Towards (Making SEQL) a Strong Model of Categorization

Praveen Paritosh and Sven Kuehne, 10/31/2001 1:22:39 AM

1 The very general issues

Markman has been saying that structural alignment could explain categorization. His key ideas are –

M1. Categories within a superordinate are more comparable, an evidence for alignable structure in our categories.

M2. When an exemplar is observed, (a MAC-like) reminding provides us with similar exemplars *as well as* generalizations from our LTM, and information highlighted by structural alignment of the exemplar and the reminding is incorporated into the category structure.

M3. For nascent categories, we pay more attention to commonalities and alignable differences than non-aligned differences; and as the categories mature, there is enhanced learning of the non-aligned differences increases. (Zhang and Markman, 1998, 2000) [Implication: Initial assimilation gives more weightage to the common structure, and with increased maturity, we have increased assimilation of non-alignable bits of knowledge from exemplars. Question: how could the category data structure ever maintain non-alignable parts of representation?]

M4. Referential communication (labels in language, like "bottles", "mammals") effects the categories we form. (Markman and Magin, 1998)

SEQL already embodies the key ideas, and the new changes suggested here will let us tinker with M2 through M4.

Ramscar and Pain (CogSci, 1996) argued about the similarity of analogy and categorization. The support of the argument in that paper was the story categorization task, where the (preferred) dimension of categorization was structural similarity. For a lot of common categories (like different types of computers, cars, etc) it seems like categorization is more like literal similarity than true analogy, with attributes playing stronger role. The following are what we need to add to a purely structural-similarity account of categorization[1] to explain such categorization -

C1. Categorization often deals with things that have considerable structural overlap (e.g., cars) as well as surface features (entities, attributes and functions) play a stronger role in determining category membership. Specifically, there are to be richer ways of comparing the surface features (red is more similar to orange than black, or 1 meter length is more similar to 1.2m than 5m) than just identical/ different, which might be more vital to categorization than analogy.

C2. SMT tells us the similarity between two cases. Seems like that our data structure for a category (or generalization or abstraction) has to have more category specific information than an exemplar. Think of any category (an NU graduate student, a Chinese restaurant, etc). Our knowledge of a category matures as we see more and more instances of it, we also have a notion of how much variable/well-defined a category is (see discussion of Rips' studies below), whether it is multi-modal (so an NU grad student might be multi-modal with peaks corresponding to engineering, humanities and sciences graduate students, each of which are pretty typical and there isn't much in-between). So, we need to maintain the `maturity` (number of exemplars assimilated) and `variability` (the structural variability of the exemplars assimilated in a generalization. One metric could be 1 – NSES, averaged over the exemplars encountered) parameters[2] for each of the generalizations/categories. And maybe there are more.

My personal concerns are –

---

[1] Currently the match between a generalization and an exemplar is no different from the match between two exemplars – so implicitly we are saying that *test for category membership* and *similarity* are equivalent, which has been the cause of innumerable arguments in psychological literature, one of which are Rips' arguments, which we will try to explain in this paper.

[2] These parameters will be used in the generalization process/algorithm. `maturity`, simple to compute, might play a more complicated role in conjunction with `variability` to compute thresholds for categories, role of non-alignable knowledge (as in M3).

By exposure to quantitative dimensions like length, price, weight, we build a sense of the quantity. The sense of quantity in a dimension (like "power of computers") has two distinct components –

P1. The landmarks in the dimension (possibly corresponding to (the power of) handhelds, laptops, desktops, servers, supercomputers). These are arrived at by looking at the value of the alignable dimensions along structurally different generalizations.

P2. Intuitive distributions within each of the generalizations – this captures the knowledge that the power of laptops (because of size constraints) is relatively less variable than desktops which is way lesser than those of supercomputers (extremely wide range of performance). These distributions might also be providing a finer sense of quantity, useful in coming with categorical estimates. Also, if within a generalization we find that a lot of quantitative attributes have multi-modal distributions, that might trigger considering breaking up that category into sub-categories.

I'll just throw this one in as a far off question to think about – does *typicality* arise out of systematicity/ structural considerations, or frequency/distributional information (the exemplar closest to the central tendency is typical)? A strong model of categorization should be able to account for the fact that we can readily come up with typical exemplars of the categories that we have.[3]

2 The effect of prior knowledge (of categories) on categorization

Clearly, one's existing knowledge of categories that exists influences the generalizations that s/he makes when presented with a sequence of new exemplars. The effect could go both ways -

a. An expert (in knowing the categories associated with the specific examples) might make more finer grained distinctions, hence more generalizations.
b. An expert might be economic in creating a new category for something as s/he is aware of the essence of that category.

And a really rough way in which this will work is that SEQL calls MAC with the new exemplar, and adds the strongest generalizations/exemplar matches to the stored generalizations/exemplars to match against. The MAC phase can retrieve both exemplars and generalizations.[4] Now this raises the question. If the generalizations are stored just as they are now, MAC will more likely retrieve exemplars than generalizations as they have richer content vectors with exemplar specific information. That might be fine. But in the case we have built prior categories in that domain, and we are looking at the effect of that on the new categories, we should be able to retrieve them from the KB. This will depend upon how we are storing generalizations and if we need to tweak MAC to be able to take into account this new data structure for generalizations. (For example, a stronger/mature[5] generalization should be more likely retrieved[6] than a poorer one).

3 The diversity in a category, or, are all categories equal?

In a generalization task, SEQL uses the same threshold with all of the generalizations. It might be possible that the certain generalizations are more spread/diverse and some others are more well-defined. For example, the category of stealth bombers is more well defined (and might have a stronger threshold for an
exemplar to be included in it) than commercial people carrier planes (which are more diverse).

Rips' (1989) study sparked some of the ideas in this paper, although what is said here is qualitatively different from the Rips' arguments, and a subset of these ideas could explain Rips' results. Below we briefly introduce the Rips experiments, as that helps to make some connections to our arguments.

---

[3] This becomes important when I would want to make a categorical estimate – How long does an NU CS PhD take? What does a typical CS grad student (or investment banker, for contrast) wears to work?

[4] As suggested in today's meeting; in the combined SEQL-MAC/FAC model, this is the stage where we could tweak the thresholds based on the `variability` (and `maturity`) of the category. This seems to be a reasonable answer to Dedre's apprehensions about arbitrary timing choices for these processes.

[5] More number of exemplars have been assimilated to that generalization.

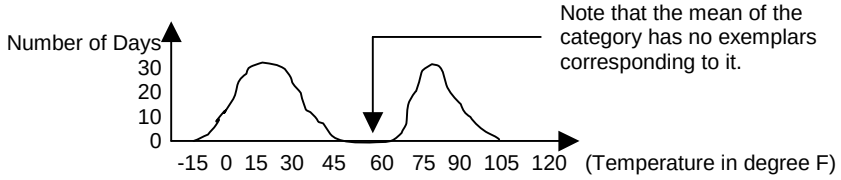
[6] As Dedre pointed out, there are also *recency effects*, but we might ignore them for time being.

I. Consider the categories of Pizzas and US-Quarters. Imagine an object that is 3-inches (which happens to be the mean of the largest quarter and the smallest pizza as the subject has seen) in diameter, and that's all you know.

a. Is it more likely to be a pizza or a quarter?
b. b. Is it more similar to pizzas or quarters?

The idea is people categorize this unknown 3-inch object as a pizza but think it's more similar to quarters. (Some other examples from his study: duration of basketball games and dinner parties, number of members in the US Senate and rock groups, size of basketball hoops and grapefruits, height of stop signs and size of cereal boxes, etc. All of the examples had the property that one of them was relatively fixed and the other more variable).

II. Also, certain categories are multi-modal. So imagine the category of daily high temperatures in Chicago in January and July. It is a bimodal distribution with one peak in 15-30 and another in 75-90 degree F (see figure below). Although the mean of the entire distribution, which is in the range 45-60 degree F, is similar to the distribution, there are almost no elements in the distribution corresponding to that. Thus, there is a dissociation between similarity and categorization.



(Some other examples from his study: weight of 100 children, half of which are 5-year-olds and half 15-year-olds; hair length of teenagers, half of whom are boys and half girls. All of the examples had the property that the dimension in question had a bimodal distribution for the entire category[7]).

What is Rips getting at? A dissociation between the *test for category membership* and *similarity*. There seems to be more to an exemplar being included in a category than the average similarity of that exemplar with all the other exemplars in that category (or the similarity between that exemplar and some averaged representation of all the exemplars in that category). One could explain his results by having the data structure for the abstraction of size of quarters, pizzas, or rock groups not be mere quantitative values (as they are for individual instances of any of those), but be distributions. And the test for category membership take into account the spread/variance of the distribution, or properties like multimodality[8].

But the really really cool point Rips never made. The above dissociation between similarity and categorization is much more fundamental than just dimensional. Are there structural equivalents of variability and multimodality? We think so.

1. Structural variability – Consider the categories: commercial people carrier planes, private jets, helicopters, stealth bombers. The commercial people carriers are quite variable as opposed to stealth bombers which are more well defined, and do not allow for much structural variability. Thus with the knowledge of variability, we could tighten our thresholds for saying some new exemplar is a stealth bomber as opposed to the commercial people carriers.
2. Structural multimodality – The category of NU grad students might be multi-modal with peaks corresponding to engineering, humanities and sciences graduate students, each of which are pretty typical and there isn't much in-between (Or, the graduate students in 1890 Maple, which are Computer Science or Learning Science). The category of human beings is bimodal with peaks corresponding to males and females. Frequently, such multimodality leads to forming

[7] Although, most of his categories seem to be adhoc in the sense that more than a single category that's bimodal, they look more like two categories forcibly glued together.
[8] So, if the distribution of a dimension obtained from a bunch of exemplars is multimodal, one should test a new exemplar with each of the peaks and not just the average of the entire distribution.

subcategories and could help explain the hierarchical nature of our categories. In such cases, for something to be a human being, it has to strongly match to either male or female, and not weakly match to something in between (just as in Rips' temperatures in Chicago which are either close to 15 or 90 degree F and not to 50 degree F). Our current data structure for generalizations cannot capture structural multimodality at all (one generalization cannot represent two different structural representations, maybe this means that we have one generalization of NU Grad student with a lower threshold which points to three new generalizations corresponding to engineering, sciences and humanities, each of which have higher thresholds[9]).

4 Summary

Inspired by dissociation between similarity and categorization, we present ideas to make SEQL a stronger model of categorization. The implications of the arguments presented here suggest two classes of changes to SEQL –

1. Changes to the data structure for generalization –
    a. Every generalization has `maturity` and `variability` parameters.
    b. A generalization has pointers to sub-generalizations (follows from structural multimodality, and helps us explain hierarchical nature of categories).
    c. For quantitative dimensions, the generalization has distributions (one might need to keep the numeric values around for a while before we can start computing distributions).
    d. For non-quantitative attributes (symbolic surface features), our generalizations have frequency information. (For example, we are able to say that most laptops are black in color while most desktops are white, or most of the electrical engineering grad students at NU are Chinese).
2. Changes to the generalization algorithm –
    a. `maturity` and `variability` help tweak thresholds.
    b. As we are no longer modeling sequence effects, SEQL will possibly run multiple passes over the same set of exemplars, allowing us to test for a larger number of strategies[10].
    c. Orchestrating SEQL and MAC/FAC – with all the changes to the generalization data structure suggested above, are we able to retrieve the relevant generalizations from LTM when presented with new exemplars? And in what ways does our previous knowledge effect our current generalization task?

This paper is not complete, and will hopefully generate both psychological and computational implementation issues that will lead to a stronger structural account of categorization.

[9] The fact that these new subgeneralizations of engineering, humanities and sciences grad students will have higher thresholds automatically follows from the fact that now each of these will have way higher average NSES scores.

[10] To give an example, we have noticed that people form the most well-defined categories (with high match score) the first. Studies in which the order in which people make categories when they have access to all the exemplars will help us in formulating various strategies.