Category Learning and Acquisition with Connectionist Networks: A Proportional Difference Approach

Carolyn Saund
COMP150-03
Computational Models of Cognitive Science
Spring 2015

May 1, 2015

1 Introduction

Category learning is the formation of new object categories in the context of human knowledge. Category learning is typically broken into two different tasks: classification, and category formation (Ashby 2005). That is, either classifying an object into previously known categories or make a judgement that it is sufficiently different from exisiting exemplars to break into a new category. Using a classical classification task developed by Shepard (1961), Kruschke (1992) uses a model which closely models human performance in category learning. I attempt to use this data and model to predict category fragmentation.

The ALCOVE model is an exemplar-based feed-forward connectionist network with three layers. This network works by learning both weights between hidden and output units, but also an attentional learning weight, which serves to indicate which features within examples are most influential on categorization. These learned attention strengths allow categorization, but may also lead to a way to classify something as an unkonwn category. The model I put forth in this paper uses both attentional weights as well as the probabilistic nature of judgement exploited in the ALCOVE model to find new category spaces after it has been trained. Using the architecture developed in the training phase, this model is able to classify new categories as unknown while still retaining its proven ability to perform classification tasks on categories on which it is trained.

In this paper, I will discuss the ALCOVE-CAdd (Category Addition) algorithm and how it uses the strengths of the ALCOVE model to form new categories. I will begin by explaining mathematically how the model judges new categories and how these parameters may be manipulated based on different cognitive tasks. Then, I will discuss the practical applications of this model, and the disadvantages of its implementation, including the assumptions and cognitive factors which this model fails to take into account. Finally, I will propose future experiments and alterations which could prove the effectiveness of this model in describing cognitive processing.

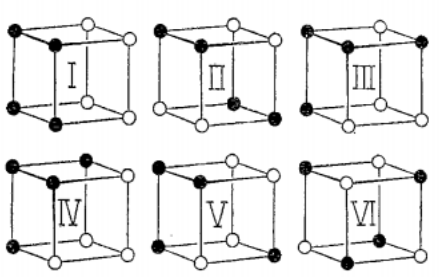

Figure 1: The six types of basic classification represented abstractly by coloring corners of a cube.

2 Task Description

Using data from the classical experiment of Shephard (1961) illustrated in Figure 1, input to the model was a vector of scalar feature activations, followed by the input category. Overall, there were six categories, each with eight features. So an input for the category five example in Figure 1, for example, would be $\langle 0, 1, 1, 0, 1, 0, 1, 0, 5\rangle$. However, although the data uses binary inputs, any input may be a scalar value.

The ALCOVE model uses a classification technique to decide categories by

$$Pr(K) = exp(\phi a_k^{out}) / \sum_{out} exp(\phi a_k^{out}$$

What this equation describes is the idea that the selected category is based on its activation as a proportion to total overall activation in all possible categories. Using this proportion, I instate a proportional threshold γ which acts as a lower bound proprotion to break into a new category. That is, if the maximally activated category is not sufficiently more active than all over category nodes, the algorithm will split into a new category, described as either

$$a_k^m ax / \sum_{out} exp(a_k^{out}) > \gamma \quad \text{or as} \quad a_k^m ax / \sum_{out} exp(a_k^{out}) < \gamma$$

In the first case, when the activation is greater than gamma, it is allowed to be classified. In the second case, when the activation is not sufficiently proportional, it will break into a new category.

Importantly, the purpose of this algorithm is to keep in-category classification high, while simultaneously breaking apart new, untrained categories. With these goals in mind, the way this model is tested is by training on two or more categories, and then at test exposing it to multiple new categories. The ideal model would, when trained on two categories and presented four, would return the information that in the end, there were four total categories, and retain its extremely high classification rate on category one and two. It is important to note that we do not measure category accuracy within new categories, but only that the correct number of categories were identified, and old category exemples are classified as falling within one of the existing categories. We call the categories on which the model was trained *in-categories* and the categories that are used in testing that are not already represented *out-categories*.
In this experiment, I train on varying numbers of categories, but always test on a group of 40 samples which contains six categories.
A major limitation in this method is that you must know the number of categories that is to be expected by the test cases. However, although this does not mimick natural category separation in human behavior, it is a testable method and does still attempt to model human discovery in organic

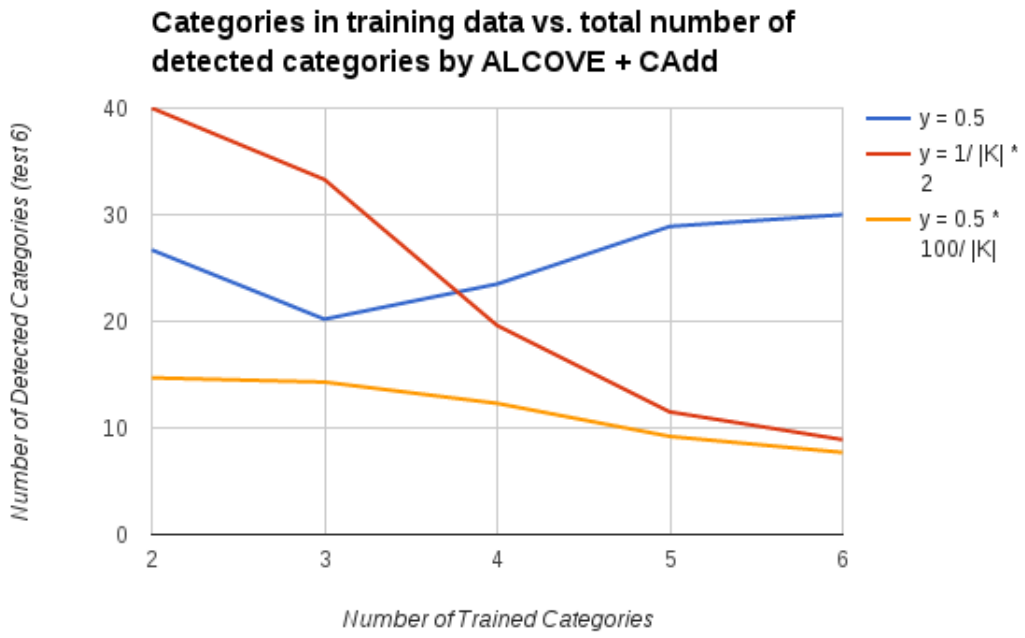

Figure 2: results of training the model at each γ value. The model trained on varying number of categories, but results indicate tests when all six categories were presented.

category learning.

3 Solution

3.1 Description

Category formation is dependent on γ in this model, making this the key parameter to select. If the γ value is too loose - that is, γ is closer to 1 making the threshold harder to acheive - many items will not have sufficient in-category activation proportions to fall neatly into any category, and so there will be many new categories. Contrarily, if γ is too strict, meaning it is close to a perfectly proportioned in-category activation (then very few exemplars will fall between categories so evenly, and thus will not form many new categories.

3.2 Experiments and Results

Three different γ values were used in this experiment.

The first is $\gamma = 0.5$, meaning that for the example to have a clear category activation, one category had to have at least 50% of the total activation. This simple majority metric proved to be overeager to form new categories. Additionally, because of the constant rate, even when more categories means lower activation, the constant threshold meant that it became very difficult for the

model to keep classifications. That is, although the ratio and porportions changes depending on the number of categories on which the model has been trained, the constant threshold means that it becomes more and more difficult for the model to accurately classify. As a result, performance did not significantly change with more presented categories (Figure 2).

The next attempted value was $\gamma = 1/|K| * 2$, meaning that whenever any node has twice as much activation as it would if each were perfectly proportional, the algorithm will assume a new category. Although this proved far superior to a flat cut-off, it still was ineffective. For instance, when trained on two categories, there is no way to acheive the classification threshold because each node will always have some activation, but the thresold is 1. Barring this edge case, the algorithm performs well compared to the constant threshold (Figure 2).

The final value tried was $\gamma = 1/|K| + 0.5 * (1/|K|)$, which means that the model judges to create a new category whenever any node is more active than others by 1.5x if all were perfectly proprotional. This proved to be the most effective threshold by which to define new categories (Figure 2). In the final case, when presented six categories after only training on the original two, the model averaged finding 7.5 categories, correctly recovering four, while simultaneously continuing to classify objects in the in-categories.

3.3 Analysis

A major benefit to this model is it reuses the existing architecture of the trained net. That is, when a new category is encountered, the structure is dynamically updated without having to change the structure of the learned exemplars. Since the new exemplar is used as a perfect category representation, and each input vector is the same, all attentional weights can be used as they are, and the weights between the new category and exemplars of that category are calculated at run-time. This means that no re-training has to occur, and all of the benefits of learning exemplars previously are retained even as categories are added.

This original goal of this algorithm was to break categories up by a lack of certainty. This achieves that goal - when new objects fall between distinct categories using the same features, they are decidedly neither category and should be reclassified.

Although this model is able to recover categories at appropriate thresholds, the current mechanism (the γ parameter) by which categories are split proves impreceise. Categories were split far more often than predicted, even with the best treshold. My prediction is that relative activation proportions alone are not enough to judge category formation. Because this uses the total activation, it fails to take into account the amount of activation relative to other examples. So, low levels of all-around activation may lead to gross miscategorizations. Additionally, even if one category is strongly activated, the total activation may be small compared to better exemplars of given categories. So, if an out-category falls adjacent to one of the in-categories, items in that category may always be incorrectly classified as one of the in-categories by virtue of relative activation.

Another problem with this method is it does not use any manner of sequential or temporal presentation to influence results. For example, if it is shown 10 perfect examples of an in-category and then 10 perfect examples of an out-category, it performs the same as if they are interwoven

in presentation. There is strong reason to believe presentation matterns immensely in the realm of human category learning (Carlhavo and Goldstone 2013) so going forward, this model will likely need to be altered to correctly fit human data in this respect.

It is important to note that although this model does have a method for learning new categories, the data used was originally paired with the ALCOVE model to test human category learning for a classification task. As such, there is no human data using these categories to reflect new category learning, as the goal of this algorithm was exploratory: to discover if this algorithm can be modified to successfully break up categories.

4 Future Work

There are many simple ways to build on this model as it is. The easiest, and likely that which would provide the largest improvement, is to build in an average overall activation variable which can be used at run-time to discern whether an example activates categories enough to even be considered for proportional classification. That is, if an example activates categories significantly less than any training example, it will be judged as a new category. This will avoid the problem described above of adjacent categories.

Another way to alter the categorization that was unexplored is by varying the specificity of hidden nodes receptive fields in the ALCOVE model. With a lower or higher specificity, categories have higher or lower activation thresholds, and consequently may have more narrow activation regions (and require more specific exemplars). This means that, depending on the range of activation sufficient to choose a category, new exemplars may be readily classified or broken into new categories at wildly different thresholds than these γ values predict. Variations of pairs of these parameters may lead to an even more ideal category learning pattern.

The most important aspect of this experiment is now to gather human data on category learning to discover human patterns of learning new categories so that all future models can be accurate attempts to fit this data.

5 Conclusion

Although the model developed is overeager to find new categories compared to distinct categories it has been shown to learn, it does successfully retain classification ability while simultaneously learning new categories. It is possible that with better parameter thresholds and new parameter pairings, this model can be vastly improved. The high performance of ALCOVE's ability to model human category learning and classification is optimistic in predicting human's ability to learn new categories.

6 References

Ashby, F. G., & Maddox, W. T. (2005). Human category learning. Annu. Rev. Psychol., 56, 149-178.

Carvalho, P. F., & Goldstone, R. L. (2014). Putting category learning in order: category structure and temporal arrangement affect the benefit of interleaved over blocked study. Memory & cognition, 42(3), 481-495.

Kruschke, J. K. (1992). ALCOVE: an exemplar-based connectionist model of category learning.Psychological review, 99(1), 22.

Love, B. C. (2002). Comparing supervised and unsupervised category learning. Psychonomic Bulletin & Review, 9(4), 829-835.

Nosofsky, R. M., Gluck, M. A., Palmeri, T. J., McKinley, S. C., & Glauthier, P. (1994). Comparing modes of rule-based classification learning: A replication and extension of Shepard, Hovland, and Jenkins (1961). Memory & cognition, 22(3), 352-369.

Shepard, R. N. (1964). Attention and the metric structure of the stimulus space. Journal of mathematical psychology, 1(1), 54-87.

Shepard, R. N., Hovland, C. I., & Jenkins, H. M. (1961). Learning and memorization of classifications.Psychological Monographs: General and Applied,75(13), 1-42.

Simeon, M., & Hilderman, R. (2008, November). Categorical proportional difference: A feature selection method for text categorization. In Proceedings of the 7th Australasian Data Mining Conference-Volume 87 (pp. 201-208). Australian Computer Society, Inc..