A Note on Tractability and Artificial Intelligence

Tarek Richard Besold[1] and Robert Robere[2]

[1] Institute of Cognitive Science, University of Osnabrück
tbesold@uos.de
[2] Department of Computer Science, University of Toronto
robere@cs.toronto.edu

Abstract. The recognition that human minds/brains are finite systems with limited resources for computation has led researchers in Cognitive Science to advance the Tractable Cognition thesis: Human cognitive capacities are constrained by computational tractability. As also artificial intelligence (AI) in its attempt to recreate intelligence and capacities inspired by the human mind is dealing with finite systems, transferring the Tractable Cognition thesis into this new context and adapting it accordingly may give rise to insights and ideas that can help in progressing towards meeting the goals of the AI endeavor.

Two famous ideas conceptually lie at the heart of many endeavors in computational cognitive modeling and artificial intelligence (AI): The "computer metaphor" of the mind, i.e. the concept of a computational theory of mind , and the Church-Turing thesis. The former bridges the gap between humans and computers by advocating the claim that the human mind and brain can be seen as an information processing system and that reasoning and thinking corresponds to processes that meet the technical definition of computation as formal symbol manipulation, the latter gives an account of the nature and limitations of the computational power of such a system.

But the computer metaphor and the Church-Turing thesis also had significant impact on cognitive science and cognitive psychology. One of the primary aims of cognitive psychology is to explain human cognitive capacities which are often modeled in terms of computational-level theories of cognitive processes (i.e., as precise characterizations of the hypothesized inputs and outputs of the respective capacities together with the functional mappings between them). Unfortunately, computational-level theories are often underconstrained by the available empirical data, allowing for several different input-output mappings and corresponding theories. To mitigate this problem, different researchers over the last decades have proposed the use of mathematical complexity theory, and namely the concept of NP-completeness, as an assisting tool (see, e.g., [1]), bringing forth the so called "*P-Cognition thesis*": Human cognitive capacities are hypothesized to be of the polynomial-time computable type.

An immediate objection concerns the finite capacity of human intelligence (and, indeed, of all computing systems). The analysis of algorithms — and complexity theory in general — is primarily interested in the *asymptotic* growth in the requirement of resources for problems, which is reflected in the fact that our preferred computational model (the Turing Machine) is theoretically allowed an unlimited amount of time and space which can be used during computation. In contradiction to this theoretical model, one could claim that since all "real" computing systems have a strictly limited amount of resources then they have no more power than that of a finite state machine.

K.-U. Kühnberger, S. Rudolph, and P. Wang (Eds.): AGI 2013, LNAI 7999, pp. 170–173, 2013.

This, however, ignores the purpose of complexity theory and algorithmic analysis, which is to provide tools to study the *rate of growth* of computational resources. In this light, one should read negative results provided by complexity theory as giving an *upper* bound on the size of instances that finite computational systems can solve comfortably, where the upper bound provided depends on the given problem at hand and the complexity of both the reduction used and the problem that was reduced to. Therefore, when the P-Cognition thesis states that "human cognitive capacities are of the polynomial-time computable type", our interpretation is that "humans can comfortably solve non-trivial instances of this problem, where the exact size depends on the problem at hand".

Accepting this, it seems that using "polynomial-time computable" as synonymous with "efficient" may even be too restrictive: In fact, it is often the case that we as humans are able to solve problems which may be hard in general but suddenly become feasible if certain parameters of the problem are restricted. This idea has been formalized in the field of *parametrized complexity theory*, in which "tractability" is captured by the class FPT.[1] Now, identifying "problems with an FPT algorithm" as "efficiently solvable" cannot be done lightly and requires some judgement using the problem definition. To this effect, we remark that most known problems with *FPT*-algorithms have reasonable parametrized growth functions — we believe that even if $f(\kappa) = 2^{O(\kappa)}$ then the algorithm is still effective on instances where κ is guaranteed to be small.

From this line of thought, [4] introduces a specific version of the claim that cognition and cognitive capacities are constrained by the fact that humans basically are finite systems with only limited resources for computation. This basic idea is formalized in terms of the so called "FPT-*Cognition thesis*", demanding for human cognitive capacities to be fixed-parameter tractable for one or more input parameters that are small in practice (i.e., stating that the computational-level theories have to be in FPT). But whilst the P-Cognition thesis also found its way into AI (cf., e.g., [5]), the FPT-Cognition thesis to the best of our knowledge this far has widely been ignored.

Therefore, we propose a way of (re)introducing the idea of tractable computability for cognition into AI and cognitive systems research by rephrasing the FPT-form of the Tractable Cognition thesis into a "*Tractable AGI thesis*" (Tractable Artificial and General Intelligence thesis): As not only humans, but also all of the currently available computing systems are ultimately finite systems with limited resources (and thus in this respect are very similar to human minds and/or brains), in close analogy it seems highly recommendable to also demand for computer models of AI and complex cognitive capacities to be of the (at least) fixed-parameter tractable type.

Tractable AGI thesis
Models of cognitive capacities in artificial intelligence and computational cognitive systems have to be fixed-parameter tractable for one or more input parameters that are small in practice (i.e., have to be in FPT).

In the rest of the note, we will work through an example of an application of the TAGI thesis to a modern area of research.

For a long time, attempts at modeling and reproducing human rationality and reasoning in artificial systems had been based on logical formalisms. Whilst originally al-

[1] A problem P is in FPT if P admits an $O(f(\kappa)n^c)$ algorithm, where n is the input size, κ is a parameter of the input constrained to be "small", c is an independent constant, and f is some computable function. For an introduction to parametrized complexity theory see, e.g., [2,3].

most exclusively classical (i.e., monotonic, homogeneous) logics were used, researchers rather quickly faced the challenge of having to deal with the defeasibility of common sense reasoning, resulting in the development of non-monotonic logics. Unfortunately, the latter formalisms exhibit one severe drawback: To the best of our knowledge thus far all prominent approaches to logic-based non-monotonic reasoning have been shown to be highly intractable, i.e., NP-hard or worse. Under the computational assumption that $\mathsf{P} \neq \mathsf{NP}$ it follows that the existence of an "algorithmic cure-all" is unlikely — there is no efficient algorithm which will be able to exactly compute the processes of reasoning exactly for any set of input parameters. This observation significantly contributed to a progressing abandonment of the logical approach to modeling cognitive capacities and processes within cognitive psychology and later also AI. In reaction a theory rivaling logics as standard tools has gained almost overwhelming acclaim: Bayesian probabilistic modeling. This conceptually seemingly orthogonal view claims to offer several advantages over the classical logic approaches; within a short time, Bayesianism has had considerable successes in modeling human behavior (see, e.g., [6]).

However we think that unconditional optimism is not justified, as — despite being seemingly ignored — the switch of formalisms has not made the intractability issue disappear. As well, probabilistic (Bayesian) inference of the most probable explanation (MPE) of a set of hypothesis given observed phenomena has been shown to be similar to logic-based non-monotonic reasoning in that it is NP-hard and stays computationally intractable even under approximation [7]. So also here, for growing input sizes the feasibility of completing computation within a reasonably short period of time, and without using an unrealistic amount of memory, may become questionable dramatically quickly. Still, an entire abandonment of the overall approach would most likely be premature, as there seem to be ways of avoiding the fall into the abyss of intractability: For instance in [8], a way of salvaging parts of the potential cognitive plausibility of Bayesian approaches has been sketched by proposing that instead of the original problem domains some form of restricted input domains might be used in human cognition, possibly allowing to avoid the impending complexity explosion.

So how may the Tractable AGI thesis be of help here? It could serve as a principle for deciding whether a particular Bayesian-style model is worth investing further effort or should be abandoned. Suppose the cognitive modeler or AI system designer is able to prove that his model at hand is — although in its most general form NP-hard — at least fixed-parameter tractable for some set of parameters κ. This implies that if the parameters in κ are fixed small constants for problem instances realized in practice, then it is possible to efficiently compute a solution. However, there is also a non-trivial corollary of this: any instance of the problem can be reduced to a *problem kernel*.

Definition 1. Kernelization
Let P be a parameterized problem. A kernelization of P is an algorithm which takes an instance x of P with parameter κ and maps it in polynomial time to an instance y such that $x \in P$ if and only if $y \in P$, and the size of y is bounded by $f(\kappa)$ (f a computable function).

Theorem 1. Kernelizability [9]
A problem P is in FPT if and only if it is kernelizable.

This theorem on the one hand entails that any positive FPT result obtainable for the model in question essentially implies that there is a "downward reduction" for the underlying problem to some sort of smaller or less-complex instance of the same problem, which can then be solved — whilst on the other hand (assuming $\mathsf{W}[1]^2 \neq \mathsf{FPT}$) any negative result implies that there is no such downward reduction.

This (formal) correspondence between complex instances and simpler manifestations of a problem seems to match well with an observation from problem-solving and reasoning experiments with human participants: Different forms of reductions from complex to simpler (but still solution-equivalent) problems are reported to be pervasive and crucial in many human problem solving scenarios (see, e.g., [10]).

And also on the more theoretical side, the described equivalence might form a connecting point to recent, much-noticed developments in cognitive science and cognitive psychology. A growing number of researchers in these fields argues that humans in their common sense reasoning do not apply any full-fledged form of logical or probabilistic (and thus intractable) reasoning to possibly highly complex problems, but rather rely on mechanisms that reduce the latter to equivalent, simpler ones (see, e.g., [11]). But although the number of supporters of these and similar ideas is constantly growing, a commonly accepted formal account of how this reduction process might work (or even technically be characterized and reconstructed) thus far has not been given. Recognizing this as a serious deficit, Theorem 1 and its interpretation can provide inspiration and first hints at hypothesizing a specialized cognitive structure capable of computing the reduced instance of a problem, which then might allow for an efficient solving procedure — with the overall hypothesis in turn possibly serving as foundation and starting point for the development of computational accounts and (in the long run) a computational recreation of heuristics in computational cognitive models of reasoning, problem-solving and AI.

References

1. Frixione, M.: Tractable competence. Minds and Machines 11, 379–397 (2001)
2. Flum, J., Grohe, M.: Parameterized Complexity Theory. Springer (2006)
3. Downey, R.G., Fellows, M.R.: Parameterized Complexity. Springer (1999)
4. van Rooij, I.: The tractable cognition thesis. Cognitive Science 32, 939–984 (2008)
5. Nebel, B.: Artificial intelligence: A computational perspective. In: Brewka, G. (ed.) Principles of Knowledge Representation, pp. 237–266. CSLI Publications (1996)
6. Oaksford, M., Chater, N.: Précis of bayesian rationality: The probabilistic approach to human reasoning. Behavioral and Brain Sciences 32, 69–84 (2009)
7. Kwisthout, J., Wareham, T., van Rooij, I.: Bayesian intractability is not an ailment that approximation can cure. Cognitive Science 35(5), 779–784 (2011)
8. van Rooij, I.: Rationality, intractability and the prospects of “as if” explanations. In: Szymanik, J., Verbrugge, R. (eds.) Proc. of the Logic & Cognition Workshop at ESSLLI 2012. CEUR Workshop Proceedings, vol. 883, CEUR-WS.org (August 2012)
9. Downey, R.G., Fellows, M.R., Stege, U.: Parameterized complexity: A framework for systematically confronting computational intractability. In: Contemporary Trends in Discrete Mathematics: From DIMACS and DIMATIA to the Future, AMS (1997)
10. Anderson, J.R.: Cognitive Psychology and Its Implications. W. H. Freeman and Company (1985)
11. Gigerenzer, G., Hertwig, R., Pachur, T. (eds.): Heuristics: The Foundation of Adaptive Behavior. Oxford University Press (2011)

---

[2] $\mathsf{W}[1]$ is the class of problems solvable by constant depth combinatorial circuits with at most 1 gate with unbounded fan-in on any path from an input gate to an output gate. In parameterized complexity, the assumption $\mathsf{W}[1] \neq \mathsf{FPT}$ can be seen as analogous to $\mathsf{P} \neq \mathsf{NP}$.