Semantic Cognition using Question Answering

Abhijeet Gusain (agusain1@stevens.edu)

Abstract

This report explores the foundations of cognitive processes, emphasizing their emergence from neuronal interactions and synaptic connections within the brain. Our approach centers on the idea that knowledge in interactive and distributed processing systems is encoded in connection strengths, evolving gradually through experiential learning. The acquisition and generalization of properties across objects are intricately linked to the patterns of neural activity probing the stored knowledge in these connections. We investigate how semantic knowledge degrades with the decline of neural activity patterns, shedding light on the impact of brain disorders on cognitive functions. With the advancement of transformer-based models, we want to see how well they perform in semantic cognition. For this purpose, we formatted a question-answering dataset to train a transformer and related its understanding to human semantic cognition.

Keywords: Semantic Cognition, Transfomer, DistilBERT

Introduction

There have been many experiments to understand human semantic abilities which are focused on understanding human cognition. The mechanism of semantic cognition is still an open question. Recent advancements in transformer-based (Vaswani et al., 2017) AI models with a growing interest in Large Language Models (LLMs) have opened another avenue of generating additional data for model training. State-of-the-art pre-trained LLMs (e.g. Llama (Touvron et al., 2023), ChatGPT (Ray, 2023), Falcon (Penedo et al., 2023)) can engage with humans in fluent conversations often working as a virtual assistant in solving complex tasks including Medical Exams questions (Gilson et al., 2022), code snippet generation (Chen et al., 2021), and multi-session conversations (Shuster et al., 2022). The outputs of these LLMs can be optimized by modifying the input prompts. These are a set of instructions given as input to the LLMs in textual format detailing the desired outputs as well as any additional conditions or criteria. Interaction with LLM raises the question of whether these models can represent and understand the information in the same as humans do.

Semantic cognition defined in (Ralph, Jefferies, Patterson, & Rogers, 2017) refers to our ability to use, manipulate, and generalize knowledge that is acquired over the lifespan to support innumerable verbal and non-verbal behaviors. Semantic cognition relies on two principal interacting neural systems: representation and control. This is referred to as the controlled semantic cognition framework. Control here refers to how the learning progresses; which things are learned first and which things are learned later.

I am redefining the problem as mentioned Data section and applying transformer-based DistilBERT to visualize the representation learned by them. I compared these representations to a previous experiment done by Rumelhart (D. Rumelhart, Zornetzer, Davis, & Lau, 1990).

Related Work

There are two foundational models approach to understanding semantic cognition.

The hierarchical propositional approach

Back when scientists were first using computer models to understand how our minds work, they thought that our understanding of things is based on categories and statements. Quillian Quillian (1968) suggested that if we organize concepts in a hierarchy from very specific to very general, we could save space in our brains. For example, if a statement is true for all living things, we could just store it at the very top of our mental hierarchy. Other statements true for only certain groups, like animals but not plants, could be stored at the next level, and so on. Specific facts about one thing could be stored right with that thing. To figure out if a statement is true for something, you could check that thing and move up the hierarchy until you find where the statement is stored or reach the very top. It's like a mental tree structure to help us organize information. Quillian thought that if we organize concepts in a tree structure from very general (like ”living things”) to very specific (like ”pine tree”), it would be efficient. But, there are problems with this idea when we look at how our minds actually work.

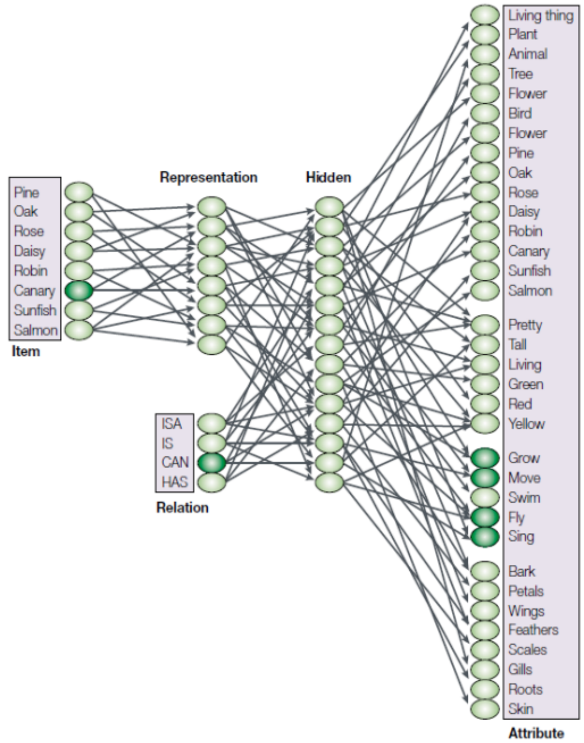

Figure 1: Data representation of Rumelhart model

Firstly, Quillian's model predicted that we'd be quicker to remember specific things about objects rather than general things. However, real-life experiments didn't show that. It turns out we're not faster at recalling specific details just because they're stored closer in our mental hierarchy.

Another issue is that Quillian suggested storing information at the top level, like "living things." But that raises questions. Which general categories should we use? When should we start using them? What criteria do we follow to create these categories? Also, it's tricky because not all members of a category share the same traits. For instance, not all plants have leaves; some, like pine trees, have needles. If we say 'has leaves' is true for all plants, we need a way to say it's not true for those that don't have leaves. But if we only store it with plants that do have leaves, we lose the benefit of generalizing that knowledge. So, the model faces challenges in accurately representing how our minds organize and retrieve information.

Parallel distributed processing In Parallel Distributed Processing (PDP) (D. E. Rumelhart, McClelland, & PDP Research Group, 1986) models, information processing happens through the spread of activation among simple units that function like neurons. Unlike traditional models, semantic information isn't directly stored; instead, it's reconstructed when prompted, a process known as pattern completion.

Let's take an example from Hinton's (Hinton & Anderson, 2014) early model. Imagine you're given two parts of a three-item proposition, like 'canary ISA —' and your task is to fill in the missing part ('bird'). In this model, filling in happens as activation spreads among units through their connections. The final result depends on the strengths (or weights) of these connections, and these weights are shaped by experience.

When Hinton first introduced the model, there were only basic algorithms available for learning these connection weights. This means that the model learns and refines its understanding of relationships between concepts based on experience and the patterns of activation among its units.

The model, introduced by Rumelhart (D. Rumelhart et al., 1990), features a feedforward structure, allowing activation to flow in a single direction—from units representing items and relations through hidden layers to an output layer with potential completions of three-constituent propositions. This design simplifies Hinton's model, where activation could flow in all directions.

Training the network involves multiple epochs, or sweeps through a set of training examples. Small adjustments to connection weights are made after processing each example, leading to gradual learning, resembling the developmental process in children who gradually acquire knowledge through day-to-day experiences. While the training set isn't fully representative of children's experiences, it captures two crucial aspects: the hierarchical similarity structure observed by Quillian and the way children learn similarities and exploits them through exposure to object examples.

Data

Data used for this experiment has 8 items named Pine, Oak, Rose, Daisy, Robin, Canary, Sunfish, Salmon and possible relation using ISA, Is, Can, Has and 36 properties naming Living thing, Plant, Animal, Tree, Flower, Bird, Fish, Pine, Oak, Rose, Daisy, Robin, Canary, Sunfish, Salmon, Pretty, Big, Living, Green, Red, Yellow, Grow, Move, Swim, Fly, Sing, Skin, Roots, Leaves, Bark, Branch, Petals, Wings, Feathers, Gills, Scales.
In the previous experiment and shown in Figure 1 we have an item layer and representation layer separately for items and a relation layer attached to the relation layer. This is explicitly segregating items and relations for the experiment. In contrast to this, my experiment follows a more sophisticated approach to treating item-relation-property

as a single sentence. I represented the same data in question-answer format by which we can learn whether the property attached to the item is valid or not.

For example: Can a canary sing? Answer: Yes

This reformatting of data seems simple but it does not explicitly categorize items, relations, and properties. The final reformation resulted in 1152 combinations of the whole dataset.

Mehtod

Our experiment aims to learn the same representations as by feedforward network even though the data representation is changed. Since our data is represented as question answering, I decided to use BERT (Devlin, Chang, Lee, & Toutanova, 2019) which is designed to pre-train deep bidirectional representations from the unlabeled text by joint conditioning on both the left and right context in all layers. As a result, the pre-trained BERT model can be fine-tuned with just one additional output layer to create state-of-the-art models for a wide range of tasks, such as question answering and language inference, without substantial task-specific architecture modifications. Pre-trained DistilBERT (Sanh, Debut, Chaumond, & Wolf, 2020) is perfect for generating representation for this task.

- Tokenization: I used WordPiece (Song, Salcianu, Song, Dopson, & Zhou, 2021) for tokenization of sentences

- Padding: Pad all lists to the same size, so they can represent the input as one 2-d array, rather than a list of lists (of different lengths).

- Attention Masking: If padded data is directly sent to DistilBERT, that would slightly confuse it. We need to create another variable to tell it to ignore (mask) the padding added when it's processing its input

- Output: Distilbert is given both padded data and attention masking to generate the final embedding for the sentence.

- Evaluation: The evaluation is done by added logistic regression model to the result of DistilBERT to calculate whethe these representations can reach the same semantic understanding as the previous experiment mentioned in Section 2 (Related work)

Challenge: How to visualize? As we have removed the representation layer and item layers from the previous experiment, I need to come up with a new method to visualize the "learning" from DistilBert. I noticed that each item has 144 (4*36) embedding related to them which can be used to describe their item-relation-properties together. I summed these embeddings for each item and then compared the Euclidean distance between them.

Results

By comparing the Euclidean distance between the summed embeddings, the dendrogram learned is similar to the previous experiments as shown in Figures 2 and 3. We can notice that these representations are similar to human semantic abilities to categorize objects given properties associated with them.

Learning can be controlled by selecting the layers that can demonstrate how learning is being progressed and control which things to learn and forget. The experiment shows the second last layer has not learned all clustering in comparison to the last layer where all heriarchical representations are learned correctly.

Figure 2: The hierarchical clustering of DistilBERT using second last layer

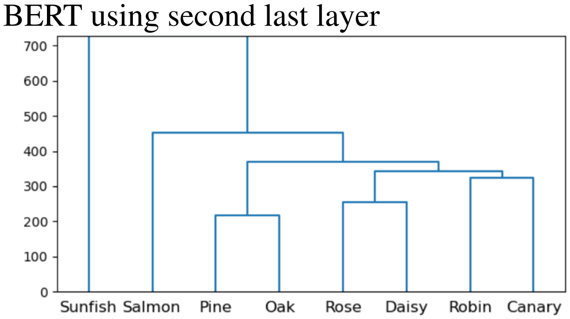

Figure 3: The hierarchical clustering of DistilBERT using final layer

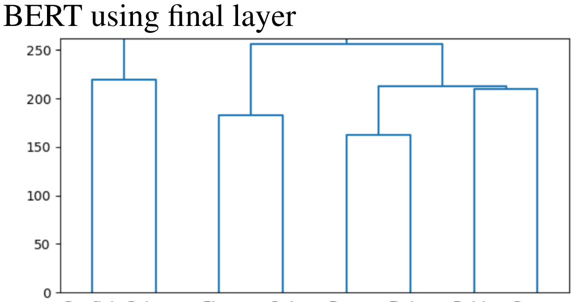

Table 1: Accuracy chart for DistilBERT-Logistic Regression Train-Test split 80-20%

Data	Accuracy
Train	94.51
Test	94.09

Conclusion

This experiment emphasizes the success of utilizing DistilBERT to model semantic cognition and highlights the potential for transformer-based models in capturing a human-like understanding of concepts and their properties without the need to explicitly segregate items and properties.

Acknowledgments

I extend my deepest appreciation to Professor Mark Ho for his unwavering support, mentorship, and the invaluable lessons I have gained under his guidance. His influence has left an indelible mark on this project, and I am profoundly grateful for the privilege of having him as a mentor.

References

Chen, M., Tworek, J., Jun, H., Yuan, Q., de Oliveira Pinto, H. P., Kaplan, J., ... Zaremba, W. (2021). *Evaluating large language models trained on code.*

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). *Bert: Pre-training of deep bidirectional transformers for language understanding.*

Gilson, A., Safranek, C., Huang, T., Socrates, V., Chi, L., Taylor, R. A., & Chartash, D. (2022). How well does chatgpt do when taking the medical licensing exams? the implications of large language models for medical education and knowledge assessment. *medRxiv*, 2022–12.

Hinton, G. E., & Anderson, J. A. (2014). *Parallel models of associative memory: updated edition*. Psychology press.

Penedo, G., Malartic, Q., Hesslow, D., Cojocaru, R., Cappelli, A., Alobeidli, H., ... Launay, J. (2023). *The refinedweb dataset for falcon llm: Outperforming curated corpora with web data, and web data only.*

Quillian, M. (1968). Semantic memory. m minsky, ed, semantic information processing. *The MIT-Press, Cambridge, MA*, *22*.

Ralph, M. A. L., Jefferies, E., Patterson, K., & Rogers, T. T. (2017). The neural and computational bases of semantic cognition. *Nature reviews neuroscience*, *18*(1), 42–55.

Ray, P. P. (2023). Chatgpt: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. *Internet of Things and Cyber-Physical Systems*.

Rumelhart, D., Zornetzer, S., Davis, J., & Lau, C. (1990). An introduction to neural and electronic networks. In *An introduction to neural and electronic networks* (pp. 405–420). Academic Press.

Rumelhart, D. E., McClelland, J. L., & PDP Research Group, C. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition, vol. 1: Foundations*. MIT press.

Sanh, V., Debut, L., Chaumond, J., & Wolf, T. (2020). *Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter.*

Shuster, K., Xu, J., Komeili, M., Ju, D., Smith, E. M., Roller, S., ... Weston, J. (2022). *Blenderbot 3: a deployed conversational agent that continually learns to responsibly engage.*

Song, X., Salcianu, A., Song, Y., Dopson, D., & Zhou, D. (2021). *Fast wordpiece tokenization.*

Touvron, H., Martin, L., Stone, K., Albert, P., Almahairi, A., Babaei, Y., ... others (2023). Llama 2: Open foundation and fine-tuned chat models. *arXiv preprint arXiv:2307.09288*.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, *30*.