On Super-Turing Computing Power and Hierarchies of Artificial General Intelligence Systems

Jiří Wiedermann
Institute of Computer Science
Academy of Sciences of the Czech Republic
Pod Vodárenskou věží 2, 182 07 Prague

Abstract

Using the contemporary view of computing exemplified by recent models and results from non-uniform complexity theory we investigate the computational power of artificial general intelligence systems (AGISs). We show that in accordance with the so-called Extended Turing Machine Paradigm such systems can be seen as non-uniform evolving interactive systems whose computational power surpasses that of classical Turing machines. Our results shed light to the question asked by R. Penrose concerning the mathematical capabilities of human mathematicians which seem to go beyond classical computability. We also show that there is an infinite hierarchy of AGISs each of which is capable to solve strictly more problems than its predecessors in the hierarchy.

Characterizing the Computational Properties of AGISs

According to its definition artificial general intelligence is a form of intelligence at the human level and definitely beyond. Implicitly, this statement alone evokes an idea of (partially?) ordered "intelligence levels", one of which should correspond to human intelligence, with still some levels of "superhuman" intelligence above it. The point in time when the "power" of AGISs will reach and trespass the level of human intelligence has obtained a popular label: the Singularity (cf. Kurzweil, 2005). Nevertheless, it seems that in the AI literature there has not been much explicit attention paid to the formal investigation of the "power" and the "levels of intelligence" (in the sense mentioned above) of AGISs. It is the goal of this short notice to present an approach based on the recent developments in the computational complexity theory answering certain questions related to the computational power of AGISs.

Artificial general intelligence systems must clearly be (i) *interactive* — in order to be able to communicate with their environment, to reflect its changes, to get the feedback, etc.; (ii) *evolutionary* — in order to develop over generations, and (iii) potentially *time-unbounded* — in order to allow for their open-ended development.

Therefore AGISs cannot be modelled by classical Turing machines — simply because such machines do not possess the above mentioned properties. The AGISs must be modelled by theoretical computational models capturing interactivness, evolvability, and time-unbounded operation of the underlying systems. Such models have recently been introduced by van Leeuwen & Wiedermann, (2001) or (2008).

Definition 1 *An* interactive Turing machine with advice *is a Turing machine whose architecture is changed in two ways:*

- *instead of an input and output tape it has an* input port *and an* output port *allowing reading or writing potentially infinite streams of symbols;*
- *the machine is enhanced by a special, so-called* advice tape *that, upon a request, allows insertion of a possibly non-computable external information that takes a form of a finite string of symbols. This string must not depend on the concrete stream of symbols read by the machine until that time; it can only depend on the number of those symbols.*

An advice is different from an oracle also considered in the computability theory: an oracle value can depend on the current input (cf. Turing, 1939). The interactive Turing machines with advice represent a *non-uniform model of interactive, evolving, and time-unbounded computation*. Such machines capture well an interactive and time-unbounded software evolution of AGISs.

Interactive Turing machines with advice are equivalent to so-called *evolving automata* that capture well hardware evolution of interactive and time-unbounded computations (Wiedermann & van Leeuwen, 2008).

Definition 2 *The* evolving automaton *with a schedule is an infinite sequence of finite automata sharing the following property: each automaton in the sequence contains some subset of states of the previous automaton in that sequence. The schedule determines when an automaton has to stop processing of its inputs and thus, when is the turn of the next automaton.*

The condition that a given automaton has among its states a subset of states of a previous automaton captures one important aspect: it is the persistence of data in the evolving automaton over time. In the language of finite automata this condition ensures that some information available to the current automaton is also available to its successor. This models passing of information over generations.

On an on–line delivered potentially infinite sequence of the inputs symbols the schedule of an evolving automaton

determines the *switching times* when the inputs to an automaton must be redirected to the next automaton. This feature models the (hardware) evolution.

An evolving automaton is an infinite object given by an explicit enumeration of all its elements. There may not exist an algorithm enumerating the individual automata. Similarly, the schedule may also be non-computable. Therefore, also evolving automata represent a non-uniform, interactive evolutionary computational model.

Note that at each time a computation of an evolving automaton is performed by exactly one of its elements (one automaton) which is a finite object.

Based on the previous two models van Leeuwen & Wiedermann (2001) have formulated the following thesis:

Extended Turing Machine Paradigm *A computational process is any process whose evolution over time can be captured by evolving automata or, equivalently, by interactive Turing machines with advice.*

Interestingly, the paradigm also expresses the equivalence of software and hardware evolution.

In Wiedermann & van Leeuwen (2008) the authors have shown that the paradigm captures well the contemporary ideas on computing. The fact that it also covers AGISs adds a further support to this paradigm.

Thesis 3 *From a computational point of view AGISs are equivalent to either evolving automata or interactive Turing machines with advice.*

The Super-Turing Computing Power of AGISs

The power of artificial general intelligent systems is measured in terms of sizes of sets of different reactions (or behaviors) that those systems can produce in potentially infinite interactions with their environment.

The super-Turing power of AGISs is shown by referring to super-Turing computing power of interactive Turing machines with advice.

Namely, in van Leeuwen & Wiedermann (2001) it was shown that such machines can solve the halting problem. In order to do so they need an advice that for each input of size n allows to stop their computation once it runs beyond a certain maximum time. This time is defined as the maximum, over computations over all inputs of size n and over all machines of size n that halt on such inputs.

Proposition 4 *The artificial general intelligence systems have super-Turing computational power.*

Roger Penrose (1994) asked about the power of human thoughts: how to explain the fact that mathematicians are able to find proofs of some theorems in spite of the fact that in general (by virtue of Gödel's or Turing's results) there is no algorithm that would always lead to a proof or refutation of any theorem. In our setting the explanation could be that the mathematicians discover a "non-uniform proof", i.e., a way of proving a particular theorem at hand and probably nothing else. This proof is found in a non-predictable potentially unbounded interaction of mathematicians (among themselves and also in the interaction with others and with their environment) pondering over the respective problems.

Hierarchies of AGISs

For interactive Turing machines with advice or for evolving automata one can prove that there exist infinite proper hierarchies of computational problems that can be solved on some level of the hierarchy but not on any of the lower levels. Roughly speaking, the bigger the advice, the more problems can be solved by the underlying machine.

Proposition 5 *There is infinity of infinite proper hierarchies of artificial general intelligence systems of increasing computational power.*

Among the levels of the respective hierarchies there are many levels corresponding formally (and approximately) to the level of human intelligence (the Singularity level) and also infinitely many levels surpassing it in various ways.

Common Pitfalls in Interpretations of the Previous Results

Our results are non-constructive — they merely show the existence of AGISs with super-Turing properties, but not the ways how to construct them. Whether such systems will find solutions of non-computable problems depends on the problem at hand and on getting a proper idea at due time stemming from the sufficient experience, insight and a lucky interaction.

Whether a Singularity will ever be achieved cannot be guaranteed; from our results we merely know that in principle it exists. Our results give no hints how far in the future it lies. Moreover, we have no idea how far apart are the levels in the respective hierarchies. It is quite possible that bridging the gap between the neighboring "interesting" levels of intelligence could require an exponential (or greater) computational effort. Thus, even an exponential development of non-biological intelligence of the AGISs may not help to overcome this gap in a reasonable time.

Acknowledgment

This research was carried out within the institutional research plan AV0Z10300504 and partially supported by a GA ČR grant No. P202/10/1333

References

Kurzweil, R. (2005). The Singularity is Near. Viking Books, 652 pages

Penrose, R. (1994). Shadows of the Mind (A Search for the Missing Science of Consciousness). Oxford University Press, Oxford, 457 p.

Turing, A. M. 1939). Systems of logic based on ordinals, Proc. London Math. Soc. Series 2, Vol. 45, pp. 161-228

van Leeuwen, and Wiedermann, J. (2001). The Turing machine paradigm in contemporary computing, *Mathematics unlimited - 2001 and beyond*, Springer-Verlag, pp. 1139–1155

Wiedermann, J., and van Leeuwen, J. (2008). How We Think of Computing Today. (Invited Talk) Proc. CiE 2008, LNCS 5028, Springer, Berlin, pp. 579-593