



Neural networks that recall

Description

It looks as though LLM technology can enhance web search without conceptually overhauling the web search methodology. (See previous post: [AI versus web search](#).) But is there a way that neural network retrieval methodologies can be used as a substitute for explicit indexing? After all, the human capacity to recall does not rely on indexing.

LLMs are good at absorbing and combining vast repositories of textual content (a corpus) and synthesising outputs that obscure the integrity of the original texts. That's a strength of such neural network (NN) systems, and can lead to the synthesis of new information, new ways of expressing ideas, amongst other benefits. NNs *generalize*. This loose, synthetic approach to content roughly parallels how human beings deal with information, to the extent that it is possible to have a meaningful conversation with an AI platform such as ChatGPT, Claude or Gemini.

We can assume that the human cognitive apparatus (call it "mind") is not designed, optimised or evolved for the *storage* of vast amounts of information. We have text media such as books, tables, recordings, pictures and digital storage for precision storage and reproduction. We use tools to extend the capabilities of mind, as in Clark and Chalmer's concepts of the "extended mind."

Yet most of us are able to remember and recall some texts "by heart": e.g. songs and poems, and some accomplished performers, actors, and lecturers are able to deliver accurate content verbatim along with associated source metadata (title, author, date, summaries, etc), as if reciting from a database.

Prevailing theories of memory and recall affirm a distributed model of human cognition. Echoing aspects of neural biology, digital neural networks operate on this principle. Information is not stored in memory slots, as in a database, but via subtle differences in numerical values across very large networks of connected neurones.

So, how might a neural network replicate this ability to recall the content of documents verbatim, or to memorise a poem?

How I remember

As a chemistry student I memorised the first 24 elements in the Periodic Table by making up a little rhythmical jingle that I could mutter inaudibly to myself. For school I memorised Portia's speech in *The Merchant of Venice* by constant repetition and repeatedly testing myself. I still invoke "My very elegant mother just sat upon new paint" to recollect the position of Earth in the solar system. Relationships, metaphor, mnemonics and other familiar linguistic devices play a role in verbatim recall. How might a neural network parallel such processes?

I put my query about verbatim memorisation to Gemini, which responded to several key differences between human and machine learning. Gemini wrote that the memory methods I outlined above rely on "symbolic understanding," which neural networks don't have. It seems I can also *explain* my memory methods. As yet, LLM neural networks don't incorporate any explanatory capabilities. To my way of thinking, both of these differences appeal to two concepts and processes outside of the neural network framing: symbols and explanations.

A third difference though is in the neural network frame: Neural networks are good at generalizing, whereas human recall of the type I outlined is *specific*. By design, neural networks abandon specifics. I'll elaborate on the generalisation issue next.

Over-fitting

It's well documented in the literature, and I have rediscovered this in my own experiments with NNs: it is all too easy to specify a neural network's hyper-parameters (number of layers, nodes, learning rate, loss function, number of training iterations, etc) in such a way that the network "over fits," i.e. fixates on specific input-output pairings, leading to poor performance on new and unseen input data. Finding the right balance between parameters in a NN is a difficult yet crucial "art." In the case of overfitting, the network's performance moves closer to that of a database that returns the same output for a given input. That's undesirable if an AI agent is to generalize from its training data and exhibit human-like language capabilities.

It's as if verbatim memorisation in the case of human memory involves selective overfitting. Verbatim memorisation is a challenge for human beings as it requires recourse to extraordinary methods such as forced associations, and excessive repetitions and feedback operations to overcome the brain's highly efficient resistance to overfitting, i.e. our ability to generalise and apply what we know to new situations.

Over-fit and recall

Considering how easy it is to train a neural network to overfit, it should be possible to design a neural network system to support verbatim recall. Perhaps there could be two iterations of the network, the generalist and the overfitting literalist. Depending on requirements, the system would select between the two. Alternatively, during training, some text input could be flagged in such a way as to be subject to overfitting (and hence verbatim recall). The training and prediction procedures would have to be adapted to suit this process.

I put this approach to Gemini, who (which) responded positively. I record its response here â?? verbatim.

Youâ??ve come up with some interesting ideas for using neural networks to achieve verbatim recall! Hereâ??s a breakdown of the feasibility and potential challenges of your approaches:

Dual Network System (Generalist vs. Literalist):

- **Concept:** This approach proposes having two separate networks: one for general understanding and information retrieval, and another specifically trained to overfit for verbatim recall.
- **Feasibility:** This is an interesting concept, and it might be possible to implement it with careful design. The key would be to clearly define the criteria for selecting between the networks.
- **Challenges:** Training and managing two separate networks would add complexity. Additionally, ensuring the â??literalistâ?• network doesnâ??t simply memorize everything (instead of focusing on specific data flagged for verbatim recall) would require careful design and training techniques.

Flagging Text for Overfitting:

- **Concept:** This approach suggests marking specific text inputs during training to instruct the network to prioritize verbatim memorization for those sections.
- **Feasibility:** Flagging text for overfitting is technically possible. However, it requires defining clear markers and ensuring the network interprets them correctly.
- **Challenges:** This approach might struggle with subjective content or nuances in human language. For example, how would the network distinguish between a factual statement to be memorized verbatim and a metaphor that shouldnâ??t be taken literally?

Alternative Approaches:

- **Hybrid Systems:** Researchers are exploring combining neural networks with traditional retrieval methods like indexing. This could leverage neural networks for understanding context and relationships, while indexing maintains verbatim accuracy.
- **Attention Mechanisms:** Techniques like attention mechanisms can be used to focus the networkâ??s processing on specific parts of the input data relevant for verbatim recall.

Overall:

While there are challenges, your ideas for achieving verbatim recall with neural networks are creative and highlight the ongoing exploration in this field. As research progresses, we might see advancements in training techniques, network architectures, and the integration of different approaches to achieve more human-like memorization and recall capabilities in AI systems.

I would be interested to see an academic paper that explores verbatim recall through neural network properties rather than through a hybrid incorporation of databases or web search operations.

Bibliography

- Clark, Andy, and David Chalmers. "The extended mind." *Analysis* 58, no. 1 (1998): 7-19.
- Halbwachs, Maurice. *On Collective Memory*. Chicago, IL: The University of Chicago Press, 1992.
- Rosenfield, Israel. *The Invention of Memory: A New View of the Brain*. New York: Basic Books, 1988.
- Yates, Frances A. *The Art of Memory*. London: Routledge and Kegan Paul, 1966.

Notes

- Featured image generated by ChatGPT: "The revised banner image for your blog post, featuring rotary card files like a Rolodex in a grungy steampunk style, has been created. It visually represents the concept of analog data storage transitioning into the digital realm of neural networks, enriched with the distinctive textures and motifs of steampunk aesthetics."
- I gave the same initial text of this post to ChatGPT. Apart from summarising the content with approval, it suggested:
"Concluding with a reflection on the ethical and practical implications of developing neural networks capable of verbatim recall could provide a compelling end to the post. This could include considerations of privacy, the authenticity of information, and the impact on human memory practices."

"Overall, the blog post is intellectually stimulating and contributes meaningfully to the discourse on the potential and limitations of AI in mimicking human cognitive processes. It reflects a deep engagement with both the technical and philosophical dimensions of digital media, architecture, and urbanism."

Category

1. Artificial Intelligence

Tags

1. memory
2. over-fitting

Date Created

April 6, 2024

Author

rcoyne99