Hamming Distance Behaviour as an Analytical Tool of a Pattern Recogniser

Soo-Dong Lee
Department of Electrial Engineering and Electronics
(Received December 30, 1981)

〈Abstract〉

In this paper, it is shown that after a period of training, the output of the digital learning net with decision feedback loops enters a stable state within three cycles of state space activity related to an unknown input pattern.

Thus the single-layer nets can make a strong decision from the preprocessed stable output.

Each response of the nets, both the digital learning net and the single-layer nets, is estimated in terms of Hamming distance behaviour, and this provides an important analytical tool for an unknown system.

해밍거리(Hamming Distance)를 이용한 패턴認識시스템의 解析

李 秀 東
電氣 및 電子工學科
(1981. 12. 30접수)

〈요 약〉

본 논문에서는 판정·피드백 루프를 가지는 디지탈학습망(Digital Learning Net)은 일정기간의 학습후, 미지의 외부입력에 대해 3싸이클의 상태공간 동작후에는 안정상태에 도달함을 보였다. 따라서 단층망군(Single-layer Nets)은 안정상태로부터 오차없이 외부 입력을 판정한다.

각 망의 응답은 해밍거리(Hamming Distance)에 의해 통계적으로 예측할 수 있음을 밝혔으며, 이 해석법은 미지의 시스템을 해석할 수 있는 강력한 방법을 제공한다.

I. Introduction

It is known that a reandom feedback in the digital learning network for the artificially intelligent pattern recognition inherently produces few and short cycles in its state space, and as a result the net enters stable states and performs a clustering operation as one of its natural properties.[1]

Thus the digital learning network(DLN) of random access memories (RAMS) with random feedback connections could be trained to enter stable cycles of patterns on their feedback loops as a response to unknown patterns.[2]

It is also known that the behaviour of the single-layernets is heavily dependent on the Hamming distance of an unknown input pattern to the patterns in the training set.[3]

In this paper, a prototype pattern recognition system shown in Fig. 2 in the preceding paper of the same author[4] is proposed and analysed in terms of Hamming distance behaviour to investigate whether it can make strong decisions after a few cycles of state space activity.

In the example the prototype system is restricted to four-category system, however this system can be easily extended to any multi-category system by merely increasing the number of descriminators in the single-layer nets.

Let us say that the input, output and training patterns shown in Fig. 2 in the preceding paper of the same author[4] are described as follows.

Input pattern (A_i); $a_{ik} \in \{A_i \subset A\}$, A being the universal set of patterns, and a_{ik} being the k-th element of an ordered set A_i.

Output pattern (A_0), $a_{0k} \in \{A_0 \subseteq T\}$, a_{0k} being the k-th element of an ordered set A_0.

Prototype pattern (A_{pi}); $a_{pk}' \in \{A_p' \subseteq A_p \subset T\}$, a_{pk}' being the k-th element of an ordered set A_p', and A_p' being a member of the prototype patterns set A_p.

Training pattern (T_j'); $t_{jk}' \in \{T_j' \subseteq T' \subset T\}$, t_{jk}' being the k-th element of an ordered set T_j', T_j' being a member of i-th category training set T_i', and T being the training set, where $|A_i| = |A_0| = |A_p'| = |T_j'| = 256$, and a_{ik}, a_{0k}, a_{pk}' and t_{jk}' are all one bit binary number.

Before any training has taken place, all initial states enter the all 0s state directly. During training, for each $\{T_j' \subseteq T'\}$ on I matrix and each $\{A_p' \subseteq A_p\}$ on Z matrix, A_p' is impressed on the Z matrix by means of data-in terminals and decision feedback address terminals, which are imposed by A_p' and binary $\{i\}$ respectively.

For each A_p' on Z matrix, an output all 1s is impressed on the corresponding descriminator by means of data-in terminals which are set to all 1s.

In the system, the training procedure creates four stable-states re-entrant for all members of four categories.

When the system is faced with an unknown pattern on its I matrix and Z matrix respectively, the response of the nets (both the digital learning net and the single-layer nets) can be estimated. The most important estimation factor is the nets' property to respond in a similar way to patterns that are close in Hamming distance to those seen during training step.

II. Training as a Function of Hamming Distance

Each training pattern T_j' in the training set is an ordered set where $|T_j'| = 256$ and j, i are distinguishing labels for each of the T patterns and N categories respectively, thus $1 \leqslant j \leqslant T$ and $1 \leqslant i \leqslant N$. If we perform an exclusive OR operation of an unknown pattern A_i with T_j', element for element, we obtain a new ordered set, say $D_{T_j'}$, and the number of 1s in $D_{T_j'}$ is the Hamming distance $H_{IT_j'}$.

$$H_{IT_j'} = \sum_{k=1}^{256} (a_{ik} \oplus t_{jk}').$$

When the effects of pattern feedback addresses and decision feedback addresses are ignored in the digital learning net, that is, only the effect of input addresses is considered, the probability of k-th RAM giving an output a_{pk}', considering A_i and T_j', is the probability of 3-tuple sampling of 0s in $D_{T_j'}$. Thus

$$P(T_{jk}) = \binom{R-n}{H_{TT_j'}} \Big/ \binom{R}{H_{IT_j'}}.$$

where $R = |T_j'| = |A_i| = 256$ and the sampling size $n = 3$.

One can now calculate the probability of a_{pk}' being generated at the output of k-th RAM after the occurrence of all T training patterns belonging to one category. After two training patterns, the probability is

$$P(T_{1,2}') = P(T_1') + P(T_2') - P(T_{(1,2)}'),$$

where $P(T_{(i,j)}')$ is the probability of sampling 0s in $D_{T'_{(i,j)}}$, ($D_{T'_{(i,j)}} = D_{T_i'} + D_{T_j'}$), corresponding to the commonality between $D_{T_i'}$ and $D_{T_j'}$, which can be readily calculated by using OR operation. Thus for T training patterus. the probability of k-th RAM responding to A_i with a_{pk}' is

$$P(T^i)=P(T_{1,2,\cdots,}T^i)$$

$$=\sum_{j=1}^{T}P(T_j^i)+(+1)\sum_{i,j=1\,\mathrm{to}\,T}^{\binom{T}{2}}P(T_{(i,j)}^i)$$

$$+(-1)^2\sum_{i,j,k=1\,\mathrm{o}\,T}^{\binom{T}{3}}P(T_{(i,j,k)}^i)$$

$$+(-1)^{T-1}\sum_{i,j,\;=1\,\mathrm{to}\,T}^{\binom{T}{T}}P(T^i_{(i,j,\cdots)})$$

where $P(T^i_{(i,j,k)})$ is the probability of sampling 0s in $D_{T^i(i,j,k)}$,

$$D_Ti_{(i,j,k)}=D_{T_i}i+D_{T_j}i+D_{T_k}i,$$

and$P(T^i_{(i,j,\;)})$ is the probability of sampling 0s in D_Ti,

$$D_Ti=D_{T_1}i+D_{T_2}i+\cdots+D_{T_T}i=\underset{j=1\,\mathrm{to}\,T}{OR}(D_{T_j}i).$$

Now when the effects of the input address and the decision feedback address are ignored, that is only the effect of the pattern feedback address is considered in the digital learning net, the Hamming distance between A_i on the I matrix and a prototype pattern A'_p is given by

$$H_{IA'_p}=\sum_{k=1}^{256}(a_{ik}\ominus a'_{pk}),$$

and the probability of k-th RAM giving an output a'_{pk} is

$$P(A'_p)=\binom{R-n}{H_{IA'_p}}\Big/\binom{R}{H_{IA'_p}},$$

where $R=|A_i|=|A_p'|=256$ and the sampling size $n=3$.

In the digital learning net of the system, a set of training patterns belonging to any one category and their corresponding prototype pattern are trained simultaneously with a proper decision address. Thus these two sets of patterns are mutually inclusive in terms of address during training.

Therefore, after all the T training patterns belonging to a set of training patterns, say T', and their corresponding prototype pattern, say A_p', are occurred on the I and Z matrices respectively, that is, when one category training is finished, the probability of k-th RAM within the digital learning net responding with an output a'_{pk} on the Z matrix is

$$P^i=P(T^i)\times P(A_p^i).$$

After all the $T\times N$ training patterns and N prototype patterns are occurred on the I and Z matrices respectively, the probability of k-th RAM giving an output $\bigcup_{i=1}^{N}a'_{pk}$, ignoring the effect of the decision feedback address, is given by

$$P=1-\prod_{i=1}^{N}(1-P^i).$$

Ⅲ. Estimation of an Unknown Pattern with Hamming Distance

After training, assume that the system is faced with an unknown pattern A_i on the I matrix and Z matrix at the same time. At the first recognition step, the digital learning net will respond its output to the Z matrix before the decision made at the output of the single-layer nets reaches its input. Thus uhe output pattern generated on the Z matrix, say A'_0, entirely depends on the input and output addresses of the digital learning net, and is obtained from the information about $T\times N$ training patterns and N prototype patterns. Therefore,

$$A'_0\subseteq\bigcup_{i=1}^{N}A'_p.$$

There is no variaton of an unknown pattern from the prototype patterns in A'_0, thus the single-layer nets can be enforced to make a strong decision during the second recognition step.

At the first recognition step, single-layer nets faced with an unknown pattern on the Z matrix receive their 4-tuple input addresses from the Z matrix, then respond the contents of the addressed store locations on the corresponding data-out terminals. The matched data

(1s) of each descriminator are then summated respectively and transferred to the maximum response detector to be classified.

When the descriminator having the maximum response has been classified. the encoder generates the binary number corresponding to the classification on the decision bits. Thus decision is fed back to the decision feedback address terminals of the digital learning net. This brings the second recognition step.

After the first recognition step, the Z matrix changes its pattern from A_0 $(=A_i)$ to A_0', while the I matrix is holding A_i, and the first decision appears on the decision feedback address terminals of each RAM of the digital learning net.

At the second recognition step, the digital learning net, faced with A_i, A_0' and the first decision at the corresponding address terminals, generates the output pattern which consists of the elements belonging to any one prototype pattern selected by the decision feedback.

Thus after the second recognition step, the probability of k-th RAM within the digital learning net, considering A_i, A_0' and the first decision bits, giving an output a_{pk}' on the Z matrix, that is. the probability of generating any one prototype pattern A_p', say A_0'', on the Z matrix is

$$P'' = P(T') \times P'(A_p'),$$

where

$$P'(A_p') = \binom{R-n}{H_{O'A'_p}} \Big/ \binom{R}{H_{O'A'_p}},$$

the Hamming distance $H_{O'A'_p} = \sum_{k=1}^{256} (a_{o'k} \oplus a'_{pk})$, $a_{o'k} \in \left\{ A'_0 \subseteq \bigcup_{i=1}^{N} A'_p \right\}$ and the sampling size $n=3$.

The Hamming distance $H_{0'A'_i}$ can be calculated, in terms of the number of bits by which A'_0 and A'_p differ each other, as follows. After the first recognition step, the probability of generating A'_0 on the Z matrix is given by P, and the probability of A'_0 including only the elements in A'_p is

$$P'_0 = P - \sum_{\substack{j=1 \text{ to } N \\ (j \neq i)}}^{\binom{N-1}{1}} P^j + P^i \Bigg[\sum_{\substack{j=1 \text{ to } N \\ (j \neq i)}}^{\binom{N-1}{1}} P^j + (-1) \sum_{\substack{j,k=1 \text{ to } N \\ (j,k \neq i)}}^{\binom{N-1}{2}} P^j \cdot P^k + \cdots + (-1)^{N-2} \sum_{\substack{j,k,l,\ldots=1 \text{ to } N \\ (j,k,l,\ldots \neq i)}}^{\binom{N-1}{N-1}} P^j \cdot P^k \cdot P^l \cdots \Bigg],$$

where $P^i = P(T^i) \times P(A'_p)$ and the category size $N=4$.

Therefore after the first recognition step, the probability of generating m elements in A_p on the Z matrix is

$$P_n = \binom{R}{m} \cdot P^m \cdot (1-P)^{R-m},$$

where the most likelihood response of the digital learning net

$$m = R \cdot P.$$

After the first recognition step. the probability of generating m' elements only in A_p' on the Z matrix is

$$P'_{om'} = \binom{R}{m'} \cdot (P_o')^{m'} \cdot (1-P_o')^{R-m'},$$

where the most likelihood response, $m' = R \cdot P_o'$.

Thus the Hamming distance $H_{o'}A_p'$ is given by

$$H_{o'}A_p' = R \cdot (1 - P_m \cdot P'_{om'} - (1-P_m)(1-P'_{om'})) = R \cdot (P_m + P'_{om'} - 2P_m \cdot P'_{om'}).$$

At the second recognition step. the single-layer nets faced with A_0', classify A_0' by assigning the classification to the descriminator which has the strongest response, and make the second decision on the decision bits. This decision is very strong one because A_0' consists of the prototype patterns.

When the two decisions, which were made during two previous recognition steps, are same the system makes a strong final decision from the output pattern A_0'', hence the output pattern enters the stable-state from the third recognition step. But when the two decisions are different the rejection occurs. This happens when an unknown pattern A_i is far more similar to another category training patterns than to its own training patterns, although it is more similar to its own prototype pattern. Thus dues to the fact that the output pattern generated after the third recognition step, say A_0''', consists of the elements belonging to both A_0', and A_0'', about which the different decisions were made. However, this is the worst case and can be removed by choosing proper training patterns and prototype patterns.

When the system is faced with an unknown pattern on the I matrix we can now calculate the most likelihood response of the descriminator from which the final decision is made.

After two recognition cycles, the probability of a RAM in the i-th descriminator responding with a 1 is given by

$$P_R i = \binom{R-n'}{H_0''{}_{A'_p}} \Big/ \binom{R}{H_0''{}_{A'_p}},$$

where the Hamming distance $H_0''{}_{A'_p}$ is

$$H_0''{}_{A'_p} = \sum_{k=1}^{R} (a_0'' \oplus a'^d{}_k)$$

$$= R \cdot (1 - P'').$$

for any i $(1 \leq i \leq N)$, $a_0''k \in A_0'' \subseteq A'_p$ and the sampling size $n' = 4$.

The probability of generating m Is at the output of the i-th descriminator is given by

$$P_R i_m = \binom{k}{m} (P_R i_m)^m (1 - P_R i_m)^{k-m},$$

where $k = R/n' = 64$.

The most likelihood response of the i-th descriminator regarding to A_i is

$$m = k \cdot P_R i.$$

Ⅳ. Conclusions

From the analysis the followings can be estimated.

(1) After two cycles a decision can be made, and this decision is relatively very strong one because the final decision is made from the pattern which is a subset of any one prototype pattern.

(2) An input pattern can be preprocessed to a prototype-like pattern during recognition cycles.

(3) The most likelihood response of the system can be estimated in terms of Hamming distance.

Refernces

1. I. Aleksander, "Some psychological properties of digital learning nets", Int. Journ. Man-Machine Studies, 1970, 2, pp. 189—212.
2. I. Aleksander, and E. H. Mamdani, "Microcircuit learning nets: improved recognition by means of pattern feedback", Electron. Lett., 1968, 4, pp. 425—426.
3. I. Aleksander, "Micorocircuit learning nets: Hamming distance behaviour", Electron. Lett., 1970, 6, p. 134.
4. S. D. Lee, "Artificially Intelligent Dynamic Pattern Recognition," U. I. T. Report, 1982