ELSEVIER

Neurocomputing 32–33 (2000) 973–978

NEUROCOMPUTING

www.elsevier.com/locate/neucom

Associative symmetry vs. independent associations

Daniel S. Rizzuto*, Michael J. Kahana

Volen Center for Complex Systems, Brandeis University, 415 South street, Waltham, MA 02454, USA

Accepted 13 January 2000

Abstract

We develop a neural network model of paired-associate learning based upon an auto-associative learning mechanism. We show that this relatively simple neural network can replicate complex human behavioral data, but only when the correlation between forward and backward learning is highly correlated. This network-based analysis is used to constrain psychological theories of association in humans. © 2000 Elsevier Science B.V. All rights reserved.

Keywords: Associations; Memory; Neural networks; Symmetry

1. Introduction

In the human memory literature, there are two competing theoretical views regarding the nature of associative formation. The independent associations hypothesis (IAH) maintains that if two symbols, A and B are encoded successively, the forward association between A and B will tend to be stronger than the backward association between B and A. Additionally, the strengths of forward and backward associations are hypothesized to be independent [8]. In contrast to this position, representatives of Gestalt psychology viewed symbolic associations as composite representations, incorporating elements of each to-be-remembered symbol into a new entity [1,6]. We refer to this view as the associative symmetry hypothesis (ASH). According to this position, the strengths of forward and backward associations are highly correlated with one another.

* Corresponding author. Tel.: + 1-781-736-3292; fax: + 1-781-736-2398.
E-mail address: neuro@volen.brandeis.edu, rizzuto@brandeis.edu (D.S. Rizzuto).

PII: S0925-2312(00)00268-X

Early research addressing this issue was concerned with the relative recall probabilities of forward and backward tests. The experimental paradigm used in these studies was the classic method of paired-associate learning. A pair is studied in the order A then B and tested either in the forward order ($A - ?$) or the backward order ($B - ?$). Evidence for a forward recall advantage, if obtained, would support the independent associations view. The surprising failure to find any such advantage in a number of experimental paradigms seemed to lend support for the associative symmetry hypothesis [1,7].

However, even if forward and backward recall are equal on average, this would not demonstrate associative symmetry [8]. It is easy to conceive of two independent processes producing results which are equal on average. A more direct approach to solve this problem requires an examination of the correlation between forward and backward recall for a pair of items studied by a given subject. This item-by-item correlational analysis involves testing subjects on the same pair of items twice and measuring the dependency between forward and backward recall. Here we present a neural network formalism for analyzing and modeling such paired associate learning data.

2. A neural network model

We developed an auto-associative neural network model to simulate the successive forward and backward recall of symbol pairs. Although an auto-associative architecture is typically used to encode a single neural representation, here we store the concatenation of a symbol pair, storing both auto-associative *and* hetero-associative information in different quadrants of the memory matrix. In our model, the general form of the storage equation for a list of L symbol pairs is given by,

$$W = \sum_{v=1}^{L} (\mathbf{a}^v \oplus \mathbf{b}^v)(\mathbf{a}^v \oplus \mathbf{b}^v)^{\mathrm{T}}, \tag{1}$$

where $\mathbf{a}^v$ and $\mathbf{b}^v$ are binary (± 1), N-element vectors representing the items to be associated. The symbol $\oplus$ denotes the concatenation of two vectors, and W represents the $2N \times 2N$ weight matrix. Two quadrants of the matrix contain auto-associative information ($\mathbf{aa}^{\mathrm{T}}$ and $\mathbf{bb}^{\mathrm{T}}$), while the other two quadrants contain hetero-associative information ($\mathbf{ab}^{\mathrm{T}}$ and $\mathbf{ba}^{\mathrm{T}}$).

In our model, not every weight in the resulting matrix is correctly stored. We utilize a probabilistic encoding algorithm which enables the network to account for the effect of repetition on performance (i.e. learning). The hetero-associative quadrants of the matrix drive the associative recall process and we introduce two random variables, γ_{f} and γ_{b}, which control learning in the forward and the backward directions, respectively.

For an item in the list, the probability of storing each hetero-associative weight in the quadrant that mediates forward recall (i.e., $\mathbf{ba}^{\mathrm{T}}$) is given by

$$\Delta W_{ij} = \begin{cases} s_i^v s_j^v & \text{with prob. } \gamma_{\mathrm{f}}, \\ 0 & \text{with prob. } 1-\gamma_{\mathrm{f}}, \end{cases} \tag{2}$$

where $\mathbf{s}^v = (\mathbf{a}^v \oplus \mathbf{b}^v)$ and $\gamma_f \sim \mathcal{N}(\mu, \sigma)$. Similarly, the probability of storing each hetero-associative weight which mediates backward recall is given by $\gamma_b \sim \mathcal{N}(\mu, \sigma)$. The remaining two quadrants have their learning probabilities set exactly to μ. The parameter μ represents the mean probability of encoding at a particular level of learning in the associative learning experiment. To simulate the effects of repeated presentations on recall, we fit μ separately for each of the three levels of learning, thereby quantitatively fitting the learning data without constraining the model to a particular mechanism.

If γ_f and γ_b are perfectly correlated with one another, the model implements the ASH. If γ_f and γ_b are independent of one another, the model implements the IAH. Rather than pitting the two hypotheses against each other, we can allow the model to determine the correlation between γ_f and γ_b, denoted $\rho(\gamma_f, \gamma_b)$, that best fits the data. The power of this parameter lies in its ability to change the behavior of the model from that approximating the IAH when ρ approaches zero, to that approximating the ASH when ρ approaches unity.

After each iteration the state of the network is compared to the target symbol. If the cosine of the angle between the two vectors is greater than a constant criterion then the target is "recovered". If not, the process repeats until a maximum number of iterations has been reached. At this time, if the network has not reached the criterion level of item recovery, that item is considered non-recallable.

3. Experiment

We applied our model to a dataset on paired associate learning [4]. Subjects studied a list of 12 unique word pairs. Each pair was presented visually for two seconds each followed by a one second inter-stimulus interval. Subjects were instructed to read the words aloud from left to right in order to ensure that they processed the two words in temporal succession. To assess learning, equal numbers of word pairs were presented either one, three, or five times in each list. The order of presentation was random, subject to the constraint that identical pairs were never repeated successively. After studying the list of word pairs, subjects performed a distractor task (pattern matching) in order to minimize the role of recency sensitive retrieval processes [3]. In the first test phase, subjects were tested on each of the studied pairs — half in the forward direction ($A - ?$) and half in the backward direction ($? - B$). In the second test phase, half of those pairs that were first tested in the forward order were tested in the backward order, and the other half were again tested in the forward order. The same was true of pairs that were tested in the backward order in the first test phase. This produced a 2×2 factorial of test 1 — test 2 possibilities (Forward–Forward, Forward–Backward, Backward–Forward, and Backward–Backward).

A contingency table was constructed for each of the four conditions, allowing us to measure probabilities of recall as well as the correlation between the two tests, as measured by Yule's Q. For a contingency table given by

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix},$$

Table 1
Experimental results compared to the best model fit (in parentheses) at each presentation level. The last three columns contain the best fitting values of ρ, σ, and μ for each presentation level

	Yule's Q		Accuracy (% Correct)	Best fitting parameters		
	Same	Reversed		μ	σ	ρ
1p	0.99 (0.99)	0.92 (0.89)	34.6 (33.2)	0.42	0.25	0.92
3p	0.99 (0.99)	0.96 (0.93)	64.3 (65.0)	0.54	0.27	1.00
5p	0.99 (0.99)	0.96 (0.91)	73.3 (75.6)	0.62	0.23	0.95

Yule's Q is given by the equation, $Q = (ad - bc)/(ad + bc)$, and varies from -1 to 1 [2]. According to both the IAH and the ASH, the correlation between successive identical tests (e.g., forward on test one and forward on test two) should be near unity. However, according to the IAH, the correlation between reversed tests (e.g., forward on test one and backward on test two) should be near zero. The ASH predicts the correlation between tests in opposite directions should be near unity.

Two primary findings are consistent with the ASH. First, there were no significant differences between forward and backward recall probabilities within any presentation level (see Table 1). Second, comparable correlations were observed between identical tests and between reversed tests. Averaged across different numbers of presentations, Yule's Q was 0.99 for identical successive tests (i.e., Forward–Forward or Backward–Backward) and 0.95 for reverse successive tests (i.e., Forward–Backward and Backward–Forward).

4. Simulating the experimental results

Applying our network model to the experimental results, we fit μ, σ, and ρ separately for one, three, and five presentations. For each learning level we fit the model to the contingency tables for the *Same* and *Reversed* conditions. By fitting each (normalized) quadrant of the contingency tables, the model was able to simultaneously fit correlations and accuracy on tests one and two. In each case, we fail to reject the null hypothesis that the simulated results are generated from the same underlying distribution as the empirical data (for one presentation, $\chi^2 = 0.57$, $df = 3$, $p > 0.5$, for three presentations, $\chi^2 = 0.60$, $df = 3$, $p > 0.5$, and for five presentations, $\chi^2 = 0.80$, $df = 3$, $p > 0.5$). The best fitting parameters for each presentation level are shown in Table 1. As expected, the mean level of encoding (μ) increases with learning. Most importantly, the best fitting values of ρ indicate that very strong correlations between forward and backward storage are necessary to fit the human behavioral data at every level of learning.

Once the optimal mean learning probability for each presentation level was found, we completed a comprehensive search of the local parameter space while keeping the mean level of encoding fixed. Fig. 1 plots the correlation between simulated successive

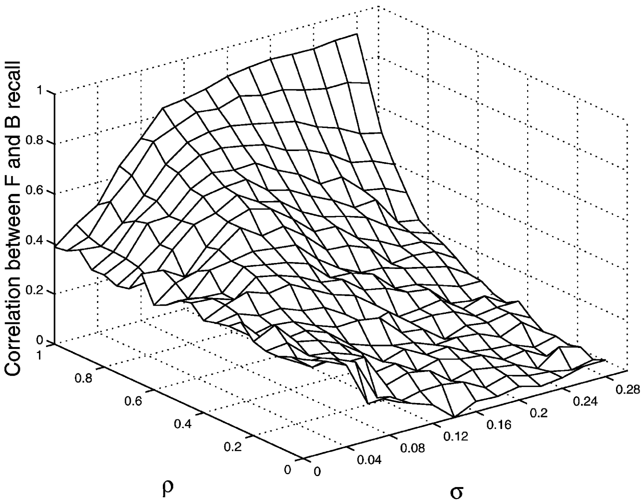

Fig. 1. Simulated correlation between forward and backward recall as a function of σ and ρ.

tests in opposite directions for words presented five times as a function of both ρ and σ. It is apparent that the only way of producing high correlations in the retrieval of forward and backward associations is to have extremely high correlations between γ_f and γ_b. In addition, there is a monotonic relationship between the correlation of learning the forward–backward associations and the correlation in recall between successive tests in opposite directions.

5. Conclusions

Our model's strength relies upon its ability to modulate its behavior between the two distinct hypotheses (ASH vs. IAH) of associative formation. Via the correlational parameter, ρ, the behavior of the auto-associative neural network was able to modulate intermediately between the two boundary conditions and settle into the best fitting region of the parameter space. The fact that the best fitting parameter sets for all levels of learning included a $\rho \simeq 1$ is strong evidence for symmetric paired-associate learning. It is useful to note that symmetry of retrieval probabilities is not always the rule. Learning of longer lists produces an asymmetric advantage for forward retrieval both in free recall [3] and in serial recall [5].

Our auto-associative network model of hetero-associative memory implements a stochastic learning algorithm acting at the level of the "synapse" and quantitatively fits human accuracy and correlation data from a paired associate learning task. In addition, data on the correlations between successive forward and backward recall tests support the notion that an auto-associative mechanism may underly at least some forms of hetero-associative learning.

References

[1] S.E. Asch, S.M. Ebenholtz, The principle of associative symmetry, Proc. Am. Philos. Soc. 106 (1962) 135–163.

[2] Y.M.M. Bishop, S.E. Fienberg, P.W. Holland, Discrete Multivariate Analysis Theory and Practice, MIT Press, Cambridge, MA, 1975.

[3] M.W. Howard, M.J. Kahana, Contextual variability and serial position effects in free recall, J. Exp. Psychol. Learn. Mem. Cog. 25 (1999) 923–941.

[4] M.J. Kahana, Associative symmetry and memory theory, submitted for publication.

[5] M.J. Kahana, J. Caplan, Associative symmetry in probed recall of serial lists, submitted for publication.

[6] W. Köhler, Gestalt Psychology, Liveright, New York, 1947.

[7] B.B. Murdock, Direction of recall in short term memory, J. Verbal Learn. Verbal Behav. 1 (1962) 119–124.

[8] G. Wolford, Function of distinct associations for paired-associate performance, Psychol. Rev. 78 (1971) 303–313.