Unfortunately, the Universal Predictor Cannot Be Made Constructive

Julio C. Urenda, Olga Kosheleva, and Vladik Kreinovich

Abstract A recent article in the Notices of the American Mathematical Society reminded the mathematics community that, under the Axiom of Choice, it is possible to have a universal predictor: if we input, into this predictor, the values of a function for all moments $t < t_0$ for some t_0, then, for almost all t_0, this predictor correctly predicts the next values of this function on some interval $[t_0, t_0 + \varepsilon)$. This predictor cannot be used for actual predictions: it is based on the Axiom of Choice and is, therefore, not constructive. A natural question is: maybe it is possible to have another universal predictor, which *is* constructive? In this paper we show that, unfortunately, it is not possible to have a constructive universal predictor. In other words, the above universal predictor result cannot be used for actual predictions.

1 Formulation of the problem

A universal predictor result: a brief reminder. A recent article [1] attracted the attention of the mathematics community to an interesting result – that first appeared in [2, 3]. According to this result, once we assumed Axiom of Choice – which is a normal practice in working mathematics – it is possible to have what is reasonable to call a *universal predictor* P. The input to this predictor is a function $f(t)$ defined on open lower half-line $(-\infty, t_0)$ for some t_0. Based on this function, the predictor

Julio C. Urenda
Department of Mathematical Sciences, University of Texas at El Paso, 500 W. University
El Paso, Texas 79968, USA, e-mail: jcurenda@utep.edu

Olga Kosheleva
Department of Teacher Education, University of Texas at El Paso, 500 W. University
El Paso, Texas 79968, USA, e-mail: olgak@utep.edu

Vladik Kreinovich
Department of Computer Science, University of Texas at El Paso, 500 W. University
El Paso, Texas 79968, USA, e-mail: vladik@utep.edu

returns an extension $\overline{f}(t)$ of the given function to the whole real line. The magical property of this predictor is that it is almost always locally correct.

In precise terms, for every function $F(t)$ from real numbers to real numbers, for almost every t_0 – namely, for every t_0 except for a countable nowhere dense set of real numbers – there exists an $\varepsilon > 0$ such that:

- if we apply the predictor P to the restriction $f \stackrel{\text{def}}{=} F_{(-\infty,t_0)}$ of the function $F(t)$ to the interval $(-\infty,t_0)$,
- then for the predicted function $\overline{f} = P(f)$, we have $\overline{f}(t) = F(t)$ for all

$$t \in [t_0, t_0 + \varepsilon).$$

In other words, if we know all the previous values of the function $F(t)$, i.e., its values for all $t < t_0$, then the universal predictor allows us to correctly predict the values of this function for the whole following interval $[t_0, t_0 + \varepsilon)$.

Main idea behind the universal predictor: a brief reminder. The main idea behind the universal predictor is very straightforward: due to the Axiom of Choice, any set can be well-ordered, i.e., there is a linear (total) strict order $<$ on this set for which every descending sequence $x_1 > x_2 > \ldots$ has to end – so that there are no infinite descending sequences.

Then, once we are given a function $f(t)$, as $\overline{f}(t)$, we select, among all the functions that extend $f(t)$, the function which is the smallest in terms of this order. It can then be proven that thus defined predictor P indeed has the universal predictor property.

A natural question. A natural question is: can this exciting result help with the actual prediction?

Of course, we cannot use the above-described predictor: it is based on the Axiom of Choice and is, therefore, not constructive in any reasonable sense. So, the actual question is: while the above-described predictor is not constructive, maybe it is possible to have a different – and constructive – universal predictor, i.e., a predictor that would actually help us predict?

What we do in this paper. In this paper, we prove that, unfortunately, a universal predictor is not possible. Specifically, we first describe, in precise terms, what is meant by a constructive universal predictor, and then we prove that such a predictor is not possible.

2 Formulation of the Problem in Precise Terms and First Negative Result

Discussion. In practice, we cannot know all infinitely many values $f(t)$ corresponding to all possible values $t \leq t_0$: at any given moment of time, we can only store finite number of measurement results. Let us order the moments of time corresponding to these measurements into an increasing sequences $t_1 < \ldots < t_n$. In these terms, all we know are the moments t_i and the corresponding values $f_i \stackrel{\text{def}}{=} f(t_i)$.

The resulting formulation of the prediction problem. In this case, the prediction problem takes the following form:

- given the values $t_1 < \ldots < t_n$, the values $f_1 = f(t_1), \ldots, f_n = f(t_n)$, and the value $t_{n+1} > t_n$,
- produce the value $\overline{f}_{n+1}$.

The prediction is correct if $\overline{f}_{n+1} = f(t_{n+1})$ and wrong otherwise.

Discussion. We want to prove that, in some reasonable sense, the prediction property is false for almost all functions $f(t)$. A natural way to describe "almost all" is to have some natural measure on the corresponding set.

For real numbers and for tuples of real numbers, we have a natural measure: Lebesgue measure. However, on the set of all functions there are many different measures – such as Wiener measure corresponding to random walk – none of which is very natural. So, to make the result natural, we reduce the problem from functions to tuples.

We can do this since. Indeed, strictly speaking, this problem is about a general function $f(t)$, it reality, this problem only takes into account $n+1$ values of this function. So, instead of a function, we can consider the tuple $\mathscr{F} = (f_1, \ldots, f_n, f_{n+1})$. Now, we are ready to formulate the first negative result.

Definition 1. *Let $t_1 < \ldots < t_n < t_{n+1}$ be an increasing sequence of real numbers.*

- *By a* predictor *P, we mean a mapping that maps n-tuples of real numbers into a real number $P(f_1, \ldots, f_n)$.*
- *We say that predictor P is* correct *on a tuple $\mathscr{F} = (f_1, \ldots, f_n, f_{n+1})$ if $P(f_1, \ldots, f_n) = f_{n+1}$..*
- *We say that predictor P is* wrong *on a tuple $\mathscr{F} = (f_1, \ldots, f_n, f_{n+1})$ if it is not correct on this tuple.*
- *We say that a prediction is wrong* for almost all functions *if for every n-tuple $(f_1, \ldots, f_n)$, the set of all the values f_{n+1} for which the tuple $(f_1, \ldots, f_n, f_{n+1})$ leads to the correct prediction has Lebesgue measure 0.*

Proposition 1. *For every sequence $t_1 < \ldots < t_n < t_{n+1}$, every predictor is wrong for almost all functions.*

Proof. By definition, the predictor is correct is $f_{n+1} = P(f_1, \ldots, f_n)$. For each tuple $(f_1, \ldots, f_n)$, the set of all the tuples for which this equality is true consists of a single tuple – i.e., has Lebesque measure 0. The proposition is proven.

Comment. Of course, this result holds if we consider all possible functions $f(t)$, and thus, all possible tuples $(f(t_1), \ldots, f(t_n), f(t_{n+1}))$.

The result will be different if we limit ourselves to some m-parametric family of functions $f(t, c_1, \ldots, c_m)$, e.g., the family of all polynomials of degree not exceeding $m-1$. Then for $n > m$, we will be able to determine all m values from m equations

$f(t_i,c_1,\ldots,c_m) = f_i$, $1 \le i \le m$ and thus, we will be able to predict all future values of the function $f(t,c_1,\ldots,c_m)$.

Discussion. One may argue that this setting is not fully realistic: it only takes into account predictions that are exact, but in reality, reasonably accurate predictions, when $|f_{n+1} - P(f_1,\ldots,f_n)| \le \varepsilon$ for some small pre-set threshold value $\varepsilon > 0$, are also good.

If we change the criterion for prediction correctness from exact equality to this inequality, we no longer get the above strong statement about predictors being almost always wrong. However, we will get a statement that is almost as strong; that for every sequence $t_1 < \ldots < t_n$, for each predictor P, and for each tuple $(f_1,\ldots,f_n)$:

- the values f_{n+1} for which the desired inequality holds form a narrow interval $[P(f_1,\ldots,f_n) - \varepsilon, P(f_1,\ldots,f_n) + \varepsilon]$ of width 2ε, while
- for the rest of the real line – of infinite total width – this inequality does not hold.

3 Second Negative Result

Discussion. Since we *cannot* always predict the value $f(t_{n+1})$ at the given next values t_{n+1}, maybe we can predict the value $f(t_{n+1})$ for *some* value t_{n+1} that depends on the input?

In this section, we show that this is not possible either.

Definition 2. *Let $t_1 < \ldots < t_n$ be an increasing sequence of real numbers.*

- *By a* predictor (P,T), *we mean two mapping $T(f_1,\ldots,f_n) > t_n$ and $P(f_1,\ldots,f_n)$ that map n-tuples of real numbers into real numbers.*
- *We say that predictor (P,T) is* correct *on a function $f(t)$ if for $t_{n+1} = T(f_1,\ldots,f_n)$, we have $P(f_1,\ldots,f_n) = f(t_{n+1})$.*
- *We say that predictor (P,T) is* wrong *on a function $g(t)$ is it is not correct on this function.*
- *We say that the predictor (P,T) is* wrong for almost all functions *if for every n-tuple $(f_1,\ldots,f_n)$, the set of all the values f_{n+1} for which, for $t_{n+1} = T(f_1,\ldots,t_n)$, we have $f_{n+1} = P(f_1,\ldots,f_n)$, has Lebesgue measure 0.*

Proposition 2. *For every sequence $t_1 < \ldots < t_n$, every predictor is wrong for almost all functions.*

Proof is similar to the proof of Proposition 1.

Acknowledgments

This work was supported in part by the National Science Foundation grants 1623190 (A Model of Change for Preparing a New Generation for Professional Practice in

Computer Science), HRD-1834620 and HRD-2034030 (CAHSI Includes), EAR-2225395 (Center for Collective Impact in Earthquake Science C-CIES), and by the AT&T Fellowship in Information Technology.

It was also supported by a grant from the Hungarian National Research, Development and Innovation Office (NRDI), and by the Institute for Risk and Reliability, Leibniz Universitaet Hannover, Germany.

References

1. J. D. Hamkins, "The nearly perfect prediction theorem", *Notices of the American Mathematical Society*, 2025, Vol. 72, No. 3, pp. 308–309.
2. C. S. Hardin and A. D. Taylor, "A peculiar connection between the Axiom of Choice and predicting the future", *American Mathematical Monthly*, 2008, Vol. 115, No. 2, pp. 91–96.
3. C. S. Hardin and A. D. Taylor, "The mathematics of corredinated inference: A study of generalized hat problem", In: *Developments in Mathematics*, Vol. 33, Springer, Cham, Switzerland, 2013, https://wwdoi.org/10.1007/978-3-319-01333-6.