Relations between Variables

Jonas Peters, Philipp Haack

July 3, 2025

Introduction

In chapter 3 *Relations between Variables* of the book *The Algebraic Mind* by Marcus (2001), the author analyzes to what extent we can implement *algebraic rules* with neural networks, specifically *multilayer perceptrons*.

An algebraic rule is a formal definition to compute operations such as multiplying two numbers, or even constructing past tense of verbs, as it includes appending `-ed` to the verb stem.

One important aspect of these algebraic rules is that we humans seem to be able to freely generalize these operations. For example, you can calculate $42 \cdot 9$ despite never having encountered this exact arithmetic task before, and you can create novel sentences which are syntactically correct and semantically meaningful.

The discussion in Marcus' book comprises two aspects: First, there is the question of how the human mind is capable of performing tasks like cognition, language, and learning. Second, he argues which architectures he believes are most suited for specific tasks, and what properties these neural networks should have. The hope is that advancements in one area facilitates progress in the other.

Line of Argument

In the presentation, we will thoroughly analyze Marcus' line of argument. In order to give you a quick overview, we will introduce you to his main arguments.

First, he claims that many functions we humans can compute are one-to-one mappings. Marcus refers to them as *universally quantified one-to-one mappings*, or *UQOTOM* for short. One important aspect of these mappings is *injectivity*, i.e. no two distinct inputs map to the same output. This includes functions like adding by one, forming simple past of verbs, or even the identity function which just returns back the input.

To Marcus' mind, the models ability to represent and learn these UQOTOM is a central aspect of the models capability to implement algebraic rules, since injectivity enforces the model to generalize and not rely on memorization. Based on this premise, he argues that MLPs that are trained by *backpropagation* are insufficient for generalizing UQOTOM beyond their training space.

He proposes a solution for enforcing MLPs to learn UQOTOM: Using a single input node and only linear activations. This way, the network can only learn a linear mapping (or map everything to zero). However, having multiple input nodes remains a challenge. Thus, he argues that we should focus on using *registers* instead, and how to implement and coordinate elementary instructions on them.

Marcus' argument in a nutshell

- **Premise**: Many algebraic rules are UQOTOM
- **Premise**: Humans can learn/ generalize these rules, so our models should too
- **Conclusion**: Models should be able to represent and learn UQOTOM (even with restricted training data)
- **Claim**: MLPs don't generalize UQOTOM well outside the training space (based on observation, argued via training independence)
- **Conclusion**: Vanilla MLPs are not well suited for tasks that require a high amount of symbol manipulation
- **Suggestion**: Motivated by computer architecture and supported by case studies, he suggests the use of registers

Contemporary Language Models

As Marcus released his book in 2001, we can evaluate his claims in hindsight. Are feedforward networks really fundamentally flawed when it comes to symbol manipulation tasks? To what extent do successful models encode higher order logic in their architecture, and do they use some kind of registers?

In the recent paper *Attention Is All You Need* by Vaswani et al. (2017), the authors introduced a new neural network architecture called *Transformer* to process human language. This technology now serves as the foundational architecture for the large language models that power virtually all modern AI systems, from ChatGPT to Google's Gemini.

We will give a high level overview of their functionality without going into detail. This allows us to assess Marcus claims based on contemporary research. To this end, we introduce the paper *The Illusion of Thinking* by Shojaee et al. (2025).

Discussion

We will discuss the validity of the arguments made by Marcus', and to what extent alternative conclusions are possible. Furthermore, we also discuss some open ended questions regarding the debate *symbolism vs connectionism.* Some questions might be:

- Should we assess the models capability to generalize algebraic rules by its ability to learn UQOTOM?
- Does the recent success of connectionist models challenge the role of symbolic models in future AI research?

References

Marcus, Gary F (2001). *The algebraic mind: Integrating connectionism and cognitive science*. Cambridge, MA: The MIT Press.

Shojaee, Parshin et al. (2025). *The Illusion of Thinking: Understanding the Strengths and Limitations of Reasoning Models via the Lens of Problem Complexity*. arXiv: `2506.06941 [cs.LG]`.

Vaswani, Ashish et al. (2017). "Attention is all you need". In: *Advances in neural information processing systems 30 (NIPS 2017)*, pp. 5998–6008. URL: `https://proceedings.neurips.cc/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf`.