A Two-stage Categorization Model of Family Resemblance Sorting

Woo-Kyoung Ahn and Douglas L. Medin[1]
Department of Psychology
University of Illinois at Champaign-Urbana
Champaign, Il 61820

ABSTRACT

A two-stage model is applied to category construction. The first stage of the model involves looking for a defining feature among exemplars and creating initial categories based on the defining features. In the second stage, overall similarity is calculated to categorize the remaining exemplars that were not classified by the defining feature. For some types of exemplar structures, family resemblance sorting emerges as a product of the two-stage model. A series of experiments was carried out to contrast the two-stage model with Anderson's induction model (Anderson, 1988) and CLUSTER/2 (Michalski & Stepp, 1983). The results showed that the two-stage model is a better predictor of when family resemblance sorting will or will not occur.

INTRODUCTION

Categories in the real world are known to have a family resemblance structure (Rosch & Mervis, 1975). Family resemblance categories are fuzzy categories where the members are generally similar to each other, but where there is no set of defining properties that any and all examples have. Rosch (1975) predicted that when asked to sort examples linked by overall similarity, people would tend to create categories in a way that potential prototypes are at the centers of the categories.

However, Medin, Wattenmaker, and Hampson (1987), in their Experiments 1, 2, and 3, found that people rarely constructed categories based on overall similarity. Instead, subjects typically sorted exemplars on the basis of a single dimension. In one of their experiments, Medin et al. also used trinary-valued dimensions coupled with the requirement that exactly two categories be created, which made subjects unable to create uni-dimensional categories. Under this condition, a few family resemblance sortings were obtained but none of the subjects' descriptions was consistent with a family resemblance explanation. Instead, they simply used a primary dimension plus either a conjunction or a disjunction of features. These results suggest a two-stage model of categorization, in which the first stage involves looking for a defining feature among given exemplars and the second stage involves computing similarity of remaining exemplars to the initially created categories.

Several other alternative models for category construction have also been proposed. For example, Michalski and Stepp (1983) developed CLUSTER/2 which forms a class only if it is describable by a concept from a predefined concept class. Recently several iterative algorithms have been developed by Anderson(1988) and Fisher (1987), which try to maximize the inferential potential of categories. This paper compares these three classes of recent models. The predictions made by each model will be compared with the results obtained in an experiment.

DESCRIPTION OF MODELS

Two-stage Model

The two-stage model is developed to capture people's goal of finding simple structure in the world. The idea is that people may impose more structure than is objectively present. Given that the world is not organized in such a simple manner, people are forced to deal with exceptions. Therefore, two stages seem to be involved in creating categories; In the first stage, the most

---

[1] The address of both authors will be Department of Psychology, University of Michigan, Perry Building, 330 Packard Rd, Ann Arbor, MI 48104, as of August, 1989.

	D1	D2	D3	D4		D1	D2	D3	D4
E1	0	0	0	0	E6	1	1	1	1
E2	0	0	0	1	E7	1	1	1	0
E3	0	0	1	0	E8	1	1	0	1
E4	0	1	0	0	E9	1	0	1	1
E5	1	0	0	0	E10	0	1	1	1

Figure 1. The structure of Exemplars used in Medin et al's experiment

important dimension is selected as a primary dimension for each category and exemplars are classified along this dimension. In the second stage, the exceptions are classified into initially created categories through various strategies.

A specific version of the two-stage model is developed, in which the model tries to construct two categories from given exemplars. According to this model, subjects will first select the most salient dimension in the exemplars (e.g. size). Then the subjects will divide the exemplars into two groups according to the two extreme values along the selected dimension (e.g. small vs. large). The second stage involves classifying the remaining exemplars which do not have the extreme values (e.g. medium). These exemplars are categorized into either of the two initially created groups depending on their overall similarity to each group. The judgment of the overall similarity involves all the dimensions in the exemplar as in the conventional models.

The two-stage model produces uni-dimensional sorting from the stimuli used in Medin et al's experiment. The structure of the examples used in their experiments was shown in Figure 1. In this figure, D1, D2, D3, and D4 indicate each dimension in the example, 0, 1's indicate different values in the same dimension, and E1, E2,... E10 indicate examples to be categorized.

Given the task demand of creating two categories, the second stage is not even necessary because all the dimensions have only two values and therefore, there cannot be any remaining exemplars after the first stage. No matter which dimension is selected as the most salient one, the model predicts uni-dimensional sorting (e.g. E1, E2, E3, E4, and E10 in one category and the rest in the other when D1 is selected as the most salient dimension).

The two-stage model may seem to be unable to produce family resemblance sorting because the model looks for defining features for each category. However, in some cases, the two-stage model does generate the family resemblance categories as a by-product of the process carried out in the second stage. Shortly we shall see how the model actually produces the family resemblance sorting with concrete examples.

CLUSTER/2

Unlike conventional clustering models, CLUSTER/2 does not use a measure of similarity as a basis for categorization. Instead, it uses a measure based on descriptions of candidate clusterings. The main goal of CLUSTER/2 is to generate categories with a minimum number of attributes used in a description, maximum number of attributes that singly discriminate among all classes, and maximum number of attributes that take different values in different classes. This goal is used as a criterion to judge quality of clustering.

The system goes through several iterations of the following steps. First, the system chooses initial seeds randomly or according to some criterion (e.g. values that are most distant from each other). For each seed, it generates a set of all maximally general descriptions of the seed which do not intersect with the set of remaining seeds. The descriptions are modified according to the clustering quality criterion. New seeds are selected and the entire steps are repeated until there is no improvement in clustering.

One of the most important criteria used in CLUSTER/2 is simplicity of descriptions, which is similar to the first stage of the two-stage model. However, since CLUSTER/2 does not have an additional way of handling exceptions, the system cannot generate family resemblance categories, which cannot be described in simple terms.

Iterative Algorithms

Recently several iterative algorithms of category construction have been developed in which new examples are entered incrementally and classified to the category that maximizes the inferential potential of the resulting partition. As a clustering criterion, Fisher (1987) used category utility developed by Gluck and Corter (1986), which is a product of a base rate of each feature, cue validity, and category validity. For each object, the system calculates category utilities of that object coming from each of existing categories and a category utility of that object coming from a new category. The object is placed into a category which maximizes the category utility.

Similarly, Anderson's iterative algorithm calculates two kinds of probabilities (i.e. the probabilities of a new object coming from old categories, P_k, and the probabilities of the new object coming from a new category, P_o). These two probabilities are operationalized in terms of the equations as follows.

$$P_k = \frac{cn_k}{(1-c)+cn} \prod_{i=1} \frac{C_{ki}+1}{n_k+m_i}$$

$$P_o = \frac{1-c}{(1-c)+cn} \prod_{i=1} \frac{1}{m_i}$$

where n is the number of objects so far, n_k is the number of objects in category K so far, C_{ki} is the number of objects in category so far with the same valued on the ith dimension as the object to be classified, m_i is the number of values on dimension i, and c is a cohesion parameter, which is the probability that any two objects will be in the same category.

This model cannot explain Medin et al's results because it produces family resemblance categories from the examples used in their experiments (i.e. E1, E2, E3, E4, and E5 in one category and the rest in the other category).

TEST OF MODELS

To examine different predictions of each model on concrete examples, three sets of examples (Sets A, B, and C) were developed. The abstract notation of the structure of the examples is shown in Figure 2.

Predictions of the Two-stage Model

For Set A, no matter which dimension was chosen for the most salient dimension, the model categorizes E1, E2, E3, E4, and E5 into one group and the rest into another group. This sorting turns out to be the family resemblance sorting. To illustrate more specifically what the model does, suppose D1 was chosen as the most salient dimension. Then E1, E2, E3, and E4 are classified as one category and E6, E7, E8, and E9 are classified as another category. Then which category E5 and E10 each belong to should be decided. Since E5 has greater overall similarity to E1, E2, E3, and E4 than E6, E7, E8, and E9, it is categorized with the former group. Similarly, E10 is

	Set A				Set B				Set C			
	D1	D2	D3	D4	D1	D2	D3	D4	D1	D2	D3	D4
E1	0	0	0	0	0	0	0	0	0	0	1	0
E2	0	0	0	1	0	0	0	1	0	0	1	1
E3	0	0	1	0	0	0	2	0	0	1	0	0
E4	0	1	0	0	0	1	0	0	1	1	0	0
E5	1	0	0	0	2	0	0	0	1	0	0	1
E6	2	2	2	2	2	2	2	2	1	2	2	1
E7	2	2	2	1	2	2	2	0	2	2	1	2
E8	2	2	1	2	2	2	1	2	2	2	1	1
E9	2	1	2	2	2	0	2	2	2	1	2	2
E10	1	2	2	2	1	2	2	2	1	1	2	2

Figure 2. Three Sets of Exemplars Used to Test Models

Table 1. Predictions of the two-stage model

	Set A				Set B				Set C			
Category A	0	0	0	0	0	0	0	0	0	0	1	0
	0	0	0	1	0	0	0	1	0	0	1	1
	0	0	1	0	0	0	2	0	0	1	0	0
	0	1	0	0	0	1	0	0	1	1	0	0
	1	0	0	0					1	0	0	1
Category B	2	2	2	2	2	0	0	0	1	2	2	1
	2	2	2	1	2	2	2	2	2	2	1	2
	2	2	1	2	2	2	2	0	2	2	1	1
	2	1	2	2	2	2	1	2	2	1	2	2
	1	2	2	2	2	0	2	2	1	1	2	2
					1	2	2	2				

categorized with E6, E7, E8, and E9. Therefore, this test of the two-stage model on Set A shows that the model can also generate family-resemblance categories.

For Set B, the model generates uni-dimensional sorting where the defining dimension of a category is the one specified as the most salient dimension. For example, if D1 is selected as the most salient one, then E1, E2, E3, and E4 are grouped together and E5, E6, E7, E8, and E9 are grouped together. In the second stage, E10, the remaining example, is judged to be more similar to E5, E6, E7, E8, and E9 group and is classified into this group, resulting in a uni-dimensional category of E1, E2, E3, and E4 along D1.

For Set C, the model generates the family resemblance categories. For example, if D1 is entered as the most salient dimension, E1, E2, and E3 will be grouped together and E7, E8, and E9 will be grouped together. As in Set A, in the second stage, E4, E5, E6, and E10 are each classified according to overall similarity, resulting in family resemblance sorting.

Table 1 shows the summary of predictions made by the two-stage model.The two categories generated are arbitrarily named Category A and B. The categories generated for Set B are the ones when D1 is selected as the most salient one.

Predictions of CLUSTER/2

For CLUSTER/2, types of dimensions had to be specified. We used the dimensions actually used in the experiment: linear for D1 and D2, and nominal for D3 and D4. The parameters entered were as follows; Mink = 2, Maxk=4, covertype = disjoing, H1=3, H2=2, H3=3, Cbase = 2, probe = 2, NIDspeed= fast, maxheight=99, minsize=4, beta=3.0, LEF = ((sparseness=0.3) (simplicity=0.3)). (See Michalski & Stepp, 1983 for more details on the parameters.) These parameters were used as default values in the current program and we have not yet fully explored the parameter space. With these parameters, CLUSTER/2 generated three clusterings on each set of exemplars, differing in the number of clusters in each clustering. Since the system does not have any preference among those three clusterings, only those clusterings with two clusters were used for comparison with the results obtained in the experiment, in which subjects were asked to categorize the exemplars into two.

Table 2 shows the categorization made by CLUSTER/2 for each set of exemplars. As mentioned earlier, CLUSTER/2 did not generate family resemblance categories from any of the three sets. Instead, all the categories generated are uni-dimensional. If the parameter for the simplicity criteria is lowered, it may produce family resemblance categories but it seems to be against the main idea behind the development of the system (i.e. generating meaningfully describable categories).

Table 2. Predictions of CLUSTER/2

	Set A	Set B	Set C
Category A	0 0 0 0	0 0 0 0	0 0 1 0
	0 0 1 0	0 0 2 0	2 2 1 2
	1 0 0 0	0 1 0 0	0 1 0 0
	0 1 0 0	2 0 0 0	1 1 0 0
	2 2 1 2	2 2 2 2	2 1 2 2
	2 2 2 2	2 2 2 0	1 1 2 2
	2 1 2 2	2 2 1 2	
	1 2 2 2	1 2 2 2	
		2 0 2 2	
Category B	2 2 2 1	0 0 0 1	1 2 2 1
	0 0 0 1		0 0 1 1
			2 2 1 1
			1 0 0 1

Predictions of Anderson's Algorithm

A simulation program of Anderson's iterative algorithm was written in GCLISP. However, the current version of the algorithm has an obvious limitation to be compared to the results of the experiment which will be described in the next section. Since the probabilities are based on matching and mismatching features on the same dimension, the model does not consider the degree of mismatch on continuous dimensions. More specifically, the probability of an object coming from an old category depends on the number of objects in the category so far with the same value on the ith dimension as the object to be classified (C_{ki}). Therefore, for example, in the current algorithm, the difference between 1 cm and 2 cm is same as the difference between 1 cm and 10 cm. Only exact matches can increase the probability.

To extend the model to handle continuous dimensions, the simulation program is written in a way to increase the probability by a certain amount if the two values are similar along a continuous dimension. For example, if there are three values on a continuous dimension (e.g. 3 cm, 4 cm, and 5 cm), the exact match will increase C_{ki} by one, the moderate match (e.g. 3 cm and 4 cm, or 4 cm and 5 cm) will increase it by 0.5, and the extreme mismatch (e.g. 3 cm and 5 cm) will increase it by 0.

Since the algorithm has potential to be order-sensitive, two different presentation orders were tried for each set; one with the lowest variability between two consecutive exemplars (e.g. the order of 0000, 0001, 0010, etc.) and the other with the highest variability (e.g. the order of 0000, 2222, 0001, 2221, etc.). The presentation order affected the categorization of Set B and C, and the two different clusterings are each presented under "with low var" and "with high var" in Table 3.

To compare the predictions with the results obtained in our experiment, the parameter c was adjusted to generate two categories. The range of the value of c which generated the two categories is specified in the last row.

EXPERIMENT

To test which model describes human behavior better, an experiment was conducted where people were asked to construct categories from examples.

Method

Each subject received a set of 10 cards on which examples were drawn. The order of the cards within each set was randomized and one set of cards was given to each subject all at once. Then they were asked to categorize the instances into two groups in a way that seemed natural to them. They were also told that there could be different number of examples in the two groups and that there was no one correct answer.

Table 3. Predictions of Anderson's model

	Set A	Set B with low var	Set B with high var	Set C with low var	Set C with high var
Category A	0 0 0 0 0 0 0 1 0 0 1 0 0 1 0 0 1 0 0 0	0 0 0 0 0 0 0 1 0 0 2 0 0 1 0 0 2 0 0 0 2 0 2 2 2 2 2 0	0 0 0 0 0 0 0 1 0 0 2 0 0 1 0 0 2 0 0 0	0 0 1 0 0 0 1 1 0 1 0 0 1 1 0 0 1 0 0 1 1 2 2 1 1 1 2 2	1 1 2 2 1 0 0 1 2 1 2 2 1 1 0 0 2 2 1 1 0 1 0 0 2 2 1 2 0 0 1 1 0 0 1 0
Category B	2 2 2 2 2 2 2 1 2 2 1 2 2 1 2 2 1 2 2 2	2 2 2 2 2 2 1 2 1 2 2 2	2 2 2 2 2 2 2 0 2 2 1 2 2 0 2 2 1 2 2 2	2 2 1 2 2 2 1 1 2 1 2 2	1 2 2 1
c value	0.3-0.8 for low var 0.3-0.7 for high var	0.3-0.5	0.3-0.7	0.3	0.5-0.6

There were three groups of subjects depending on which set of exemplars they received. We used three sets of exemplars specified in Figure 2. The actual dimensions used were size, number of arms, types of line, and color. Each dimension has three values such as small, medium, and large for the size dimension and green, red, and yellow for the color dimension. Based on these dimensions and values, outline drawings of cartoonlike starfish were developed. A pilot study was also conducted to create roughly equal intervals between two adjacent values on the same dimension and to attempt to equate saliency among dimensions.

Sixty undergraduate students at the University of Illinois participated in the experiment in partial fulfillment of a course requirement for introductory psychology. There were 20 subjects in each condition.

Results and Discussion

For Set A, 55% of the subjects produced family resemblance categories, and 45% of the subjects produced uni-dimensional sorting. For Set B, 100% of the subjects produced uni-dimensional sorting. For Set C, 35% of the subjects produced family-resemblance categories, 55% produced uni-dimensional sorting, and 10% produced other responses. Table 4 summarizes the results from the present experiment and Medin et al's experiments, and the predictions made by each model for comparison. The numbers in parenthesis indicate the percentage of the subjects' sorting predicted by each model. Overall, the two-stage model was the best predictor of the subjects' sorting.

Two-stage model.

Overall, the two-stage model seemed to give the best account of sorting. The two-stage model predicted 55% of the subjects' response on Set A, 100% on Set B and 35% on Set C. The reason why the two-stage model did not predict 45% of the response on Set A and 65% on Set C can be explained in terms of different strategies used in the second stage. At first, it was assumed that people judge overall similarity of exceptions to initially created categories in the second stage. However, subjects could have also classified the remaining examples based on the similarity of the

Table 4. Summary of results and predictions made by each model

	Medin et al.	Set A	Set B	Set C
Subjects	1-D 100%	FR 55% 1-D 45%	FR 0% 1-D 100%	FR 35% 1-D 55% others 10%
Two-stage	1-D (100%)	FR (55%)	1-D (100%)	FR (35%)
CLUSTER/2		1-D along D4 (0%)	1-D along D4 (0%)	1-D along D4 (0%)
Anderson				
with low var	FR (0%)	FR (55%)	other (0 %)	1-D (55%)
with high var	FR (0%)	FR (55%)	FR (0%)	other (0%)

value on the salient dimension to the value on the same dimension in each category created initially. For example, suppose the subjects created small vs. large categories in the first stage and the remaining examples had medium size. Then subjects might compare the similarity of medium to large and the similarity of medium to small. Then they might place the remaining examples in the category with the higher similarity. In this case, uni-dimensional categories were created from Set A and Set C.

To further test this idea, in the follow-up study, we asked subjects to create two categories of equal size. This task presumably prevents subjects from using the strategy that was just described because this strategy creates two unequal sized categories. In this experiment using Set A only, 100% of subjects created family resemblance categories. This result strongly suggests that the uni-dimensional sorting obtained in the current study is due to this strategy difference.

CLUSTER/2.

CLUSTER/2 could not predict any of the family resemblance sorting obtained in this experiment as shown in Table 4. Furthermore, the uni-dimensional categories predicted by the system does not have the same structure as the subjects' uni-dimensional categories. While the system used a nominal dimension (D4) to divide the examples into two, subjects preferred continuous dimensions presumably because they want to use two extreme values to create contrasting categories.

Anderson's model.

Regardless of input ordering, the model correctly predicted family resemblance sorting from Set A, which was 55% of subjects' response. Also, when the low variability ordering was used, the model could predict subjects' uni-dimensional sorting from Set C, which was 55% of subjects' response. However, when the high variability ordering was used, Anderson's model failed to predict any sorting from Set C. Also, the model was not a good predictor for Set B and the set used in Medin et al's experiments, as mentioned earlier.

The predictions of Anderson's model is hard to compare because it is not clear how the effect of different input ordering should be interpreted. Although the subjects received each set of cards all at once, it is possible that the subjects might have selected and classified the examples in the low variability order. Additional assumptions on how subjects handle each example incrementally seem to be necessary. In addition, his model does not seem to be able to explain why all subjects generated family resemblance categories from Set A when asked to create two equal-sized groups as mentioned earlier. The effect of various task demands seems to be outside the boundary conditions for this model.

CONCLUSION

So far, we have shown that the two-stage model predicts the experimental data the best among the three clustering models considered. People seem to like structures with more organization than is present in the examples (i.e. defining features) but then given the demands of the task (assigning all examples to one of two categories), they have to figure out what to do with the examples that do not fit.

In a similar view, Michalski has proposed a two-tiered concept representation (Michalski, in press). In this representation, concepts consist of the first tier, called the Base Concept Representation (i.e. typical properties of a concept in an explicit, comprehensible, and efficient form) and the second tier, called the Inferential Concept Interpretation (i.e. inference rules and metaknowledge that define allowable transformations of the concept under different contexts, and handle special cases and exceptional instances).

Fillmore (1982) and Lakoff (1987) argued that concepts consist of idealized cognitive models with clear boundaries and necessary and sufficient conditions. According to them, the reason why natural concepts have fuzzy boundaries is simply because the background conditions for the idealized model do not exactly fit the real world situations.

In our two-stage model, the first stage represents ideal rather than typical features. Processes associated with the second stage yield categories where these idea or defining features become converted to typical features. In the same way, family resemblance categories may represent a compromise between a preference for highly structured concepts and the necessity of mapping concepts onto real world examples.

Acknowledgement

We thank Brad Whitehall for running CLUSTER/2 to test on the exemplars. This work was supported by NSF grant BNS88-12193 to the second author.

REFERENCES

Anderson, J.R. (1988). The place of cognitive architectures in a rational analysis. Proceedings of the Tenth Annual Conference of the Cognitive Science Society.

Fillmore, C. (1982). Towards a descriptive framework for spatial deixis. In R. Jarvella & W. Klein (Eds.), Speech, place, and action. London: Wiley.

Fiehser, D. (1987). Knowledge acquisition via incremental conceptual clustering, Machine Learning 2, 139-172.

Gluck, M., & Corter, J. (1985). Information, uncertainty, and the utility of categories. Proceedings of the Seventh Annual Conference of the Cognitive Science Society, Irvine, CA. 283-287.

Lakoff, G. (1987). Cognitive models and prototype theory. In U. Neisser (Ed.), Concepts and conceptual development. (pp 63-100). Cambridge, NY: Cambridge University Press,.

Medin, D.L., Wattenmaker, W.D., & Hampson, S.E. (1987). Family resemblance, conceptual cohesiveness, and category construction. Cognitive Psychology, 19, 242-279.

Michalski, R.S. (in press). Two-tiered concept meaning, inferential matching and conceptual cohesiveness. In S. Vosniadou and A. Ortony (Eds.) Similarity and analogy.

Michalski, R.S., & Stepp, R. E. (1983). Learning from observation: Conceptual clustering. In R. S. Michalski, J.G. Carbonnell, & T.M. Mitchell (Eds.) Machine learning: An artificial intelligence approach. Palo Alto, CA: Tioga Publishing.

Rosch, E. (1975). Universals and cultural specifics. In R.Brislin, S.Bochner, & W. Lonner (Eds.) Cross-cultural perspectives on learning. New York: Halsted Press.

Rosch, E., & Mervis, C.B. (1975). Family resemblance: Studies in the internal structure of categories. Cognitive Psychology, 7, 573-605.