Time-Bounded Kolmogorov Complexity May Help in Search for Extra Terrestrial Intelligence (SETI)

Martin Schmidt

Department of Computer Science, University of Texas at El Paso
El Paso, TX 79968, USA, email mschmidt@cs.utep.edu

Abstract

One of the main strategies in Search for Extra Terrestrial Intelligence (SETI) is trying to overhear communications between advanced civilizations. However, there is a (seeming) problem with this approach: advanced civilizations, most probably, save communication expenses by maximally compressing their messages, and the notion of a maximally compressed message is naturally formalized as a message x for which Kolmogorov complexity $C(x)$ is close to its length $l(x)$, i.e., as a "random" message. In other words, a maximally compressed message is indistinguishable from the truly random noise, and thus, trying to detect such a message does not seem to be a working SETI strategy.

We show that this argument does not take into consideration the *time* necessary for compression and decompression of message. If we take this time into consideration, and therefore consider time-bounded versions of Kolmogorov complexity, then the above "problem" disappears. We also show which version of time-bounded Kolmogorov complexity is most appropriate for formalizing SETI strategies.

According to modern physics, extra terrestrial intelligence is very probable. According to modern physics, our Universe is largely homogeneous; hence, in many locations, it is possible to have the same physical conditions as the ones that led to the emergence of intelligent life on Earth. It is therefore reasonable to conclude that in many of these places, intelligent life did emerge.

A contact with an extra terrestrial intelligence could be highly beneficial for our civilization; hence, SETI. The processes which lead to the emergence and development of intelligent life are far from being deterministic; there are many random factors involved. As a result, the rates with which different civilizations appear and progress are expected to be drastically different. Hence, we can expect both civilizations which are largely behind us in technological and scientific progress, as well as civilizations which are way ahead of ours. A contact with such an advanced civilization can bring new knowledge, help us solve our problems, and give a big boost to our civilization.

In view of this potential benefit, a lot of effort goes into Search for Extra Terrestrial Intelligence (SETI); see, e.g., [2, 8, 9] and references therein.

A (seeming) problem with the current SETI strategy. If there are advanced civilizations out there, they, most probably, know about each other and therefore, intensely communicate with each other. In view of this communication, one of the main current SETI strategies consists of watching the radio-signals in the sky and trying to determine which of them correspond to communications between advanced civilizations.

There is, however, a (seeming) problem in this approach. Of course, it is possible to distinguish between a meaningful signal and a random noise which typically comes from different celestial objects like galaxies, quasars, etc. However, it would be a waste of energy to send a pure signal (and saving energy is extremely important for long-distance inter-civilization communications which require a lot of energy). The signal will be, most probably, sent compressed. The more advanced the civilization, the more it knows how to compress its signals, and the more perfect this compression will be. As a result, if we try to overhear the "conversation" between different civilizations, what we will see is a signal x which cannot be further compressed.

This un-compressibility admits a natural formalization. Since most modern measuring devices have binary output, we can assume that the signal x is a binary sequence, i.e., a word in a binary alphabet $\{0, 1\}$. By a *compression*, intuitively, we mean a pair consisting of a compressed signal c and an instruction i (for some universal computer) which enables us to reconstruct the original signal x from its compression c. In other words, a compression is a program (c, i) for a universal computer which generates x. Therefore, the idea that the word x cannot be further compressed can be formalized as follows: for some universal computer

U, every program p which generates x has the same length $l(p)$ as x (i.e., $l(p) \approx l(x)$), or the length which is larger than the length of x (i.e., $l(p) > l(x)$). The shortest length of a program which generates x is called its *Kolmogorov complexity* $C(x)$ [1, 7]. In terms of Kolmogorov complexity, the idea that the word x cannot be further compressed can be thus formalized as $C(x) \approx l(x)$.

This approximate equality, however, is exactly the idea of formalizing of the intuitive notion of a "random" sequence x. Thus, we come to a conclusion that signals exchanged by advanced civilizations must be "random", i.e., such signals are indistinguishable from random noise. This indistinguishability seems to invalidate the main current SETI strategy.

In this paper, we show how this problem can be overcome.

How to resolve this problem: enter time-bounded Kolmogorov complexity. The above negative conclusion is correct if we deal with chit-chat. However, civilizations communicate not only to simply exchange news; one of the main reasons for their communication is that they want to help each other solve problems by relaying known solutions. This is true for an exchange between a less advanced and a more advanced civilization (this is exactly what we hope for when we spend efforts on SETI). This is also true for an exchange between equally advanced civilizations: these civilizations will encounter many similar problems; so all of them will benefit if instead of duplicating their efforts in each civilization solving all these problems, they will divide these problems between themselves, solve their share of the problems, and exchange the solutions.

This only makes sense for problems for which the solution is difficult to get but easy to check, and for which the solution has a reasonable length. In other words, we are talking about the problems of the following type: given a word w, find a new word s whose length $l(s)$ is bounded by some reasonable function of the length $l(w)$ of the input w, and for which, some easily checkable property $P(w, s)$ holds. Most problems solved by engineering and science are of this type: e.g., in mathematics, given a formulation w of a statement, we want to find a proof s of either this statement w, or of its negation $\neg w$ (checking whether a given step-by-step proof is correct is easy; computers could do it even in the 1960s); in engineering, we formulate the requirements w for, e.g., a bridge, and we want to find a design s which satisfies all these requirements; in physics, we have data w, and we want to find a simple law s which describes all this data, etc. If we follow the standard formalizations of "reasonable function of length" as "polynomial function of length", and "easily checkable" as "checkable in polynomial time", we conclude that a general "problem" is a problem from the class NP (or, to be more precise, *search problem* in the sense of L. A. Levin) [3].

For such problems, maximal compression (for which the length of a compressed message is equal to its Kolmogorov complexity) does not make much sense. Indeed, we want to send a solution s to the problem w. However, we can always compress this solution into the formulation w, because for problems from the class NP there is a simple algorithm for finding a solution: exhaustive search. So, if we follow the idea of sending the maximally compressed message, then we could as well send the original problem w instead of its solution, and thus not send the solution at all.

The main idea of sending solutions is to save time on solving the corresponding problem, and there is no sense in wasting all the saved time on de-compressing the message. When we formalized the incompressibility in terms of Kolmogorov complexity, we neglected computation time; with this time-saving application in mind, we must take this time into consideration: e.g, consider a *time-bounded* version of Kolmogorov complexity instead of the the original one. In other words, instead of considering $C(x) = \min l(p)$ (where min is taken over all programs generating x), we must consider $C(x) = \min f(t(p), l(p))$, where $t(p)$ is the running time of a program p, and $f(t, l)$ is a function of two variables.

The less time and/or memory compression-decompression requires, the better. Therefore, we should require that $f(t, l)$ be a non-decreasing function of t and l.

The choice of an appropriate time-bounded version of Kolmogorov complexity. Which function $f(t, l)$ should we choose? We talk about transmitting solutions s to problems w (from the class NP). The decompression time $t(p)$ takes the smallest possible value if we simply send the (uncompressed) solution s. Realistically, the only processing time $t(p)$ for the received signal is the time which is necessary to check that s is indeed a solution (i.e., that the desired property $P(w, s)$ is true). Alternatively, we can send all the bits of s but one; then, to reconstruct the desired solution, we must check both possible completions of the sent signal: completion by 0 and completion by 1. Then, the length of the sent signal decreased by one bit $l'(p) = l(p) - 1$, while the processing time increases twice $t'(p) = 2t(p)$. Similarly, we can skip 1 bit from any message at the cost of increasing the processing time twice.

This skipping a single bit should not change the fact that what we are considering is an intelligent message; therefore, it is reasonable to require that the function $f(t, l)$ used in our distinction between a message and a random noise be invariant under such transformation. In other words, we require that the desired function $f(t,l)$ satisfy, for every two integers t and l, the equation $f(2t,l-1)=f(t,l)$.

This equation has also been obtained in [4, 5] (where it appeared in a slightly different situation); in these papers, it is shown that a function $f(t,l)$ satisfies this equation if and only if $f(t,l)=F(t\cdot 2^l)$ for some function $F(z)$ of one variable.

Since we required that the function $f(t,l)$ is monotonically non-decreasing in both variables, we can conclude that the function $F(z)$ is non-decreasing too. So, looking for the best compression means looking for a program p for which the product $t(p)\cdot 2^{l(p)}$ takes the smallest possible value. The corresponding modified Kolmogorov complexity $C'(x)=\min(t(p)\cdot 2^{l(p)})$ (where min is taken over all programs p which generate x) was introduced by L. A. Levin who proved (see [6, 7]) that if we are looking for an optimal (asymptotically fastest) universal algorithm for solving different search problems, then this optimal algorithm should check all possible designs in the increasing order of $C'(x)$.

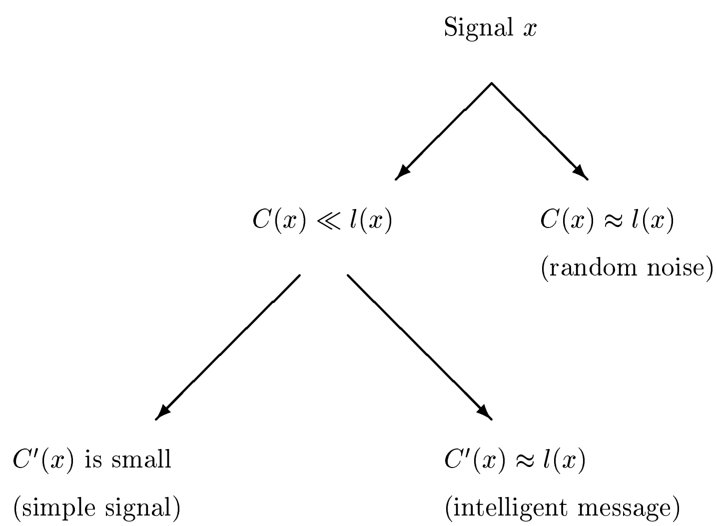
Conclusion. A short conclusion is that a signal x is a possible communication between two advanced civilizations if $C'(x)\approx l(x)$.

A longer conclusion is as follows. By using Kolmogorov complexity $C(x)$ alone, we can only classify signals into:

- simple signals, with $C(x)\ll l(x)$, and
- complicated (random) signals (with $C(x)\approx l(x)$).

Now, with a new complexity measure $C'(x)\ge C(x)$, we can get a more detailed classification by comparing *both* $C(x)$ and $C'(x)$ with the length $l(x)$:

- if $C(x)\approx l(x)$, then a signal x is a random noise;
- if $C(x)\ll l(x)$ and $C'(x)\approx l(x)$, then the signal x is an intelligent message;
- if $C'(x)\ll l(x)$, then the signal x is (truly) simple.



So, the SETI strategy of looking for intelligent message among random noises is no longer paradoxical.

Comment. Of course, this idea is not yet a working SETI algorithm, because Kolmogorov complexity is not algorithmically computable (see, e.g., [1, 7]). (Levin's modification of Kolmogorov complexity is actually computable but computing it requires too long time, so for all practical purposes, it is not computable at all.) So, to actually check for SETI, we must use feasibly computable approximations to $C'(x)$.

Acknowledgments. This work started when the author was working with the National Solar Observatory, Sunspot, New Mexico, USA, supported by the NSF grant "Electrical Engineering and Computer Science Projects to Support Research at the National Solar Observatory". The author is thankful to Drs. Vladik Kreinovich, Jeff Kuhn, and Luc Longpré for their supervision and help.

References

[1] C. Calude, *Information and randomness: An algorithmic perspective*, Springer-Verlag, Berlin, 1994.

[2] C. B. Cosmovicic, S. Bowyer, and D. Werthimer, "Astronomical and Biochemical Origins and the Search for Life in the Universe", *Proceedings of the 5th International Conference on Bioastronomy*, Bologna, Italy, 1996.

[3] M. R. Garey and D. S. Johnson, *Computers and intractability: a guide to the theory of NP-completeness*, Freeman, San Francisco, 1979.

[4] M. Koshelev, "Towards The Use of Aesthetics in Decision Making: Kolmogorov Complexity Formalizes Birkhoff's Idea", *Bulletin of the European Association for Theoretical Computer Science (EATCS)*, 1998, Vol. 66, pp. 166–170.

[5] V. Kreinovich, L. Longpré, and M. Koshelev, "Kolmogorov complexity, statistical regularization of inverse problems, and Birkhoff's formalization of beauty", In: A. Mohamad-Djafari (ed.), *Bayesian Inference for Inverse Problems*, Proceedings of the SPIE/International Society for Optical Engineering, Vol. 3459, San Diego, CA, 1998, pp. 159–170.

[6] L. A. Levin, "Universal sequential search problems", *Problems of Information Transmission*, 1973, Vol. 9, No. 3, pp. 265–266.

[7] M. Li and P. Vitányi, *An introduction to Kolmogorov complexity and its applications*, Springer-Verlag, N.Y., 1997.

[8] G. S. Shostak, *Sharing the Universe: Perspective on Extraterrestrial Life*, Berkeley Hills Books, Berkeley Hills, California, 1998.

[9] D. W. Swift, *Seti Pioneers: Scientists Talk About Their Search for Extraterrestrial Intelligence*, University of Arizona Press, Tucson, Arizona, 1993.