6-2020

How to Detect Future Einsteins: Towards Systems Approach

Olga Kosheleva
University of Texas at El Paso, olgak@utep.edu

Vladik Kreinovich
University of Texas at El Paso, vladik@utep.edu

Comments:

Technical Report: UTEP-CS-20-59

How to Detect Future Einsteins: Towards Systems Approach

Olga Kosheleva[1] and Vladik Kreinovich[2]
[1]Department of Teacher Education
[2]Department of Computer Science
University of Texas at El Paso
El Paso, Texas 79968, USA
olgak@utep.edu, vladik@utep.edu

Abstract

Talents are rare. It is therefore important to detect and nurture future talents as early as possible. In many disciplines, this is already being done – via gifted and talented programs, Olympiads, and other ways to select kids with unusually high achievements. However, the current approach is not perfect: some of the kids are selected simply because they are early bloomers, they do not grow into unusually successful researchers; on the other hand, many of those who later become very successful are not selected since they are late bloomers. To avoid these problems, we propose to use systems approach: to find the general formula for the students' growth rate, the formula that would predict the student's future achievements based on his current and previous achievement levels, and then to select students based on the formula's prediction of their future success.

1 Formulation of the Problem

Talent is rare. In every activity, there are people who have extraordinary talents, whose contribution to humanity is irreplaceable. Such people are rare. It is therefore important to detect and nurture such people as early as possible.

Isn't this what we are doing already? At first glance, in many countries and in many disciplines, the search for early Einsteins is going full speed ahead already: from the middle and high school levels, students are selected for gifted and talented programs; many students participate in regional Olympiads in mathematics, in physics, in chemistry, in computing, etc. Winners go to national and international competitions, they get accepted to good colleges.

Some future talents are indeed revealed this way, but not all of them, and not everyone who has been thus selected and praised becomes a new Einstein.

Thus, this process needs to be improved. Here are some examples how the current process does not always work.

Not all geniuses are fast. The first example is Einstein himself. Now, he is universally recognized as a genius, but when he was in school, he was *not* the best student in math. There were no Olympiads in his time, but there is no way he could have won even in his school. It is difficult to tell what happened when he was a student, but numerous memoirs of his later activity reveal one reason why Olympiads would not have been his strongest point: he was a deep thorough but comparatively slow thinker, and Olymiads – with their strict time limit on problem solving – are designed for fast response. The Olympiads design seem to follow the widely spread belief that you need to be fast thinker to be good in math and in other similar disciplines. Analysis of real-life geniuses shows that this belief is largely a myth; see, e.g., [1].

Sure, some geniuses are fast. For example, the famous 18/19 century mathematician Carl Friedrich Gauss was known to be so fast in math in his early years that his teachers had to all the time come up with new problems to keep him busy. On the other hand, we all know colleagues who are not very fast in solving problems, who spend a lot of time on them, but who them come up with deep and impactful solutions which are often much more valuable for time. For example, in mathematics, two winners of the Fields Medal – the top award in mathematics – Laurent Schwartz and Maryam Mirzakhani – mentioned honestly that they have always been slow in solving math problem – and that because of the widely spread myth that speed is important, they often doubted their own talent [2, 11].

Some geniuses are late bloomers. Another reason why the Olympiads system does not always reveal future Einsteins is that people mature at a different speed: some are early bloomers, some are late bloomers. The Olympiad system reveals early bloomers, but their advantage in comparison with other kids their age is sometimes caused not by a special talent, but just by the fact that they bloomed earlier. Moreover, there is evidence that the most drastic changes to the world are brought by late bloomers; see, e.g., [5].

Everyone knows the story of a little prodigy Mozart who became a great composer, but there are also numerous examples of late bloomers. For example, there is a well-known story about Feodor Chaliapin, the internationally renowned 20 century Russian opera signer. When he was a kid, with Alexei Peshkov (who later became a famous Russian writer Maxim Gorky) he wanted to join the church choir. The person in charge of the choir listened to both, accepted Alexei, but dismissed Feodor as being unable to sign. The famous Russian pianist Sviatoslav Richter started taking lessons in piano at the usually late (for pianists) age of twenty two; see, e.g., [8]. Many famous artists started unusually late, e.g., Vincent Van Gogh [9] and Paul Gauguin [10].

Whether a child is an early bloomer or a later bloomer is partly determined by the genes – in particular, by the gender. For example, in middle school, girls pass algebra at a higher rate than boys – this does not necessarily mean that they are smarter than boys, just like the fact that in high school, boys are more successful in math does not predict their future math success; see, e.g., [12].

So basing decisions only on early success would be a mistake – we will miss late bloomers.

It is important to test everybody. We cannot just rely on those who show early success, this means that we have to test everyone. For example, a recent study [6] showed that if we test everyone, we can find quite a few previously undetected gifted students.

But how to test? We cannot just test the student's current ability, we need to take into account how this ability change with time.

How can we do it? In this paper, we do not yet propose an answer, but we show a path that will hopefully lead to an answer.

2 A Systems Approach to Solving the Problem

What we want. At any year t, we can describe a person's ability to solve problems in the corresponding discipline by some numerical characteristic x_t – e.g., by a score on some test.

We know this ability x_{t_0} at a current moment of time t_0, we hopefully know the values x_{t_0-1}, x_{t_0-2}, ... at the previous years.

Based on these known values, we want to predict the ability x_T at some future year T – and we want to select and nurture those individuals for whom this predicted values is high.

Main idea. The original problem of selecting future Einsteins sounds very unusual and complicated. However, once we have reformulated this problem in precise terms, we see that this is a common problem of predicting the future values of a time series based on its past values. In is therefore reasonable to apply the usual ways of solving this general problem.

The general idea of such prediction is based on the fact that, with good accuracy, the current value x_t significantly depends only of a few past values, i.e., that

$$x_t = f(x_{t-1}, \ldots, x_{t-k}) \tag{1}$$

for some small k.

Which k should we use? We cannot have $k = 1$, since this would mean that the current value x_t uniquely determines the next value x_{t+1}; this value, in turn, uniquely determines the next-to-next value x_{t+2}, etc. – and at the end, we conclude that the future value x_T is uniquely determined by the current value x_t.

This is, in effect, the current approach to selecting future Einsteins – based on their current success rate x_t at a standardized test or on an Olympiad, and we have already argued that this approach missed late bloomers.

So, for prediction, we need to use at least one more past moment of time. When we consider dependence on two past moments of time, we can describe different maturing speeds. For example, to describe the general linear dependence of achievement on time $x_t = c_0 + c_1 \cdot t$, we can use the formula $x_t - x_{t-1} = x_{t-1} - x_{t-2}$, i.e., equivalently,

$$x_t = 2x_{t-1} - x_{t-2}.$$

What dependencies should we try? The general dynamics is described by a function $f(x_{t-1}, \ldots, x_{t-k})$ of k variables. From the commonsense viewpoint, it is reasonable to assume that when the previous values x_{t-i} change a little bit, then the predicted value should also change a little bit, i.e., that the function $f(x_{t-1}, \ldots, x_{t-k})$ is continuous, is smooth, and probably even analytical.

As we have mentioned, prediction of time series is a usual problem in applications to the physical world, and in physics, a usual approach to finding an unknown dependence is to expand it in Taylor series and to keep a few first terms in this expansion – i.e., to approximate the desired function by a polynomial; see, e.g., [4]. Let us thus use a similar approach. (In the Appendix, we provide a mathematical justification for the use of polynomials as opposed to other possible approximations – such as, e.g., Fourier series.)

As the first approximation, it is thus reasonable to use linear dependencies. The advantage of such a linear formula is that we know the explicit expression for a solution to a general linear dynamics:

$$x_t = a_0 + a_1 \cdot x_{t-1} + \ldots + a_k \cdot x_{t-k}.$$

Specifically, this solution is a linear combination of a constant and expressions of the type $x^p \cdot \exp(a \cdot x) \cdot \cos(b \cdot x)$ and $x^p \cdot \exp(a \cdot x) \cdot \sin(b \cdot x)$, where $\rho = a + \mathrm{i} \cdot b$ are solutions of the equation $1 = a_1 \cdot \rho + \ldots + a_k \cdot \rho^k$, and p is a non-negative integer not exceeding the multiplicity of the corresponding solution. This allows us to describe individual growth curves, but since the values a and b are fixed, we cannot describe general growth curves, with varying growth rates – and this is what we want to be able to detect future geniuses. Thus, we need to use a non-linear dependence – at least quadratic:

$$x_t = a_0 + \sum_{i=1}^{k} a_i \cdot x_{t-i} + \sum_{i \le j} a_{ij} \cdot x_{t-i} \cdot x_{t-j}. \quad (2)$$

So what do we propose. Our proposal is to select the parameters k and d, and to try to approximate the actual growth curve x_t of all the students by the d-th order dependence:

$$x_t = a_0 + \sum_{i=1}^{k} a_i \cdot x_{t-i} + \ldots + \sum_{i_1 \le \ldots \le i_d} a_{i_1 \ldots i_d} \cdot x_{t-i_1} \cdot \ldots \cdot x_{t-i_d}. \quad (3)$$

To find the coefficients $a_{i_1 \ldots}$, one can use the usual Least Squares method.

It is reasonable to start with the simplest case $k = 2$ and $d = 2$. If this does not lead to a good description of the growth rates, then it is reasonable increase either k or d by 1, to check which of the two increases leads to a better fit, etc.

Once we arrive at a good fit formula, we can start using it to predict future Einsteins.

We are optimistic. Will this approach work? We believe it will. One of the main reasons for our belief is that many similar mathematical approaches have been successful in solving problems from all the stages of the pedagogical process; see, e.g., [7].

But we are not naively optimistic. Do we believe that the very first application of our ideas will immediately lead to a success? Probably not.

Prediction of time series is a difficult problem. For example, several years ago, Vladik's father (and Olga's father-in-law), who was a professional engineer specializing in control, decided to apply known control techniques to predict the effect of different doses of medicine on his sometimes high blood pressure. This sounded like a good idea in comparison with the naive trial-and-error approach recommended by his doctor – but the resulting complex algorithm did not bring him any faster in finding the optimal dosage of medicine than for his friends in similar situations who followed doctor's recommendations – the complex algorithm turned out to be actually slower.

A similar complex situation is happening right now, when many specialists around the world are trying to predict the dynamics of the Covid-19 pandemic: so far, none of these attempts has been very successful.

Difficult, yet, but we hope that eventually, we will succeed, and there will be be no future Einstein left behind!

Acknowledgments

This work was supported in part by the US National Science Foundation grants 1623190 (A Model of Change for Preparing a New Generation for Professional Practice in Computer Science) and HRD-1242122 (Cyber-ShARE Center of Excellence).

The authors are greatly thankful to all the participants of the V International Forum on Teacher Education (Kazan, Russia, May 29–31, 2019), especially to Aydar Kalimullin and Mourat Tchoshanov, for valuable discussions, and to Branka Radulovic for her encouragement and help.

References

[1] J. Boaler, *Limitless Mind: Learn, Lead, and Live Without Barriers*, Harper One, New York, 2019.

[2] K. Bryan, "Traiblazing math genius who was first woman to win Fields Medal age 40", *Independent*, July 15, 2017.

[3] D. Epstein, *Range: How Generalists Triumph in a Specialized World*, Macmillan, London, 2019.

[4] R. Feynman, R. Leighton, and M. Sands, *The Feynman Lectures on Physics*, Addison Wesley, Boston, Massachusetts, 2005.

[5] A. Grant, *Originals: How Non-Conformists Rule the World*, Penguin, New York, 2016.

[6] L. F. Kebede, "No more 'secret handshake': Universal screening qualifies 600 more Memphis students for gifted education", *Chalkboard*, 2020, available at https://www.chalkbeat.org/posts/tn/2020/01/14/no-more-secret-handshake-universal-screening-qualifies-600-more-memphis-students-for-gifted-education-more-to-come/

[7] O. Kosheleva and K. Villaverde, *How Interval and Fuzzy Techniques Can Improve Teaching*, Springer Verlag, Cham, Switzerland, 2018.

[8] W. Moskalew, A. Richter, and D. von Reincke, *Svetik: A Family Memoir of Sviatoslav Richter*, Toccata Press, London, 2015.

[9] S. Naifeh and G. W. Smith, *Van Gogh: The Life*, Random House, New York, 2011.

[10] Quantum Publishing, *The Great Masters*, Quantum Publishing, London, 2003.

[11] L. Schwartz, *A Mathematician Grappling with His Century*, Birkhäuser, Basel, 2001.

[12] US Department of Education, Office for Civil Rights, *Gender Equity in Education*, Washington, DC, 2012, available at https://www2.ed.gov/about/offices/list/ocr/docs/gender-equity-in-education.pdf

A Which Models Should We Use to Describe the Dynamics of Student Abilities

Towards a general formulation of the problem. We want to find a general formula $f(x_{t-1}, \ldots, x_{t-k})$ – i.e., a general analytical function of k variables – that predicts the student's abilities x_t at moment t based on his/her abilities x_{t-i} at previous moments of time.

The space of all possible functions is infinite-dimensional – in the sense that to fully identify a function, we need to select infinitely many parameters. However, at any given moment of time, the number of observations is finite, and based on the finite number of observations, we can only determine finitely many parameters. Thus, we need to select a finite-parametric (finite-dimensional) class of possible functions f. A natural way to do it is to select a basis of n

analytical functions $e_1(x_{t-1}, \ldots, x_{t-k})$, …, $e_n(x_{t-1}, \ldots, x_{t-k})$, and to consider functions of the type

$$f(x_{t-1}, \ldots, x_{t-k}) = C_1 \cdot e_1(x_{t-1}, \ldots, x_{t-k}) + \ldots + C_n \cdot e_n(x_{t-1}, \ldots, x_{t-k}) \quad (4)$$

corresponding to different values of the coefficients C_i.

For such families, the dependence on the unknown parameters C_i is linear and thus, determining the coefficients C_i can be done by using well-known and easy-to-use techniques such as Least Squares.

How do we measure the student's ability: a brief reminder. Whether it is the usual IQ test or some other tests, usually, the score is obtained as follows:

- we add up the student's scores on several problems and then
- *normalize* the result – e.g., for IQ, we divide this score by the average score of all the students who took this test.

It is important to take into account that this average score changes with time, as a result of which what was 120 many years ago may now be 110 or vice versa. In other words, depending on which year we take for this normalization, we may get somewhat different result.

When we transition from one normalization to another, all the values x_{t-i} are multiplied by the same coefficient λ – e.g., in the above example this coefficient is $\lambda = 110/120$.

Related invariance requirement. Which year we select for normalization is a question of mutual agreement, we could select a different year and use values $\lambda \cdot x_{t-i}$ instead of the current values x_{t-i}. It is therefore reasonable to require that our general model not depend on this arbitrary choice, i.e., that the corresponding set of all possible approximating functions not change if we re-normalize all the values:

$$\{C_1 \cdot e_1(x_{t-1}, \ldots, x_{t-k}) + \ldots + C_n \cdot e_n(x_{t-1}, \ldots, x_{t-k})\}_{C_1,\ldots,C_n} =$$

$$\{C_1 \cdot e_1(\lambda \cdot x_{t-1}, \ldots, \lambda \cdot x_{t-k}) + \ldots + C_n \cdot e_n(\lambda \cdot x_{t-1}, \ldots, \lambda \cdot x_{t-k})\}_{C_1,\ldots,C_n}. \quad (5)$$

Our main result. We prove that the invariance condition (5) implies that all the functions $e_i(x_{t-1}, \ldots, x_{t-k})$ are polynomials – and thus, that all the resulting approximating functions (4) are polynomials as well.

Proof. Indeed, each function $e_i(x_{t-1}, \ldots, x_{t-k})$ is analytical. Thus, it can be expanded into Taylor series. Let us denote the sum all the terms of overall order d in this expansion by $P_d(x_{t-1}, \ldots, x_{t-k})$, so that

$$e_i(x_{t-1}, \ldots, x_{t-k}) = P_0(x_{t-1}, \ldots, x_{t-k}) + P_1(x_{t-1}, \ldots, x_{t-k}) + \ldots \quad (6)$$

Here, P_0 is the constant term, P_1 is the sum of all the linear terms in the Taylor expansion, P_2 is the sum of all the quadratic terms in this expansion, etc. By definition, for each d, we have

$$P_d(\lambda \cdot x_{t-1}, \ldots, \lambda \cdot x_{t-k}) = \lambda^d \cdot P_d(x_{t-1}, \ldots, x_{t-k}),$$

thus from

$$e_i(\lambda \cdot x_{t-1}, \ldots, \lambda \cdot x_{t-k}) = P_0(\lambda \cdot x_{t-1}, \ldots, \lambda \cdot x_{t-k}) + P_1(\lambda \cdot x_{t-1}, \ldots, \lambda \cdot x_{t-k}) + \ldots,$$

we can conclude that

$$e_i(\lambda \cdot x_{t-1}, \ldots, \lambda \cdot x_{t-k}) = \lambda^0 \cdot P_0(x_{t-1}, \ldots, x_{t-k}) + \lambda^1 \cdot P_1(x_{t-1}, \ldots, x_{t-k}) + \ldots \quad (7)$$

Due to invariance requirement (5), the function (7) belongs to the space (4). Let us show that for all values d for which the d-th term is not identically 0, the function $P_d(x_{t-1}, \ldots, x_{t-k})$ also belongs to the linear space (4). Indeed, let us denote such indices by $d_1 < d_2 < \ldots$, then we have

$$e_i(x_{t-1}, \ldots, x_{t-k}) = P_{d_1}(x_{t-1}, \ldots, x_{t-k}) + P_{d_2}(x_{t-1}, \ldots, x_{t-k}) + \ldots \quad (8)$$

and

$$e_i(\lambda \cdot x_{t-1}, \ldots, \lambda \cdot x_{t-k}) =$$
$$\lambda^{d_1} \cdot P_{d_1}(x_{t-1}, \ldots, x_{t-k}) + \lambda^{d_2} \cdot P_{d_2}(x_{t-1}, \ldots, x_{t-k}) + \ldots \quad (9)$$

With this function, a function $\lambda^{-d_1} \cdot e_i(\lambda \cdot x_{t-1}, \ldots, \lambda \cdot x_{t-k})$ also belongs to the space (4), and this function has the form

$$\lambda^{-d_1} \cdot e_i(\lambda \cdot x_{t-1}, \ldots, \lambda \cdot x_{t-k}) =$$
$$P_{d_1}(x_{t-1}, \ldots, x_{t-k}) + \lambda^{d_2 - d_1} \cdot P_{d_2}(x_{t-1}, \ldots, x_{t-k}) + \ldots$$

When $\lambda \to 0$, the limit of the right-hand side is equal to $P_{d_1}(x_{t-1}, \ldots, x_{t-k})$. A finite-dimensional linear space is closed, so this limit function also belongs to the linear space (4). Thus, the difference $e_i(x_{t-1}, \ldots, x_{t-k}) - P_{d_1}(x_{t-1}, \ldots, x_{t-k})$ of the two functions from the linear space (4) also belongs to this space, and for this difference, we have

$$e_i(x_{t-1}, \ldots, x_{t-k}) - P_{d_1}(x_{t-1}, \ldots, x_{t-k}) = P_{d_2}(x_{t-1}, \ldots, x_{t-k}) + \ldots$$

Due to scale-invariance (5), if we plug in $\lambda \cdot x_{t-i}$ instead of x_{t_i}, we still get a function from the linear space (4):

$$e_i(\lambda \cdot x_{t-1}, \ldots, \lambda x_{t-k}) - P_{d_1}(\lambda \cdot x_{t-1}, \ldots, \lambda \cdot x_{t-k}) = P_{d_2}(\lambda \cdot x_{t-1}, \ldots, \lambda \cdot x_{t-k}) + \ldots =$$
$$\lambda^{d_2} \cdot P_{d_2}(x_{t-1}, \ldots, x_{t-k}) + \ldots$$

Thus, if we multiply this function by a constant λ^{-d_2}, we still get a function from the linear space (4):

$$\lambda^{-d_2} \cdot (e_i(\lambda \cdot x_{t-1}, \ldots, \lambda x_{t-k}) - P_{d_1}(\lambda \cdot x_{t-1}, \ldots, \lambda \cdot x_{t-k})) = P_{d_2}(x_{t-1}, \ldots, x_{t-k}) + \ldots$$

In the limit $\lambda \to 0$, we conclude that the function $P_{d_2}(x_{t-1}, \ldots, x_{t-k})$ also belongs to the linear space (4).

Similarly, we can prove that all the non-zero terms $P_{d_m}(x_{t-1}, \ldots, x_{t-k})$ also belong to the linear space. All these terms have different degrees and are, therefore, linearly independent. Since the dimension of the linear space (4) is n, we can only have no more than n linearly independent functions in this space. Thus, no more than n terms $P_d(x_{t-1}, \ldots, x_{t-k})$ can be different from 0. So, there are finitely many terms in the formula (8) that describes the basis function $e_i(x_{t-1}, \ldots, x_{t-k})$. By definition, each of the terms $P_d(x_{t-1}, \ldots, x_{t-k})$ is a polynomial, so the function $e_i(x_{t-1}, \ldots, x_{t-k})$ is the sum of finitely many polynomials and thus, a polynomial itself.

The statement is proven.