Need for Simplicity and Everything Is a Matter of Degree: How Zadeh's Philosophy is Related to Kolmogorov Complexity, Quantum Physics, and Deep Learning

Vladik Kreinovich[1], Olga Kosheleva[1], and Andres Ortiz-Munoz[2]

[1]University of Texas at El Paso
500 W. University, El Paso, Texas 79968, USA
vladik@utep.edu, olgak@utep.edu
[2]California Institute of Technology
Pasadena, California 91125, USA, aoortiz@caltech.edu

1. Two Main Ideas of Zadeh's Philosophy

- The main principle behind Lotfi Zadeh's research is a phrase that he himself repeated many times:

  Everything is a matter of degree.

- However, he also actively pursued another principle: that there is a need for simplicity.
- These two principles underlie his ideas of fuzzy logic.
- The first principle is that everything is a matter of degree.
- So instead of 0 (false) and 1 (true) we have all possible numbers from the interval $[0, 1]$.
- However, to flesh out this argument, Zadeh used simplicity as a guiding principle.

2. Main Ideas of Zadeh's Philosophy (cont-d)

- Zadeh selected the simplest "and"- and "or"-operations (i.e., t-norms and t-conorms).
- He selected the simplest membership functions, etc.
- Many of us remember that:
  - every time we came to him with some complex developments,
  - his first reaction always was: can we make it simpler?
- At first glance, it may seem that these two principles are specific for fuzzy logic and related areas.
- However, what we show in this talk is that these principles are valid way beyond fuzzy logic.

3. Main Ideas of Zadeh's Philosophy (cont-d)

- We show how these principles can be naturally applied:
  - first, to the principles themselves,
  - then to the description of the physical world, and
  - finally, to the way we gain knowledge about the world.
- Thus, we arrive at the main ideas behind Kolmogorov complexity, quantum physics, and deep learning.
- This shows that there probably is a lot of remaining potential in these principles.
- This will enable us to solve many useful problems in the future as well.

4. Zadeh's Principles Naturally Lead to Kolmogorov Complexity

- Let us apply the principles to themselves.
- In accordance with the everything-is-a-matter-of-degree principle:
  - we should not naively divide the constructions into simple and complex;
  - instead, we should assign, to each construction, a degree of complexity.
- The complexity of an object is naturally described by how complex it is to design it.
- In particular, for objects generated by a computer:
  - a natural measure of complexity is
  - the complexity of a program that generated this object.

5. Kolmogorov Complexity (cont-d)

- What is the simplest possible measure of complexity?
- The simplest possible idea is just to count the number of symbols in a program:
  - if the program if short, it is probably rather simple;
  - if the program is long, it is probably complicated.
- Thus, we define the complexity of an object x as the length of the shortest program that generates x.
- This definition is known as *Kolmorogov complexity* and is usually denoted by $K(x)$.

6. Kolmogorov Complexity (cont-d)

- The length of a program depends on the programming language.
- However, as Kolmogorov has shown:
  - the definitions $K_1(x)$ and $K_2(x)$ of Kolmogorov complexity based on two different languages
  - differ by a constant: $|K_1(x) - K_2(x)| \le C_{12}$ for all x.
- Thus, in effect, different programming languages lead to a similar notion of Kolmogorov complexity.

7. Why Was Kolmogorov Complexity Invented – And How Is It Useful?

- We showed that Kolmogorov complexity naturally follows from the Zadeh's principles.
- However, this not how this notion was originally invented.
- This notion was invented by Andrei Kolmogorov, a 20 century Russian mathematician.
- Kolmogorov was one of the pioneers of modern (mathematically precise) probability theory.
- In addition to doing pure mathematics, Kolmogorov spent a lot of time and effort on applications.

8. Why Kolmogorov Complexity (cont-d)

- Because of his interest in applications, Kolmogorov noticed a discrepancy between:
  - the traditional mathematical description of randomness and
  - the physicists' and engineers' intuitive ideas of randomness.
- Indeed, from the purely mathematical viewpoint, when you flip a fair coin n times:
  - all resulting binary sequences – describing the n results – have the exact same probability 2^{-n} and
  - are, in this sense, completely equal.

9. Why Kolmogorov Complexity (cont-d)

- However, from the practitioner's viewpoint:
  - some sequences are random – in the sense that one can expect to get them when flipping a coin, and
  - some sequences like 000...0 or 010101...01 are *not* random and thus, not expected.
- From the purely mathematical viewpoint:
  - if the same person wins the main multi-million prize of a super-lottery several years in a row,
  - it is possible.
- However, if this would happen in real life, it would be clear that the lottery is rigged.
- To eliminate this discrepancy, Kolmogorov decided to formalize the intuitive notion of randomness.
- His idea was very straightforward.

10. Why Kolmogorov Complexity (cont-d)

- Why is the sequence 000...0 not possible?
- Because this sequence is very simple, it can be obtained by running a very simple program.
- Similarly, a sequence 010101...01 can be computed by a very simple program: just print 01 in a loop.
- On the other hand:
  - if we have a sequence of actual coin-flipping results,
  - there will be hardly any dependencies,
  - so the only way to generate this sequence 011... is to write something like `print`(011...).
- The length of such a program is about the same as the length $n = \text{len}(x)$ of the original binary sequence.

11. Why Kolmogorov Complexity (cont-d)

- In other words:
  - for a truly random x, the length $K(x)$ of the shortest program that generates x is close to $\text{len}(x)$:
  $$K(x) \approx \text{len}(x),$$
  - while for a sequence which is not random, such length is much smaller: $K(x) \ll \text{len}(x)$.
- This led Kolmogorov and other researchers to define a random sequence as a sequence for which
$$K(x) \geq \text{len}(x) - C \text{ for some } C.$$
- This definition depends on C.

12. Why Kolmogorov Complexity (cont-d)

- If we select C to be too small:
  - we may miss a sequence which is actually random and
  - which accidentally has some dependencies.
- On the other hand, if we select C to be too large, we risk naming not-very-random sequences as random; so:
  - instead of a single definition of what is random and what is not random,
  - we get, in effect, a *degree* of randomness of a sequence,
  - gauged by the smallest value C for which the above inequality holds.

13. Why Kolmogorov Complexity (cont-d)

- This smallest value, as one can easily check, is equal to the difference $C = \text{len}(x) - K(x)$:
  - when this difference is small, the sequence is truly random;
  - when this difference is large, the sequence is not so random.
- Kolmogorov complexity and the related notion of randomness have indeed been very useful, both:
  - in physics applications and
  - in the analysis of algorithms.
- Kolmogorov complexity was invented in the 1960s, at the same time as fuzzy logic.
- But if this notion was not invented by Kolmogorov, Zadeh's principles could have helped invent it.

14. Zadeh's Principles Naturally Lead to the Main Ideas Behind Quantum Physics

- Fuzzy logic originated by the observation that:
  - in most cases, we cannot say that someone is absolutely young or absolutely old;
  - for most people, "young" is a matter of degree.
- Let's apply the everything-is-a-matter-of-degree principle to the physical world.
- Then, we should not expect the system to be definitely in one state.
- In many cases, we should have a system which is:
  - with some degree, in one state, and
  - with some degree, in another state.
- How can we describe this in precise mathematical form?

15. Quantum Physics (cont-d)

- We have two original states s_1 and s_2.
- Each state can be described by, e.g., the values of the corresponding quantities, i.e., by a tuple

$$s_i = (s_{i1}, \ldots, s_{in}) \text{ of real numbers.}$$

- We need a function that transforms the two states into a new combined state.
- From the mathematical viewpoint, the simplest functions are linear functions.
- Thus, the simplest possible way to combine the two states is to have a linear combination $a_1 \cdot s_1 + a_2 \cdot s_2$.

16. Quantum Physics (cont-d)

- This is exactly what in quantum physics is called *superposition* of the two states.
- The only difference that in quantum physics, the coefficients a_i can be complex numbers.
- The existence of such superposition is one of the main ideas behind quantum physics.
- It is also the main idea behind *quantum computing* and its successes.

17. What Is the Dynamics Here?

- To fully describe a physical phenomenon:
  - it is not sufficient to describe the current state,
  - we need also to describe how this state changes in time – i.e., what is the corresponding dynamics.
- We thus need to describe:
  - how the state s of a physical system changes with time,
  - i.e., how the rate of change $\dfrac{ds}{dt}$ depends on the state: $\dfrac{ds}{dt} = F(s)$, for some function $F(s)$.
- Which function $F(s)$ should we select? For this selection, it is reasonable to use the simplicity principle.

18. What Is the Dynamics Here (cont-d)

- As we have mentioned earlier, the simplest functions are linear functions.
- So it is reasonable to use linear function $F(s)$.
- This is exactly what is happening in quantum physics.
- There, Schroedinger's equation describes the rate of change in a state as a linear function of this state.

19. Zadeh's Principles Naturally Lead to the Main Ideas Behind Deep Learning

- Let's go from the description of a physical world to the description of how we learn things about the world.
- First, according to the everything-is-a-matter-of-degree principle, at each given moment of time:
  - about each statement about the world,
  - we do not necessarily have a precise opinion.
- Mostly, we have a degree of confidence about the given statement.
- How do these degrees of confidence change?
- The simplest possible functions, as we have mentioned several times, are linear functions.
- So, a natural transformation of degrees is linear.

20. Deep Learning (cont-d)

- The problem with using only linear transformations is that:
  - by combining several linear transformations, we can only have linear functions,
  - and some real-world processes that we want to describe are not linear.
- So, we need some non-linear transformations as well.
- What are the simplest nonlinear transformations?
- First, in general, the more inputs, the more complex the transformation.
- From this viewpoint, the easiest are the functions of one variable.
- What are the simplest functions of one variable which are not linear?

21. Deep Learning (cont-d)

- A natural idea is to use functions which are *piece-wise linear*, i.e., for example, consist of 2 linear parts.
- The simplest value separating the two parts is 0, so we end up with a function $s(x)$ which is:
  - equal to $c_- \cdot x$ for $x \le 0$ and to $c_+ \cdot x$ for $x \ge 0$,
  - for some appropriate values c_- and c_+.
- What are the simplest real numbers?
- Clearly, 0 and 1, so we end up with a function which is equal to $0 \cdot x = 0$ for $x \le 0$ and to $1 \cdot x = x$ for $x \ge 0$.
- This function can be equivalently described as

$$s(x) = \max(0, x).$$

- So, we arrive at the following description of the corresponding computational schemes.

22. Deep Learning (cont-d)

- The input to a non-linear transformation comes from a linear combination of inputs.
- So, we have transformations of the type $x_1, \ldots, x_n \to \max\left(0, \sum\limits_{i=1}^{n} w_i \cdot x_i - w_0\right)$ for some weights w_i.
- This is exactly the transformation performed by each neuron in a deep neural network.
- Thus, by combining such transformations, we get the main structure of deep neural networks.

23. Deep Learning (cont-d)

- Also:
  - if instead of nonlinear functions of one variable, we allow nonlinear functions of several variables,
  - then the simplest are – like in fuzzy logic – min and max.
- Such transformations form the basis of convolutional neural networks – another type of deep learning.
- So, Zadeh's ideas are in good accordance with deep learning as well.

24. What Next?

- What next? There are two natural ideas.
- First is to continue applying Zadeh's ideas to other areas.
- An example of such application in the Appendix.
- Another idea is to take into account that a degree does not have to be a single number.
- We have already seen this in quantum physics – where the coefficients a_i are, in general, complex numbers

$$a + b \cdot \mathrm{i}, \text{ i.e., pairs of real numbers.}$$

- The use of degrees described by several numbers have been successful in fuzzy applications.
- Example: interval-valued and more general fuzzy degrees.

25. What Next (cont-d)

- And even in applications like education:
  - where traditionally the degree of student's knowledge was described by a single number,
  - the new tendency is to use several numbers,
  - e.g., describing the degree of student's knowledge of different topic,
  - and thus providing a more adequate description of the student's knowledge.

26. Acknowledgments

This work was supported in part by the National Science Foundation grants:

- 1623190 (A Model of Change for Preparing a New Generation for Professional Practice in Computer Science);
- HRD-1834620 and HRD-2034030 (CAHSI Includes).

It was also supported by the program of the development of the Scientific-Educational Mathematical Center of Volga Federal District No. 075-02-2020-1478.

27. Appendix: Causality as a Matter of Degree

- *Causal relation* $e \preccurlyeq e'$ between space-time events is one of the fundamental notions of physics.
- *In Newton's physics*, it was assumed that influences can propagate with an arbitrary speed:
$$e = (t, x) \preccurlyeq e' = (t', x') \Leftrightarrow t \le t'.$$
- In *Special Relativity*, the speeds of all the processes are limited by the speed of light c:
$$e = (t, x) \preccurlyeq e' = (t', x') \Leftrightarrow c \cdot (t' - t) \ge d(x, x').$$
- In the *General Relativity Theory*, the space-time is curved, so the causal relation $\preccurlyeq$ is even more complex.
- Different theories, in general, make different predictions about the causality $\preccurlyeq$.
- So, to experimentally verify fundamental physical theories, we need to experimentally check whether $e \preccurlyeq e'$.

28. Defining Causality: Challenge

- Intuitively, $e \preccurlyeq e'$ means that:
  - what we do in the vicinity of e
  - changes what we observe at e'.
- If we have *two* (or more) copies of the Universe, then:
  - in one copy, we perform some action at e, and
  - we do not perform this action in the second copy.
- If the resulting states differ, this would indicate $e \preccurlyeq e'$:

World 1	World 2
rain $* e'$	$e' *$ no rain
rain dance $* e$	$e *$ no rain dance

- Alas, in reality, we only observe *one* Universe, in which we either perform the action or we do not.

Two Main Ideas of . . .
Kolmogorov Complexity
Why Was Kolmogorov . . .
Quantum Physics
What Is the Dynamics . . .
Deep Learning
What Next?
Acknowledgments
Appendix: Causality . . .
Home Page
Title Page
Page 29 of 39
Go Back
Full Screen
Close
Quit

29. Kolmogorov Complexity: A Brief Reminder

- If we flip a coin 1000 times and still get all heads, common sense tells us that this coin is not fair.
- Similarly, if we repeatedly flip a fair coin, we cannot expect a periodic sequence $0101\ldots01$ (500 times).
- Traditional probability theory does not distinguish between random and non-random sequences.
- Kolmogorov, Solomonoff, Chaitin: a sequence $0\ldots0$ isn't random since it can be printed by a short program.
- Let an integer $C>0$ be fixed. We say that a string x is *random* if $K(x)\ge \mathrm{len}(x)-C$, where

$$K(x)\stackrel{\text{def}}{=}\min\{\mathrm{len}(p): p \text{ generates } x\}.$$

30. The Corresponding Notion of Independence

- If y is independent on x, then knowing x does not help us generate y.
- If y depends on x, then knowing x helps compute y; example:
  - knowing the locations and velocities x of a mechanical system at time t
  - helps compute the locations and velocities y at time $t + \Delta t$.
- Let an integer $C > 0$ be fixed. We say that a string y is *independent* of x if $K(y \mid x) \geq K(y) - C$, where

$$K(y \mid x) \stackrel{\text{def}}{=} \min\{\text{len}(p) : p(x) \text{ generates } y\}.$$

- We say that a string y is *dependent* on the string x if

$$K(y \mid x) < K(y) - C.$$

31. How to Define Space-Time Causality: First Seeming Reasonable Idea

- At first glance, we can check whether $e \preccurlyeq e'$ as follows:
  - First, we perform observations and measurements in the vicinity of the event e, and get the results x.
  - We also perform measurements and observations in the vicinity of the event e', and produce x'.
  - If x' depends on x, i.e., if $K(x' \mid x) \ll K(x')$, then we claim that e can casually influence e'.
- If $e \preccurlyeq e'$, then indeed knowing what happened at e can help us predict what is happening at e'.
- However, the inverse is not necessarily true.
- We may have $x \approx x'$ because both e and e' are influenced by the same past event e''.
- Example: both e and e' receive the same signal from e''.

32. Towards a Working Definition of Causality

- According to modern physics, the Universe is quantum in nature; for microscopic measurements:
  - we cannot predict the exact measurement results,
  - we can only predict probabilities of different outcomes; the actual observations are truly random.
- For each space-time event e:
  - we can set up such a random-producing experiment in the small vicinity of e, and
  - generate a random sequence r_e.
- This random sequence r_e can affect *future* results.
- So, if we know the random sequence r_e, it may help us predict future observations.
- Thus, if $e \preccurlyeq e'$, then for some observations x' performed in the small vicinity of e', we have $K(x' \mid r_e) \ll K(x')$.

33. Discussion and Resulting Definition

- *Reminder:* when $e \preccurlyeq e'$, then the random sequence r_e can affect the measurement results at e':

$$K(x' \,|\, r_e) \ll K(x').$$

- If $e \not\preccurlyeq e'$, then the random sequence r_e cannot affect the measurement results at e': $K(x' \,|\, r_e) \approx K(x')$.
- So, we arrive at the following semi-formal *definition:*
  - For a space-time event e, let r_e denote a random sequence generated in the small vicinity of e.
  - We say $e \preccurlyeq e'$ if for some observations x' performed in the small vicinity of e', we have

$$K(x' \,|\, r_e) \ll K(x').$$

- Our definition follows the ideas of casuality as *mark transmission*, with the random sequence as a mark.

Two Main Ideas of . . .
Kolmogorov Complexity
Why Was Kolmogorov . . .
Quantum Physics
What Is the Dynamics . . .
Deep Learning
What Next?
Acknowledgments
Appendix: Causality . . .
Home Page
Title Page
Page 34 of 39
Go Back
Full Screen
Close
Quit

34. This Definition Is Consistent with Physical Intuition

- If $e \preccurlyeq e'$, then we can send all the bits of r_e from e to e'.
- The signal x' received in the vicinity of e' will thus be identical to r_e.
- Thus, generating x' based on r_e does not require any computations at all: $K(x' \mid r_e) = 0$.
- Since the sequence $x' = r_e$ is random, we have
$$K(x') \geq \text{len}(x') - C.$$
- When $r_e = x'$ is sufficiently long ($\text{len}(x') > 2C$), we have $K(x') \geq \text{len}(x') - C > 2C - C = C$, hence
$$0 = K(x' \mid r_e) < K(x') - C \text{ and } K(x' \mid r_e) \ll K(x').$$
- So, our definition is indeed in accordance with the physical intuition.

35. Randomness Is a Matter of Degree

- *Reminder*: a sequence x is random if $K(x) \geq \text{len}(x) - C$ for some $C > 0$.
- For a given sequence x, its *degree of randomness* $d(x)$ can be defined by the smallest integer C for which

$$K(x) \geq \text{len}(x) - C.$$

- One can check that this smallest integer is equal to the difference $d(x) = \text{len}(x) - K(x)$.
- For *random* sequences, the degree $d(x)$ is small.
- For sequences which are *not random*, the degree $d(x)$ is large.
- In general, the smaller the difference $d(x)$, the more random is the sequence x.

36. Space-Time Causality Is a Matter of Degree

- Our definition of causality is that $K(x' \,|\, r_e) < K(x') - C$ for some large integer C.
- The larger the integer C, the more confident we are that an event e can causally influence e'.
- It is therefore reasonable to define a *degree of causality* $c(e, e')$ as the largest integer C for which
$$K(x' \,|\, r_e) < K(x') - C.$$
- One can check that this largest integer is equal to the difference $c(e, e') = K(x') - K(x' \,|\, r_e) - 1$.
- The larger this difference $c(e, e')$, the more confident we are that e can influence e'.
- In other words, just like randomness turns out to be a matter of degree, causality is also a matter of degree.

37. Remaining Open Problems

- It is desirable to explore possible physical meaning of such "degrees of causality" $c(e, e')$.
- Maybe this function $c(e, e')$ is related to relativistic metric – the amount of proper time between e and e'?
- Another open problem: the above definition works for objects in a small vicinity of one spatial location.
- In quantum physics, not all objects are localized in space-time.
- We can have situations when the states of two spatially separated particles are entangled.
- It is desirable to extend our definition to such objects as well.

38. Conclusions

- We propose a new operationalist definition of causality $e \preccurlyeq e'$ between space-time events e and e'.
- Namely, to check whether an event e can casually influence an event e', we:
  - generate a truly random sequence r_e in the small vicinity of the event e, and
  - perform observations in the small vicinity of the event e'.
- If some observation results x' (obtained near e') *depend* on r_e, then we claim that $e \preccurlyeq e'$.
- On the other hand, if all observation results x' are *independent* on r_e, then we claim that $e \not\preccurlyeq e'$.
- This new definition naturally leads to a conclusion that space-time causality is a matter of degree.