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

Karen Villaverde
Department of Computer Science
New Mexico State University
Las Cruces, NM 88003, USA
kvillave@cs.nmsu.edu

Sergio D. Cabrera
Department of Electrical and Computer Engineering
University of Texas at El Paso
500 W. University
El Paso, TX 79968, USA
sergioc@utep.edu

Abstract—To enhance education, it is important to keep track of the students’ progress and to use appropriate teaching strategies based on the current state of the students’ knowledge. From this viewpoint, the problem of selecting an optimal control of the state of student’s knowledge becomes a particular case of the general problem of controlling systems. Control theory has shown that in many situations, the optimal way to control a complex system is through a decentralized control, when, in addition to (and sometimes instead of) a centralized controller, each subsystem itself controls its own interactions with other subsystems. This is how the Internet operates, this is how many biological systems operate. In relation to education, this means that to enhance education, in addition to an instructor teaching students, we also need students teaching and helping each other.

On the qualitative level, this conclusion sounds very non-controversial: collaborative learning is a well-known successful educational technique. On the quantitative level, however, the resulting recommendations become more non-trivial: the recommendation is that students be informed about the level of knowledge of fellow students. At first glance, this may sound like a proposal to go back to the old days when student privacy was not a concern and all the grades were posted. The difference between this old practice and our proposal is both in motivation and in implementation. While the main motivation behind the old system was to encourage a (stimulating) competition, our objective is to encourage collaboration. In terms of implementation, since collaboration cannot be forced, we propose to post the grades only with the students’ consent, when the students become convinced that this indeed helps all of them to succeed.

I. INTRODUCTION: CONTROL IS NEEDED IN EDUCATION

Teaching is not easy. Education is one of the oldest human activities. For centuries and millennia, people have been inventing and developing different teaching techniques, techniques intended to help students move from their original state—in which they only know the basics of the studied material—to the desired state, in which they have mastered the corresponding knowledge. One of the main difficulties with teaching is that students are different: they have different starting knowledge, they have different learning styles. So, different teaching techniques work differently for different students.

The differences between the students do not remain the same, they changes with time: during one week, a student may encounter an emergency situation and start lagging behind, while during another week a student may catch up. It is therefore desirable to take the present state of a student into account when selecting an appropriate teaching method for this student.

How do we gauge the student’s state of knowledge: fuzzy techniques are needed. In order to select a technique which is most appropriate for a student, we need to know the current state of each student’s knowledge. How can we gauge this state of knowledge? At the end of a course, it is easier to gauge this state of knowledge: we usually have well-defined learning objectives, we can formulate problems and tasks that show how well a student achieved these objectives, and we can gauge how well a student performed on the resulting exam(s).

Gauging a state of knowledge is not so easy on the intermediate stages, when even the advanced students have not yet fully mastered the required skills. While it is difficult to get an objective measure of the student’s state of knowledge on these stages, skilled educators often have a good grasp of where each student stands, by properly assigning partial credit for relevant tasks. Even students themselves usually have a good intuitive understanding on where they stand on different topics. For example, a student studying Introduction to Computer Science may say that he has mastered data types, that he has a good grasp of if-then statements, that he is struggling with for-loops, and that he is completely lost in objects.

In all such evaluations, the instructor’s and the student’s estimates are usually formulated in terms of words from a natural language (“good grasp”, “struggling”, etc.). It is desirable to design automatized computer-based tools that would use these natural-language evaluations to give appropriate advice to students and instructors. Computers cannot easily process words from natural language; so, we need to use techniques that translate these natural-language evaluations into computer-understandable terms.

To deal with situations in which words from natural language describe degree (in our case, degree of knowledge),
such technique has been invented by L. Zadeh under the name of fuzzy logic; see, e.g., [3], [5], [12]. Thus, we need to use fuzzy techniques to generate estimates for students’ levels of knowledge.

Control is needed in education. Once we know the current state (and how different this state is from the desired state), we need to decide on the best strategy of reaching the desired state.

This is a typical engineering problem; techniques for solving this problem are known as control techniques; see, e.g., [1], [6]. Thus, we conclude that we need to use control techniques in education.

II. From Traditional Control to Decentralized Control

General idea of a decentralized control. Traditional control theory (as well as traditional decision making theory) assumes that there is a single deciding agent. This assumption makes perfect sense in simple situations, when there are few parameters to controls. In such situations, a centralized controller can control all these parameters. However, for a complex system, the number of parameters can be huge, so it becomes difficult for a centralized controller to control the values of all these parameters. Good news is that many of these parameters describe local subsystems, so for such parameters, there is often no need to invoke a central authority: corresponding decisions can be made on the spot.

For example, when we control a single ship, in many cases, we need to make centralized decisions. However, when we control a fleet of ships, many decisions are better left to the ship captains; they have a better knowledge of the current situation, and in most cases, they are best qualified to make decisions. Unnecessary centralization creates a decision bottleneck, resulting in decision delays. Excessive centralization also decreases the system reliability, because the system becomes vulnerable if a connection to the central authority is temporarily disabled.

Many successful complex systems are decentralized; the Internet is one good example. On the other hand, communist attempts of over-centralized economic control in Eastern Europe, where almost all local industrial decisions had to be approved at the center, led to economic disasters.

Modern control theory operates in line with this conclusion: it prescribes decentralized control wherever appropriate; see, e.g., [2], [9].

Let us provide a quantitative explanation for such a decentralization, on an example of a simple control situation.

Control: reminder. In general, a current state of a system is characterized by one or several parameters \( x = (x_1, \ldots, x_n) \). For some of these parameters, the user has selected desired values. For example, for air conditioning, there is a desired temperature set up by the user; for the cruise control of a car on a freeway, the desired velocity should be equal to the speed with which all the cars travel on this freeway, etc.

It is convenient to use, as the parameters \( x_i \), the differences between the actual and the desired value. For this choice of parameters, the desired values correspond to 0.

We usually know the dynamics of the system. In other words, we know how the state of the system changes with time. In precise terms, we usually know the differential equations that describe the system’s dynamics:

\[
\dot{x}_i = f_i(x, u),
\]

where, as usual, \( \dot{x}_i \overset{\text{def}}{=} \frac{dx_i}{dt} \) denotes time derivative, and \( u \) denotes the control applied to the system. Once we fix the control strategy, i.e., once we decide which control \( u(x) \) to apply for each state \( x \), the above equation simplifies into

\[
\dot{x}_i = F_i(x),
\]

where \( F_i(x) \overset{\text{def}}{=} f_i(x, u(x)) \).

This equation can be often further simplified, by using the following two arguments.

First, we can take into account that once we have reached the desired state \( x = 0 \), we should stay in this state, i.e., we should have \( F_i(0) = 0 \).

Second, we can take into account that when control is efficient, we never deviate too far away from the desired state, i.e., that the differences \( x_i \) between the actual and the desired state remain reasonably small. In this case, terms quadratic (and higher order) in \( x_i \) can safely ignored. Thus, we can expand the function \( F(x) \) into Taylor series in \( x_i \)

\[
F_i(x) = F_i(0) + \sum_{j=1}^{n} F_{ij} x_j + \frac{1}{2} \sum_{j=1}^{n} \sum_{k=1}^{n} F_{ijk} x_j x_k + \ldots
\]

where \( F_{ij} \overset{\text{def}}{=} \frac{\partial F_i}{\partial x_j}, \ F_{ijk} \overset{\text{def}}{=} \frac{\partial^2 F_i}{\partial x_j \partial x_k} \), etc., and keep only linear terms in this expansion. Taking into account that \( F_i(0) = 0 \), we get the following system of linear differential equations:

\[
\dot{x}_i = \sum_{j=1}^{n} F_{ij} x_j.
\]

This is the system that we will consider in this section.

Simple example of a control situation. In general, a system consists of several subsystems. We will consider a simple case when the state of each subsystem is characterized by a single parameter.

In this case, \( n \) parameters \( x_1, \ldots, x_n \) which describe the state of a system mean that we have \( n \) subsystems, each of which is described by the corresponding parameter \( x_i \). Each parameter \( x_i \) is the difference between the actual state and the desired state of the \( i \)-th subsystem.

Case of centralized control. In the case of centralized control, only the central authority can influence the state of the \( i \)-th system. Each system independently communicates with the central authority, and the control decision for each system is made based on the state of this system. In the education example, this means that the teacher provides feedback to each
For such centralized control, the rate $F_i(x)$ with which the state of the $i$-th system changes depends only on the state $x_i$ of this system and does not depend on the state $x_j$ of any other subsystem $j \neq i$. Thus, for $j \neq i$, we get $F_{ij} = \frac{\partial F_i}{\partial x_j} = 0$, and the above general equation takes the form $\dot{x}_i = F_{ii} \cdot x_i$.

If $F_{ii}$ is positive, then any deviation from the ideal state will increase in time; so, for an efficient control, we must have $F_{ii} < 0$, i.e., $F_{ii} = -k_i$ for some $k_i > 0$. If we start with a deviation $x_i(0) = \Delta_i \neq 0$, then the solution to the corresponding differential equation $\dot{x}_i = -k_i \cdot x_i$ takes the form $x_i(t) = \Delta_i \cdot \exp(-k_i \cdot t)$. The larger $k_i$, the faster we get back to the ideal state; since our goal is to keep the system as close to the ideal state as possible, we should use the largest possible value of $k_i$.

There are usually some physical limitations on the values of $k_i$; e.g., a car can accelerate or decelerate only so much. If we denote the corresponding limit by $k$, then we conclude that for each subsystem, we should use $k_i = k$. The resulting dynamics takes the form

$$\dot{x}_i = -k \cdot x_i,$$

and the resulting solution is

$$x_i(t) = \Delta_i \cdot \exp(-k \cdot t).$$

**Case of decentralized control: description.** In the decentralized control, in addition to the centralized control, each subsystem can also be influenced by other subsystems. In the education example, a student not only gets information from the instructor, he or she also gains information from other students.

The effect of other students depends on the difference between their levels of knowledge:

- if the students are at the same level of knowledge, then they cannot learn much from each other;
- however, if one of the students knows much more than the other one, then this additional knowledge can be transferred.

In other words, the effect of $j$-th student on the $i$-th student is proportional to the difference $x_j - x_i$: this effect adds a term $d \cdot (x_j - x_i)$ (where $d$ is a proportionality coefficient) to the right-hand side $\dot{x}_i = \ldots$ of the corresponding dynamic equation. Thus, the equation becomes

$$\dot{x}_i = -k \cdot x_i + d \cdot \sum_{j \neq i} (x_j - x_i).$$

The values $k$ and $d$ reflect the ability of an instructor and of a fellow student to convey information. Of course, an instructor is usually skilled in teaching while the students are still learning themselves, so we should expect that $k \gg d$.

**Case of decentralized control: solving the resulting system of equations.** In the case of centralized control, the behavior of each subsystem was described by a separate easy-to-solve differential equation. In the decentralized case, we have a more complex system, in which each variable depends on the others. We will show, however, that this seemingly more complex system of differential equations still allows a simple explicit solution.

To show this, let us add the zero term $d \cdot (x_i - x_i)$ to the right-hand side of the above equation (7). After this addition, the equation (7) takes the form

$$\dot{x}_i = -k \cdot x_i + d \cdot \sum_{j=1}^{n} (x_j - x_i).$$

By representing the sum of differences as the difference between the sums, we conclude that

$$\dot{x}_i = -k \cdot x_i + d \cdot \sum_{j=1}^{n} x_j - d \cdot n \cdot x_i,$$

i.e., equivalently, that

$$\dot{x}_i = -(k + d \cdot n) \cdot x_i + d \cdot \sum_{j=1}^{n} x_j.$$  

By adding the equations (10) corresponding to all $i = 1, \ldots, n$, we get

$$\sum_{i=1}^{n} \dot{x}_i = -(k + d \cdot n) \cdot \sum_{i=1}^{n} x_i + d \cdot n \cdot \sum_{j=1}^{n} x_j,$$

Since the derivative of the sum is equal to the sum of the derivatives, we conclude that

$$\frac{d}{dt} \sum_{i=1}^{n} x_i = -(k + d \cdot n) \cdot \sum_{i=1}^{n} x_i + d \cdot n \cdot \sum_{j=1}^{n} x_j,$$

i.e., that for $s \overset{\text{def}}{=} \sum_{i=1}^{n} x_i$, we get the equation

$$\dot{s} = -(k + d \cdot n) \cdot s + d \cdot n \cdot s = -k \cdot s.$$  

When we have an initial deviation in only one of the subsystems, i.e., when $x_i(0) = \Delta_i$ and $x_j(0) = 0$ for all $j \neq i$, then the initial value of the sum $s(0)$ is also equal to $\Delta_i$. Thus, from the equation (12), we can conclude that

$$s(t) = \sum_{j=1}^{n} x_j(t) = \Delta_i \cdot \exp(-k \cdot t).$$

The equation (10) now takes the form

$$\dot{x}_i = -(k + d \cdot n) \cdot x_i + \Delta_i \cdot \exp(-k \cdot t).$$

According to the general formulas for solving linear differential equations with constant coefficients and exponential right-hand sides, a general solution to (14) is the sum of a solution in the absence of the exponential term (i.e., of a term proportional to $\exp(-(k + d \cdot n) \cdot t)$) and a term proportional to the exponential one. In other words, we have

$$x_i(t) = A \cdot \exp(-(k + d \cdot n) \cdot t) + B \cdot \exp(-k \cdot t).$$
To determine the coefficients $A$ and $B$, we do two things:

- first, we take into account that for $t = 0$, the expression (15) should be equal to $x_i(0) = \Delta_i$;
- second, we substitute the expression (15) into the equation (14) and equate the resulting expressions on the left- and right-hand sides.

First, we substitute $t = 0$ into the expression (15) and take into account that $x_i(0) = \Delta_i$; we then conclude that

$$\Delta_i = A + B.$$  \hfill (16)

Second, we substitute the expression (15) into the equation (14). The left-hand side of (14) takes the form

$$\dot{x}_i(t) = -A \cdot (k + d \cdot n) \cdot \exp(-(k + d \cdot n) \cdot t) - B \cdot k \cdot \exp(-k \cdot t),$$  \hfill (17)

while the right-hand side of (14) takes the form

$$-(k + d \cdot n) \cdot x_i + \Delta_i \cdot \exp(-k \cdot s) = -A \cdot (k + d \cdot n) \cdot \exp(-(k + d \cdot n) \cdot t) - B \cdot (k + d \cdot n) \cdot \exp(-k \cdot s) + \Delta_i \cdot \exp(-k \cdot s) = -A \cdot (k + d \cdot n) \cdot \exp(-(k + d \cdot n) \cdot t) - (B \cdot (k + d \cdot n) - d \cdot \Delta_i) \cdot \exp(-k \cdot s).$$  \hfill (18)

In (17) and (18), the coefficients at $\exp(-(k + d \cdot n) \cdot t)$ already coincide. By comparing the coefficients at $\exp(-k \cdot t)$, we conclude that

$$B \cdot k = B \cdot (k + d \cdot n) - d \cdot \Delta_i.$$  \hfill (19)

By subtracting $B \cdot k$ from both sides and moving the negative term to the other side, we get

$$B \cdot d \cdot n = d \cdot \Delta_i.$$  \hfill (20)

Dividing the resulting equality by the common factor $d$, we conclude that $B \cdot n = \Delta_i$ and thus, that

$$B = \frac{\Delta_i}{n}.$$  \hfill (21)

Using (16), we get

$$A = \Delta_i - B = \Delta_i - \frac{\Delta_i}{n} = \Delta_i \cdot \left(1 - \frac{1}{n}\right).$$  \hfill (22)

Thus, the solution takes the form

$$x_i(t) = \Delta_i \cdot \left(1 - \frac{1}{n}\right) \cdot \exp(-(k + d \cdot n) \cdot t) + \frac{\Delta_i}{n} \cdot \exp(-k \cdot t).$$  \hfill (23)

**Comparison: decentralized control is better.** Let us compare the formula (6) corresponding to the centralized control with the formula (23) corresponding to decentralized control.

Asymptotically, for large $t$, the term $\exp(-(k + d \cdot n) \cdot t)$ decreases much faster than $\exp(-k \cdot t)$. (Actually, as we have mentioned, the coefficient $d$ is much smaller than $k$, but the product $d \cdot n$ can be similar in size or even larger than $k$.) Thus, asymptotically:

- in the centralized case, we have
  $$x_i(t) \sim \Delta_i \cdot \exp(-k_1 \cdot t);$$
- in the decentralized case, we have
  $$x_i(t) \sim \frac{\Delta_i}{n} \cdot \exp(-k_1 \cdot t).$$

So, for large $t$, we get an $n$ times smaller deviation. In other words, with decentralized control, we reach the ideal state much faster than for the centralized one.

This conclusion, however, comes from the simplified case when only one of the variables deviates from the ideal state, i.e., in education terms, when only one student lags behind, while all other students show perfect knowledge. In this case, the fact that this lagging-behind student can get an additional help from all other students definitely helps breach the gap faster.

What if we have a more realistic situation, in which several students lag behind and some other students may be ahead? It is reasonable to assume that the initial deviations $x_i(0)$ corresponding to different students are random and independent, with 0 average — since we are trying to set up a path of learning which is, on average, appropriate for the students. It is known that the sum $s(0) = \sum_{i=1}^{n} x_i(0)$ of many independent identically distributed random variables has a distribution which is close to Gaussian; this is known as the Central Limit Theorem; see, e.g., [7].

The variance of this sum is equal to the sum of the variances, i.e., is equal to $n \cdot \sigma^2$, where $\sigma^2$ is the variance of the distribution of deviations $x_i$. Thus, the standard deviation for the sum is equal to $\sqrt{n} \cdot \sigma$. In other words, the sum $s(0)$ is approximately equal to $\sqrt{n} \cdot \sigma$, for $\sigma \approx \Delta_i$. Thus, similar to the above, we get

$$s(t) = s(0) \cdot \exp(-k \cdot s),$$  \hfill (24)

and hence,

$$x_i(t) = (\Delta_i - s(0)) \cdot \exp(-(k + d \cdot n) \cdot t) + s(0) \cdot \exp(-k \cdot t).$$  \hfill (25)

Therefore, asymptotically:

- in the centralized case, we have
  $$x_i(t) \sim \Delta_i \cdot \exp(-k_1 \cdot t);$$
- in the decentralized case, we have
  $$x_i(t) \sim \frac{\Delta_i}{n} \cdot s(0) \cdot \exp(-k_1 \cdot t),$$

with $s(0) \approx \Delta_i / \sqrt{n}$.

So, for large $t$, we get an $\sqrt{n}$ times smaller deviation. In other words, even when we take into account that we may have deviations for all the subsystems, still, with decentralized control, we reach the ideal state much faster than for the centralized one.
III. HOW TO APPLY THESE RESULTS TO EDUCATION

Recommendations from advanced control theory. Our simple analysis has shown that in general, to get a better control, we need to switch from a centralized control (in which the central controller controls all the objects) to a decentralized control (in which objects also control each other).

In education terms, this means that a typical centralized arrangement in which an instructor is the only one who teaches, the only one who knows the state of knowledge of all the students – is clearly not the best arrangement. We can achieve much better results if, in addition to the instructor teaching, students also teach each other.

On the qualitative level, these ideas are already used in advanced education. The idea that students should also teach each other is well known in pedagogy. This idea is an important part of many successful advanced teaching strategies such as collaborative learning; see, e.g., [4] and references therein.

In Computer Science, a similar idea is known as pair programming, where several students help each other while working on the same programming assignment. This idea originated in industry, but it turned out that it is useful in learning as well [4], [8], [11]; it is known to be one of the most efficient learning strategies [4]. One of us (KV) has successfully used this idea in teaching Introduction to Computer Science.

The corresponding techniques are welcome by students themselves. Many authors noticed that students themselves like teamwork; see, e.g., [10]. Moreover, it has been noticed that the Millennials – the students of the new generation – are much more accustomed to teamwork that the previous generations of students.

These methods are already used, so what is missing? While the collaborative learning methods are actively used, they are only used on the qualitative level. To get better results, to be able to use the advanced control techniques, we need to use these methods on quantitative level.

How can we use these methods on quantitative level? To be able to use the advanced control methods on the quantitative level, we need to provide the corresponding quantitative information to all those who are involved in control. Since the main idea is that students themselves should be in control, they should help each other, they should select the ways to help each other, we thus arrive at the need to provide all the students with the information about the state of knowledge of other students from their group.

This seems like a return to the past – but it is actually a similar idea but on a new level. At first glance, the above suggestion seems like a proposal to go back to the past. Several decades ago, when student privacy was not viewed as a concern, all the grades were posted, students knew each other’s grades. While there were some positive consequences of this – e.g., a healthy competition was often stimulated, encouraging students to study harder – current pedagogy believes that the possible side effect of lower self-esteem is too negative to go back to this old practice.

However, what we propose is not to return to the old practice, what we propose is similar but very different. The main drawback of the old system was that usually, it used to report an overall grade on a test; the only possible positive consequence of this information was that a student who may be otherwise happy with his or her average grade would be challenged to study harder if this student knew that many other students in the class were ahead of him/her.

Our objective is exactly the opposite: instead of encouraging competition, we want to encourage collaboration. What we propose is posting, within each class section, levels of knowledge of different students corresponding to different topics, so that students be able to team together and improve their knowledge in all the topics. An exceptional student with perfect knowledge in all the topics may not benefit much from this collaboration, but students who have deficiencies in some areas will definitely benefit from having others help them with these topics – at the exchange of being helped in other topics.

Of course, as every other collaboration, this cannot be forced. We cannot just post the grades and order the students to work together. We need to convince them that this works, we need to convince them that posting and the resulting collaboration are beneficial for everyone – and the simple mathematical model presented in this paper is one of the ways how we can (hopefully) convince students (and instructors too) that such collaboration is a good idea.

ACKNOWLEDGMENTS
The authors are thankful to the anonymous referees for valuable suggestions.

REFERENCES