ISRN SICS-R--95/11--SE

Some Comments on the Information Stored in Sparse Distributed Memory

by

Jan Kristoferson

SICS Research Report
R95:11
ISSN 0283-3638

Report R95:11

ISRN : SICS-R--95/11-SE
ISSN : 0283-3638

Some Comments on the Information Stored in Sparse Distributed Memory

Jan Kristoferson

RWCP[1] Neuro SICS[2] Laboratory
Box 1263, S-164 28 Kista, Sweden
Email: janke@sics.se

December 1995

Abstract

We consider a sparse distributed memory with randomly chosen hard locations, in which an unknown number T of random data vectors have been stored. A method is given to estimate T from the content of the memory with high accuracy. In fact, our estimate is unbiased, the coefficient of variation being roughly inversely proportional to $\sqrt{MU}$, where M is the number of hard locations in the memory and U the length of data, so the accuracy can be made arbitrarily high by making the memory big enough. A consequence of this is that the good reading methods in [5] and [6] can be used without any need for the special extra location introduced there.

Keywords: Sparse distributed memory, SDM.

Contents

[1] Real World Computing Partnership
[2] Swedish Institute of Computer Science

1 Introduction

We consider a Sparse Distributed Memory, either of Kanerva's original design, see Kanerva [2], or of Jaeckel's selected-coordinates design, see Jaeckel [1]. Let us give a short explanation of the concept:

An *address* is a binary string (vector) of 1s and 0s. A *datum* is a binary string (vector) of 1s and -1s. Let N be the length (dimension) of the addresses, U the length (dimension) of the data. A *hard location* is a place in the memory where to store data. The *content* of a hard location is a U-dimensional vector of integers. The coordinates of the content vectors are called *positions*. In conventional neural-net terms, hard locations are hidden units, their input weights define the position, and there output weights define the contents. In Kanerva's original design a hard location is given by an address, so there are totally 2^N possible hard locations. In Jaeckel's design a hard location is given by a *mask* (a subset of K coordinates out of the N coordinates in the address strings) and one bit for each coordinate in the mask, so there are totally $\binom{N}{K}2^K$ possible hard locations in this case. In Kanerva's design an address X *activates* a hard location h if the Hamming distance $d(X, h)$ is at most equal to the *activation radius* R. In Jaeckel's design an address X *activates* a hard location h if all bits in X lying in h's mask match the corresponding bits in h. *Storing a datum at the address* X means adding (as a vector) the datum to the contents of all hard locations activated by X. In the sequel we assume that T data have been stored, the contents of all hard locations initially being zero. *Reading at the address* X means calculating by some method, from the contents of the hard locations activated by X, a datum to be read at X. (In the standard reading method we sum the contents and choose, for each position, the reading 1 if the sum is ≥ 0, and -1 if the sum is < 0.)

Now let M be the number of hard locations, and p the *probability of activation*[3]. If we store T data vectors in the memory, we could ask how much (in the sense of information theory) of the information in the given data can be retrieved from the memory by using a suitable reading procedure. To specify the question we make the following assumptions. We assume that the hard locations and the storage addresses are randomly and independently chosen (with uniform distribution over all possible locations and addresses, respectively). The data are also assumed to be randomly chosen, but the databits don't need to be uniformly distributed over $\{-1, 1\}$; instead we have for each position a separate (unknown) probability for the databits in that position being 1.

In Sjodin [5] and [6] good reading methods are given. There an extra hard location is used to save the number T of stored data. Information on the value of T exists

[3] The probability that a randomly chosen hard location is activated by a randomly chosen (write or read) address. It is, in fact, $2^{-N}\sum_{k=0}^{R}\binom{N}{k}$ for Kanerva's design and 2^{-K} for Jaeckel's design.

in the original memory (without any special extra location), however. Here we will show that we can compute an accurate estimate of T from the content of the original memory.

2 The stochastic variable Q_u

The only values that really "exist" in the memory are the $C_{m,u}$s, where $C_{m,u}$ is the integer saved in hard location m, position u. We will show how those can be used to give a good estimate of T. We will now drop the subscript u in the notation, considering the position u to be fixed until further notice. For instance, we will write C_m instead of $C_{m,u}$. We also need some new notation.

Thus, let W_t be the databit (1 or -1) in position u in the tth stored datum. We don't give any notation for the unknown probability $\Pr(W_t = 1)$, since we shall not need to consider it, curiously enough.

Furthermore, define $Y_{m,t}$ to be equal to 1 if the storage address for the tth datum activates hard location m, and equal to 0 otherwise. We have $\Pr(Y_{m,t} = 1) = p$ and $\Pr(Y_{m,t} = 0) = 1 - p$.

Observe that

$$C_m = \sum_t W_t Y_{m,t} \tag{1}$$

and that the W_ts and $Y_{m,t}$s make up a collection of independent stochastic variables. Now we introduce the stochastic variable $Q = Q_u$:

$$Q = \frac{1}{M}\sum_m C_m^2 - (\frac{1}{M}\sum_m C_m)^2 \tag{2}$$

Considering the experiment of drawing a hard location at random (i.e., with equal probability for each one), Q is the variance of the content (in position u) of the drawn location.

In the Appendix the expected value and the variance of Q are calculated. The results are as follows:

$$E(Q) = \frac{(M-1)\gamma}{M}T \tag{3}$$

$$\sigma^2(Q) = T\frac{(M-1)^2}{M^3}(\gamma + \frac{TM - 3(M-1)}{M-1}2\gamma^2) \tag{4}$$

where $\gamma = p(1-p)$. Using these equations we get a good possibility to estimate the value of T.

3 An estimate of T

We now reinstate the subscript u and consider the stochastic variables

$$R_u = \frac{M}{(M-1)\gamma} Q_u \tag{5}$$

According to equations (3) and (4) we have

$$\begin{aligned} E(R_u) &= T & (6) \\ \sigma^2(R_u) &= \frac{T}{M}(\frac{1}{\gamma} + \frac{2TM}{M-1} - 6) & (7) \end{aligned}$$

As our estimate of T we now take

$$S = \frac{1}{U} \sum_u R_u \tag{8}$$

From equations (6) and (7) and the independence of the R_us we get the following final results (approximating $\frac{M}{M-1}$ by 1)

$$\begin{aligned} E(S) &= T & (9) \\ \sigma(S) &= T\frac{1}{\sqrt{MU}}\sqrt{2 + \frac{\frac{1}{p(1-p)} - 6}{T}} & (10) \end{aligned}$$

We remark that the last square root in equation (10) is near $\sqrt{2}$ for "normal" values of the parameters (cf. e.g., Kristoferson [4]).

4 Karlsson's design

In Karlsson [3] a modification of Jaeckel's design is introduced, having great implementation advantages for simulations on a sequential computer. Karlsson's design differs from Jaeckel's thus: for each used mask all the 2^K possible bit combinations are used for hard locations. So $M = r2^K$ if r is the total number of used masks. The former assumption on randomly chosen hard locations is now replaced by the assumption that the masks used for hard locations are randomly (uniformly and independently) distributed among all possible masks.

We define the stochastic variable Q as before, see equation (2). The fact that $\sum_m Y_{m,t} = pM$ for every t facilitates the calculation of the expected value of Q, whereas the fact that, for a fixed value of t, the $Y_{m,t}$s are not independent seems to make the calculation of the variance of Q very messy. The expected value is calculated in the Appendix, giving

$$E(Q) = \gamma T \tag{11}$$

We refrain from the calculation of $\sigma^2(Q)$, taking for granted that it will be of the same magnitude as before, which has been confirmed by running (on a Sparc-station) a simulation program written in C .

It is now obvious how to get a good estimate of T also for Karlsson's design.

Appendix. Calculation of the expected value and the variance of Q_u

We first look at Kanerva's and Jaeckel's designs. We have $E(Y_{m,t}) = p$. To facilitate the calculations we introduce the centralized version $Z_{m,t}$ of $Y_{m,t}$: $Z_{m,t} = Y_{m,t} - p$. Letting $\gamma = p(1-p)$, we easily find that

$$\begin{aligned} E(Z_{m,t}) &= 0 && (12)\\ E(Z_{m,t}^2) &= \gamma && (13)\\ E(Z_{m,t}^4) &= \gamma - 3\gamma^2 && (14) \end{aligned}$$

Observe that the W_t's and $Z_{m,t}$'s make up a collection of independent stochastic variables. We introduce the variables

$$D_m = \sum_t W_t Z_{m,t} \tag{15}$$

Using equations (1) and (2), we find that

$$Q = \frac{1}{M}\sum_m D_m^2 - (\frac{1}{M}\sum_m D_m)^2 \tag{16}$$

or, slightly rewritten,

$$Q = (\frac{1}{M} - \frac{1}{M^2})\sum_m D_m^2 - \frac{1}{M^2}\sum_{m\neq n} D_m D_n \tag{17}$$

From this we get $E(Q) = M(\frac{1}{M} - \frac{1}{M^2})E(D_m^2) - \frac{1}{M^2}\sum_{m\neq n} E(D_m D_n)$. Using equation (15), independence, and equation (12), and observing that $W_t^2 = 1$, we deduce that $E(Q) = (1 - \frac{1}{M})TE(Z_{m,t}^2)$. Equation (13) then gives equation (3).

We now want to calculate $E(Q^2)$. Expanding the square of the right member of equation (17) we will get terms of the form $aD_m D_n D_q D_r$, where a is a constant and some of m, n, q, r are identical. Considering again equation (15), independence, and equation (12), we find that only the terms containing D_m^4 for some m and the terms containing $D_m^2 D_n^2$ for some distinct m and n give nonzero contributions to $E(Q^2)$. So we find that[4]

$$E(Q^2) = \alpha E(D_m^4) + \beta E(D_m^2 D_n^2) \tag{18}$$

[4]In the following equation and in similar situations below it is understood that m and n are distinct.

where $\alpha = M(\frac{1}{M} - \frac{1}{M^2})^2 = \frac{(M-1)^2}{M^3}$
and $\beta = M(M-1)(\frac{1}{M} - \frac{1}{M^2})^2 + 2M(M-1)\frac{1}{M^4} = \frac{(M-1)(M^2-2M+3)}{M^3}$.

Once again using equation (15), independence, equation (12), and $W_t^2 = 1$, we find that the terms giving nonzero contributions to $E(D_m^4)$ are T terms of the form $Z_{m,t}^4$ and $3T(T-1)$ terms of the form $Z_{m,s}^2 Z_{m,t}^2$ $(s \neq t)$, and we find that the terms giving nonzero contributions to $E(D_m^2 D_n^2)$ are T terms of the form $Z_{m,t}^2 Z_{n,t}^2$ and $T(T-1)$ terms of the form $Z_{m,s}^2 Z_{n,t}^2$. Using this and equations (13) and (14) in equation (18), and using equation (3), we can calculate the variance $\sigma^2(Q) = E(Q^2) - (E(Q))^2$. The result is equation (4).

Turning to Karlsson's design, we first observe that equations (12) - (16) still hold. Since $\sum_m Z_{m,t} = \sum_m Y_{m,t} - pM = 0$ we have $\sum_m D_m = \sum_t W_t \sum_m Z_{m,t} = 0$, so equation (16) is simplified to

$$Q = \frac{1}{M} \sum_m D_m^2 \tag{19}$$

Using equation (15), independence, equation (12), and the fact that $W_t^2 = 1$, we get $E(Q) = TE(Z_{m,t}^2)$. Equation (13) then gives equation (11).

References

[1] L. A. Jaeckel. An Alternative Design for a Sparse Distributed Memory. RIACS Technical Report 89.28, Research Institute for Advanced Computer Science, NASA Ames Research Center, 1989.

[2] P. Kanerva. Sparse distributed memory and related models. In Mohamad H. Hassoun, editor, *Associative Neural Memories*. Oxford University Press, 1993.

[3] R. Karlsson. Evaluation of the Fast Activation Mechanism for the Kanerva SDM Memory. SICS Research Report R95:10, Swedish Institute of Computer Science, 1995.

[4] J. Kristoferson. Best Probability of Activation and Performance Comparisons for Several Designs of Sparse Distributed Memory. SICS Research Report R95:09, Swedish Institute of Computer Science, 1995.

[5] G. Sjödin. Convergence and New Operations in SDM. SICS Research Report to appear, Swedish Institute of Computer Science, 1995.

[6] G. Sjödin. Improving the Capacity of SDM. SICS Research Report to appear, Swedish Institute of Computer Science, 1995.