Paper submitted to the Cognitve Science Meeting, Seattle, 1987.
Comments appreciated.

PROCESS AND CONNECTIONIST MODELS OF PATTERN RECOGNITION

DOMINIC W. MASSARO & MICHAEL M. COHEN

Program in Experimental Psychology, University of California,
Santa Cruz, California 95064 U.S.A.

Abstract

The present paper explores the relationship between a process/mathematical model and a connectionist model of pattern recognition. In both models, pattern recognition is viewed as having available multiple sources of information supporting the identification and interpretation of the input. The results from a wide variety of experiments have been described within the framework of a fuzzy logical model of perception. The assumptions central to this process model are 1) each source of information is evaluated to give the degree to which that source specifies various alternatives, 2) the sources of information are evaluated independently of one another, 3) the sources are integrated to provide an overall degree of support for each alternative, and 4) perceptual identification and interpretation follows the relative degree of support among the alternatives. Connectionist models have been successful at describing the same phenomena. These models assume interactions among input, hidden, and output units that activate and inhibit one another. Similarities between the frameworks are described, and the relationship between them explored. A specific connectionist model with input and output layers is shown to be mathematically equivalent to the fuzzy logical model. It remains to be seen which framework serves as the better heuristic for psychological inquiry.

Introduction

A growing consensus in pattern recognition is that there are multiple sources of information that the perceiver evaluates and integrates to achieve perceptual recognition. Consider recognition of the word *performance* in the spoken sentence

The actress was praised for her outstanding performance.

Recognition of the word is achieved via a variety of bottom-up and top-down sources of information. Top-down sources include semantic, syntactic, and phonological constraints and bottom-up sources include audible and visible features of the spoken word (Massaro, in press a, b).

A Fuzzy Logical Framework for Pattern Recognition

According to the this framework, well-learned patterns are recognized in accordance with a general algorithm, regardless of the modality or particular nature of the patterns (Massaro, 1979; 1984a, 1984b, in press b; Oden, 1978, 1981). The model has received support in a wide variety of domains and consists of three operations in perceptual (primary) recognition: feature evaluation, feature integration, and pattern classification. Continuously-valued features are evaluated, integrated, and matched against prototype descriptions in memory, and an identification decision is made on the basis of the relative goodness of match of the stimulus information with the relevant prototype descriptions. The model is called a fuzzy logical model of perception (abbreviated FLMP).

Central to the FLMP are summary descriptions of the perceptual units (Oden & Massaro, 1978). These summary descriptions are called prototypes and they contain a conjunction of various properties called features. A prototype is a category and the features of the prototype correspond to the ideal values that an exemplar should have if it is a member of that category. The exact form of the representation of these properties is not known and may never be known. However, the memory representation must be compatible with the sensory representation resulting from the transduction of the input. Compatibility is necessary because the two representations must be related to one another. To recognize an object, the perceiver must be able to relate the information provided by the object itself to some memory of the object category.

Prototypes are generated for the task at hand. The sensory systems transduce the physical event and make available various sources of information called features. During the first operation in the model, the features are evaluated in terms of the prototypes in memory. For each feature and for each prototype, feature evaluation provides information about the degree to which the feature in the speech signal matches the corresponding feature value of the prototype.

Given the necessarily large variety of features, it is necessary to have a common metric representing the degree of match of each feature. Two features must share a common metric if they eventually are going to be related to one another. To serve this purpose, fuzzy truth values (Zadeh, 1965) are used because they provide a natural representation of the degree of match. Fuzzy truth values lie between zero and one, corresponding to a proposition being completely false and completely true. The value .5 corresponds to a completely ambiguous situation whereas .7 would be more true than false and so on. Fuzzy truth values, therefore, not only can represent continuous rather than just categorical information, they also can represent different kinds of information. Another advantage of fuzzy truth values is that they couch information in mathematical terms (or at least in a quantitative form). This allows the natural development of a quantitative description of the phenomenon of interest.

Feature evaluation provides the degree to which each feature in the stimulus matches the corresponding feature in each prototype in memory. The goal, of course, is to determine the overall goodness of match of each prototype with the stimulus. All of the features are capable of contributing to this process and the second operation of the model is called feature integration. That is, the features (actually the degrees of matches) corresponding to each prototype are combined (or conjoined in logical terms). The outcome of feature integration consists of the degree to which each prototype matches the stimulus. In the model, all features contribute to the final value, but with the property that the least ambiguous features have the most impact on the outcome.

The third operation during recognition processing is pattern classification. During this stage, the merit of each relevant prototype is evaluated relative to the sum of the merits of the other relevant prototypes. This relative goodness of match gives the proportion of times the stimulus is identified as an instance of the prototype. The relative goodness of match could also be determined from a rating judgment indicating the degree to which the stimulus matches the category. The pattern classification operation is modeled after Luce's (1959) choice rule. In pandemonium-like terms (Selfridge, 1959), we might say that it is not how loud some demon is shouting but rather the relative loudness of that demon in the crowd of relevant demons. Two important predictions of the model are 1) two features can be more informative than just one and 2) a given feature has a greater effect to the extent a second feature is ambiguous.

Relationship to Connectionist Models

The framework provided by the FLMP anticipated many of the distinguishing properties of new connectionism (Massaro, 1986a, 1986b; Oden & Rueckl, 1986). A connectionist model of perception (CMP) also is an information-processing system having and manipulating information (McClelland & Rumelhart, 1986). The information is represented in terms of the activations and inhibitions of neural-like units. The units are assumed to exist at different levels; for example, the TRACE model of speech perception (McClelland & Elman, 1986) consists of units at the feature, phoneme, and word levels. The units interact with one another via connections among the units. The connectivity is implemented by positive and negative weights that are either specified in advance or learned through feedback.

A prototypical connectionist framework shares several fundamental properties with the current theoretical framework as instantiated in the FLMP. First, both frameworks assume continuous rather than discrete representations; the fuzzy truth values of the FLMP are analogous to the continuous levels of activation and inhibition of connectionist models. Second, both frameworks acknowledge the existence of multiple simultaneous constraints on human performance. Both frameworks provide an account of the evaluation and integration of multiple sources of information in pattern recognition. Third, there is the parallel assessment of multiple candidates or hypotheses at multiple levels in both models. Fourth, both frameworks provide a common metric for relating qualitatively different sources of information. In the FLMP, each source of information is represented by fuzzy truth values representing the degree to which alternative hypotheses are supported. Activation level plays the analogous role in connectionist models. Fifth, the automatic categorization of a novel instance can be accomplished in both frameworks. Finally, both frameworks conceptualize pattern recognition as finding the best fit between the relevant constraints and the pattern that is perceived.

The close fit between the present framework and connectionism dictates an exploration of their similarities and differences. Although the two frameworks appear to agree on important theoretical criteria, the specific models to date differ in terms of the amount of connectivity in the system. The FLMP assumes no top-down influences of a higher-level unit on activation of a lower-level unit and no inhibition among units at a given level. Connectionist models, such as the interactive activation models of written word recognition and speech perception, usually make both of these assumptions. As presently formulated, many of the connectionist models with two-way connections among different levels of units and connectivity among units at a given level are too powerful. They are capable of predicting not only observed results but also results that do not occur (Massaro, 1986a). That is, some connectionist models can simulate results that have not been observed in psychological investigations and results generated by incorrect process models of performance (Massaro, in preparation).

Mathematical Equivalence of Two Models

It can be shown that the FLMP makes mathematically equivalent predictions to those made by a two-layer CMP, with input and output units. As in all instantiations of a theory, particular assumptions must be made about the description of the results of interest. Different assumptions would probably change the relationship between the two models. The models are compared in an expanded factorial designs in which two or more dimensions of information are varied independently of one another in a pattern recognition task. Each of the dimensions is also presented alone. Labeling the dimensions as X and Y, X_i would correspond to the ith level of the X dimension. Similarly, Y_j would correspond to the jth level of the Y dimension. A given stimulus composed of a single dimension would be labeled X_i or Y_j, and a given combination would be represented by $X_i\ Y_j$.

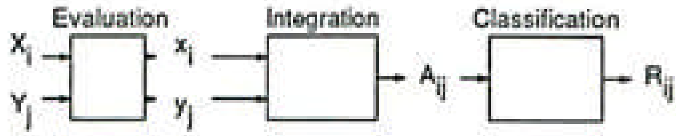

Figure 1. Schematic representation of the three operations involved in perceptual recognition, according to the fuzzy logical model of perception.

Figure 1 illustrates the three stages involved in pattern recognition. The sources of information are represented by uppercase letters. The evaluation process transforms these into psychological values (indicated by lowercase letters) that are then integrated to give an overall value. The classification operation maps this value into some response, such as a discrete decision or a rating.

The FLMP assumes three operations between presentation of a pattern and its categorization, as illustrated in Figure 1. Feature evaluation gives the degree to which a given dimension supports each test alternative. The physical input is transformed to a psychological value, and is represented in lowercase. For a given response alternative A_{ij}, X_i would be transformed to x_i, and analogously for dimension Y_j. Each dimension provides a feature value at feature evaluation. Feature integration consists of a multiplicative combination of feature values supporting a given alternative A_{ij}. If x_i and y_j are the values supporting alternative A_{ij}, then the total support for the alternative A_{ij} would be given by the product $x_i\ y_j$.

$$A_{ij} \quad : \quad x_i\, y_j.$$

The third operation is pattern classification, which gives the relative degree of support for each of the test alternatives. In this case, the probability of an A_{ij} response given $XiYj$ is

$$P(A_{ij} \mid X_i Y_j) = \frac{x_i\ y_j}{\sum} \tag{1}$$

where $\sum$ is equal to the sum of the merit of all relevant alternatives, derived in the same manner as illustrated for alternative A_{ij}.

The CMP is assumed to have an input layer and an output layer, with all input units connected to all output units. It is assumed that each level of each dimension is represented by a unique unit at the input layer. Each response alternative is represented by a unique unit at the output layer. Figure 2 gives a schematic representation of two input units connected to a single output unit.

An input unit has zero input, unless its corresponding level of the stimulus dimension is presented. Presentation of an input unit's target stimulus gives an input of one. The activation of an output unit by an input unit is given by the multiplicative combination of the input activation and a weight w. With two active inputs X_i and Y_j, the activation entering output unit A would be

$$A_{ij} \quad : \quad x_i + y_j$$

where $x_i = w_i\ X_i$ and $y_j = w_j\ Y_j$. The total activation leaving an output unit is given by the sum of the input activations, passed through a sigmoid squashing function (McClelland & Rumelhart, 1986).

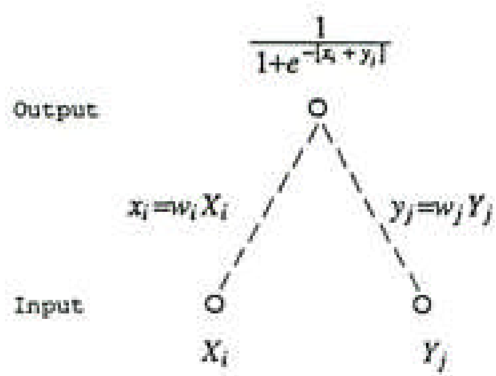

Figure 2. Illustration of connectionist model with two input units and one output unit.

$$A_{ij} \quad : \quad \frac{1}{1+e^{-[x_i + y_j]}}$$

A connectionist model does not specify completely the input-output relationship. The output activations have to be mapped into a response, and Luce's choice rule is usually assumed to describe this mapping (McClelland & Rumelhart, 1985). Taking this tack, the activation A_{ij} transformed into a response probability by Luce's choice rule gives

$$P(A_{ij} \mid X_i \; Y_j) = \frac{\frac{1}{1+e^{-[x_i + y_j]}}}{\sum} \tag{2}$$

where $\sum$ is equal to the sum of the activations of all relevant outputs, derived in the manner illustrated for alternative A_{ij}.

The FLMP does not specify the psychophysical relationship between the physical stimulus and its sensory transformation. Neither does a connectionist model; both models require free parameters to specify this relationship. The free parameters are weights in the connectionist model and truth values in the FLMP. A unique weight is assumed for each level of each dimension in the CMP, and a unique truth value is required for each level of each dimension in the FLMP. Thus, the same number of free parameters is required by the two models. The number of free parameters is equal to the number of levels of the X dimension plus the number of levels of the Y dimension. Although a threshold unit is sometimes assumed in connectionist models, no such unit is assumed here. We also have Luce's choice rule operating for both the CMP and the FLMP. In this case, a formal equivalence between the two models exists if adding the weighted activations at input and transformed by the sigmoid squashing function is mathematically equivalent to multiplying fuzzy truth values. Given that the CMP's activated X_i and Y_j input units are equal to one, the activations entering an output unit are equal to w_i and w_j. It follows that the activation of an output unit in the CMP is predicted to be $\frac{1}{1+e^{-[w_i + w_j]}}$. The degree of support for a given test alternative for the FLMP is equal to $x_i \; y_j$.

The truth values in the FLMP are constrained between zero and one, following the assumption of fuzzy logic (Zadeh, 1965). Accordingly, $x_i \; y_j$ must lie between zero and one. The sigmoid squashing function also takes on values only between zero and one, even though the weights are unbounded. It follows that the models can make mathematically identical predictions because 1) for every x_i, there is a w_i, and 2) for every y_j, there is a w_j such that $\frac{1}{1+e^{-[w_i + w_j]}}$ equals $x_i \; y_j$. It can be shown that there exists a correspondence between these predictions such that equivalent predictions can be made by the two models.

$$x_i\, y_j = \frac{1}{1+e^{-[w_i + w_j]}} \tag{3}$$

The proof of the above equivalence is most obvious for the single-dimension conditions of the expanded factorial design. There exists a unique relationship between the weights in the CMP and the truth values in the FLMP if an expanded factorial design is used. In this case, it can be proved that $\frac{1}{1+e^{-[w_i]}}$ equals x_i and $\frac{1}{1+e^{-[w_j]}}$ equals y_j. Equivalently, weight w_i equals $-\ln\left(\frac{1}{x_i} - 1\right)$. Data from an expanded factorial design always give only one set of parameters for the FLMP, and also force the CMP to come up with a unique set of mathematically equivalent weights. Given this equivalence, we can translate directly between the two kinds of parameters. We might argue also that the truth values are more informative in the FLMP analysis because it is easy to conceptualize values between 0 and 1, and the truth value gives the contribution of a source of information uncontaminated by other sources. This latter feature is another value of independence models relative to models with high interconnectivity, in which the contribution of one source can not be pulled apart from the contribution of other sources.

Similar predictions exist for three or more stimulus dimensions and three or more response alternatives. Increasing the number of response alternatives does not change the relationship between the two models because this increase only affects the number of outputs, and these are handled equivalently by Luce's choice rule in both models. Increasing the number of dimensions adds the same number of terms to both models, preserving the equivalence shown in Equation 3. In the FLMP, the three dimensions of support for alternative A_{ij} would be

$$A_{ij} \quad : \quad x_i\, y_j\, z_k$$

In the CMP, the activation of three input units would give

$$A_{ij} \quad : \quad \frac{1}{1+e^{-[x_i + y_j + z_k]}}$$

where $x_i = w_i\, X_i$, $y_j = w_j\, Y_j$, and $z_k = w_k\, Z_k$. The total activation of an output unit is given by the sum of the three input activations passed through the sigmoid squashing function, and so on for a larger number of inputs.

The FLMP specifies mathematically evaluation and integration processes. The CMP implements evaluation and integration by activations and inhibitions between input units and output units. Evaluation corresponds to the activation along a single connection between an input unit and an output unit. Integration in the CMP corresponds to the sum of all the activations entering a given output unit, and transformed by the sigmoid squashing function. The correspondence between the FLMP and CMP reveals that the two models, couched in different theoretical frameworks, can make identical predictions in practice. A remaining issue is how process and connectionist models differ from one another, and whether there is an advantage of one over the other.

References

Luce, R. D. (1959). *Individual choice behavior*. New York: Wiley.

Massaro, D. W. (1979). Reading and listening (Tutorial paper). In P. A. Kolers, M. Wrolstad, & H. Bouma (Eds.), *Processing of visible language: Vol. 1* (pp. 331-354). New York: Plenum.

Massaro, D. W. (1984a). Building and testing models of reading processes. In P. D. Pearson (Ed.), *Handbook of reading research* (pp. 111-146). New York: Longman.

Massaro, D. W. (1984b). Time's role for information, processing, and normalization. *Annals of the New York Academy of Sciences, Timing and Time Perception, 423*, 372-384.

Massaro, D. W. (1986a, November). *Connectionist models of mind.* Paper given at the twenty-seventh annual meeting of the Psychonomic Society, New Orleans.

Massaro, D. W. (1986b). The computer as a metaphor for psychological inquiry: Considerations and recommendations. *Behavior Research Methods, Instruments, & Computers, 18*, 73-92.

Massaro, D. W. (in press a). Integrating multiple sources of information in listening and reading. In D. A. Allport, D. G. MacKay, W. Prinz, & E. Scheerer (Eds.), *Language Perception and Production: Shared Mechanisms in Listening, Speaking, Reading and Writing.* London: Academic Press.

Massaro, D. W. (in press b). *Speech perception by ear and eye: A paradigm for psychological inquiry.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Massaro, D. W., & Cohen, M. M. (1983). Evaluation and integration of visual and auditory information in speech perception. *Journal of Experimental Psychology: Human Perception and Performance, 9*, 753-771.

Massaro, D. W., & Oden, G. C. (1980). Speech perception: A framework for research and theory. In N. J. Lass (Ed.), *Speech and language: Advances in basic research and practice: Vol. 3* (pp. 129-165). New York: Academic Press.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18*, 1-86.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part I. An account of basic findings. *Psychological Review, 88*, 375-407.

McClelland, J. L., & Rumelhart, D. E. (1985). Distributed memory and the representation of general and specific information. *Journal of Experimental Psychology: General, 114* 159-188.

McClelland, J. L., & Rumelhart, D. E. (1986). *Parallel distributed processing* Cambridge: MIT press.

Oden, G. C. (1978). Semantic constraints and judged preference for interpretations of ambiguous sentences. *Memory & Cognition, 6*, 26-37.

Oden, G. C. (1981). A fuzzy propositional model of concept structure and use: A case study in object identification. In G. W. Lasker (Ed.), *Applied systems and cybernetics: Vol. VI* (pp. 2890-2897). Elmsford, NY: Pergamon Press.

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

Oden, G. C., & Rueckl, J. G. (1986, November). *Taking language by the hand: Reading handwritten words.* Paper given at the twenty-seventh annual meeting of the Psychonomic Society, New Orleans.

Selfridge, O. G. (1959). Pandemonium: A paradigm for learning. In *Mechanization of thought processes* (pp. 511-526). London: Her Majesty's Stationery Office.

Zadeh, L. A. (1965). Fuzzy sets. *Information and Control, 8*, 338-353.

Acknowledgements

The research reported in this paper and the writing of the paper were supported, in part, by grants from the National Science Foundation (BNS-83-15192) and the graduate division of the University of California, Santa Cruz. Bruce Bridgeman, Brian Fisher, Ray Gibbs, Alan Kawamoto and other members of the PEP Program made helpful comments on this enterprise.