Network Analyses: The Case of First and Second Person Pronouns*

Yoshio Takane, Yuriko Oshima-Takane, & Thomas R. Shultz
Department of Psychology, McGill University
Montreal, Quebec H3A 1B1
Canada

Abstract

Feed-forward neural network models may be viewed as approximating nonlinear functions connecting inputs to outputs. We analyzed the mechanism of function approximations underlying learning of first and second person pronouns by the cascade-correlation (CC) network. The CC network dynamically grows nets to approximate increasingly more complicated functions. It starts as a net without hidden units, but as soon as it "perceives" that it can no longer improve its performance within the limit of current net topology, it automatically recruits a new hidden unit. This process is repeated until a satisfactory degree of function approximation is achieved. Learning of first and second person pronouns presents an interesting problem in psychology. When the mother talks to her child, *me* refers to herself, and *you* to the child. However, when the child talks to the mother, *me* refers to the child, and *you* to the mother. Learning of the shifting reference of these pronouns can be regarded as a special kind of nonlinear function learning, where the function to be learned stipulates *me* if the speaker and the referent agree, and *you* if the addressee and the referent agree. We investigated how this function is approximated by the CC network using graphic techniques. The function approximation typically depends on the sample of input-output patterns used in training, which is called the problem of environmental bias. We examined the effects of environmental bias in two conditions: the addressee condition in which the addressee was always the child, and the nonaddressee condition in which the child was neither the speaker nor the addressee. It was found that exposures to nonaddressee patterns were crucial for networks' learning of the target function underlying the correct use of pronouns, and that a more variety of nonaddressee patterns facilitate the learning.

1 Introduction

Feed-forward neural network (NN) models can be viewed as approximating nonlinear functions that connect inputs to outputs. Networks' approximation capabilities have been investigated by such authors as Barron (1993), Cybenko (1989), Hornik, Stinchcombe & White (1989), and White (1990). We have been analyzing the functions learned by a generative network algorithm, cascade-correlation (Takane, Oshima-Takane & Shultz, 1994; Oshima-Takane, Takane, & Shultz, 1995). In this paper we analyze the mechanism of function approximations underlying cascade-correlation's learning of first and second person pronouns.

2 Cascade-Correlation (CC) Learning Networks

NN models consist of a set of units, each performing a simple operation. Units receive contributions from other units, compute activations by summing the incoming contributions and applying prescribed (nonlinear) transformations to the summed contributions, and send out their contributions according to the activations and strengths of output connections. A network of such units interconnected with each other can produce interesting effects. It can capture almost any kind of nonlinear effects of input variables and interactions among them without explicitly so instructed (e.g., Hornik, et al., 1989). Recent results by Barron (1993) also indicated that networks' approximation is particularly attractive for functions with high dimensional inputs.

The CC learning network is capable of dynamically growing nets (Fahlman & Lebiere, 1990). It starts as a net without hidden units, and it adds hidden units to improve its performance until a satisfactory degree of performance is reached. Thus, no a priori net topology has to be specified. Hidden units are added one at a time in such a way that all pre-existing units are connected to the new one. Input units are directly connected to output units (cross connections) as well as to all hidden units. The cross connections capture linear effects of input variables. Hidden units, on the other hand, produce nonlinear and interaction effects among the input variables, necessary for specific tasks. When a new hidden unit is recruited, the weights (representing connection strengths) associated with input connections are determined so as to maximize the correlation between residuals from network predictions at the stage and projected outputs from the recruited hidden unit, and are fixed throughout the rest of the learning process. This avoids the necessity of back-propagating error across different levels of the network, and leads to faster and more stable convergence. The weights associated with output connections are, however, re-estimated after a

[1]In the *Proceedings of the 1995 IEEE-SMC International Conference*, pp. 3594–3599

new hidden unit is recruited.

The CC architecture provides an interesting perspective on human development and learning (Shultz, Schmidt, Buckingham & Mareschal, 1995). Adjusting output weights assimilates new information into existing knowledge structures. Unassimilable information requires accommodation by changing existing knowledge structures. Changes in network topology can define distinct developmental stages. Fixed input weights imply that knowledge once learned cannot be erased, but can only be overridden by adding more hidden units.

3 The Learning of First and Second Person Pronouns

Learning first and second pronouns presents a psychologically interesting problem. When the mother talks to the child, *me* refers to the mother and *you* to the child. However, when the child talks to the mother, *me* refers to the child, and *you* to the mother. How children learn the shifting reference of these pronouns has extensively been studied by Oshima-Takane (1988, 1992) and her collaborators (Oshima-Takane, Goodz, & Derevensky, in press). The problem can be regarded as a special type of concept learning, where the concepts (or rules) to be learned are: Use *me* when the speaker and the referent agree, and use *you* when the addressee and the referent agree.

To analyze the task more closely, let us look at Table 1. There are three input variables: Speaker (S), Addressee (A) and Referent (R), and one output variable (Y) indicating the pronoun to be used. Let us assume that there are only three persons involved: Child, Mother and Father. The three input variables can take either one of these three values. There are two constraints, however: (1) S and A can never agree, and (2) either S and R should agree or A and R should agree; other patterns require pronouns other than *me* or *you*. These constraints limit the number of meaningful combinations to 12, which are shown in the table. It can be verified that the rules mentioned in the previous paragraph indeed hold for all the 12 patterns listed in the table. For example, when the father is talking to the child and refers to himself, he uses *me* (pattern 1), while when the mother is taking to the child and refers to the child, she uses *you* (pattern 4), etc. The child has to learn the three relevant input variables, and be able to identify which two of the three variables take identical values in particular situations.

The problem is equivalent to finding a function that connects the three input variables to the output variable. We assigned the values of 0, 2 and -2, respectively, to Child, Mother and Father on the three input variables, and the value of .5 to *me* and of $-.5$ to *you* on Y. The target function in this case is given by

$$Y = (A - R)/(A - S) - 0.5 \qquad (1)$$

It can easily be verified that when $S = R$, $Y = .5$, and when $A = R$, $Y = -.5$. Figs. 1a & 1b present graphical displays of the target function. Fig. 1a depicts the *me* surface, where the x-axis represents the addressee dimension, and the y-axis both the speaker and the referent which should agree. Fig. 1b depicts the *you* surface, where the x-axis now represents both the addressee and the referent which should agree, and the y-axis only the speaker. The z-axis in both figures represents the output variable, Y, that is .5 for *me* and $-.5$ for *you*. The surfaces were drawn for the values of S, A and R between -3.5 and 3.5 inclusive in steps of .5. Only the grid points defined by combinations -2, 0 and 2 on the x- and y-axes (excluding the diagonal points where the S and A agree) are used as training points. The remaining points are not used in the training. The four addressee patterns, two in each of the two figures, are on the lines labelled 0 on the x-axis, and the four child-speaking patterns are on the lines labelled 0 on the y-axis, orthogonal to the addressee patterns. The four nonaddressee patterns are on the lines 45 degrees apart from both the addressee and the child-speaking patterns.

4 Simulation Studies

What is the purpose of simulation, if the target function is known as in the present case? To explain why, we will have to look at Table 1 again. There are three distinct groups of input patterns in the table. The first group, called addressee patterns, consists of patterns in which the addressee is always Child. The second group, called nonaddressee patterns, consists of patterns in which Child is neither the speaker nor the addressee. The third group, called child-speaking patterns, consists of patterns in which the speaker is always Child. Oshima-Takane (1988) hypothesized that relevant information necessary for learning the correct use of the pronouns is not provided in the speech addressed to the child, and that the child has to pay attention to overheard speech to learn their correct use. Her hypotheses have been empirically verified (Oshima-Takane, 1988, 1992; Oshima-Takane, et al., in press) in both experimental and observational studies. However, for obvious ethical reasons children cannot be tested under the pure addressee or nonaddressee condition. This is where simulation studies will be particularly important, because nets can be trained under these pure conditions. According to Oshima-Takane's hypotheses, nets will learn an incorrect function when trained with only addressee patterns, but arrive at a correct function when trained with nonaddressee patterns. The child-speaking patterns provide a test of whether the correct function is learned or not.

Two simulation studies were conducted. In the first study, nets were trained under the pure addressee condition in Phase 1, followed by the addressee patterns plus the child-speaking patterns in Phase 2. If indeed the nets learn an incorrect function under the pure addressee condition as hypothesized by Oshima-Takane (1988), Phase 2 requires further training to deal with the child-speaking patterns. In the second simulation study, nets were trained under the pure nonaddressee condition in Phase 1, followed by the nonaddressee patterns plus the child-speaking patterns. If indeed the pure nonaddressee patterns are the necessary and sufficient condition, Phase 2 training is not necessary; training could stop immediately without changing the function learned in Phase 1. We can depict the functions arrived at in each phase of each simulation study graphically, and see what sort of functions (correct or incorrect) were produced under what conditions.

5 Network Analyses

The top figures in Figs. 2a & 2b show the *me* and the *you* surfaces constructed under pure addressee training. Both surfaces correctly discriminate the two *me* addressee patterns from the two *you* addressee patterns. However, they do not correctly discriminate the four child-speaking patterns. As expected, Phase 2 training was necessary to deal with the child-speaking patterns. The bottom portions of Figs. 2a & 2b show the *me* and the *you* surfaces obtained during Phase 2 training. They now correctly discriminate the four child-speaking patterns (as well as the four addressee patterns). However, the derived surfaces are still quite disparate from the corresponding target surfaces. Figs. 3a & 3b display the surfaces obtained from the second simulation study. Here, Phase 1 training was done under the pure nonaddressee condition, followed by the nonaddressee patterns plus the child-speaking patterns in Phase 2. The two top surfaces correctly discriminate the four nonaddressee patterns, but do not correctly discriminate the four child-speaking patterns or the four addressee patterns. The *me* surface was already correct in Phase 1, and few changes were made in Phase 2. However, the *you* surface required Phase 2 training to accommodate the child speaking patterns. The addressee patterns were still incorrect for *you* even after Phase 2.

In order to further examine the hypothesis about the importance of nonaddressee speech, we included two additional persons in the simulation studies. The two additional persons were coded as 1 and -1. With five persons, the learning environment is substantially richer than before; there are eight addressee patterns and the same number of child-speaking patterns, but there are 24 nonaddressee patterns. The richer environment may facilitate learning. Figs. 4a & 4b show the two surfaces obtained under the 5-person pure nonaddressee condition. They correctly discriminate not only the nonaddressee patterns, but also the child-speaking patterns and the addressee patterns. Generalizations (function values at untrained points) also seem quite good. The whole surfaces look quite similar to the corresponding target surfaces. Few changes occur in the approximated surfaces in Phase 2, because they are already quite good in Phase 1. The corresponding surfaces obtained under the 5-person addressee condition (not shown here) exhibited similar characteristics to those obtained under the 3-person addressee condition.

The nonaddressee patterns are crucial for pronoun learning, indicating importance of overheard speech. However, three persons are not sufficient for immediate correct generalization. A richer environment involving more people helps to ensure that the child can produce pronoun correctly as soon as he starts to use them. Developmental changes in function approximations underlying the pronoun learning have been more closely investigated by Oshima-Takane, et al. (1995).

6 References

Barron, A. (1993). Universal approximation bounds for superpositions of a sigmoidal function. *IEEE Transactions on Information Theory*, **39**, 930-945.

Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. *Mathematical Control Signals Systems*, **2**, 303-314.

Fahlman, S.E., & Lebiere, C. (1990). The cascade correlation learning architecture. In D.S. Touretzky (Ed.), *Advances in neural information processing systems 2* (pp. 524-532). San Mateo: Morgan Kaufmann.

Hornik, M., Stinchcombe, M., & White, H. (1989). Multilayer feed-forward networks are universal approximators. *Neural Networks*, **2**, 359-366.

Oshima-Takane, Y. (1988). Children learn from speech not addressed to them: The case of personal pronouns. *Journal of Child Language*, **15**, 94-108.

Oshima-Takane, Y. (1992). Analysis of pronomial errors: A case study. *Journal of Child Language*, **19**, 111-131.

Oshima-Takane, Y., Goodz, E., & Derevensky, J.L. (in press). Birth order effects on early language development: Do second born children learn from overheard speech? *Child Development.*

Oshima-Takane, Y., Takane, Y., & Shultz, T.R. (1995). The learning of personal pronouns: Network models and analysis. Submitted to *NIPS 96.*

Shultz, T.R., Schmidt, W.C., Buckingham, D., & Mareshal, D. (1995). Modelling cognitive development with a generative connectionist algorithm. In T. Simon & G. Halford (Eds.), *Developing cognitive competence: New approaches to process modelling* (pp. 205-261). Hillsdale, NJ: Lawrence Erlbaum.

Takane, Y., Oshima-Takane, Y., & Shultz, T.R. (1994). Approximations of nonlinear functions by feed-forward neural networks. In N. Ohsumi (Ed.), *Proceedings of the Annual Meeting of the Japan Classification Society* (pp. 26-33). Tokyo: Japan Classification Society.

White, H. (1990). Connectionist nonparametric regression: Multilayer feedforward networks can learn arbitrary mappings. *Neural Networks*, **3**, 535-549.

(a) Me (b) You

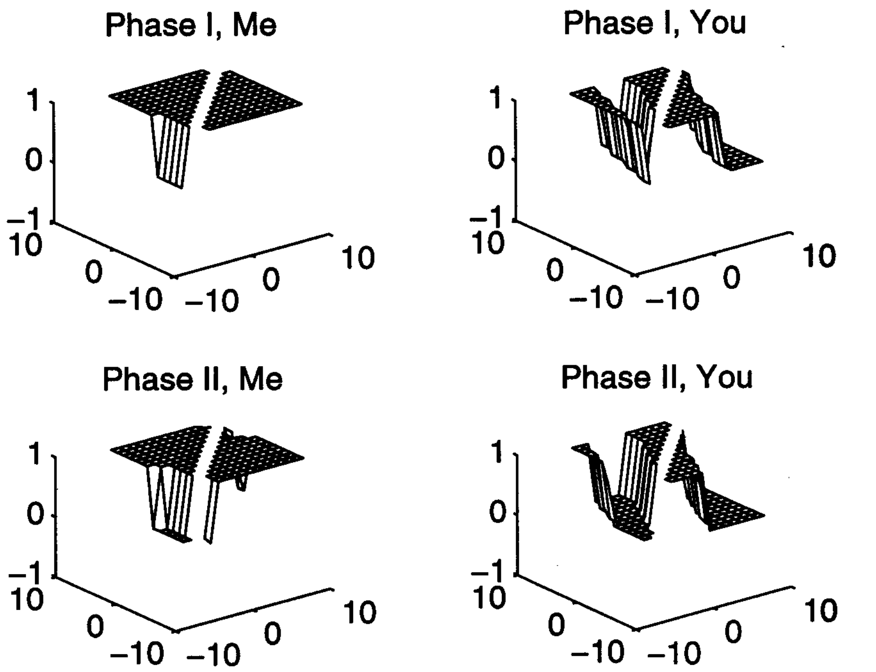

Figure 3: Approximated Functions in the Three-Person Nonaddressee Condition

(a) Me (b) You

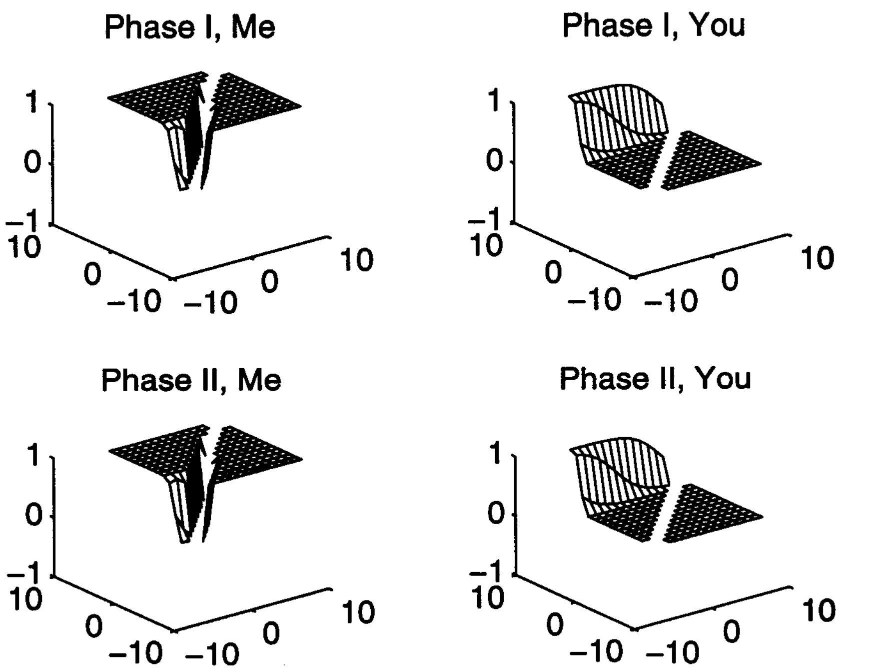

Figure 4: Approximated Functions inthe Five-Person Nonaddressee Condition

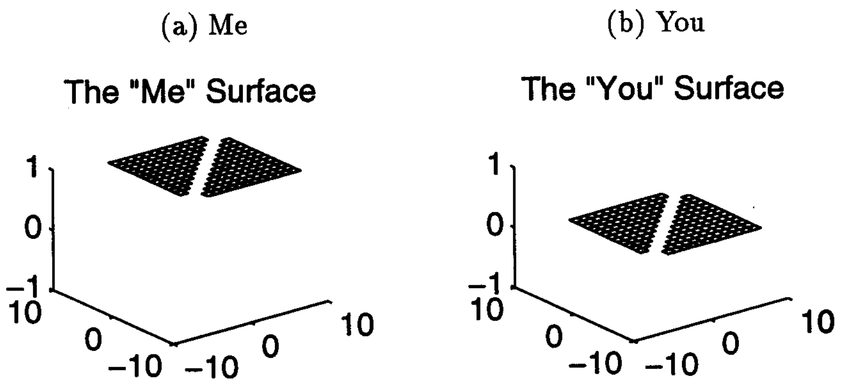

Figure 1: The Target Function

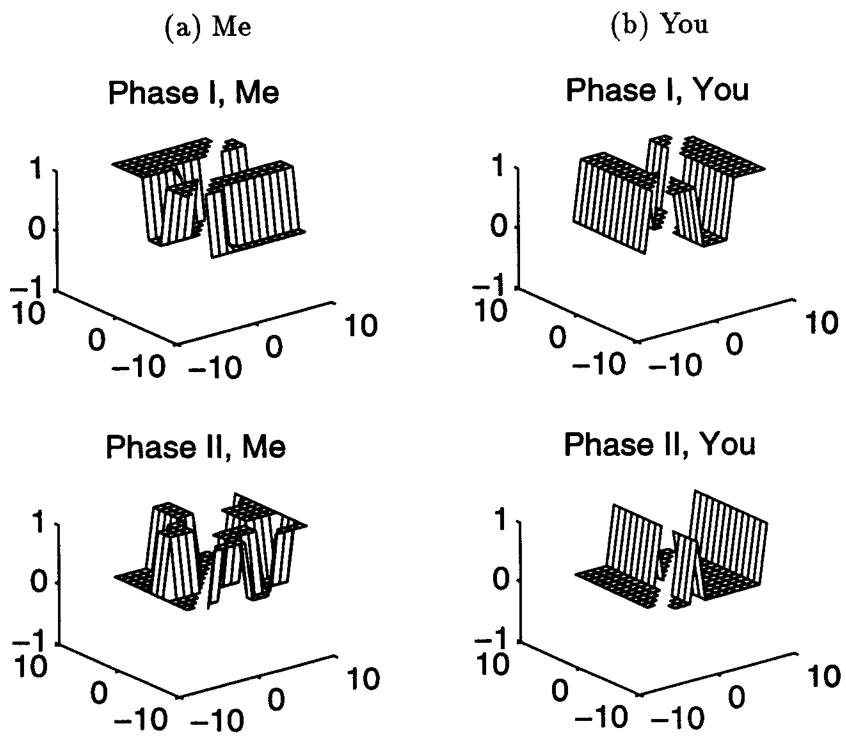

Figure 2: Approximated Functions in the Three-Person Addressee Condition

Table 1: Training Patterns in the 3-Person Situation

Condition		Input Variables			Output Variables
		Speaker	Addressee	Referent	Pronoun
Phase 1:					
Pure addressee (AD)	1)	Father	Child	Father	me
	2)	Father	Child	Child	you
	3)	Mother	Child	Mother	me
	4)	Mother	Child	Child	you
Pure nonaddressee (NA)	5)	Father	Mother	Father	me
	6)	Father	Mother	Mother	you
	7)	Mother	Father	Mother	me
	8)	Mother	Father	Father	you
Phase 2:					
Child-speaking patterns (CS)	9)	Child	Father	Child	me
	10)	Child	Father	Father	you
	11)	Child	Mother	Child	me
	12)	Child	Mother	Mother	you