1 Combining PACKER and POLR

1.1 Selecting a Source

The Path of Least Resistance (POLR) model assumes that people select as a source a representation (for Ward, this is a subcategory, for us it will be an exemplar) according to it's representativeness: more representative exemplars are more easily retrieved, and thus more likely to be used as the source.

Within the exemplar view, the typicality of a stimulus within a category (a good proxy for representativeness) has been measured as the total similarity between the stimulus and the exemplars assigned to the category (see Nosofsky, 1988). The typicality t of a stored example x belonging to category h is computed as the summed similarity between x and the members of category h, $x_j \in h$.

$$t(x, h) = \sum_{x_j \in h} s(x, x_j) \tag{1}$$

If only one category is known, the probability that x will be retrieved is a function of its relative typicality among all members of the category:

$$p(x|h) = \frac{\exp\{\theta_{source} \cdot t(x, h)\}}{\sum_i \exp\{\theta_{source} \cdot t(x_i, h)\}} \tag{2}$$

In Ward's view, the probability a given category of exemplars will be retrieved is also based on representativeness. Thus, when there are multiple categories in memory, $p(x) = p(x|h)p(h)$. We will assume that all categories are equally representative in the artificial domain, however, so $p(h)$ is uniform across all categories.

1.2 Generating the Next Item

Once an item is generated, PACKER's rules kick in. The probability a stimulus y will be generated is a function of its relative similarity to the source x:

$$p(y|x) = \frac{\exp\{\theta_{generate} \cdot s(y, x)f(x)\}}{\sum_i \exp\{\theta_{generate} \cdot s(y_i, x)f(x)\}} \tag{3}$$

where $f(x)$ is parameterized as in PACKER. To obtain predictions not depending on any given source, the new item can be drawn from a combined distribution:

$$p(y) = \sum_j p(y|x_j)p(x_j) \tag{4}$$

In doing so, this model differs from PACKER only in that the contribution of exemplars is weighted according to representativeness. When $\theta_{source} = 0$, the two models are identical.

1.3 Feature Weighting

An additional claim made by Ward is that people do not tweak along features entirely at random. Instead, they attempt to maintain commonalities along features that are viewed as characteristic of the category, only changing features that are viewed as unimportant to the source.

This can be implemented as a change to the feature weights on similarity judgments:

$$s(x_i, x_j) = exp(-c \sum_k |x_{ik} - x_{jk}|w_k) \tag{5}$$

Specifically, the feature weights w_k can be set according to each feature's characteristicness in the source category, measured as the relative amount of variability of that feature within the category: characteristic features should be less variable.

$$w_k = \frac{\exp\{-\theta_{wts} \cdot \text{range}(k)\}}{\sum_i \exp\{-\theta_{wts} \cdot \text{range}(k_i))\}} \tag{6}$$

figure out if that produces the ward-predicted results later...

References

Nosofsky, R. M. (1988). Exemplar-based accounts of relations between classification, recognition, and typicality. *Journal of Experimental Psychology: learning, memory, and cognition*, *14*(4), 700.