Degree-Based (Interval and Fuzzy) Techniques in Math & Science Education

Olga Kosheleva

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

http://www.cs.utep.edu/vladik/olgavita.html

1. Introduction

- In education, evaluations of the student's knowledge, skills, and abilities are often subjective.
- Teachers often make these evaluations by using words from natural language like "good", "excellent".
- Traditionally, these evaluations are first transformed into exact numbers.
- This transformation, however, ignores the uncertainty of the original estimates.
- We show that taking this uncertainty into account helps on all stages of education process:
  - in planning education,
  - in teaching itself, and
  - in assessing the education results.

2. Applications to Planning Education and to Teaching Itself

Here, interval and fuzzy techniques help us:

- to better plan the order in which the material is presented and the amount of time allocated for each topic;
- to find the most efficient way of teaching inter-disciplinary topics;
- to stimulate students by explaining historical (informal) motivations behind different concepts and ideas.

3. Interval and Fuzzy Techniques in Assessment

In assessment, interval and fuzzy techniques help:

- to design a better grading scheme for test and assignments – that stimulates more effective learning,
- to provide a more adequate individual grading of contributions to group projects – by taking into account
  - subjective estimates of different student contributions, and
  - the uncertainty of these estimates;
- to provide a more adequate description of the student knowledge and of the overall teaching effectiveness.

Planning the Order in Which the Material Is Presented. I

4. Planning the Order in Which the Material is Presented

- In general, it is not clear what is the best order of presenting the material.
- The change in order often drastically changes the learning efficiency, sometimes in a counter-intuitive way.
- E.g.: it is usually assumed that students learn math concepts better if *concrete* examples come *first*.
- However, empirically, the *abstract-first* approach often enhances learning.
- We describe a simple model explaining why presentation order affects the learning efficiency.
- We then show how this explanation can be used:
  - to avoid inhibition of learning
  - and to enhance the student learning.

5. Learning: A Natural Geometric Representation

- The process of learning means that we change the state of a student:
  - from a state in which the student did not know the material (or does not have the required skill)
  - to a state in which the student has (some) knowledge of the required material.
- Let s_0 denote the original state of a student.
- Let S denote the set of all the states corresponding to the required knowledge or skill:
  - we start with a state $s_0 \notin S$, and
  - we end up in a state s which is in the set S.
- It is natural to define a metric $d(s, s')$ as the difficulty (time, effort, etc.) needed to go from state s to state s'.

6. Geometric Interpretation (cont-d)

- Our objective is to help the students learn in the easiest (fastest, etc.) way.
- In terms of the metric d, this means that we want to go:
  - from the original state $s_0 \notin S$
  - to the state $s \in S$ for which the effort $d(s_0, s)$ is the smallest possible.
- In geometric terms, the smallest possible effort means the shortest possible distance.
- Thus, our objective is to find the state $s \in S$ which is the closest to s_0.
- Such closest state is called the *projection* of the original state s_0 on the set S.

7. Learning Complex Material: Geometric Interpretation

- Let S_i, $1 \le i \le n$, denote the set of states in which a student has learned the i-th part of the material.
- Our objective: reach a state which belongs to the intersection $S \stackrel{\text{def}}{=} S_1 \cap \ldots \cap S_n$.
- In these terms, if we present the material in the order S_1, S_2, ..., S_n, this means that:
  - we first project s_0 onto the set S_1, resulting is a state $s_1 \in S_1$ which is the closest to s_0;
  - then, we project s_1 onto the set S_2, resulting is a state $s_2 \in S_2$ which is the closest to s_1; etc.
- By the time the students have learned S_n, they have somewhat forgotten S_1 – so we must repeat.
- Thus, starting from the state s_n, we again sequentially project onto the sets S_1, S_2, etc.

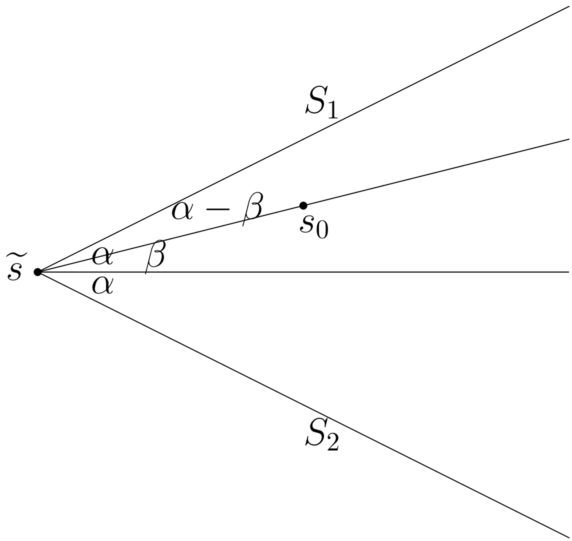
8. The Above Geometric Interpretation Makes Computational Sense

- The above "sequential projections" algorithm is actually actively used in many applications.
- For convex sets S_i:
  - we get a known Projections on Convex Sets (POCS) method;
  - POCS guarantees (under reasonable conditions) convergence to a point from $S_1 \cap \ldots \cap S_n$;
  - in our terms, this means that the students will eventually learn all parts of the necessary material.
- In the general (not necessarily convex) case:
  - the convergence is not always guaranteed,
  - but the method is still efficiently used, and often converges.

9. The Simplest Case: Two-Part Knowledge

- In this case, there are only two options:
  - we begin by studying S_1, then, we study S_2, then, if needed, we study S_1 again, etc.
  - we begin by studying S_2, then, we study S_1, then, if needed, we study S_2 again, etc.
- The amount of knowledge is reasonably small – otherwise, we would have divided into more than 2 pieces.
- In geometric terms, this means that the original state s_0 is close to the desired intersection set $S_1 \cap S_2$.
- Since all the states are close to each other, we can approximate the borders of S_i by linear expressions.
- Thus, these borders are straight lines (or planes in 3-D space).

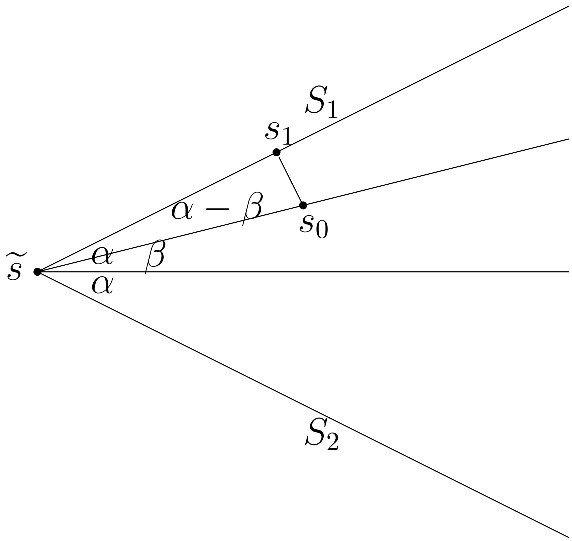
10. Resulting Geometric Configuration



Here:

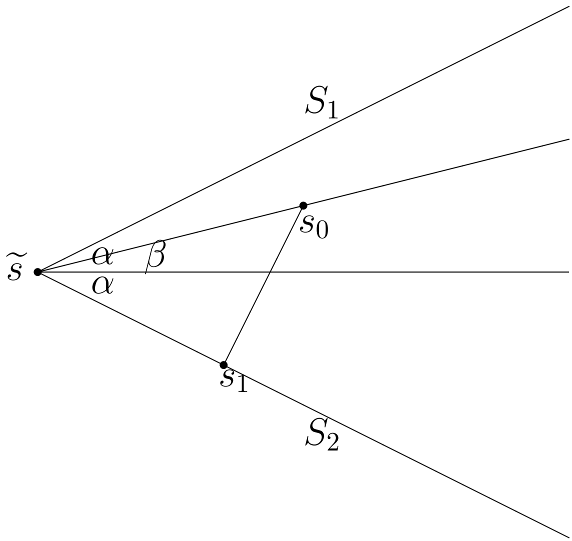
- 2α is the angle between the borders of S_1 and S_2;
- β is the angle between the direction $\widetilde{s}s_0$ and the mid-line.

11. First Option: S_1 then S_2



- Here, $s_0 s_1 \perp S_1$, so $d_1 \stackrel{\text{def}}{=} d(\widetilde{s}, s_1)$ is $d_1 = d_0 \cdot \cos(\alpha - \beta)$.
- On the next step, the angle is 2α, so $d_2 = d_1 \cdot \cos(2\alpha) = d_0 \cdot \cos(\alpha - \beta) \cdot \cos(2\alpha)$.
- In general, $d_k = d(s_k, \widetilde{s}) = d_0 \cdot \cos(\alpha - \beta) \cdot \cos^{k-1}(2\alpha)$.

12. Second Option: S_2 then S_1



- Here, $s_0 s_1 \perp S_1$, so $d_1 \stackrel{\text{def}}{=} d(\widetilde{s}, s_1)$ is $d_1 = d_0 \cdot \cos(\alpha + \beta)$.
- On the next step, the angle is 2α, so $d_2 = d_1 \cdot \cos(2\alpha) = d_0 \cdot \cos(\alpha + \beta) \cdot \cos(2\alpha)$.
- In general, $d_k = d(s_k, \widetilde{s}) = d_0 \cdot \cos(\alpha + \beta) \cdot \cos^{k-1}(2\alpha)$.

13. Analysis and Recommendations

- If we start w/S_1, we get $d_k = d_0 \cdot \cos(\alpha - \beta) \cdot \cos^{k-1}(2\alpha)$.
- If we start w/S_2, we get $d_k = d_0 \cdot \cos(\alpha + \beta) \cdot \cos^{k-1}(2\alpha)$.
- In general, $\cos(\alpha - \beta) \neq \cos(\alpha + \beta)$.
- This explains why the effectiveness of learning depends on the order in which the material is presented.
- Starting w/S_1 is better iff $\cos(\alpha - \beta) < \cos(\alpha + \beta)$, i.e., iff $\alpha - \beta > \alpha + \beta$.
- *Resulting recommendation:* start with the material that we know the least.
- This ties in with a natural commonsense recommendation to concentrate on one's deficiencies.
- This explains why studying more difficult (abstract) ideas first enhances learning.

Planning the Order in Which the Material Is Presented. II

14. Outline

- In general, human being are rational decision makers.
- However, in many situations, they exhibit unexplained "inertia", reluctance to switch to a better decision.
- We show that this seemingly irrational behavior can be explained if we take uncertainty into account.
- We also explain how this phenomenon can be utilized in education.

15. Traditional Approach to Human Decision Making: A Brief Reminder

- *Situation:* we have alternatives $A_1, \ldots, A_n$.
- *Idea:* alternatives are characterized by their "utility values" $u(A_1), \ldots, u(A_n)$.
- *Preference:* A_i is preferable to A_j if and only if

$$u(A_i) > u(A_j).$$

- *Empirical testing:* we need to compare
  - empirically "testable" behavior (such as preferring one alternative A_i to another alternative A_j) and
  - difficult-to-test comparison between the (usually unknown) utility values.
- *Conclusion:* empirical testing is difficult.

16. A Testable Consequence of the Traditional Approach to Decision Making

- *Fact:* for every two alternatives A_i and A_j:
  - either $u(A_i) > u(A_j)$, i.e., the alternative A_i is better,
  - or $u(A_j) > u(A_i)$, i.e., the alternative A_j is better.
- *Comment:* exact equality of $u(A_i)$ and $u(A_j)$ is highly improbable.
- In the first case $u(A_i) > u(A_j)$,
  - if we originally only had A_i, and then we add A_j, then we stick with A_i;
  - on the other hand, if we originally only had A_j, and then we add A_i, then we switch our choice to A_i.
- Similarly, in the second case $u(A_j) > u(A_i)$.

17. The Above Testable Consequence is in Perfect Agreement with Common Sense

- *Claim:* the above behavior is in perfect agreement with common sense.
- *Case 1:* the alternative A_i is preferable to the alternative A_j.
- *Expected behavior:* choose A_i irrespective of whether we started with only A_i or only A_j.
- *Case 2:* the alternative A_j is preferable to the alternative A_i.
- *Expected behavior:* choose A_j irrespective of whether we started with only A_i or only A_j.

18. For Close Alternatives, Decision Makers Do Not Behave in This Rational Fashion

- *Empirical result:* when the alternatives are close in value, decision maker exhibit "inertia".
- *Example:* selecting between two similar retirement plans A_i and A_j.
- *Case 1:* we start with the plan A_i and then add A_j.
- *Typical behavior:* stick to A_i.
- *Case 2:* we start with the plan A_j and then add A_i.
- *Typical behavior:* stick to A_j.
- *Why this is counter-intuitive:*
  - if A_i is better, then in Case 2, people should switch to A_i;
  - if A_j is better, then in Case 1, people should switch to A_j.

19. Maybe Human Behavior Is Irrational?

- How can we explain this seemingly irrational behavior?
- One possible explanation is that many people do often make bad (irrational) decisions:
  - waste money on gambling,
  - waste one's health or alcohol and drugs, etc.
- However, the above inertial behavior occurs among the most successful (otherwise rational) people.
- It is therefore reasonable to look for an explanation of this seemingly irrational behavior.
- It turns out that
  - we can come up with such an explanation
  - if we take into account uncertainty related to decision making.

20. How to Take Into Account Uncertainty in Decision Making Situations

- In practice, we can predict the consequences of alternatives only approximately, with some accuracy ε.
- So, instead of the exact values $u(A_i)$ and $u(A_j)$, we only know approximate values $\widetilde{u}_i$ and $\widetilde{u}_j$.
- The actual utility values can be within intervals $\mathbf{u}_i = [\widetilde{u}_i - \varepsilon, \widetilde{u}_i + \varepsilon]$ and $\mathbf{u}_j = [\widetilde{u}_j - \varepsilon, \widetilde{u}_j + \varepsilon]$.
- If the estimates are close, i.e., if $|\widetilde{u}_i - \widetilde{u}_j| < 2\varepsilon$, then
  - there exist values $u_i \in \mathbf{u}_i$ and $u_j \in \mathbf{u}_j$ s.t. $u_i < u_j$; and
  - there exist values $u_i \in \mathbf{u}_i$ and $u_j \in \mathbf{u}_j$ s.t. $u_i > u_j$.
- Thus, switching may decrease utility.
- So, it is prudent not to switch (especially since often switching comes with a penalty).

21. Another Case when Inertia is Beneficial: Control of a Mobile Robot

- We change direction based on the moment-by-moment measurements of the robot's location and/or velocity.
- Measurements are never 100% accurate.
- The resulting measurement noise leads to random deviations – shaking and "wobbling".
- Each change in direction requires that energy from the robot's battery go to the robot's motor.
- So, this wobbling drains the batteries and slows down the robot's motion.
- *Natural idea:* only change if it's clear (beyond uncertainty) that this will improve the performance.
- *Result:* UTEP robot's 1st place at 1997 AAAI competition.

22. Asymmetric Paternalism: Practical Application of Present-Biased Preferences

- *Fact:* the decision-making inertia is used in practice, to encourage desirable behavior.
- *Example:* a kid can drink either a healthy fruit juice or a soda drink which has no health value.
- *Traditional paternalism:* prohibit undesirable choices.
- *Problem:* this enforcement rarely works.
- *More efficient idea:*
  - at first provide only the desired alternative,
  - and then introduce all the other alternatives.
- *Example:* have only healthy drinks for the first few weeks of school, but then allow all the choices.
- *Result:* due to inertia, kids tend to stick to their original healthier choice.

23. How Does Our Explanation Help?

- *Fact:* asymmetric paternalism works.
- *Natural question:* do we need any explanation to make it work?
- *Problem:* sometimes this approach works, and sometimes it does not.
- *Additional problem:* it is not known how to predict when it will work.
- *Our solution:* this approach works when $|\widetilde{u}_i - \widetilde{u}_j| < 2\varepsilon$.
- *Comment:* for fuzzy numbers,
  - we can get a similar answer for "not switching with a given confidence",
  - if we similarly compare the intervals (α-cuts) for $u(A_i)$ and $u(A_j)$ corresponding to this level α.

24. Potential Applications to Education

- *Current applications:* in economy and in health.
- *Our idea:* use it in education.
- *Example:*
  - when the students just come to class from recess or from home, it is difficult to get their attention;
  - once they get engaged in the class, it is difficult for them to stop when the bell rings.
- *Objective:* prevent students from switching to a passive state A_j.
- *How to use this phenomenon:*
  - to start a class with engaging fun material, to get them into the studying state A_i;
  - they will (hopefully) remain in A_i even when a somewhat less fun necessary material is presented.

What is the Best Way to Distribute Efforts Among Students: Towards Quantitative Approach to Human Cognition

25. Deciding Which Teaching Method Is Better: Formulation of the Problem

- Pedagogy is a fast developing field.
- New methods, new ideas and constantly being developed and tested.
- New methods and new idea may be different in many things:
  - they may differ in the way material is presented,
  - they may also differ in the way the teacher's effort is distributed among individual students.
- To perform a meaningful testing, we need to agree on the criterion.
- Once we have selected a criterion, a natural question is: what is the optimal way to teaching the students.

26. How Techniques Are Compared Now: A Brief Description

- The success of each individual student i can be naturally gauged by this student's grade x_i.
- So, for two different techniques T and T', we know the corresponding grades $x_1, \ldots, x_n$ and $x'_1, \ldots, x'_{n'}$.
- In pedagogical experiments, the decision is usually made based on the comparison of the average grades

$$E \stackrel{\text{def}}{=} \frac{x_1 + \ldots + x_n}{n} \text{ and } E' \stackrel{\text{def}}{=} \frac{x'_1 + \ldots + x'_{n'}}{n'}.$$

- *Example:* we had $x_1 = 60$, $x_2 = 90$, hence $E = 75$. Now, we have $x'_1 = x'_2 = 70$, and $E' = 70$. In T':
  - the average grade is worse, but
  - in contrast to T, no one failed.

27. Towards Selecting the Optimal Teaching Strategy: Possible Objective Functions

- *Fact:* the traditional approach – of using the average grade as a criterion – is not always adequate.
- *Conclusion:* other criteria $f(x_1, \ldots, x_n)$ are needed.
- *Maximizing passing rate:* $f = \#\{i : x_i \geq x_0\}$.
- *No child left behind:* $f(x_1, \ldots, x_n) = \min(x_1, \ldots, x_n)$.
- *Best school to get in:* $f(x_1, \ldots, x_n) = \max(x_1, \ldots, x_n)$.
- *Case of independence:* decision theory leads to $f = f_1(x_1) + \ldots + f_n(x_n)$ for some functions $f_i(x_i)$.
- *Criteria combining mean E and variance V* to take into account that a larger mean is not always better:

$$f(x_1, \ldots, x_n) = f(E, V).$$

- *Comment:* it is reasonable to require that $f(E, V)$ is increasing in E and decreasing in V.

28. Towards Selecting the Optimal Teaching Strategy: Formulation of the Problem

- Let $e_i(x_i)$ denote the amount of effort (time, etc.) that is needed for i-th student to achieve the grade x_i.
- Clearly, the better grade we want to achieve, the more effort we need.
- So, each function $e_i(x_i)$ is strictly increasing.
- Let e denote the available amount of effort.
- In these terms, the problem of selecting the optimal teaching strategy takes the following form:

$$\text{Maximize } f(x_1, \ldots, x_n)$$

under the constraint

$$e_1(x_1) + \ldots + e_n(x_n) \le e.$$

29. Explicit Solution: Case of Independent Students

- *Maximize:* $f_1(x_1) + \ldots + f_n(x_n)$ under the constraint

$$e_1(x_1) + \ldots + e_n(x_n) \le e.$$

- *Observation:* the more efforts, the better results, so we can assume $e_1(x_1) + \ldots + e_n(x_n) = e$.
- *Lagrange multiplier:* maximize

$$J = \sum_{i=1}^{n} f_i(x_i) + \lambda \cdot \sum_{i=1}^{n} e_i(x_i).$$

- Equation $\dfrac{\partial J}{\partial x_i} = 0$ leads to $f_i'(x_i) + \lambda \cdot e_i'(x_i) = 0$.
- Thus, once we know λ, we can find all x_i.
- λ can be found from the condition $\sum\limits_{i=1}^{n} e_i(x_i(\lambda)) = e$.

30. Explicit Solution: "No Child Left Behind"

- In the No Child Left Behind case, we maximize the lowest grade.
- There is no sense to use the effort to get one of the student grades better than the lowest grade.
- It is more beneficial to use the same efforts to increase the grades of all the students at the same time.
- In this case, the common grade x_c that we can achieve can be determined from the equation

$$e_1(x_c) + \ldots + e_n(x_c) = e.$$

- Students may already have knowledge $x_1^{(0)} \le x_2^{(0)} \le \ldots$
- In this case, we find the largest k for which $e_1(x_k^{(0)}) + \ldots + e_k(x_k^{(0)}) \le e$ and then $x \in [x_k^{(0)}, x_{k+1}^{(0)})$ s.t.

$$e_1(x) + \ldots + e_{k-1}(x) + e_k(x) = e.$$

31. Explicit Solution: "Best School to Get In" Case

- Best-School-to-Get-In means maximizing the largest possible grade x_i.
- The optimal use of effort is, of course, to concentrate on a single individual and ignore the rest.
- Which individual to target depends on how much gain we will get:
  - first, for each i, we find x_i for which $e_i(x_i) = e$, and then
  - we choose the student with the largest value of x_i as a recipient of all the efforts.

32. Need to Take Uncertainty Into Account

- We assumed that:
  - we know *exactly* the benefits $f(x_1, \ldots, x_n)$ of achieving knowledge levels x_i;
  - we know *exactly* how much effort $e_i(x_i)$ is needed for each level x_i, and
  - we know *exactly* the level of knowledge x_i of each student.
- In practice, we have *uncertainty*:
  - we only know the *average* benefit $u(x)$ of grade x to a student;
  - we only know the *average* effort $e(x)$ needed to bring a student to the level x; and
  - the grade $\widetilde{x}_i$ is only an approximate indication of the student's level of knowledge.

33. Average Benefit Function

- *Objective function:* $f(x_1, \ldots, x_n) = u(x_1) + \ldots + u(x_n)$.
- Usually, the benefit function is reasonably smooth.
- In this case, if (hopefully) all grades are close, we can keep only quadratic terms in the Taylor expansion:

$$u(x) = u_0 + u_1 \cdot x + u_2 \cdot x^2.$$

- So, the objective function takes the form

$$f(x_1, \ldots, x_n) = n \cdot u_0 + u_1 \cdot \sum_{i=1}^{n} x_i + u_2 \cdot \sum_{i=1}^{n} x_i^2.$$

- *Fact:* $E = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i$ and $M = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i^2 = V + E^2$.
- *Conclusion:* f depends only on the mean E and on the variance V.

34. Case of Interval Uncertainty

- *Situation:* we only know intervals $[\underline{x}_i, \overline{x}_i]$ of possible values of x_i.
- *Fact:* the benefit function $u(x)$ is increasing (the more knowledge the better).
- *Conclusion:*
  - the benefit is the largest when $x_i = \overline{x}_i$, and
  - the benefit is the smallest when $x_i = \underline{x}_i$.
- *Resulting formula:* $[\underline{f}, \overline{f}] = \left[\sum\limits_{i=1}^{n} u(\underline{x}_i), \sum\limits_{i=1}^{n} u(\overline{x}_i)\right]$.
- *Reminder:* for quadratic $u(x)$ and exactly known x_i, we only need to know E and M.
- *New result:* under interval uncertainty, we need all n intervals.

35. Case of Fuzzy Uncertainty

- In many practical situations, the estimates $\widetilde{x}_i$ come from experts.
- Experts often describe the inaccuracy of their estimates in terms of imprecise words from natural language.
- A natural way to formalize such words is to use fuzzy logic:
  - for each possible value of $x_i \in [\underline{x}_i, \overline{x}_i]$,
  - we describe the degree $\mu_i(x_i)$ to which x_i is possible.
- Alternatively, we can consider α-cuts $\{x : \mu_i(x_i) \geq \alpha\}$.
- For each α, the fuzzy set $y = f(x_1, \ldots, x_n)$ has α-cuts

$$\mathbf{y}(\alpha) = f(\mathbf{x}_1(\alpha), \ldots, \mathbf{x}_1(\alpha)).$$

- So, the problem of propagating fuzzy uncertainty can be reduced to several interval propagation problems.

Back to the Future: Advanced Control Techniques Justify–on a New Level–Traditional Education Practices

36. Teaching Is Not Easy

- Education is one of the oldest human activities.
- We need 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.
- Students have different starting knowledge, different learning styles.
- The differences between the students change with time: a student may lag behind or catch up.
- It is desirable to take the present state of a student into account when selecting a teaching method.

37. How to Gauge the Student's State of Knowledge: Fuzzy Techniques Are Needed

- At the end of a course, we can gauge this state of knowledge against well-defined learning objectives.
- Gauging a state of knowledge is not so easy on the *intermediate* stages.
- Skilled educators often have a good grasp of where each student stands.
- Even students themselves usually have a good intuitive understanding on where they stand on different topics.
- These estimates are usually formulated by words from a natural language ("good grasp", "struggling", etc.).
- It is thus reasonable to use fuzzy techniques to estimate students' levels of knowledge.

38. Control Is Needed in Education

- Once we know the current 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.
- Thus, we conclude that we need to use control techniques in education.

39. From Traditional Control to Decentralized Control

- Traditional control theory assumes that there is a single deciding agent.
- This assumption makes perfect sense in simple situations, when there are few parameters to control.
- In such situations, a centralized controller can control all these parameters.
- However, for a complex system, the number of parameters can be huge.
- 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.
- In such cases, decisions can be made locally.

40. From Traditional Control to Decentralized Control (cont-d)

- 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.
- Unnecessary centralization creates a decision bottleneck, resulting in decision delays.
- Excessive centralization decreases reliability: if center fails, the system fails.
- Many successful complex systems are decentralized; the Internet is one good example.
- On the other hand, over-centralized economic control in Eastern Europe led to economic disasters.

41. Control: Reminder

- In general, a current state of a system is characterized by one or several parameters $x = (x_1, \ldots, x_n)$.
- It is convenient to use, as the parameters x_i, the *differences* between the actual and the desired value.
- We usually know how the state of the system changes with time: $\dot{x}_i = f_i(x, u)$.
- Once we fix the control strategy $u(x)$, we get $\dot{x}_i = F_i(x)$, where $F_i(x) \stackrel{\text{def}}{=} f_i(x, u(x))$.
- Once we have reached the desired state $x = 0$, we should stay in this state, i.e., we should have $F_i(0) = 0$.
- When control is efficient, the differences x_i are small.
- In this case, terms quadratic (and higher order) in F_i can safely ignored, so $\dot{x}_i = \sum\limits_{j=1}^{n} F_{ij} \cdot x_j$.

42. Simple Example of a Control Situation

- Let us consider a simple case when the state of each subsystem is characterized by a single parameter.
- In this case, n parameters x_i mean that we have n subsystems, each of which is described by the corr. x_i.
- In the centralized control, only the central authority can influence the state of the i-th system.
- In the education example, this means that only the teacher provides feedback to each student.
- For such centralized control, the rate $F_i(x)$ depends only on the state x_i: $\dot{x}_i = F_{ii} \cdot x_i$.
- If F_{ii} is positive, then any deviation from the ideal state will increase in time, so $F_{ii} = -k_i$ for some $k_i > 0$.
- If we start with a deviation $x_i(0) = \Delta_i \neq 0$, then $x_i(t) = \Delta_i \cdot \exp(-k_i \cdot t)$.

43. Case of Centralized Control (cont-d)

- In general, we have $x_i(t) = \Delta_i \cdot \exp(-k_i \cdot t)$.
- The larger k_i, the faster we get back to the ideal state.
- Our goal is to keep the system as close to the ideal state as possible.
- Thus, we should use the largest possible value of k_i.
- There are usually some physical limitations on the values of k_i.
- For example, 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 $x_i(t) = \Delta_i \cdot \exp(-k \cdot t)$.

44. Case of Decentralized Control

- In the decentralized case, each subsystem can also be influenced by other subsystems.
- In the education example, a student also gains information from other students.
- The effect of other students depends on the difference $x_j - x_i$ between their levels of knowledge: $d \cdot (x_j - x_i)$.
- Thus, the equation becomes $\dot{x}_i = -k \cdot x_i + d \cdot \sum\limits_{j \neq i} (x_j - x_i)$.
- 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 $k \gg d$.

45. Decentralized Control Is Better

- We have $\dot{x}_i = -k \cdot x_i + d \cdot \sum\limits_{j \ne i} (x_j - x_i)$.
- The solution takes the form

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

$$\frac{\Delta_i}{n} \cdot \exp(-k \cdot t).$$

- For centralized control, we have $x_i(t) = \Delta_i \cdot \exp(-k \cdot t)$.
- Asymptotically, for large t, the term $\exp(-(k+d \cdot n) \cdot t)$ decreases much faster than $\exp(-k \cdot t)$.
- Thus, decentralized control is better.

46. Decentralized Control Is Also Better In More Realistic Situations

- We considered the case when only one of the variables deviates from the ideal state.
- In education terms, only one student lags behind, while all other students show perfect knowledge.
- What if several students lag behind – and other students are ahead?
- It is reasonable to assume that $x_i(0)$ are random and independent; then, for $s(0) \approx \Delta_i/\sqrt{n}$, we have

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

$$s(0) \cdot \exp(-k \cdot t).$$

- For large n, we have $\sqrt{n}$ times smaller deviation than for centralized case, when $x_i(t) \sim \Delta_i \cdot \exp(-k_i \cdot t)$.

47. How to Apply These Results to Education

- In education, decentralization means that students should also teach each other.
- This idea is well known in pedagogy, e.g., in collaborative learning.
- In Computer Science, a similar idea is known as *pair programming*, where several students help each other.
- It was noticed that students themselves like teamwork, especially the new generation, "Millennials".
- While the collaborative learning methods are actively used, they are only used on the *qualitative* level.
- To use the advanced control techniques, we need to use these methods on *quantitative* level.

48. How Can We Use These Methods on Quantitative Level?

- Students should effectively help each other.
- For that, they should know where other students stand.
- This may seem a return to pre-privacy times when all the grades were publicly posted.
- However, what we propose is different.
- The main drawback of the old system was that usually, it used to report an overall grade on a test.
- The old system encouraged competition.
- Instead of encouraging *competition*, we want to encourage *collaboration*.
- We propose to post level of knowledge of each student on each topic.

49. How Can We Use These Methods on Quantitative Level (cont-d)

- We propose to post level of knowledge of each student on each topic.
- So, students will be able to team together and improve their knowledge in all the topics.
- Students with deficiencies in some areas will benefit from help – and will help others in other topics.
- Of course, as every other collaboration, this cannot be forced.
- We need to convince students that this idea works.
- A simple mathematical model presented in this talk is one of the ways to convince students and instructors.

Degree-Based Ideas and Techniques Can Facilitate Inter-Disciplinary Collaboration and Education

50. Need for Interdisciplinary Collaboration and Education

- *Need for collaboration:* a successful computational research requires an intensive collaboration between
  - domain scientists – who provide the necessary information and metadata, and
  - computer scientists who provide the corresponding computations.
- Moreover:
  - since we combine data obtained by different subdomains,
  - we also need collaboration between representatives of these subdomains.
- *Need for education:* for the collaboration to be successful, we need to *educate* each other.

51. Inter-Disciplinary Collaboration and Education: Typical Communication Situations

- *Collaboration:*
  - *Situation:* a computer scientist has a new idea on how to better organize the geosciences' data.
  - This idea, if properly understood and jointly implemented, can benefit the geosciences.
  - How to convey the computer science idea to a geoscientist?
- *Education:*
  - *Situation:* a computer scientist wants to teach, to a geoscientist, a few existing computer science ideas.
  - In the long run, this will benefit the geosciences.
  - How to convey the computer science idea to a geoscientist?

52. First Possibility: Just Convey This Idea in Computer Science Terms

- *Idea:* simply describe this idea in computer science terms.
- *Problem:* many of these terms are usually very specific.
- Even many computer scientists may be not very familiar with these terms.
- *The only serious way* for a geoscientists to understand these terms is to take several CS courses.
- *It is unrealistic* to expect such deep immersion in routine inter-disciplinary collaboration.

53. Second Possibility: Try to Illustrate This Idea in the Domain Science Terms

- *Alternative approach:* explain the idea on the example of a geosciences problem.
- *Problem:* a computer scientist is usually not a specialist in geosciences.
- *Result:* his/her description of the problem is, inevitably, flawed: e.g., oversimplified.
- *Consequence:* the problem as described is often not meaningful to a geoscientist.
- Since the motivation is missing, it is difficult to understand the idea.

54. A Fight Club

- As a result of the above problems, our weekly meetings were, for a while, not very productive.
- For a while, they turned into what we called "fight club", when
  - a geoscientist would find flaws in a geosciences model used by a computer scientist to describe the ideas;
  - a computer scientist would find flaws in the way a geoscientist would describe his/her problem.
- And then we, serendipitously, found a solution to our struggles.
- After we found this solution, we started thinking why it worked.
- And we discovered an explanation via the matter-of-degree ideology.

55. Our Successful Empirical Approach to the Inter-Disciplinary Collaboration Problem

- *What did not work:* trying to describe ideas in purely computer science terms or on a geosciences example.
- *New approach:* described these ideas by their application to a complete different area: solar astronomy.
- *Fact:* none of us is a specialist in solar astronomy.
- *Result:* this description was inevitably less technical – and therefore, much more understandable.
- *Result:* we got a much better understanding of the original computer science idea.
- *Recommendation:* illustrate a message on the domain in which both parties have equal knowledge.

56. Explanation in Terms of Degrees: General Case

- Let d_1, d_2 denote participants' degrees of knowledge.
- In principle, there are $\max(d_1, d_2)$ levels at which at least one participant has a correct understanding.
- Among these levels, only at $\min(d_1, d_2)$ levels, there is a mutually correct understanding.
- Knowledge is more or less uniformly distributed across different levels of sophistication.
- Thus:
  - of all the correct statements that could be used by one of the participants,
  - the fraction of those that will be correctly understood by both participants is equal to $d = \dfrac{\min(d_1, d_2)}{\max(d_1, d_2)}$.
- *Conclusion:* this ratio is the largest when $d_1 = d_2$.

57. An Alternative Idea: Using an "Interpreter"

- *Alternative idea:* use an "interpreter", who has a reasonable understanding in both fields.
- *First:* a describer uses the terms of his/her domain to convey the idea (or problem) to the interpreter.
- In this transaction, the degree of understanding $d_{1i} = \dfrac{\min(d_1, d_i)}{\max(d_1, d_i)}$ is reasonably high.
- *Second:* the interpreter translates the message into the respondent's domain and informs the respondent.
- Here, also, the degree of understanding d_{i2} is reasonably high.
- This strategy, by the way, works well too.
- We hope that the above formulas will help to optimize this approach as well.

Stimulating Students by Explaining Motivations Behind Concepts and Ideas

58. Stimulating Students by Explaining Motivations Behind Concepts and Ideas

- Often, students do not understand why the material is important.
- This is especially true in mathematics.
- Good teachers explain applications, so students understand the need to use formulas.
- However, they still do not understand why we need proofs – the essence of mathematics.
- Good news: we do not have to invent new reasons why proofs are important.
- There are reasons why rigorous mathematics was designed in the first place.
- What we need is convincingly convey these reasons to students of mathematics.

59. Paradoxes an as Explanation of Why Proofs Are Needed

- *Main reason for rigor:* otherwise we get paradoxes.
- *Historically first mathematical paradox:* heap paradox.
- *Interesting:* this paradox became one of the motivations for fuzzy logic.
- *Why ε-δ definitions in calculus:* otherwise problems with series like $1 + (-1) + 1 + (-1) + \ldots$:

  $$(1+(-1))+(1+(-1))+\ldots = 0 \neq 1 = 1+(-1)+1)+\ldots$$

- *What is needed:* explain paradoxes to the students before explaining the new material.
- *Fuzzy can help:* because fuzzy logic provides a natural explanation of these paradoxes.

How to Make Sure that the Grading Scheme Encourages Students to Learn All the Material: Fuzzy-Motivated Solution and Its Justification

60. Formulation of the Problem

- The material taught in a typical semester-long class consists of *several parts*.
- In many cases, it is important that a student gets reasonable knowledge of *all* the *parts* of the material.
- For example, we want a medical doctor to have basic knowledge of *all* types of diseases.
- It is desirable that the grading scheme:
  - not only gauge how well the students learn the material;
  - the grading scheme should also encourage the students to learn *all* the parts of the material.

61. Towards a Formal Description of How a Student Plans His or Her Studies

- A student has a limited time t that can be allocated to learning the material.
- The student must select, for each part $i = 1, 2, \ldots, n$, the time $t_i \geq 0$ allocated for studying this part, so that

$$t_1 + t_2 + \ldots + t_n = t.$$

- The student's knowledge can be gauged by a proportion of the material that the student learned.
- Let us assume that for each $t \geq 0$, we know the amount of knowledge $a(t)$ learned after study time t.
- For i-th part of the material, we have a grade $a_i = a(t_i)$.
- We need to select a method F to combining grades a_i into an overall grade:

$$a = F(a_1, \ldots, a_n).$$

62. The Problem Reformulated in Precise Terms

- *How:* a student allocates times t_i, $\sum t_i = t$, so as to maximize his/her overall grade $F(a(t_1), \ldots, a(t_n))$.
- *Situation:* we want the student to achieve level $\geq a_0$ in all topics.
- *We want to select* $F(a_1, \ldots, a_n)$ so that:
  - if it is possible to find time allocation for which $a(t_i) \geq a_0$ for all i,
  - then the allocation selected by the student will satisfy this property.
- *Usually:* the overall grade is computed as the weighted average of grades a_i:

$$F(a_1, \ldots, a_n) = \sum_{i=1}^{n} w_i \cdot a_i.$$

- *In this case:* selecting F means selecting weights.

63. The Desired Property Is Not Always Satisfied for the Current Grading System: Example

- *We want* the same level of knowledge a_0 for all parts of the material.
- Thus, it is reasonable to take *equal weights* $w_i = 1/n$.
- *E.g.:* a (steep) learning curve: $a(t) = t^2$ when $t \le 1$.
- *Ideal case:* a student spends time t/n on each topic.
- If $(t/n)^2 \ge a_0$, we get good knowledge on all topics.
- *Resulting grade:* the overall grade is $(t/n)^2$.
- *Another strategy:* spend time 1 on each of t topics and 0 on all $n - t$ others.
- *Result:* perfect knowledge $1 > a_0$ on selected t topics, no knowledge $0 < a_0$ of others.
- *Resulting grade:* $\dfrac{1 \cdot t + 0 \cdot (n - t)}{n} = \dfrac{t}{n}$.

64. The Desired Property Is Not Always Satisfied for the Current Grading System (cont-d)

- *Reminder:* we have two strategies:
  - in the first, the student gets good knowledge of all topics, and grade $(t/n)^2$;
  - in the second, the students gets no knowledge of some topics, and grade t/n.
- *Problem:* since $t/n < 1$, we have $(t/n)^2 < (t/n)$.
- *Conclusion:* students prefer the new strategy to the ideal one.
- *Result:*
  - even when the students have resources to attain good knowledge of all topics,
  - the grading system discourages such learning.

65. Heuristic Idea Motivated by Fuzzy Logic

- We want the student to know:
  - the 1st part of the material *and*
  - the second part *and* ...
  - the n-th part.
- For each i, we know the degree a_i to which the student knows the i-th part of the material.
- Thus, according to fuzzy methodology, we should applying a fuzzy "and"-operation (t-norm) to degrees a_i.
- A natural requirement that $F(a_1, a_1) = a_1$ is satisfied only by one fuzzy "and"-operation: $\min(a_1, a_2)$.
- If we use this "and"-operation, we get the grading scheme

$$a = F(a_1, \ldots, a_n) = \min(a_1, \ldots, a_n).$$

66. The New Grading Scheme Is Better for the Above Example

- *Ideal strategy:* the student spends time t/n on each topic, gaining knowledge $a_1 = \ldots = a_n = (t/n)^2$.
- *Resulting overall grade:*

$$F(a_1, \ldots, a_n) = \min(a_1, \ldots, a_n) = (t/n)^2.$$

- *Alternative strategy:* the student spends time 1 on each of n topics and time 0 on all other topics.
- *Resulting knowledge:* $a_1 = \ldots = a_t = 1$, $a_{t+1} = \ldots = a_n = 0$.
- *Resulting overall grade:*

$$F(a_1, \ldots, a_n) = \min(a_1, \ldots, a_n) = \min(1, \ldots, 1, 0, \ldots, 0) = 0.$$

- *Conclusion:* students will now prefer to attain good knowledge of all topics.

67. What We Do in This Talk

- In this talk, we show the above-described behavior of the min grading scheme is not accidental.
- First, we prove that:
  - if we use the fuzzy-motivated min grading scheme,
  - then the student would always prefers to equally distribute effort between different topics.
- This is exactly what we want to achieve.
- Second, we prove that min grading scheme is the only one for which students study as desired.
- To describe these results in precise terms, let us first define the problem in precise terms.

68. Formal Definitions

- We say that a function $a(t_1, \ldots, t_n)$ is *(non-strictly) increasing* if $t_1 \le t'_1$, ..., and $t_n \le t'_n$ imply
$$a(t_1, \ldots, t_n) \le a(t'_1, \ldots, t'_n).$$
- By a *learning curve*, we mean a continuous increasing function $a(t) : \mathbb{R}_0 \to [0, 1]$.
- We say that a function $F(a_1, \ldots, a_n)$ is *idempotent* if for every a, we have $F(a, \ldots, a) = a$.
- For $n \ge 2$, by a *n-grading scheme*, we mean a continuous non-strictly increasing idempotent function
$$F : [0, 1]^n \to [0, 1].$$
- Let $t > 0$ and $n \ge 2$. By a (t, n)*-learning strategy*, we mean a tuple of values $t_1 \ge 0, \ldots, t_n \ge 0$ for which
$$t_1 + \ldots + t_n = t.$$

69. Formal Definitions (cont-d)

- Let $\mathcal{S}$ be a set of (t, n)-learning strategies, and let $(t_1, \ldots, t_n) \in \mathcal{S}$.
- We say that the learning strategy is *uniformly a_0-successful* if $a(t_i) \geq a_0$ for all i.
- By an *overall grade*, we mean the value $F(a(t_1), \ldots, a(t_n))$.
- We say that the learning strategy is *$(\mathcal{S}, F)$-optimal* if its overall grade is $\geq$ than for all other strategies $\in \mathcal{S}$.
- We say that a grading scheme *encourages students to learn all the material* if for every $a(t)$, t, a_0, $\mathcal{S}$,
  - if, in the set $\mathcal{S}$, there exists a uniformly a_0-successful (t, n)-learning strategy,
  - then every $(\mathcal{S}, F)$-optimal (t, n)-learning strategy is uniformly a_0-successful.

70. Main Result

Theorem. *For every integer $n \geq 2$:*

- *the* min *grading scheme*

$$F(a_1, \ldots, a_n) = \min(a_1, \ldots, a_n)$$

  encourages students to learn all the material;

- *vice versa, if an n-grading scheme $F(a_1, \ldots, a_n)$ encourages students to learn all the material, then*

$$F(a_1, \ldots, a_n) = \min(a_1, \ldots, a_n).$$

71. Resulting Recommendations Are Not That Unusual

- *Resulting recommendation:* an overall grade for the class is the smallest of the grades for each module.
- *At first:* this may sound like a very radical idea.
- *However:* it is in line with what is usually done.
- *Example:* in our university, for a student to pass Calculus I, s/he need to pass *every* module.
- *This* corresponds to minimum.
- *In* some *computer science* classes, the student has to pass *both* the tests and the labs.
- *Similarly*, to get a degree:
  - it is not sufficient for a student to have a good GPA,
  - the student must get satisfactory grades on *all* required classes.

72. Proof of the Theorem: Part 1

- Let us first prove that the min grading scheme encourages students to learn all the material, i.e., that
  - if there exists a uniformly a_0-successful (t, n)-learning strategy,
  - then every min-optimal learning strategy is uniformly a_0-successful.
- Indeed, for a uniformly a_0-successful strategy, by definition, we have $a_i = a(t_i) \geq a_0$ for all i.
- Thus, the overall grade $a = F(a_1, \ldots, a_n) = \min(a_1, \ldots, a_n)$ corresponding to this strategy is also $a \geq a_0$.
- For the optimal strategy s, the grade is $\geq a$ thus $\geq a_0$:

$$\min(a(t_1), \ldots, a(t_n)) \geq a_0.$$

- $\forall i : a(t_i) \geq \min(a(t_1), \ldots, a(t_n))$, so $a(t_i) \geq a_0$ – i.e., the strategy s is indeed uniformly a_0-successful.

73. Part 2: Reduction to Case $a_i > 0$

- Let us now assume that a grading scheme $F(a_1, \ldots, a_n)$ encourages students to learn all the material.
- Let us prove that $F(a_1, \ldots, a_n) = \min(a_1, \ldots, a_n)$.
- It is sufficient to prove the above formula for the case when all the values a_i are positive.
- Indeed:
  - once we prove this formula for all positive a_i,
  - we can use continuity to extend it to the case when some of the values a_i are equal to 0.
- In view of this observation, in the remaining part of this proof, we will assume that $a_i > 0$ for all i.

74. Part 2, Lemma 2

- Let us prove that for all $m > 0$, $\varepsilon \in (0, m)$, and i:

$$F(1, \ldots, 1\ (i-1 \text{ times}), m-\varepsilon, 1 \ldots, 1) < m.$$

- Let us take $a_0 = m$ and a piece-wise linear f-n $a(t)$ s.t.:

$$a(0) = 0,\ \ a(1-\varepsilon) = m-\varepsilon,\ \ a(1) = m,\ \ a\left(1 + \frac{\varepsilon}{n-1}\right) = 1.$$

- For $t_i = 1$, we get $a(t_1) = \ldots = a(t_n) = m \geq a_0$.
- For this successful strategy, grade is $F(m, \ldots, m) = m$.
- For $t'_i = 1 - \varepsilon$ and $t'_j = 1 + \dfrac{\varepsilon}{n-1}$ for $j \neq i$, $t'_1 + \ldots = t$, $a(t'_i) = a(1-\varepsilon) = m - \varepsilon < m$, $a(t'_j) = 1$, and grade is

$$F(1, \ldots, 1\ (i-1 \text{ times}), m-\varepsilon, 1 \ldots, 1).$$

- For a student to prefer the successful strategy, this grade must be $< m$. Q.E.D.

75. Part 2 (cont-d)

- We know: $F(1,\dots,1\ (i-1 \text{ times}), m-\varepsilon, 1\dots,1) < m$.
- In the limit $\varepsilon \to 0$, we get

$$F(1,\dots,1\ (i-1 \text{ times}), m, 1\dots,1) \le m.$$

- For any a_i, let us denote $m = \min(a_1,\dots,a_n)$, and let i be the index for which $a_i = m$.
- By monotonicity, $F(a_1,\dots,a_{i-1},a_i,a_{i+1},\dots,a_n) \le$

$$F(1,\dots,1\ (i-1 \text{ times}), a_i, 1,\dots,1) =$$

$$F(1,\dots,1\ (i-1 \text{ times}), m, 1,\dots,1) \le m.$$

- Similarly, since $m = a_i \le a_j$ for all j, by monotonicity:

$$m = F(m,\dots,m) \le F(a_1,\dots,a_{i-1},a_i,a_{i+1},\dots,a_n).$$

- These two inequalities prove that

$$F(a_1,\dots,a_{i-1},a_i,a_{i+1},\dots,a_n) = m = \min(a_1,\dots,a_n).\ \text{Q.E.D.}$$

What is Wrong with Teaching to the Test: Uncertainty Techniques Help in Understanding the Controversy

76. What Is “Teaching to the Test”?

- In the last few decades, in the US school education, state-wide math tests have been developed.
- Student performance on these tests is very important:
  - Funding of individual schools is largely determined by the test results.
  - Schools are disbanded and teachers are fired if the test results are unsatisfactory several years in a row.
- So schools make sure that the students pass these tests.
- As a result:
  - instead of spending most of time teaching the material – as it was in the past –
  - teachers now spend a significant amount of time teaching “to the test”.

77. The Results of Teaching to the Test Are Not As Spectacular As the Proposers Hoped

- The main idea behind the tests sounds reasonable:
  - if we do not gauge how well students are doing,
  - then how will we know which schools are doing better and which schools need improvement?
- The authors of this idea expected that with testing, the students' knowledge will drastically improve.
- Alas, these expectations turned out to be too optimistic:
  - In some states and some school districts, there has been some improvement.
  - However, overall, this program has not been a spectacular success as its proponents hoped.
  - In some cases, with the introduction of state-wide testing, the students' knowledge actually decreased.

78. Teaching to the Test: A Current Controversy

- On the one hand, many politicians believe that tests are a good idea.
- On the other hand, most teachers believe that the entire approach is flawed.
- In the media, this controversy gets personal and nasty:
  - politicians accuse the teacher community of defending weak under-preforming teachers;
  - teachers accuse politicians of ignorance-motivated interference with a complex teaching process.
- The situation is more complex than the simplified media picture:
  - several knowledgable politicians, with successful teaching experience, are in favor of the tests;
  - many very good teachers are strongly against the current emphasis on these tests.

79. Population Is Somewhat Confused

- One of the frustrating aspects of the current controversy is that the general population is confused.
- On the one hand:
  - it is reasonable to require accountability, and
  - this accountability logic naturally leads to the current testing program.
- On the other hand:
  - respected teachers are against this program, and
  - empirical evidence also shows that it has not led to spectacular successes –
  - contrary to natural expectations motivated by accountability.

80. What We Do in This Talk

- In this talk, we argue that:
  - the confusion – and, to some extent, the controversy itself –
  - is largely due to the simplification of the complex pedagogical process.
- Specifically, we argue that:
  - if properly take uncertainty into account,
  - then the situation becomes much clearer.

81. The Background of Our Main Idea

- In general, it is assumed that learning comes from repetitions:
  - once a student has repeated a certain procedure certain number of times,
  - the student have mastered it.
- This is why an important part of learning each idea of high school mathematics is practice. For example:
  - unless students do a lot of exercises where they have to add fractions,
  - they will master this skill well enough to be able to easily add two fractions, and
  - this will hinder their progress in the following mathematical topics like dealing with polynomials.

82. The Background of Our Main Idea (cont-d)

- In general:
  - the only way to learn to write is to practice writing,
  - the only way to learn a foreign language is to practice it, etc.
- The required number of repetitions depends:
  - on the complexity of the topic,
  - on the match between this particular topic and the student's individual interests and prior skills, etc.
- However, the fact remains:
  - for every topic and for every student,
  - there is a number of iterations after which the student will master this topic.
- From this viewpoint, let us analyze both the traditional teaching process and teaching to the test.

83. Analysis of the Traditional Teaching Process

- The main objective of school math is that after graduation, students should have certain skills.
- These skills often build on each other, so that one skill requires another one.
- For example, to be able to solve quadratic equations, we need to know how to add, how to subtract, etc.
- Let us consider two skills A and B, s.t. B requires that the student also have learned skill A.
- Let us assume that the student needs n_A iterations to master skill A, and n_B iterations to master skill B.
- Let us denote by r the proportion of problems of type B that involve using skill A.
- Then, during n_B exercises needed to master skill B, the student, in effect, performs $r \cdot n_B$ exercises of A.

84. Analysis of the Traditional Teaching Process (cont-d)

- *Reminder:*
  - during n_B exercises needed to master skill B,
  - the student, in effect, performs $r \cdot n_B$ exercises of skill A.
- *Corollary:* it is sufficient to have $n_A - r \cdot n_B$ exercises in skill A in Year 1.
- *Fact:* this number $n_A - r \cdot n_B$ is smaller than n_A.
- *Corollary:* by the end of Year 1, the students have not yet fully mastered skill A.
- *Comment:* this is normal in education – the skills come with practice.

85. How Situation Changes When We Teach to the Test

- According to the school program, Year 1 is devoted to teaching skill A.
- We want to test how well the students learned after this year.
- However, by the end of Year 1, the students only had $n_A - r \cdot n_B < n_A$ exercises.
- So, they have not yet mastered the skill A.
- The argument "Is this how much we want our graduates to know about A?" sounds convincing.
- So, a pressure is placed on schools to improve the score on the test at the end of Year 1.
- The only way to do it is to increase the number of skill-A-related exercises in Year 1 to n_A.

86. Teaching to the Test: A Seemingly Positive Result

- The test grades for Year 1 go up – because:
  - in the past, the students did not have enough exercises to master skill A, while
  - now, they have enough exercises, so they do master skill A at the end of Year 1.
- The progress is visible, results are good.
- But are they?

87. Teaching To The Test: School Graduates Knowledge

- The main school objective – to make sure that the graduates learn both skills A and B.
- Let us show that with respect to this criterion, we should not expect any significant improvement.
- Indeed:
  - in the past, we had a total of n_A exercises in skill A:
  - now, the students have $n_A + r \cdot n_B$ exercises in skill A.
- In both cases, we have enough exercises to master skill A.
- So, in both cases, we should have the same reasonably positive result.

88. Teaching to the Test: A Serious Problem

- The problem is that school time is limited.
- Schools have additional $r \cdot n_B$ repetitions of skill A in Year 1.
- This time has to come at the expense of something else.
- Clearly, it comes at the expense of other topics that are not explicitly included in the statewide test.
- As a result,
  - while students' knowledge of the topics included in the test (like skills A and B) does not decrease,
  - the students' mastery of some other skills will necessarily drastically decrease.
- This is what teachers object to when they object to "teaching to the test".

89. We Clarified the Problem – but What Is a Solution?

- In order to compare different schools & teachers, we need to gauge the student success.
- In the ideal world, we should design better tests – this is one of the few things with which everyone agrees.
- However, even with the existing tests, we can drastically improve the situation if we *no longer require* that
  - at the end of each school year,
  - students should have a perfect knowledge of all the topics that they learned during this year.
- This requirement comes from the "crisp" thinking.
- This thinking that does not take uncertainty into account – a student either mastered the skill or did not.

90. Towards a "Fuzzy" Solution

- In reality, after a few exercises of the skill A, a student usually achieves mastery *to a degree*.
- As a result, in the traditional approach, the student will have an imperfect score on A at the end of Year 1.
- This is OK, as long as this score is what we should expect after $n_A - r \cdot n_B$ exercises, so that: that
  - after additional $r \cdot n_B$ exercises involving skill A in Year 2
  - the student will achieve the true mastery of skill A.
- Any increase of this satisfaction level should be *discouraged* because
  - it would indicate that the teachers are over-emphasizing skill A in Year 1, while
  - they could use fewer exercises of A and spend this time teaching the students some other useful skills.

91. How Fuzzy Logic Can Help

- Fuzzy logic has been explicitly designed to handle situations in which some property is true to a degree.
- This is exactly the situation that we have encountered.
- So, fuzzy logic seems to be a perfect tool for this analysis.

92. Our Idea Is More General than Teaching-to-the-Test Controversy

- Our main objective is to help in understanding and resolving the "teaching to the test" controversy.
- However, the same idea can be applied to all levels of education as well.
- We should not aim for perfect knowledge on intermediate classes.
- For example, college students taking a computer science sequence:
  - may be somewhat shaky about programming at the end of the first class,
  - but their basic skills are reinforced in the following classes.
- We used this idea in our previous research to plan an optimal teaching schedule, and it worked.

Interval and Fuzzy Techniques in Assessment

93. Assessment is Important

- *Objective:* improve the efficiency of education.
- *Important:* to assess this efficiency, i.e., to describe this efficiency in quantitative terms.
- This is important on all education levels:
  - elementary schools
  - middle schools
  - high schools
  - universities
- Quantitative description is needed because
  - it allows natural comparison of different strategies of teaching and learning
  - and selection of the best strategy.

94. Need for Value-Added Assessment

- *Traditional assessment:* by the amount of knowledge that the students have after taking this class.
- *Example:* the average score of the students on some standardized test.
- *Comment:* this is actually how the quality of elementary/high school classes is now estimated in the US.
- *Limitation:* the class outcome depends
  - not only on the quality of the class, but
  - also on how prepared were the students when they started taking this class.
- *A more adequate assessment* should estimate the *added value* that the class brought to the students.

95. Current Approaches to Value-Added Assessment and their Limitations

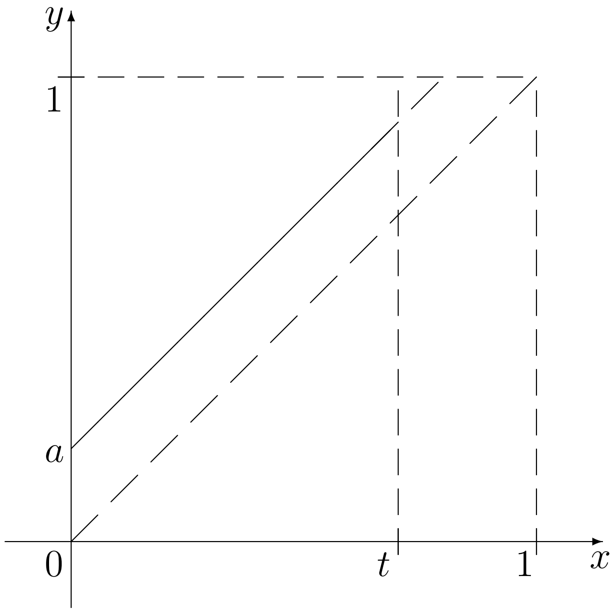
- *Main idea:* subtracting the outcome from the input.
- *Example:* subtract
  - the average grade after the class (on the post-test)
  - the average grade on similar questions asked before the class (on the pre-test).
- *Comment:* the existing techniques take into account additional parameters influencing learning.
- *Main limitation:* actually, the amount of knowledge learned depends on the initial knowledge.
- *Additional limitation:* the assessment values come from grading, and are therefore somewhat subjective.

96. Natural Idea: Using Interval and Fuzzy Techniques

- *Reminder:* assessments are subjective.
- *Conclusion:* it is natural to use interval and fuzzy techniques to process the corresponding values.
- *In this talk:* we describe how to the use fuzzy techniques.
- *Result:* interval and fuzzy techniques help us overcome both limitations of the existing value-added assessments.

97. Traditional Approach: Reminder

- *Reminder:* the post-test result y depends on the pre-test result x as $y \approx x + a$:

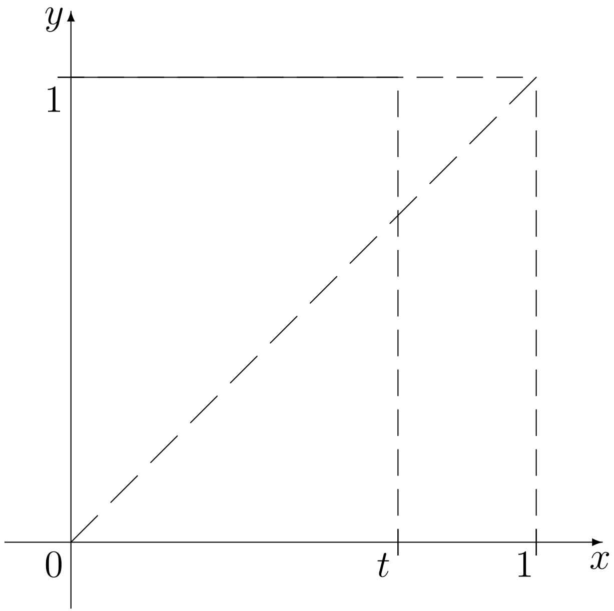


98. Linear Dependence instead of Addition: Idea

- *Problem:* the difference $y - x$ actually changes with x.
- *Natural next approximation:* $y \approx m \cdot x + a$.
- *Observation:* for f-s $f_1(x) = m_1 \cdot x + a_1$ and $f_2(x) = m_2 \cdot x + a_2$ corr. to two teaching strategies, we may have
  - $f_1(x_1) < f_2(x_1)$ for some x_1 and
  - $f_1(x_2) > f_2(x_2)$ for some $x_2 > x_1$.
- *Interpretation:*
  - for weaker students, with prior knowledge $x_1 < x_2$, the second strategy is better, while
  - for stronger students, with prior knowledge $x_2 > x_1$, the first strategy is better.
- *Conclusion:* the new model provides a more nuanced comparison between different teaching strategies.

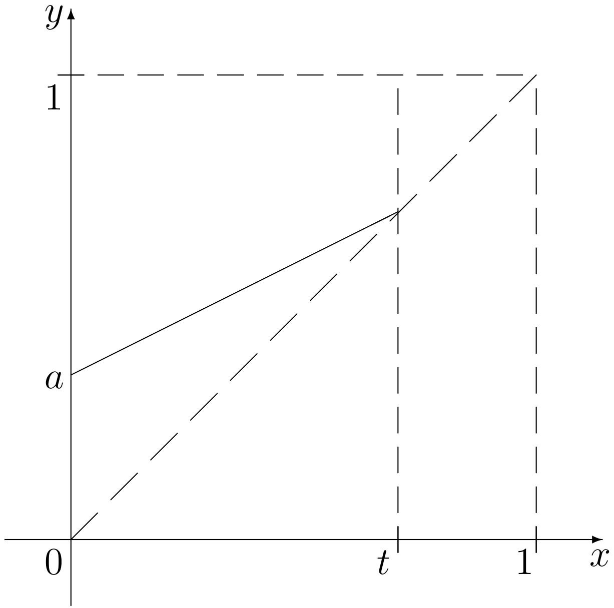
99. Ideal Case: Perfect Learning

- *Ideal case:* no matter what the original knowledge is, the resulting knowledge is perfect, $y \equiv 1$; then $m = 0$.



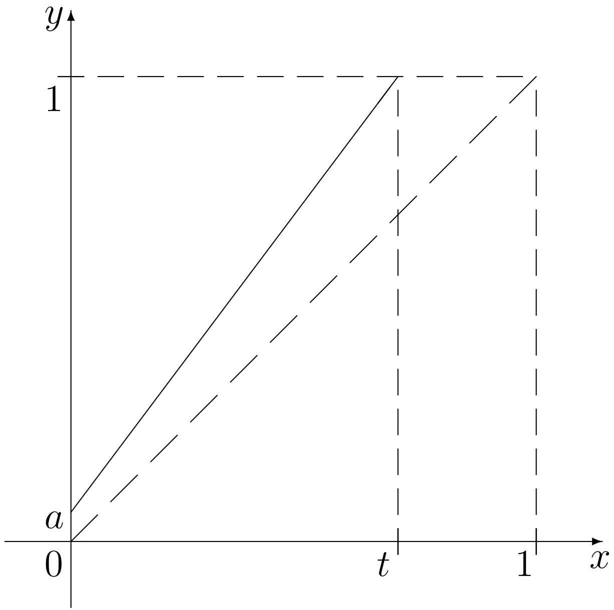
100. Example 2: Minimizing Failure Rate

- *Main idea:* to avoid failure, we concentrate on the students with low x; then $f(x) = m \cdot x + a$, with $m < 1$.



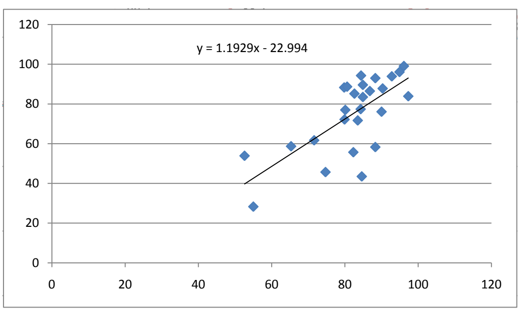
101. Example 3: Emphasis on Strong Students

- *Idea:* concentrate most of the effort on top students.
- *Result:* $f(x) = m \cdot x + a$, with $m > 1$.



102. How to Determine the Coefficients m and a: Ideal Case of Crisp Estimates

- *We know:* pre-test grades $x_1, \ldots, x_n$ and post-test grades $y_1, \ldots, y_n$.
- *Problem:* find m and a for which $y_i \approx m \cdot x_i + a$.
- *Least Squares method:* $\sum_{i=1}^{n} (y_i - (m \cdot x_i + a))^2 \to \min_{m,a}$.



103. Case of Interval Uncertainty: Analysis

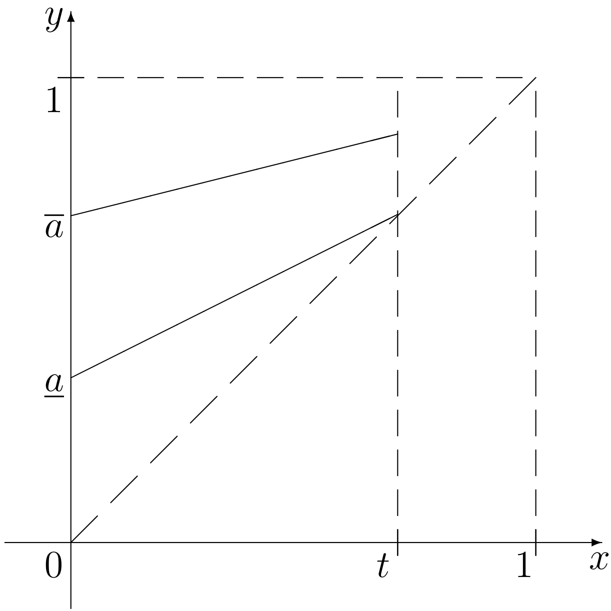
- *Fact:* the grade depends on assigning partial credit for partly correct solutions.
- *Known:* partial credit is somewhat subjective.
- *How to avoid this subjectivity:* letter grades such as A (corresponding to 90 to 100) are more objective.
- *Conclusion:* instead of the exact grade x_i, we have an interval $\mathbf{x} = [\underline{x}_i, \overline{x}_i]$ of possible grades.
- *Value-added assessment:* describe the dependence $\mathbf{y} = f(\mathbf{x})$ of the outcome grade $\mathbf{y}$ on the input grade $\mathbf{x}$:
  - we consider all the students for whom the input grade is within the interval $\mathbf{x}$;
  - then, $\mathbf{y} = f(\mathbf{x})$ is the set of all possible outcome grades for these students.

104. Which Interval-to-Interval Functions Are Reasonable

- *Example:* suppose that
  - when the pre-test grade x is in $\mathbf{x}_1 = [80, 90]$, then the post-test grade y is in $\mathbf{y}_1 = f(\mathbf{x}_1) = [85, 95]$;
  - when $x \in \mathbf{x}_2 = [90, 100]$, then $y \in \mathbf{y}_2 = f(\mathbf{x}_2) = [92, 100]$.
- *Argument:* when $x \in \mathbf{x}_1 \cup \mathbf{x}_2$, then $x \in \mathbf{x}_1$ or $x \in \mathbf{x}_2$, so $y \in \mathbf{y}_1$ or $y \in \mathbf{y}_2$.
- *Conclusion:* $f(\mathbf{x}_1 \cup \mathbf{x}_2) = f(\mathbf{x}_1) \cup f(\mathbf{x}_2)$.
- *Similar conclusion:* $f(\mathbf{x}) = \bigcup\limits_{x \in \mathbf{x}} f([x, x])$.
- *Notation:* $[\underline{f}(x), \overline{f}(x)] \stackrel{\text{def}}{=} f([x, x])$.
- *Result:* all reasonable functions $f(\mathbf{x})$ have the form $f([\underline{x}, \overline{x}]) = [\underline{y}, \overline{y}]$, where $\underline{y} \stackrel{\text{def}}{=} \min\limits_{x \in [\underline{x}, \overline{x}]} \underline{f}(x)$; $\overline{y} \stackrel{\text{def}}{=} \max\limits_{x \in [\underline{x}, \overline{x}]} \overline{f}(x)$.

105. Case of Interval Uncertainty: Algorithm

- *Idea:* based on $[\underline{x}_i, \overline{x}_i]$ and $[\underline{y}_i, \overline{y}_i]$, we use Least Squares to find values s.t. $\underline{y}_i \approx \underline{m} \cdot \underline{x}_i + \underline{a}$ and $\overline{y}_i \approx \overline{m} \cdot \overline{x}_i + \overline{a}$.



106. Case of Fuzzy Uncertainty

- *Interval assumption:* we assumed that the interval $[\underline{x}, \overline{x}]$ is guaranteed to contain the actual (unknown) value x.
- *In reality:* the bounds that we know are "fuzzy", i.e., they contain x only with some degree of confidence α.
- *Conclusion:* we have different intervals $[\underline{x}(\alpha), \overline{x}(\alpha)]$ corresponding to different degrees α.
- *Observation:* this is equivalent to knowing a fuzzy set with given α-cuts $[\underline{x}(\alpha), \overline{x}(\alpha)]$.
- *Resulting algorithm:* for each α, we find the interval-values linear function

$$[\underline{m}(\alpha) \cdot x + \underline{a}(\alpha), \overline{m}(\alpha) \cdot x + \overline{a}(\alpha)]$$

corresponding to this α.

107. How to Use the Resulting Interval and Fuzzy Estimates to Compare Different Teaching Strategies

- From the input fuzzy grades $X_1, \ldots, X_n$, we extract α-cuts corresponding to their α-cuts $X_i(\alpha)$.
- We know input-output functions corresponding $f_j([\underline{x}, \overline{x}])$ corresponding to different strategies j.
- We apply these functions to intervals $X_i(\alpha)$ and get fuzzy estimates $Y_{1,j}, \ldots, Y_{n,j}$ for post-test results.
- For each j, we apply the objective function to values $Y_{1,j}, \ldots, Y_{n,j}$.
- Thus, we get the fuzzy estimate V_j of the quality of the j-th strategy.
- We then use fuzzy optimization techniques to select the teaching strategy with the largest value V_j.

Appendix: Tastle-Wierman (TW) Dissention and Consensus Measures and Their Potential Role in Education

108. Introduction

- In many practical situations, we have to use *expert estimates* to gauge the value of a quantity.
- Expert estimates $x_1, \ldots, x_n$ rarely agree exactly:
  - sometimes, the expert estimates mostly agree with each other, so we can say that they are in consensus;
  - sometimes, the expert estimates strongly disagree.
- It is thus desirable to come up with numerical measures of dissention and consensus.
- In education, traditionally the mean grade $\bar{x} \stackrel{\text{def}}{=} \frac{\sum\limits_{i=1}^{n} x_i}{n}$ is used to gauge the results.
- Mean grades are the same if everyone gets Cs or some student fail.
- We thus need to supplement the mean with a criterion of how similar the grades are.

109. Tastle-Wierman (TW) Dissention and Consensus Measures

- W. J. Tastle and M. J. Wierman define the measure of dissention $D(x)$ as the mean value of the quantity

$$-\log_2\left(1-\frac{|x_i-\bar{x}|}{d_x}\right),$$

where and $d_x \stackrel{\text{def}}{=} x^+ - x^-$ is the width of the interval $[x^-, x^+]$ of possible values of the estimated quantity:

$$D(x) \stackrel{\text{def}}{=} -\frac{1}{n}\cdot\sum_{i=1}^{n}\log_2\left(1-\frac{|x_i-\bar{x}|}{d_x}\right).$$

- A consensus is, intuitively, an opposite to dissention; so, a consensus measure $C(x)$ is

$$C(x) = 1 - D(x).$$

110. TW Dissention and Consensus Measures: Alternative Formulas

- Often, several experts come up with the same estimate.
- In this case, we have:
  - the estimates $x_1, \ldots, x_m$, and
  - the frequency $p_1, \ldots, p_m$ of experts who come up with these estimates.
- Here, the dissention formula can be reformulated as

$$D(x) = -\sum_{j=1}^{m} p_i \cdot \log_2 \left(1 - \frac{|x_j - \bar{x}|}{d_x}\right),$$

where

$$\bar{x} \stackrel{\text{def}}{=} \sum_{j=1}^{m} p_j \cdot x_j.$$

111. Remaining Problem and What We Do

+ Wierman and Tastle show that their measure capture the intuitive meaning of dissention and consensus.

– It is not clear, from their analysis, whether these are the only possible measures that capture this intuition.

– It is also not clear what other possible measures capture this same intuition.

+ In this talk, we show that the TW measures can be naturally derived from a fuzzy logic formalization.

+ We show that the TW measures appear if we use:

  – one of the simplest t-conorms – algebraic sum – and

  – one of the simplest membership functions – a triangular one.

+ We also explain what will happen if we use more complex t-conorms and/or membership functions.

112. How to Formalize the Intuitive Idea Behind Dissention

- *Ideal case of complete consensus:* all expert estimates $x_1, \ldots, x_n$ coincide; thus, $x_i = \bar{x}$.
- *Dissention* means that some x_i are different:

$$(x_1 \text{ is different from } \bar{x}) \vee \ldots \vee (x_n \text{ is different from } \bar{x}).$$

- According to the general fuzzy methodology, to assign a degree to this statement, we must do the following:
  - first, we should assign reasonable degrees $d_{\neq}(a, b)$ to statements of the type "a is different from b";
  - then, we should select an appropriate t-conorm ("or"-operation) $t_\vee(a, b)$;
  - finally, we compute

$$d(x) = t_\vee(d_{\neq}(x_1, \bar{x}), \ldots, d_{\neq}(x_n, \bar{x})).$$

113. Let Us Use the Simplest Possible Techniques

- One of the general ideas of using fuzzy methodology is that:
  - out of all possible techniques which are consistent with our intuition,
  - we should use the computationally simplest techniques.
- Indeed, if a simple formula already captures the meaning, there is no sense in using more complex formulas.
- If our knowledge is well described by a triangular membership function, why use a more complex one?
- If our understanding of an "and"-operation is captured by $t_{\&}(a,b) = a \cdot b$, why use more complex t-norms?

114. Selecting a Membership Function $d_{\neq}(a, b)$

- First idea: $a \neq b$ if and only if $c \stackrel{\text{def}}{=} |a - b| \neq 0$.
- Thus, for a membership function $\mu_{\neq 0}(c)$, we have

$$d_{\neq}(a, b) = \mu_{\neq 0}(|a - b|).$$

- For $c = 0$, the statement "$c \neq 0$" is false, so $\mu_{\neq 0}(0) = 0$.
- For $a, b \in [\underline{x}, \overline{x}]$, the largest possible distance $c = |a - b|$ is $c = \overline{x} - \underline{x} = d_x$.
- It therefore makes sense to set $\mu_{\neq 0}(d_x) = 1$.
- Thus, the desired triangular membership function is

$$\mu_{\neq 0}(c) = \frac{c}{d_x}.$$

- Hence $d_{\neq}(a, b) = \mu_{\neq 0}(|a - b|) = \dfrac{|a - b|}{d_x}$.

115. Selecting the t-Conorm: First Try

- Computationally, the simplest t-conorm is the maximum $t_\vee(a,b)=\max(a,b)$.
- Let us consider two situations with the same range $[x^-,x^+]=[-1,1]$ (and $d_x=x^+-x^-=2$):
  1. half of the experts selected 1 and half -1;
  2. one expert selected 1, one -1, and all other experts selected 0.
- In both cases, the mean is $\bar{x}=0$, so $d_{\neq}(\pm 1,0)=0.5$ and $d_{\neq}(0,0)=0<0.5$. Thus, in both cases,

$$t_\vee\left(\frac{|x_1-\bar{x}|}{d_x},\ldots,\frac{|x_n-\bar{x}|}{d_x}\right)=\max(0.5,\ldots)=0.5.$$

- The resulting degrees are the same, but:
  - in the first case, there is a "maximal" dissention;
  - in the second case, only two experts disagree.

116. Selecting t-Conorm, Resulting Formula, and Its Relation to TW Measures

- *Reminder:* $d(x) = t_\vee(d_{\neq}(x_1, \bar{x}), \ldots, d_{\neq}(x_n, \bar{x}))$, with
$$d_{\neq}(x_i, \bar{x}) = \frac{|x_i - \bar{x}|}{d_x}.$$
- $t_\vee(a, b) = \max(a, b)$ is not adequate.
- *Conclusion:* use the next simplest t-conorm $t_\vee(a, b) = a + b - a \cdot b$:
$$d(x) = t_\vee\left(d_{\neq}\left(\frac{|x_1 - \bar{x}|}{d_x}, \ldots, \frac{|x_n - \bar{x}|}{d_x}\right)\right).$$
- *Relation with TW's $D(x)$:* $D(x) = -\dfrac{1}{n} \cdot \log_2(1 - d(x))$.
- *Proof:* uses $\log_2(1 - t_\vee(a, b)) = \log_2(1 - a) + \log_2(1 - b)$.
- *Conclusion:* we have the desired fuzzy justification of the TW measures.

117. Towards a More General Result

- The above justification is based on a rather *ad hoc* use of a special function $-\log_2(1-a)$.
- What remains unclear is how unique is this function (and thus, how unique are the TW formulas).
- We are looking for a function $z(x)$ for which, for $t_\vee(a,b) = a + b - a\cdot b$, we have

$$z(t_\vee(a,b)) = z(a) + z(b).$$

- In other words, we are looking for a "measure" $z(x)$ for which:
  - the measure that "a or b" is true is equal to
  - the sum of the measures that a is true and that b is true.

118. Example

- Vectors $x = (x_1, x_2)$ and $x' = (x'_1, x'_2)$ are different if $x_1 \neq x'_1$ or $x_2 \neq x'_2$.
- Thus, the degree to which x differs from x' equals the result of applying the "or" operation to:
  - the degree to which x_1 is different from x'_1, and
  - the degree to which x_2 is different from x'_2.
- It is thus reasonable to be able to transform these degrees into a "measure of the difference" $z(d)$ for which:
  - the measure corresponding to two-coordinate vectors should be equal to
  - the sum of the measures corresponding to both coordinates.
- Thus, we want $z(t_\vee(a, b)) = z(a) + z(b)$.

119. Main Result

Proposition. *Let* $t_\vee(a,b) = a + b - a \cdot b$. *A monotonic function* $z : [0,1] \to \mathbb{R}$ *satisfies the property*

$$z(t_\vee(a,b)) = z(a) + z(b),$$

for every a *and* b *if and only if* $z(x) = -k \cdot \log_2(x)$ *for some constant* k.

Discussion.

- We already know that the function $z(x) = -\log_2(x)$ satisfies the desired property.
- What we prove that the functions $z(x) = -k \cdot \log_2(x)$ are the only ones that satisfy this property.

120. t-Conorms: Reminder

- What if we use a different t-conorm?
- Most widely used are Archimedean t-conorms, for which, for some monotonic $f(x)$, we have

$$t_\vee(a,b) = f^{-1}(f(a) + f(b) - f(a) \cdot f(b)).$$

- A general t-conorm can be obtained:
  - by setting Archimedean t-conorms on several (maybe infinitely many) subintervals of the interval $[0,1]$,
  - by taking $t_\vee(a,b) = \max(a,b)$ when a and b are not in the same Archimedean subinterval.
- *Conclusion:* for every t-norm and for every $\varepsilon > 0$, there exists an ε-close Archimedean t-conorm.
- So, from the practical viewpoint, we can always safely assume that the t-conorm is Archimedean.

121. What If We Use a Different T-Conorm and/or a Different Membership Function?

- *Reminder:* $d(x) = t_\vee(\mu_{\neq 0}(|x_1 - \bar{x}|), \ldots, \mu_{\neq 0}(|x_1 - \bar{x}|))$, where

$$t_\vee(a, b) = f^{-1}(f(a) + f(b) - f(a) \cdot f(b)).$$

- *Resulting formulas:* for $F(z) \stackrel{\text{def}}{=} f(\mu_{\neq 0}(z))$, we get:

$$D(x) = -\log_2(1 - f(d(x))) =$$

$$-\log_2(1 - F(|x_1 - \bar{x}|)) - \ldots - \log_2(1 - F(|x_b - \bar{x}|)).$$

- *Conclusion:* for a general t-conorm and a general $\mu_{\neq 0}(c)$, it is reasonable to describe the degree of dissention as

$$D(x) = -\frac{1}{n} \cdot \sum_{i=1}^{n} \log_2(1 - F(|x_i - \bar{x}|)),$$

where $F(z) = f(\mu_{\neq 0}(z))$ and $f(z)$ is a function for which $t_\vee(a, b) = f^{-1}(f(a) + f(b) - f(a) \cdot f(b))$.

122. Corresponding Mathematical Result

Proposition. *Let*

$$t_\vee(a,b) = f^{-1}(f(a) + f(b) - f(a) \cdot f(b))$$

be an Archimedean t-conorm. A monotonic function

$$z : [0,1] \to \mathbb{R}$$

satisfies the property

$$z(t_\vee(a,b)) = z(a) + z(b),$$

for every a and b if and only if

$$z(x) = -k \cdot \log_2(1 - f(x))$$

for some constant k.

123. Conclusions

- *Problem:* estimate how close the estimates of different experts are.
- W. J. Tastle and M. J. Wierman:
  - proposed numerical measures of dissention and consensus, and
  - showed that these measures indeed capture the intuitive ideas of dissent and consensus.
- We show that the Tastle-Wierman (TW) formulas can be naturally derived from fuzzy logic.
- We also show that the TW measures can be used to gauge how different the students' grades are.