w_m$, and $w_j' = w_j$ for $j \neq m, t$.

The above equation can be rewritten as

$$\Delta = \frac{1}{2} \sum_{l=1}^{c^k} \prod_{j \neq m,t} [P(w_j | x)]^{k_j} \{S + S'\} \qquad (9)$$

where

$$S = \begin{cases} [P(w_m | x)]^{k_m} [P(w_t | x)]^{k_t} \{P(w_m | x) - P(w_t | x)\}, & \text{if } w_m \neq w_t \\ 0, & \text{if } w_m = w_t \end{cases}$$

and

$$S' = \begin{cases} [P(w_t | x)]^{k_m} [P(w_m | x)]^{k_t} \{P(w_t | x) - P(w_m | x)\}, & \text{if } w_m \neq w_t \\ 0, & \text{if } w_m = w_t. \end{cases}$$

To simplify the notation let $g_m = P(w_m | x)$ and $g_t = P(w_t | x)$. Then

$$S + S' = \begin{cases} g_m^{k_t} g_t^{k_t} (g_m - g_t)(g_m^{k_m - k_t} - g_t^{k_m - k_t}), & \text{if } w_m \neq w_t \\ 0, & \text{if } w_m = w_t. \end{cases}$$

Since, by definition,

$$k_m = \max_{j=1,c} k_j,$$

we have $k_m - k_t \geq 0$.

For arbitrary g_m and g_t we have

$$k_m - k_t \geq 0 \Rightarrow (g_m - g_t)(g_m^{k_m - k_t} - g_t^{k_m - k_t}) \geq 0$$

$$\Rightarrow S + S' \geq 0$$

$$\Rightarrow \Delta \geq 0$$

$$\Rightarrow P(e | T) \geq P(e | M)$$

and the theorem is proved.

IV. Conclusion

We have shown that if the number of training samples is large, then the probability of error of a majority rule will be no more than that of any weighted rule. In addition, experiments in a data base of small sample size indicate that, for small values of k, the majority rule (with provision to resolve ties) performs better than the distance-weighted rule proposed by Dudani. These results suggest that distance-weighted k-nearest neighbor rules offer little advantage over the majority rule.

Acknowledgment

The authors are pleased to acknowledge many helpful discussions with Richard Dubes and Karl Pettis.

References

[1] S. A. Dudani, "The distance-weighted k-nearest neighbor rule," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-6, pp. 325–327, Apr. 1976.
[2] E. A. Patrick, *Fundamentals of Pattern Recognition*. Englewood Cliffs, NJ: Prentice-Hall, 1972.
[3] R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis*. New York: Wiley, 1972.

A Study of Adaptive Control Principle Orthoses for Lower Extremities

M. KLJAJIĆ AND A. TRNKOCZY

***Abstract*—Using qualitative analysis of the stochastic model of human locomotion, and its experimental verification on the basis of gait parameter variability and energy consumption in relation to cadence, some properties relevant to normal walking and to the choice of adequate control principles of orthoses based on functional electrical stimulation (FES) are established. It is pointed out that in**

Manuscript received December 13, 1976; revised July 25, 1977. This work was supported in part by the Slovene Research Community, Ljubljana, Yugoslavia, and in part by the Rehabilitation Services Administration, Department of Health, Education, and Welfare, Washington, DC.

The authors are with the J. Stefan Institute and the Faculty of Electrical Engineering, University of Ljubljana, Ljubljana, Yugoslavia.