From the Proceedings of the14th annual conference of the Cognitive Science Society, 1992.

Direct, Incremental Learning of Fuzzy Propositions

Gregg C. Oden
Departments of Psychology and Computer Science
University of Iowa, Iowa City, IA 52242
oden@cs.uiowa.edu

Abstract

To enable the gradual learning of symbolic representations, a new fuzzy logical operator is developed that supports the expression of negation to degrees. As a result, simple fuzzy propositions become instantiable in a feedforward network having multiplicative nodes and tunable negation links. A backpropagation learning procedure has been straightforwardly developed for such a network and applied to effect the direct, incremental learning of fuzzy propositions in a natural and satisfying manner. Some results of this approach and comparisons to related approaches are discussed as well as directions for further extension.

Introduction

Over the past couple of decades, a wide array of cognitive phenomena have been successfully modeled within a fuzzy propositional theoretical framework[1] (see Massaro, 1987; Oden, 1984, Oden, Rueckl, & Sanocki, 1991 for reviews). The development of this approach to cognitive modeling was motivated by many of the same considerations that underlie current interest in neural information processing systems: in particular, the attainment of robustness and of graceful degradation under duress. In each case, the desired end is achieved largely through reliance on coarse coding, automatic generalization, compensatory information integration, and other consequences of continuous computation. Thus, although the two approaches lie on opposite sides of the symbolic/subsymbolic boundary and might thereby be supposed to be incompatible, they can in fact be seen to be members of the same, more general family of models. From this perspective, it is not surprising that instances of each class of models can be shown to be formally isomorphic under specific common conditions (see Massaro & Cohen, 1987 and Oden, 1988 for two such results). Indeed, facts such as these have been used to support the argument that the two approaches represent separate necessary levels of description of cognitive systems (Oden, 1988; see also Clark, 1989).

The present paper extends this argument by demonstrating how a connectionist learning procedure can be directly and naturally applied within the fuzzy propositional level. This speaks specifically to the common criticism made of symbolic approaches that learning must be an all-or-none process that would require a seemingly magical, external process to wire up new connections.

[1] Hereafter referred to as FuzzyProp for short. The most common FuzzyProp model is the Fuzzy Logical Model of Perception or FLMP (e.g., Massaro & Cohen, 1991; Oden & Massaro, 1978), which is based on the hypothesis of independent evaluation of conjunctive terms.

Learnable Fuzzy Propositions

A fuzzy proposition is a logical expression having component terms that may be more or less true of an object and connectives that are continuous functions of their component terms reflecting the essential logical properties of conjunction, negation, and so on. As applied to the modeling of cognitive processes, fuzzy propositions represent the knowledge that people have about patterns and categories and they provide a basis for evaluating stimuli in a way that fully exploits the information inherent in the systematic continuous variation of stimulus properties. For example, in modeling the identification of handwritten words (Oden & Rueckl, in preparation), the degree to which the initial portion of some stimulus constitutes a lower case letter 'e' involves an evaluation of the proposition that it is a loop that is not too tall:

$$t[\mathrm{loop}(x) \wedge \neg \mathrm{tall}(x)] = t[\mathrm{loop}(x)] \times \{1 - t[\mathrm{tall}(x)]\} \qquad (1$$

using (as in all of our work) multiplication to represent fuzzy conjunction[2].

[2]Multiplication is conjunctive in that (a) it yields a value of true (1.0) only if both of its terms are true and a value of false (0.0) if either or both are false, and (b) it has many (arguably the most essential) properties of conjunction such as associativity, commutativity and so on. Importantly, unlike the more common fuzzy conjunction function, $t[A \wedge B] = \min\{t[A],[B]\}$, multiplication is compensatory, meaning that it allows positive and negative errors to cancel. Much of the robustness of FuzzyProp results from the use of multiplicative conjunction and this is what distinguishes it from most other fuzzy approaches including those that are typically used in constructing fuzzy / neural systems.

It would be advantageous if the knowledge respesented by such propositions could be learned in a gradual or incremental fashion over the course of experience with instances of the concept or relation. To make such propositions incrementally learnable requires some mechanism for gradually converting conjunctions into disjunctions and vice versa. There are many possible ways to do this, but most seem ungainly and ad hoc (e.g., by defining a tunable generic connective as the weighted average of conjunctive and disjunctive expressions). The present approach is to rely on the expressability of disjunctions in conjunctive form through DeMorgan's Law: $A \vee B = \neg(\neg A \wedge \neg B)$. This converts the problem into one of devising a tunable negation operator: a variable connective that allows any term to be negated to some degree. Again, there are many possibilities. For reasons outlined below, the present work makes use of the following rule

$$t[\mathbf{N}_v\, A(x)] = \frac{a^v}{a^v + (1-a)^v} \qquad (2$$

where $a = t[A(x)]$, the degree to which predicate A is true of object x. This operator has a number of attractive properties. The most critical properties, of course, are that of reducing to the identity function for $v = 1$, to standard fuzzy negation for $v = -1$, to a nulling value — a value not dependent on $t[A(x)]$ — for $v = 0$, and to reasonable intermediate functions in between. In addition, as a bonus, the operator yields contrast intensified functions for v values beyond ± 1. Thus, the operator can better be thought of as a generalized transfer function that remaps input truth values onto output truth values. Figure 1 plots this function for several representative values of v.

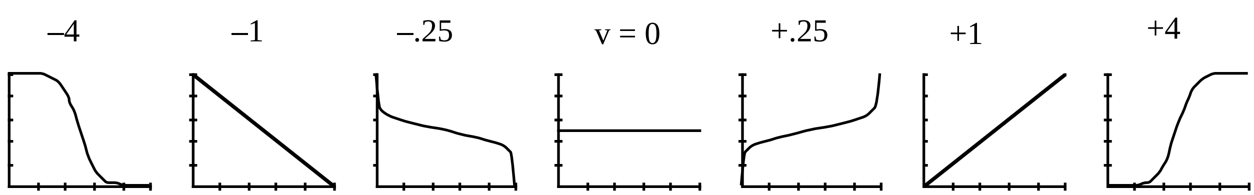

Figure 1. The $\mathbf{N}_v$ operator function for various values of v.

The operator is a natural extension of the version of fuzzy logic employed here (see footnote 2) in the sense that its algebraic structure has a direct natural interpretation in terms of basic fuzzy components (Oden, 1984)[3].

Backprop in FuzzyProp

So far, we have established the instantiability of arbitrary simple fuzzy propositions in a multilayer feedforward network having conjunctive (multiplicative) nodes with tunable generalized (negatable) links connecting nodes of successive layers. Such networks are directly analogous to standard feedforward networks (Rumelhart, Hinton, & Williams, 1986) with the v parameters of the tunable links corresponding to the weights (including those serving as node bias terms) of the standard model. Accordingly, the backward error propagation learning procedure for fuzzy propositions directly follows the form of that laid out by Rumelhart et al: Given a training set of tuples of input and desired output values,

1. evaluate the propositions on the input
2. compute a measure of error between obtained and desired output
3. adjust each v in proportion to the derivative of the error measure with respect to that v, recursively computed.

The calculation of the derivatives is just slightly more complicated in the present case compared to the standard case, in essence because the nonlinearity in the fuzzy propositional system occurs between every pair of nodes from successive layers whereas in standard backprop it occurs just once for each node in the form of the squashing function applied to the output for that node.

[3] As discussed in Oden (1992), it is also algebraically natural in the sense that it is the powering operation for the Abelian (commutative) group defined on [0..1] by the mapping $x \rightarrow x/(x+1)$ from the multiplicative group on $[0..\infty]$.

Initial tests of this learning procedure demonstrate that it, indeed, performs as it should. For example, when applied to the ever popular test case, XOR, it learns the function forthrightly as indicated by the average learning curve for a representative sample of runs shown in Figure 2.

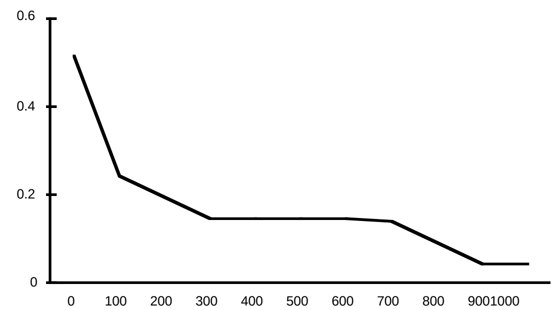

Figure 2. The course of learning XOR for 10 test runs. The root mean squared error is plotted against the number of learning epochs.

The overall form of this curve primarily reflects the fact that most of the test runs in this particular sample resulted in complete learning within a few hundred training epochs or so, a couple required around 800 epochs and one did not complete the learning until 3700 epochs (but then did learn it exactly). The median number of training epochs required to reach a criterion root mean squared error less than .01 was about 250 and the median root mean squared error after 300 epochs was .008. For the sake of comparison, note that these numbers are close to the same number of epochs reported (Rumelhart, Hinton, & Williams, 1986) to be required by standard back propagation to reach a root mean squared error of about .1 The present system needed only about 150 epochs to reach .1 root mean squared deviation.

In these tests, the system is provided with two input nodes, one output node and two intermediate layers having two and one node respectively (2-2-1-1). Thus, altogether the system is allowed seven v values to adjust, precisely the same as the minimum number (and two fewer than the typical number) of free parameters required for this problem in the original Rumelhart et al (1986) paper including both weights and bias terms. This fuzzy propositional network is sufficient for representing XOR as $(A \wedge \neg B) \vee (\neg A \wedge B)$. Sometimes, however, XOR is learnt as $(A \vee B) \wedge \neg(A \wedge B)$ by the system. This is equivalent to the former expression with respect to the input and output values used, which are all (close to) zero and one, but would not be exactly equivalent for intermediate truth values. This form of XOR is actually more compact and only really needs six parameters.

Comparisons and Extensions

The main distinctive feature of the current approach is learned representations that are directly interpretable logical functions of the input variables. In addition, as with other models having multiplicative units such as sigma-pi networks (Rumelhart, Hinton, & McClelland, 1986) or the product unit nets of Durbin and Rumelhart (1989), learning of logical functions may be faster in the present system than with standard backprop. This is due in part to the fact that the v parameters do not have to approach $\pm\infty$ in order to yield outputs close to 0 and 1. Indeed, with respect to feedforward processing in the system, inputs and outputs can take on the values of 0 and 1 exactly. (During the error backpropagation phase, these extreme values must be avoided because the relevant derivatives would be undefined.)

On theoretical grounds, the present approach is interestingly similar to and different from each of the approaches mentioned above in a number of ways. For now, let's just consider the representation of a conjunction of inputs by the standard backpropagation feedforward network in comparison with that of the present system. Figure 3 portrays this relationship in a couple of ways. On the left is a 3D plot and a contour map showing how the standard approach manages to be conjunctive, basically by applying a one dimensional nonlinear cut across the axis corresponding to the sum of the inputs (the diagonal of the 'floor' in the 3D plot). On the right are the corresponding representations for the fuzzy propositional system, which reveal that this approach more directly captures the notion of conjunction as encompassing the vicinity of the (1, 1) corner. This is a direct result of the fact that, as noted above, nonlinearity is more thoroughly ingrained in this system. It is, of course, no accident that the fuzzy propositional system is more naturally conjunctive in this sense, since it is fundamentally logic-based

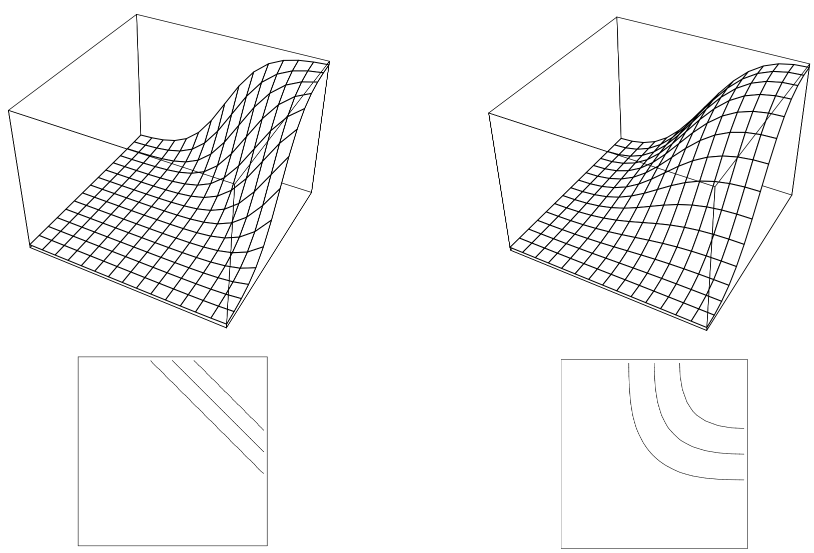

Figure 3. Plots and contour maps of conjunctive functions of two inputs for standard backprop networks (left panels) and the present fuzzy propositional system (right panels).

in structure by design. In contrast, the standard network could be thought of as only emulating the logical functions within a system that does not have an inherently logic-based structure.

The current work is related in quite a different fashion to approaches to the learning of associative relations between fuzzy terms (e.g., Jenison & Oden, 1989; Kosko, 1992). Such techniques are complementary to the one developed here and might serve, for example, to set up the analyses necessary for the evaluation of the primitive terms in the propositions of the present system.

The scheme described here can be variously extended. Inertial terms and other factors analogous to those used with standard backprop can clearly be added and other learning methods from the connectionist literature could be similarly imported into this system. More interestingly, the $\mathbf{N}_V$ operator (Equation 2) can be generalized in at least two reasonable ways: (a) by allowing separate v values for the positive and negative components of the relative ratio expression, and (b) by including an overall exponential weighting factor. Both of these generalizations have natural interpretations in fuzzy terms and yield a significant enrichment of the expressiveness of the resulting fuzzy calculus. Both can be incorporated in the backprop learning scheme with very little complication. Yet another straightforward extension of $\mathbf{N}_V$ (see Oden, 1992) enables it to perform a kind of running average of inputs in order to account for processing dynamics as in Massaro and Cohen (1991).

Conclusion

The overarching moral of this work is that one can have it all. That is, it is not necessary to choose between having symbolic expressions and direct, incremental learning procedures. The two can be combined in a natural and harmonious fashion.

References

Clark, A. 1989. *Microcognition: Philosophy, cognitive science, and parallel distributed processing*. Cambridge, MA: MIT Press.

Durbin, R., and Rumelhart, D. E. 1989. Product units: A computationally powerful and biologically plausible extension to backpropagation networks. *Neural Computation*, 1:133-142.

Jenison, R. L., and Oden, G. C. 1990. Fuzzy implication formation in distributed associative memory. *Proceedings of the Twelfth Annual Conference of the Cognitive Science Society*, 860–867.

Kosko, B. 1992. *Neural networks and fuzzy systems: A dynamical systems approach to machine intelligence*. Englewood Cliffs, NJ: Prentice-Hall.

Massaro, D. W. 1987. *Speech perception by ear and by eye: A paradigm for psychological inquiry*. Hillsdale, NJ: Erlbaum.

Massaro, D. W., and Cohen, M. M. 1987. Process and connectionist models of pattern recognition. *Proceedings of the Ninth Annual Conference of the Cognitive Science Society*, Seattle, WA.

Massaro, D. W., and Cohen, M. M. 1991. Integration versus interactive activation: The joint influence of stimulus and context in perception. *Cognitive Psychology*, 23:558-614.

Oden, G. C. 1984. Integration of fuzzy linguistic information in language comprehension. *Fuzzy Sets and Systems*, 14:29-41.

Oden, G. C. 1988. FuzzyProp: A symbolic superstrate for connectionist models. *Proceedings of the Second IEEE International Conference on Neural Networks*, Vol. 1, 293-300.

Oden, G. C. 1992. An algebra on [0..1], with application to fuzzy values. *Cognotes*, 5.

Oden, G. C., and Massaro, D. W. 1978. Integration of featural information in speech perception. *Psychological Review*, 85:172-191.

Oden, G. C., and Rueckl, J. G. In preparation. Taking language by the hand: Reading hand-written words.

Oden, G. C., Rueckl, J. G., and Sanocki, T. 1991. Making sentences make sense, or words to that effect. In G. B. Simpson (Ed.) *Understanding word and sentence*. Amsterdam: North-Holland.

Rumelhart, D. E., Hinton, G. E., and McClelland, J. L. 1986. A general framework for parallel distributed processing. In D. E. Rumelhart and J. L. McClelland (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition*, Vol. 1. Cambridge, MA: MIT Press.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). Learning internal representations by error propagation. In D. E. Rumelhart and J. L. McClelland (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition*, Vol. 1. Cambridge, MA: MIT Press.