Kai Hueske-VanCeylon

Dr. R. McKell Carter

CSCI 3702 - Cognitive Science

11/14/2021

Jack of All Trades, Master of None: Comparing Human and Machine Problem Solving

Problem-solving processes are one of the most practical benefits of intelligence; the better the processes we can come up with, the more effective we (and the machines we build) will be in most applications. Different tasks have different levels of complexity, depending on how they're structured. Solving mathematical problems is a relatively low-complexity task, since it follows a logical series of steps and has a single correct answer. Games like chess are more complex, since they often require a "search" of many possible options or solutions to the problem. Vision and language are highly complex, requiring us to identify objects and connections between ideas. Computers perform really well at low- and moderate-complexity problems, but tend to do poorly with these higher-complexity tasks. Humans, on the other hand, can paradoxically struggle more with these lower-complexity tasks, but can see and converse very naturally.

The reasons for this seem to be evolutionary. The human problem-solving process developed to be widely generalizable, because that provided the highest benefit to survival and reproduction; as the saying goes, evolution uses "whatever's in the room," meaning a problem-solving apparatus which can be reused for different tasks tends to work well (Weller, 2015). Our use of symbolic & spatial reasoning helps enable this approach, as noted by Fabry & Pantsar (2019). However, this general process tends to be less effective for a given task than a task-specific process would be, because it's based on heuristics and reformulations of the problem, which can decrease accuracy and increase the steps needed to complete the task..

In contrast, machine problem solving methods tend to be specialized for a single task, because that's been the modern approach to AI. The process of research, where the next generation of AI models are built by making small improvements to previous ones, takes the place of biological evolution in driving the development of intelligent agents. The particular pressures we apply have "artificially selected" these models to perform well at one task, rather than generalizing to many, and the results bear this out. Simple machine techniques can outperform humans at specific low-complexity tasks, like chess, and even more complex ones like simple customer assistance (Xu et al., 2020), but are still struggling with highly complex tasks. Of course, part of the widespread adoption of this specialization approach is due to the fact that generalizable machine learning is difficult to do, but the selective pressures are still there impacting the development of the field.

Essentially, while humans developed a general, heuristic-driven, approximate problem solving process to fill a generalist evolutionary niche, the processes driving the "evolution" or progress of AI select for systems specialized to particular tasks. This suggests that the pursuit of artificial general intelligence -an artificial intelligence with human-level capability - could benefit from a radical shift in perspective towards a more unified generalist model, rather than trying to combine isolated systems into one generally intelligent agent. To a certain degree, this has been borne out; recent developments in natural language processing like the BERT and GPT-3 models, which can perform better on specific tasks due to a general understanding of the structure of language, indicate that generalization is a powerful tool. Though these models require significant pre-training time & storage space (Brown, 2020), their high performance is a promising sign for the future of AI.

Human Problem-solving

From some perspectives, the human problem-solving process provides a good goal for the performance and design of AI systems. Interestingly, though, it functions in a different way to most current approaches to AI. Fabry & Pantsar argue that, particularly for mathematical problems, our problem-solving approach is largely driven by symbolic and spatial methods (for example, Gauss's method of summing consecutive integers[1]). These symbolic methods aren't what we would consider computationally optimal - they require rearranging or reformulating the problem, which can often entail more "operations" than simply performing the calculation. However, they can be considered "humanly optimal" in that they co-opt neural pathways used for other spatial tasks, allowing them to benefit from more frequent use and development. These methods also benefit from *enculturation*; they effectively use a body of skills that are culturally learned from other cognitive agents during development (Fabry & Pantsar, 2019). The combination of these factors allows symbolic methods to be more effectively employed in human problem-solving than the straightforward, sequential methods we use to build machine systems..

This co-opting of spatial neural pathways in human problem-solving is also a consequence of the broader framework of evolutionary function. Evolution "uses whatever is in the room," according to Parkinson & Wheatley (2015), repurposing existing biological/cognitive components to solve new evolutionary problems that may arise. This evolutionary repurposing is evidenced in many of our social mechanisms; for example, humans use similar neural structures to process spatial, temporal, and social distance (Parkinson et al., 2014). Evolutionary repurposing happens through genetic change over generations, but it's also possible for the brain

[1] Gauss's method entails breaking the sum into pairs of numbers summing to $n + 1$, then multiplying these, a more spatially intuitive interpretation.

to shift within a single lifetime. This shorter-term counterpart is cultural repurposing, wherein we adopt new uses for existing structures through neuroplasticity. Cultural repurposing is what enables Fabry & Pantsar's enculturated problem-solving techniques; these methods are "stored" in culture and taught during the development process, allowing them to take advantage of the existing neural structures for spatial ability (Fabry & Pantsar, 2019).

In this context, our tendency towards symbolic and spatial reasoning takes on a new aspect: it's what allowed us to utilize our powerful cognitive machinery in a way that best generalized to the various social and survival tasks we needed to perform. Having a well-developed spatial ability in a world where success depends on knowing where we are is important, and when there are other agents around us with similarly advanced cognitive abilities, success will also depend on how well we can build and understand beneficial social dynamics. Evolutionarily speaking, it "makes sense" to use the same neural equipment for analogous tasks. It also lets us generalize to tasks we haven't previously seen; if we can form an analogy to a task we've done before, we can use similar processes and pathways to solve it.

These processes together with our specialized neural structures allow us to perform well on a wide variety of complex tasks, like vision and language. Xu et al. (2020) found that humans outperform AI-based chatbots for answering complex customer questions. However, machine problem solvers tend to outperform us on more straightforward tasks like math problems, from simple addition to multivariate calculus, and even for simpler customer service tasks. Our processes of manipulating symbols generalize very well, but aren't as efficient as physical structures dedicated to performing these particular tasks.

Machine Problem-solving

In contrast to the general, heuristic-based human problem solving approach, most machine problem solving techniques - that is to say, current AI technology - are much more specialized. Machine learning techniques tend to focus on specific problems, achieving high accuracy on these tasks but often performing little better than the baseline on others. In addition to this, the tasks machines are best suited for are less complex. Machines excel at solving mathematical problems and searching well-defined problem spaces, meaning they're well suited for (and outperform humans on) problems like calculus and chess. They can achieve significant accuracy in more complex tasks as well, provided that the scope of the task is narrow enough; Xu et al. found that AI chatbots in customer service can outperform humans for simple question answering (Xu et al., 2020). Notably, this doesn't necessarily entail an *understanding* of these techniques, but rather a powerful imitative ability to generate the correct output for a task - which, for many practical purposes, is enough.

However, when tested on these complex tasks with a broader scope, machine techniques run into significant issues. Combined vision & language (V&L) tasks provide a solid basis on which to compare human and machine problem-solving; these are tasks where an AI agent has to combine visual and language inputs, "reason" about the environment depicted in the image, and produce various natural language or simple classification outputs. In other words, they're a rudimentary analog to the problems humans have been evolutionarily optimized to perform. However, the AI agents tend to achieve only superficial success on these tasks, often using "clever" methods to exploit statistical happenings and avoid truly reasoning about the image (Kafle et al., 2019). For example, an agent may be tasked with answering a question about the spatial relationship between two objects in an image - say, an apple, and a table. The agent may have seen many example questions about apples and tables in the data it was trained on, and learned that the relationship is usually that the apple rests on the table. By answering with this common relationship, it has a good chance of increasing its accuracy, *without considering any visual input*. Rather than learning to perform a complex task, machine methods tend to exploit any statistical loopholes they're given in order to increase their reported performance.

This is the crux of the difference between human and machine problem-solving: they have evolved under vastly different selective pressures. While humans evolved general, computationally suboptimal problem-solving techniques under natural selection, machine problem solving has developed under artificial selection to perform well at specific tasks, often in sterile, unrealistic environments. Human evolution was about survival in a complex environment; machine evolution is about learning to give the correct response to a question posed by researchers. This isn't to say that the state of technology doesn't play a part; obviously, problem-solving at a human level is incredibly complex, and the human brain is much more intricate than any system we can currently develop, but the selective pressure is still very real.

A Combined Approach

As Kafle et al. point out, these selective pressures - scoring AI agents on the basis of providing the correct label, rather than whether they truly perform the task we set - hinder the progress of AI as a field. A solution, then, would be to apply the correct pressure. This may be easier said than done; wrangling machine learning systems into learning in a particular way is notoriously difficult and often requires even more manipulation of datasets, problems, and formulations, which can increase the abstract, artificial nature of the task. Nevertheless, various innovative methods have been proposed, from formulating tasks which are impossible without a level of true understanding, to allowing models to answer unanswerable questions with “I don’t know”. These methods seem promising, and may yet provide significant results and advances (Kafle et al., 2019).

Another solution to the issue of differing pressures, and one which has seen significant success, is to adopt the more human generalist problem solving approach. The most recent, state-of-the-art language processing models use pretraining, a technique where a model is allowed to sift through huge language corpora from which it learns a general structure of language. Afterwards, the model can be trained on a relatively small number of examples of a particular task, or even given a single sentence as a directive (Brown et al., 2020). The broad understanding of language that the model builds, in a similar way to neural pathways shared among different human task centers, allows it to achieve remarkably high performance on multiple language-based tasks without adjustment. Recent systems based on these architectures can perform natural language tasks like question answering and detecting logical entailment with 93.6% and 73.0% accuracies, respectively, compared to a human baseline of 95.1% and 97.4% (Nangia & Bowman, 2019).

These models aren’t a silver bullet for building human-level AI, though. Since they’re trained on such large language corpora, they take significant investment of time and effort to build. This represents yet another “evolutionary pressure” on the development of AI; in situations where specialized systems have comparable performance, the use of these generalist pretrained models will be limited, since there exist good options which are less resource-intensive. Performance on these tasks was also in a relatively “sterile” environment, as the testing was done on a dataset rather than in a real-world setting. However, the use of AI agents in practical applications continues to rise, which may provide interesting information on how well these techniques generalize from a lab environment to the real world over the next few years.

Conclusion

Differing selective pressures between human and machine problem-solving have given rise to very different results. While human problem-solving operates in computationally suboptimal ways, it excels at generalizability due to a focus on heuristics and spatial/symbolic reasoning, as well as multipurpose neural structures which can be used for many different analogous tasks. Machine problem-solving, on the other hand, is excellent at specific tasks, but less effective at complex or broad ones. While it can outperform humans in some areas, in general it still lags behind. However, methods have emerged to combine the two, with powerful and promising results.

In light of this, what state is the field of AI in? It seems clear that a strategy based around combining elements of both problem-solving methods is desirable, and can enable strong performance on a breadth of traditional machine learning tasks. It may even be useful for the development of artificial general intelligence, an AI system which could perform as wide a variety of tasks just as well as - or better than - a human could. It’s entirely possible, of course, that methods of problem-solving which are vastly different from the human paradigm could be much more effective for general intelligence, but these are of course difficult for us to conceptualize, much less implement. Perhaps, given time, a problem-solving agent of our own design, with a similar thought process to our own, could innovate on its own structure and bring such a method to life. For now, though, both humans and machines will have to resign ourselves to being jacks of many trades, but true masters of none.

Works cited:

Brown, T. (2020). Language Models are Few-Shot Learners. *CoRR*. https://arxiv.org/abs/2005.14165

Fabry, R. E., & Pantsar, M. (2019). A fresh look at research strategies in computational cognitive science: The case of enculturated mathematical problem solving. *Synthese*, *198*(4), 3221–3263. https://doi.org/10.1007/s11229-019-02276-9

Kafle, K., Shrestha, R., & Kanan, C. (2019). Challenges and prospects in vision and Language Research. *Frontiers in Artificial Intelligence*, *2*. https://doi.org/10.3389/frai.2019.00028

Nangia, N., & Bowman, S. R. (2019). Human vs. Muppet: A Conservative estimate of human performance on the glue benchmark. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. https://doi.org/10.18653/v1/p19-1449

Parkinson, C., Liu, S., & Wheatley, T. (2014). A common cortical metric for spatial, temporal, and social distance. *Journal of Neuroscience*, *34*(5), 1979–1987. https://doi.org/10.1523/jneurosci.2159-13.2014

Parkinson, C., & Wheatley, T. (2015). The repurposed Social Brain. *Trends in Cognitive Sciences*, *19*(3), 133–141. https://doi.org/10.1016/j.tics.2015.01.003

Wood, R. E. (1986). Task complexity: Definition of the construct. *Organizational Behavior and Human Decision Processes*, *37*(1), 60–82. https://doi.org/10.1016/0749-5978(86)90044-0

Xu, Y., Shieh, C.-H., van Esch, P., & Ling, I.-L. (2020). AI customer SERVICE: Task COMPLEXITY, problem-solving ability, and Usage Intention. *Australasian Marketing Journal*, *28*(4), 189–199. https://doi.org/10.1016/j.ausmj.2020.03.005