Semantics Cognition Neural Network Model with Existing Knowledge from Visual Similarity

Yang Gao yg2410@nyu.edu
Siyu Shen ss14359@nyu.edu
Yuwei Wang yw1854@nyu.edu
Hengjiali Xu hx2058@nyu.edu

Abstract

The neural network model of semantics cognition by Rogers and McClelland (2003) is widely recognized for its effectiveness in explaining semantic learning behavior. In this project, we apply the model on a new data set with larger number of items and attributes to further observe its validation. We also make an extension of the model by incorporating additional information of visual similarity, to simulate the psychological case where some existing knowledge is included in the semantic learning process. Our experiments show that the original model also suits for larger data sets, and our extended model is legitimate for representing cognitive development with faster learning experience.

1. Introduction

Semantic cognition refers to human's ability to use, manipulate and generalize knowledge that is acquired over the lifespan to support innumerable behaviours. Modeling the development of semantic representations is a popular task in computational cognitive science. One of the most renowned method is by Rogers and McClelland [Mcclelland and Rogers, 2003] using fully-connected layers neural network. It has been proved effective in explaining human learning process of semantics in various ways.

In this project, we aim to apply the R&M on a new data set of CCFD [Deng et al., 2021], which contains a much larger number of items and attributes than the data used in original paper. Then, we would try to extend the model by adding an additional layer of input from visual similarities. On a high level, we try to simulate the case where the agent already has some domain knowledge with general semantic categories

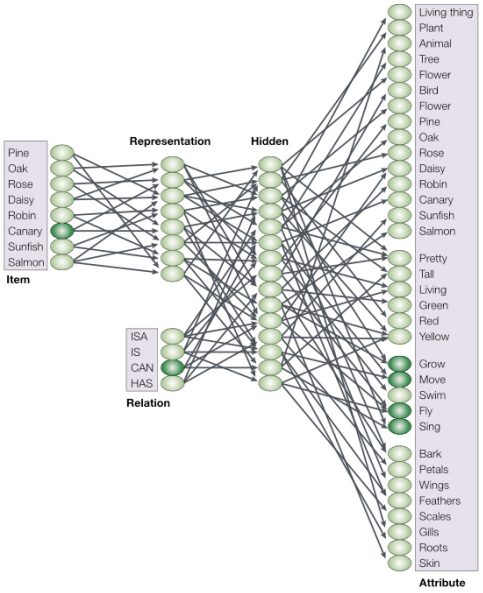

Figure 1: Neural Network Model by Rogers and McClelland

associated with the item to learn. For example, when a kid sees a picture of daisy, although he doesn't recognize it directly, it might recall him a typical image of a flower that he has seen before, and will help him to use this connection to enhance his learning of the new concept of daisy. We believe the incorporation of the additional visual similarity is valid for modeling semantic cognition, and should generate faster learning experience compared with original model.

The categorical visual similarity layer is by computing similarities between each item and general categories we choose. It is conceptually parallel to the representation layer from items, and it enters the hidden layer together with representation and relation layers. Similarities of item and categories are calculated from images selected from *ImageNet* with an initiative au-

tomatic pipeline constructed by us. More details will be explained in later sections.

In summary, our contributions in this project are as follows:

1. We applied the model by Rogers and McClelland on a larger data set CCFD, and have proved the model is also effective in simulating human cognitive learning on large space of items and attributes.

2. We extended the model by incorporating information from visual similarity and modified the neural network structure by adding one more input layer. The experiment results showed the modified model is valid in representing cognitive development with faster learning experience.

2. Related Work

Semantic cognition encompasses human performance based on knowledge about the concepts, the properties, and their relations. One approach to model semantic cognition is the parallel distributed processing (PDP) framework. Under this framework, cognitive processes arise from interactions of units (i.e., neurons) through synaptic connections. The knowledge that governs processing is stored in the strengths of the connections and is acquired gradually through experience, which simulates the conceptual development in childhood [Mcclelland and Rogers, 2003].

The neural network model of semantic cognition developed by Rogers and McClelland builds on earlier models by Hinton, Rumelhart and Williams [Rumelhart et al., 1986]. The model integrates the strengths and overcomes the weaknesses of hierarchical, categorization-based approaches and similarity-based approaches. In Hinton's model, two of the constituents of a three-item proposition (such as 'canary ISA —') could be presented with the task of filling in the third constituent ('bird'). Filling in occurs through the propagation of activation among units through their connections, and the outcome depends on the strengths of the connections, which are shaped by experience [Mcclelland and Rogers, 2003]. R&M extend the model to build a simulated neural network model that learns propositions about objects and their properties. The model is trained with propositions about eight different plant and animal concepts, including trees, flowers, fish, and birds. The input layers consists of "item layer" and "relation layer". The "hidden units" between the inputs and outputs learns the internal representations that capture semantic relationships between concepts. The model explains the tendency towards progressive differentiation of concepts observed in development. R&M also modeled the deterioration of semantic knowledge in dementia by adding noise to the learned representations.

3. Data

3.1. Data Description

We use a Chinese Conceptual Semantic Feature Dataset (CCFD) [Deng et al., 2021]. It contains 1,410 objects of various kinds with their semantic features in both Chinese and English. For our project, we use the English version only. The features include six relations, which are "is", "has", "can", "like", "need" and "other". Each object has dozens of features with all or a subset of those six relations, which leads to a total of 51,957 records. Moreover, features of each object have a score of normalized number of participants mentioning that feature.

3.2. Data Preprocessing

Considering the sparsity of our original dataset and computation limit, we subsample a subset of data. Specifically, for each item we filter out the attributes that have a normalized number of participants among 30 mentioning that feature that is smaller than 5, so that each attribute is generally reasonable for that item. Next, we count the number of records for each item, and only keep those who have equal to or more than 15 records so that each item has enough attributes. This leads to a dataset of 157 items and 527 attributes, which is fairly enough for our analysis. We utilize the LabelBinarizer function from sklearn.preprocessing package to perform one-hot encoding on items, relations and attributes respectively. We further convert the one-hot encoding of attributes into numpy array. Then, we use the groupby function to group the data by item and relation, and apply a sum function on the encoded attribute to let each row contain all the attributes of a relation of each item. Finally, we concatenate the encoding of items, relations and attributes to produce the encoded data.

For extended model part, due to the complication of constructing categorical visual similarity, we further

filtered the data set and kept only items of animals. We perform the same data preprocessing methods as described above. The final data set for training neural network with extension of visual similarity contains 77 animals and 239 attributes.

3.3. Image Processing and Category Similarity

For the extended model, we utilize similarity scores between images of items to represent visual similarity in majorly two ways. The first way is from direct human evaluation. We collect multiple images for each item, and ask participants to score similarity of each pair (ranging from 0 to 1) of items by evaluating all related images. This is the most direct and accurate way of evaluating similarity between items.

However, human evaluation is in general not scalable and hard to be applied on large data sets. Thus, we created a pipeline for automatic calculation of similarity of images. After collecting images for each item, we used another package called *Img2Vec* to automatically transform images into vectors, and then calculate cosine similarity scores between pairs of vectors as the visual similarity between items. For reference, *Img2Vec* library uses the ResNet50 model in TensorFlow Keras, pre-trained on ImageNet [Krizhevsky et al.], to generate dense image embeddings of 2048 dimensions.

To construct visual similarity between items and certain categories, we need to determine the categories from both domain knowledge and existing items of data set. Since the extended model is applied only on data of animals, we followed the 6 major groups of animals from biological definition as our categories, which are amphibians, birds, fish, invertebrates, mammals and reptiles. We randomly selected 4 images from each category as representatives. Then for each of them, we calculated the similarity score between the item image and the category image. The final visual similarity between item and given category is derived from the average of the 4 similarity scores, as demonstrated in Figure 2.

In this way, for each item of animals, we will get a 6 dimension vector, with each entry representing similarity between the item with each major animal group. These visual similarity vectors will enter neural network model as a separated additional layer.

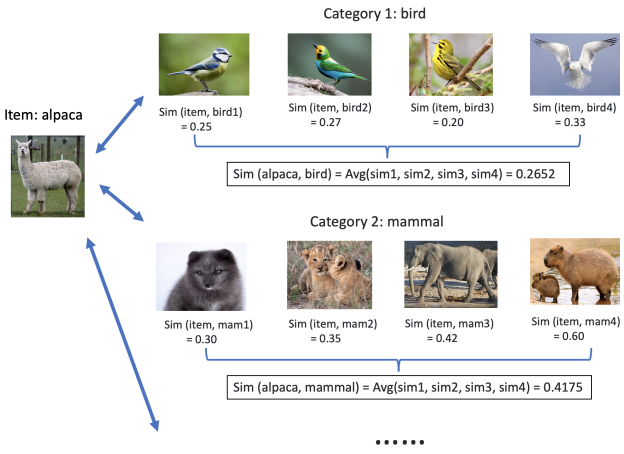

Figure 2: Calculation of Categorical Visual Similarity

4. Models and Experiments

4.1. Original Model

We mainly apply the neural network model by Rogers and McClelland [Mcclelland and Rogers, 2003] and explore the difference between the outputs from their results and our dataset. There are two input layers: item layer and relation layer, and two hidden layers: representation layer and hidden layer. In the forward function, items are passed to the representation layer, and representation layer together with the relation layer are passed to the hidden layer. Finally, the hidden layer is passed to the output which is the attribute layer. We use the ReLu activation function for the representation and hidden layers, and sigmoid activation for the attribute layer. We set representation size of eight and hidden size of fifteen, and learning rate of 0.1. The total epochs for training is 2500, we extracted representation layer learnt at epoch 500, 1000 and 2500 for analysis of learning process.

4.2. Extended Model

The extended model is by adding an additional input layer of categorical visual similarity on original model. This new input layer is parallel to representation and relation layers, and enters hidden layer directly. Categorical visual similarity is calculated from the similarity between each image of each item and of each major category, as described in Section 3.3. The dimension of categorical visual similarity layer is *num items * num categories*. An illustration of model structure is shown in Figure 3.

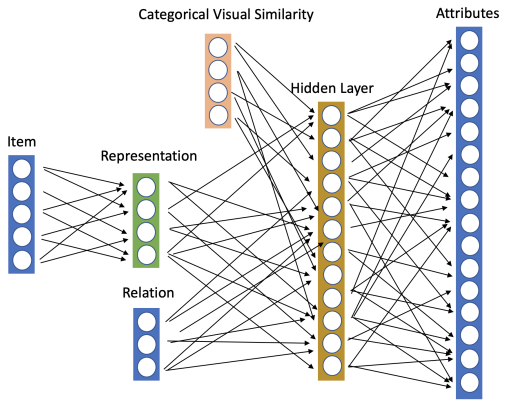

Figure 3: Extended NN Model Structure

We firstly conducted a prototype experiment with data used by [Mcclelland and Rogers, 2003] to initially observe the rationality of the extended model. The data contains 8 items - pine, oak, rose, daisy, robin, canary, sunfish and salmon. We proposed 4 categories which are tree, flower, bird and fish. We used human judgement of image similarities for this prototype experiment to get most accurate evaluations. Participants are presented with multiple images of each item and categories to participants and asked to score similarity only based on images.

Then we moved to the main experiment with data of all animal items. The categorical visual similarity for main experiment is calculated by automatic pipeline described in Section 3.3. All hyperparameters including hidden layer size, learning rate and number of epochs are kept same as original model.

5. Results and Discussion

5.1. Original Model Outcome

In order to visualize the relative locations of the different items' representation, we used PCA (Principle Component Analysis) to project the high-dimension representation space into 2-D scatter plots that preserve the relative distances. In order to discover the dynamics of differentiation in development, during modeling, we saved the representation results with epoch 500, 1000 and 2500, and drew plots for each epoch to explore the model's learning process. All the plot results are presented in Figure 4. As we can see, at epoch 500, there is no obvious differentiation of each category and most items tend to cluster together. At epoch 1000, items start to detach each other and model starts to have some differentiation. Some animals tend to cluster together (such as rabbit, fox, timber wold, mole, and otarriinae). Finally, at epoch 2500, items that belong to the same category are near each other while unrelated items tend to be far apart. There are two major clusters: animals and food. There is also a small cluster that consist of items that are not belong to animals or food. Among each cluster, similar items that belong to the same sub-category also tend to group together (such as fish, carp, and Jellyfish).

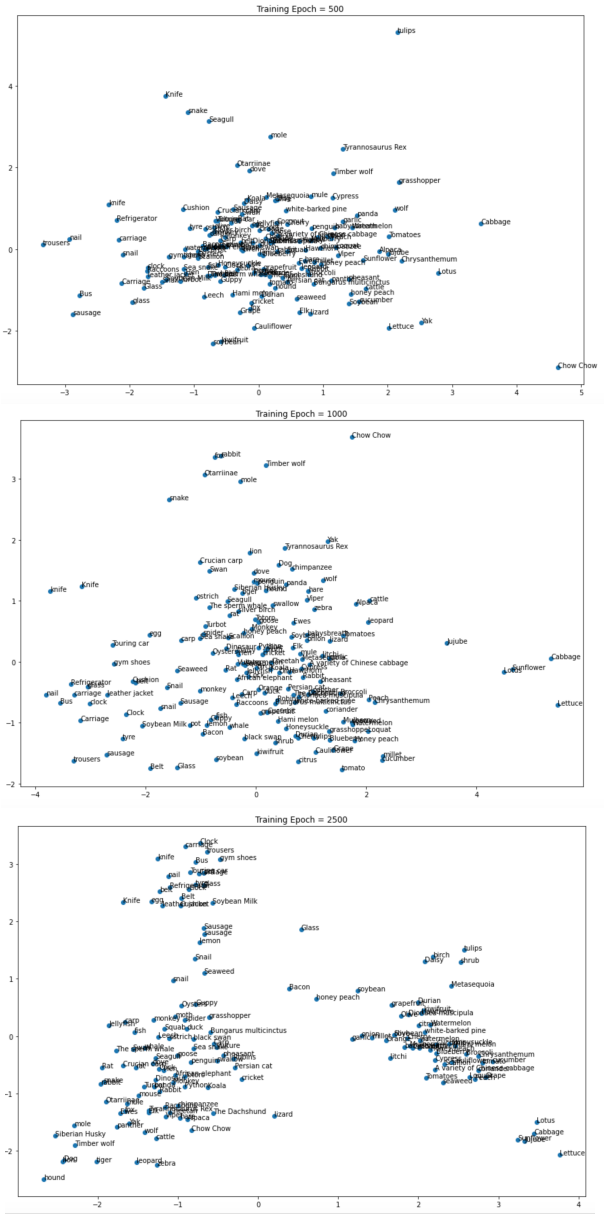

Figure 4: 2-D Graph for Representations

5.2. Extended Model Outcome

We firstly discuss the results of our prototype experiment using same items and attributes from [Mcclelland and Rogers, 2003] but adding visual similarity from human evaluation. The hierarchical clustering comparison is presented as Figure 5, based on the euclidean distance of items at training epoch 500, 1000 and 2500. As observed from the comparison, we can see the incorporation of visual similarity significantly enhance the learning experience of differentiating items. Especially, at epoch 1000, the original model could only cluster items into two major categories of plants and animals, while our extended model is already able to form 4 clusters of trees, flowers, birds and fish. It proves our implementation of visual similarity is effective in providing useful information to accelerate cognitive learning process.

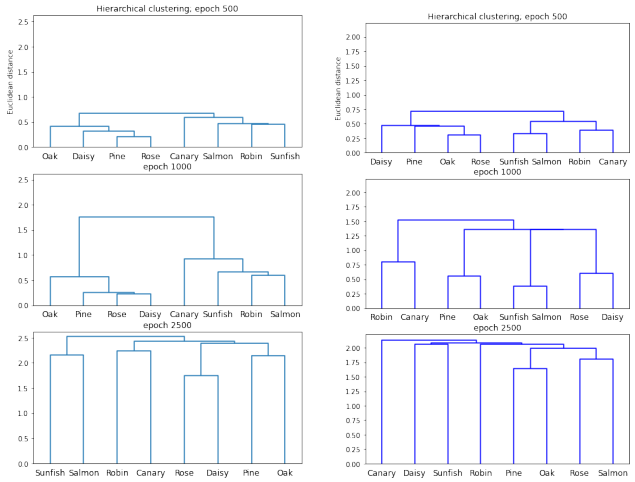

Figure 5: Hierarchical Clustering Comparison between Original (left) and Extended Model (right)

For the main experiments trained on all animals with computer-based visual similarity calculation, we extract the representation layer after training epoch 2500 and generate 2-D PCA scatter plots with similar methods as before (as shown in Figure 6). The plots are not able to imply general improvement of grouping animals. However, there exists small groups of similar animals which are much closer to each other (such as viper bungarus multicinctus, which are both snakes, and alpaca mule, which are both horse-like mammals). We think the failure of learning enhancement over the whole data set is due to inaccurate visual similarity from algorithm processing of images. Certainly, there could be many noisy and uncontrollable information for images, such as background color and positions of animal items, which demands much more complicated processing methods than *Img2Vec* which is based on extracted pixels along. The closer distances of certain groups are from the very similar sample images of such items, which still proves the effectiveness of improvement from incorporating visual similarity. But in order to generate more robust results, we would need further improvement on image collecting and processing strategy.

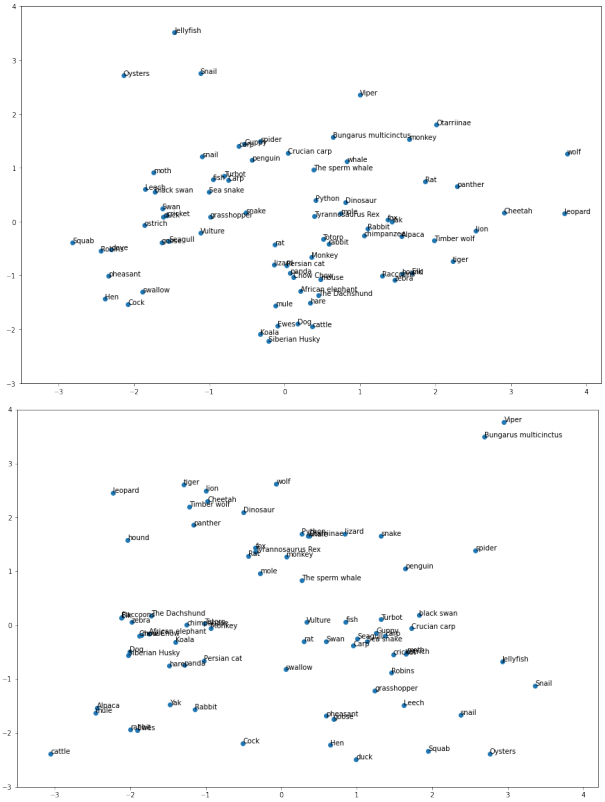

Figure 6: 2-D Graph for Animals trained on Original (above) and Extended Model (below)

6. Conclusion

In conclusion, we applied the Rogers and McClelland model [Mcclelland and Rogers, 2003] on the Chinese Conceptual Semantic Feature Dataset [Deng et al., 2021], whose number of items and attributes are much larger than the data used in original paper, and

we extended the model by adding one more input layer about categorical visual similarity from human evaluation to improve the model's learning process. For the original model outcome, items from CCFD starts to be differentiated at epoch 1000, and finished clustering well at epoch 2500. With respect to such results, we proved that the R&M model also works perfectly for larger dataset in the aspect of simulating human cognitive learning.

As for the extended model outcome, the experiment on original data by [Mcclelland and Rogers, 2003] and visual similarity evaluated by human judgement proved that incorporation of visual similarity is valid in explaining cognitive behavior, and can facilitate the learning process compared to original model. However, when training on animals from CCFD, the 2-D PCA plots failed to validate model learning enhancement in grouping animals. The results are majorly due to noisy factors for animal images and the simplicity of our image processing pipeline. Future work is needed in order to further test the robustness of our extended model.

7. Future Work

For future work, we mainly focus on improvement of categorical visual similarity embeddings. As discussed in previous sections, the quality of the additional layer plays an important role for model performance evaluation. One way for generating more accurate vectors is applying more human evaluation on data sets, or incorporate cross-reference of human judgement of the algorithm evaluation. This is the most direct strategy yet would consume much time and effort.

Another direction to improve the accuracy of estimating visual similarity embedding is to add more steps into the image processing pipeline. For example, setting bounding boxes of images by object detection tasks may effectively help us to control the size of item and remove unnecessary backgrounds. Also, applying certain filters and max pooling of images can reduce unnecessary color variances to better preserve the shape. These strategies will help us remove uncontrollable factors and produce more representative image vectors for calculating cosine similarity.

Finally, we might also generalize the existing knowledge from visual similarity into other areas, such as contextual representations from text data, or pre-trained hierarchical results from related cognitive models. The choice of domains of additional information mainly depends on the assumptions of psychological development in order to support or test learning theories in cognitive science. We believe our extended neural network structure will contribute to further improvement of semantic cognition modeling in more related fields.

References

James Mcclelland and Timothy Rogers. The parallel distributed processing approach to semantic cognition. *Nature reviews. Neuroscience*, 4, 05 2003. doi: 10.1038/nrn1076.

Yaling Deng, Ye Wang, Chenyang Qiu, Zhenchao Hu, Wenyang Sun, Yanzhu Gong, Xue Zhao, Wei He, and Lihong Cao. A chinese conceptual semantic feature dataset (ccfd). *Behavior Research Methods*, 02 2021. doi: 10.3758/s13428-020-01525-x.

David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams. Learning Representations by Back-propagating Errors. *Nature*, 323(6088):533–536, 1986. doi: 10.1038/323533a0.

Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*, page 2012.