

Creative Discussions or Memorization? Maybe Both? (on the example of teaching Computer Science)

Vladik Kreinovich¹ and Olga Kosheleva²

¹Department of Computer Science

²Department of Teacher Education

University of Texas at El Paso

500 W. University

El Paso, TX 79968, USA

vladik@utep.edu, olgak@utep.edu

Abstract

We all strive to be creative in our teaching, but there is often not enough time to make all the topics creative fun. So sometimes, we teach memorization first, understanding later. We do it, but we often do it without seriously analyzing which topics to “sacrifice” to memorization. In this talk, we use simple mathematical models of learning to come up with relevant recommendations: Namely, all the topics form a dependency graph, and if we do not have enough time to allow students to treat all topics with equal creativity, then the most reasonable topics for memorization first are the ones in the critical path of this dependency graph.

1 Formulation of the Problem

Creativity good, memorization bad. Modern pedagogical literature is very convincing: creative discussions lead to a better understanding than memorization. Gently guided by an instructor, students solve interesting problems and uncover – themselves – the desired formula, this is great. The students fell good about it, they remember it better, they use it more creatively.

And yet. And yet, and yet . . . Some students of introductory CS cannot move forward because they forgot a formula for the logarithm of the product. Some forgot even how to add fractions.

Yes, we can stop and let them recreate this formula – but:

- do we really want to teach a few weeks less computing and a few weeks more math?
- and are we, CS folks, the best teachers of math?

What we do. What many of us do is:

- have students memorize the needed math and
- use the remaining time to be creative in computing.

Even in computing, we ask students to memorize patterns corresponding to sum, maximum, etc., instead of having them re-create all these codes creatively every time; see, e.g., [10] (see also [4, 5, 9]).

Problem. We do it, but we do it shamefully. Every time an instructor resorts to memorization, it is with a feeling of guilt – should not everything in education be creative fun?

What we plan to do. Our point is: maybe we should not feel guilty, and we justify our point by analyzing simple mathematical models of teaching.

2 Informal (Qualitative) Analysis of the Problem

First argument: time is an issue. Our first argument is that:

- while creative teaching is good,
- it is often slower.

In most classes, there is a dependence between material: to study some topics, students need to know some previous ones.

In the resulting dependence, there is often a critical path – and along this path, it may be better to use memorization first – and get a deep understanding later.

Second argument: we want to optimally adjust our teaching to students. Another argument is that we want to optimally use the student's brains.

Yes, it would be nice if we could keep the brains in the permanent state of active creative fun – but brains get tired, they need rest. Here, memorization helps: to solve a non-trivial problem, we use our creative thinking to find known patterns, known ways which can help solve it – and then we “switch off” the active brain and use memorized techniques to solve the resulting subproblems.

If we end up with a quadratic equations, we do not want to recall the tricks that lead to the formulas – we just want to plug in the numbers. Meanwhile, the active brain rests and gets ready for new creative activities – and everyone benefits!

3 Formal (Quantitative) Analysis of the Problem

Formalization of the problem. Let us denote the total amount of creative effort that a student can perform during the learning period by E . Our objective is to describe what is the proper way to distribute this amount between different moments of time so as to have the best overall learning result.

Let n denote the overall number of moment of time, let e_i denote the amount of creative effort that a student uses at moment i , and let $r(e)$ denote the amount of learning that results when a student uses a creative effort e . In these terms, we want to maximize the overall results, i.e., the sum $r(e_1) + \dots + r(e_n)$, under the constraint that the overall creative effort $e_1 + \dots + e_n$ is equal to the given amount E .

In other words, we want to find the values $e_1, \dots, e_n$ that solve the following constrained optimization problem:

$$r(e_1) + \dots + r(e_n) \rightarrow \max \quad (1)$$

under the constraint

$$e_1 + \dots + e_n = E. \quad (2)$$

Analytical solution to the above problem explains the need to switch creativity on and off. The above conditional optimization problem can be easily solved by using the standard calculus technique of Lagrange multipliers, according to which the above conditional optimization problem can be reduced to an unconditional one

$$r(e_1) + \dots + r(e_n) + \lambda \cdot (e_1 + \dots + e_n - E) \rightarrow \max,$$

where λ is a constant (Lagrange multiplier).

For this unconditional optimization problem, the maximum can be computed by simply differentiating the objective function with respect to e_i and equating the resulting partial derivative to 0. As a result, we get the following equation: $r'(e_i) + \lambda = 0$. So, all the degrees e_i corresponding to the optimal degrees selection must be the solutions to the equation $F(e) = 0$, where we denoted $F(e) \stackrel{\text{def}}{=} r'(e) + \lambda$. In other words, all these degrees must be the roots of a function $F(e)$.

Intuitively, small changes in the amount of creative effort e should not drastically affect the learning result $r(e)$. Therefore, it is reasonable to assume that the function $r(e)$ is smooth and probably even analytical (i.e., can be expanded in Taylor series). In this case, the function $F(e)$ is also an analytical function. It is known that an analytical function which is not identically 0 can only have finitely many roots on an interval.

Thus, all the optimal effort amounts e_i must belong to the finite set of these solutions.

For usual analytical functions, this set of solutions is small. Indeed, an arbitrary analytical function, by definition, is equal to its Taylor series and therefore, can be approximated, with an arbitrary accuracy, by a polynomial. A polynomial of degree d can have no more than d roots; so, e.g., if a cubic polynomial is a reasonable approximation for the function $F(e)$, then, in this approximation, the function $F(e)$ has no more than 3 roots, so we use no more than three different levels of creative effort.

A more accurate approximation, e.g., by a 7-th order polynomial (which is usually enough to visually coincide for most known analytical functions on $[0, 1]$ such as $\sin$, $\cos$, etc.), would lead to no more than 7 different levels of creative effort, etc.

In other words, in the optimal learning arrangement, we should alternate between a small number of different levels of creativity. Since it is an empirical fact that it is not possible to always maintain the highest level of creativity throughout the course – in our terms, the available amount of effort E is smaller than that – this means that we do not need to alternate between higher and lower levels of student creativity.

Which topics should we ask students to memorize? All the topics form a dependency graph. Since we do not have enough time to allow students to treat all topics with equal creativity, then the most reasonable topics for memorization first are the ones in the *critical path* of this dependency graph; see, e.g., [1, 2].

4 Other Applications of This Idea

Other applications are possible. A similar argument can be used in many other situations when we want to achieve the largest overall result under restrictions on the overall effort; many such examples are described in [6]; let us mention several examples from [6].

Other applications to learning. In learning, similar arguments show the optimal distribution in learning activity is not a steadfast study, but rather periods of intense study separated by periods of relative rest.

Similarly, if we analyze the distribution of learning activity in the population, so that e_i described the amount of effort used to educate i -th person, we come to the conclusion that the optimal arrangement is not when the teaching efforts are uniformly distributed among students or when there is a continuous change from one student to another, but rather when there are a few levels and each student is assigned to a certain level of studying.

This result is in good accordance with the discrete system of university education, where the possible levels of education in a given domain are described in terms of a small list of degrees (e.g., BSc, MSc, Ph.D.).

Application to sleep. We can apply similar ideas to the description of other biological processes. For example, every biological creature has a certain level of activity. It cannot maintain the highest possible level of activity all the time, because its resources are limited. Therefore, it must distribute these resources

in such a way that the overall efficiency is the largest. Let $e_1, \dots, e_n$ denote levels of activity at different consequent moments of time. Let $r(e)$ denote the productivity of the activity with level e .

Then, the optimal levels of activity can be determined by solving the optimization problem (1), (2). We already know, from the solution to this problem, that in the optimal solution, the levels e_i cannot take arbitrary values, they should all belong to a small set of values. Therefore, the optimal activity schedule consists not of slowly changing activity from one level to another, but of switching between several discrete levels of activity.

This conclusion explains why, instead of a slow transition between high and low activity, most living creature have an abrupt transition between activity and sleep.

If we take subtler details into consideration, then we can say that the above conclusion explains why living creatures have an abrupt transition between activity and several levels of sleep (such as a REM phase and a normal sleep).

Similarity with “bang-bang” control in control theory. The above conclusion is consistent with the fact that in control theory, the optimal control often involves abrupt changes from one regime to another. For example, when driving a car, stability means, in particular, that once the car swerved, it should return to the original trajectory. The faster it returns, the more stable is the system. Therefore, from the viewpoint of stability only, the ideal (optimal) control would be the one that brings the car back on track in the shortest possible time (i.e., with the largest possible acceleration).

The non-smoothness of the optimal control is not a peculiar feature of the car example: in control theory, there are general theorems that show that under certain (reasonably general) conditions, the optimal control is indeed of the above-described “bang-bang” type (see, e.g., [7]; not incidentally, the word “bang-bang” is an “official”, well-defined and widely used term in control theory).

Application to consumption. Similarly, for a person with limited resources, the consumption schedule which leads to the largest satisfaction is not the schedule in which these resources are equally distributed, but rather the schedule in which periods of higher consumption (“feasts”) are abruptly changing to periods of lower consumptions (“fasts”). This conclusion is in good accordance with the results obtained by economists who analyzed more complicated economic models (see, e.g., [3, 8]).

Application to traffic control. In traffic, similar idea explains why the optimal traffic arrangement means that we fix a small number of speed levels, and assign (maybe dynamically) each road to one of these levels. In real life, such levels are freeway, city limits, school zone, etc.

Acknowledgments

This work was supported in part by the National Science Foundation grants HRD-0734825 and HRD-1242122 (Cyber-ShARE Center of Excellence) and DUE-0926721, and Chiang Mai University, Thailand. The authors are greatly thankful to Komasn Suriya for his encouragement.

References

- [1] Th. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms*, MIT Press, Cambridge, Massachusetts, 2009.
- [2] E. W. East, *Critical Path Method (CPM) Tutor for Construction Planning and Scheduling*, McGraw Hill, New York, 2015.
- [3] J. Elster, “Rationality and the emotions”, *The Economic Journal*, 1996, Vol. 438, pp. 1386–1397.
- [4] M. Guzdial, “What’s the best way to teach Computer Science to beginners?”, *Communications of the ACM*, 2015, Vol. 58, No. 2, pp. 12–13.
- [5] P. A. Kirschner, J. Sweller, and R. E. Clark, “Why minimal guidance during instruction does not work: an analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching”, *Educational Psychologist*, 2006, Vol. 41, No. 2, pp. 75–86.
- [6] V. Kreinovich and H. T. Nguyen, “Granularity as an optimal approach to uncertainty – a general mathematical idea with applications to sleep, consumption, traffic control, learning, etc.”, *Proceedings of the 19th International Conference of the North American Fuzzy Information Society NAFIPS’2000*, Atlanta, Georgia, July 13–15, 2000, pp. 316–320.
- [7] R. R. Mohler, *Nonlinear systems. Vol. 1. Dynamics and control*, Prentice Hall, Englewood Cliff, New Jersey, 1991.
- [8] T. Scitkovsky, *The Joyless Economy*, Oxford University Press, Oxford, 1992.
- [9] J. Sweller and G. A. Cooper, “The use of worked examples as a substitute of problem solving in learning algebra”, *Cognition and Instruction*, 1985, Vol. 2, No. 1, pp. 59–89.
- [10] A. B. Tucker, A. P. Bernat, W. J. Bradley, R. D. Cupper, and G. W. Scragg, *Fundamentals of Computing I: Logic, Problem Solving, and Computers*, McGraw Hill, New York, 1994.