Lotfi Zadeh: a Pioneer in AI, a Pioneer in Statistical Analysis, a Pioneer in Foundations of Mathematics, and a True Citizen of the World

Vladik Kreinovich

Department of Computer Science
University of Texas at El Paso, El Paso, Texas 79968, USA
vladik@utep.edu, http://www.cs.utep.edu/vladik

1. Outline

- Everyone knows Lotfi Zadeh as the Father of Fuzzy Logic.
- There have been – and will be – many talks on this important topic.
- What I want to emphasize in this talk is that his ideas go way beyond fuzzy logic:
  - he was a pioneer in AI;
  - he was a pioneer in statistical analysis; and
  - he was a pioneer in foundations of mathematics.
- My goal is to explain these ideas to non-fuzzy folks.
- I also want to emphasize that he was a true Citizen of the World.

2. How Lotfi Zadeh Became an AI Pioneer: Practical Problem

- Lotfi A. Zadeh was a specialist in control and systems.
- His textbook *Linear System Theory: The State Space Approach* (with Charles A. Desoer) was a classic.
- It provided optimal solutions to many important control problems – optimal within the existing models.
- But, surprisingly, in many practical situations, "optimal" control was worse than control by human experts.
- Clearly, something was missing from the corresponding models.
- So Zadeh asked experts what is missing.

3. What Experts Said

- Many experts explained what was wrong with the "optimal" control.
- However, these explanations were given in imprecise natural-language terms.
- For example, a expert driver can say:
  - if a car in front is close, and
  - if this car slows down a little bit,
  - then a driver should hit the breaks slightly.
- Until Zadeh, engineers would try to extract precise strategy from the expert; they would ask an expert:
  - a car is 5 m close, and
  - it slows down from 60 to 55 km/h,
  - for how long and with what force should we hit the brakes?

4. Problem with the Traditional AI Approach

- Most people cannot answer this question.
- Those who answer give a somewhat random number – and different number every time.
- If we implement exactly this force, we get a weird control – much worse than when a human drives.
- If we instead apply optimization:
  - the resulting control is optimal for the exact weight of the car;
  - but if a new passenger enters the car – the problem changes;
  - the previous optimal control is not longer optimal;
  - this control can be really bad.
- If we simply ignore expert rules, we also get a suboptimal control.

5. What Zadeh Proposed to Solve This Problem: Main Idea

- Also, what we want is imprecise:
  - e.g., for an elevator, we want a smooth ride,
  - but it is difficult to describe this in precise terms.
- Zadeh had an idea:
  - in situations when we can only extract imprecise (fuzzy) rules from the experts,
  - instead of ignoring these rules,
  - let us develop techniques that transform these fuzzy rules into a precise control strategy.
- Zadeh invented the corresponding technique – it is the technique he called *fuzzy logic*.

6. Zadeh's Idea Illustrated on a Simple Example

- Zadeh's technique can be illustrated on a simple example of a thermostat:
  - if we turn the knob to the right, the temperature T increases;
  - if we turn it to the left, the temperature decreases;
  - our goal is to maintain a comfortable temp. T_0.
- Experts can formulate rules on how the angle u to which we rotate the knob depends on T:
  - if the temperature is practically comfortable, no control is needed;
  - if the temperature is slightly higher than desired, cool the room a little bit;
  - if the temperature is slightly lower than desired, heat up the room a little bit; etc.

7. Simple Example (cont-d)

- In terms of the difference $x \stackrel{\text{def}}{=} T - T_0$:
  - if x is negligible, u should be negligible;
  - if x is small positive, then u should be small negative;
  - if x is small negative, then u should be small positive, etc.
- By using abbreviations N for “negligible”, SP for “small positive”’, and SN for “small negative”, we get:

$$N(x) \Rightarrow N(u);\ \ SP(x) \Rightarrow SN(u);\ \ SN(x) \Rightarrow SP(u); \ldots$$

- A control u is reasonable for given x ($R(x,u)$) if one of these rules is applicable: $R(x,u) \Leftrightarrow$

$$(N(x) \,\&\, N(u)) \vee (SP(x) \,\&\, SN(u)) \,\&\, (SN(x) \,\&\, SP(u)) \vee \ldots$$

Outline
AI Pioneer
Why This Is Great
Comment on Simplicity
Misunderstandings
Statistical Analysis ...
Foundations of ...
A True Citizen of the ...
A Good Person
Home Page
Title Page
◀◀ ▶▶
◀ ▶
Page 8 of 33
Go Back
Full Screen
Close
Quit

8. Three Stages of Fuzzy Control Technique

- To translate this into precise formula, we need:
  - to translate $N(x)$, $N(u)$, ... into precise terms,
  - to interpret "and" and "or", and then
  - to translate the resulting property $R(x, u)$ into a single control value $\overline{u}$.
- Since we deal with "and" and "or", this technique is related to logic.
- Since we deal with imprecise ("fuzzy") statements, Zadeh called it *fuzzy logic*.
- Let us explain all three stages of fuzzy logic technique.

9. First Stage

- First stage: how can we interpret "x is negligible"?
- For traditional (precise) properties like "$x > 5°$", the property is either true or false.
- Here, to some folks, 5 degrees is negligible, some feel a difference of 2 degrees.
- And no one can select an exact value – so that, say 1.9 is negligible but 2.0 is not.
- Same thing: there is no exact threshold separating "close" from "not close".
- At best, expert can mark the *degree* to which x is negligible on a scale from, say, 0 to 10.
- If an expert marks 7 on a scale from 0 to 10, we say that his degree of confidence that x is negligible is $7/10$.

10. Second Stage

- This way, we can find the degrees of $N(x)$, $N(u)$, $SP(x)$, etc.
- Based on these degrees, we need to estimate degrees of propositional combinations $N(x)\,\&\,N(u)$, etc.
- Ideally, we can ask the expert for degrees of all such combinations.
- However, for n basic statements, there are 2^n such combinations.
- For $n = 30$, we have $2^{30} \approx 10^9$ combinations.
- It is not possible to ask 10^9 questions.
- So, we need to be able to estimate the degree $d(A\,\&\,B)$ based on degrees $a = d(A)$ and $b = d(B)$.
- The algorithm $d(A\,\&\,B) \approx f_\&(a, b)$ for such an estimation is known as an *"and"-operation*.

11. Second Stage (cont-d)

- For historical reasons, "and"-operations are also knows as *t-norms*.
- What are natural properties of "and"-operations?
- Since $A\,\&\,B$ means the same as $B\,\&\,A$, this operation must be commutative: $f_{\&}(a,b) = f_{\&}(b,a)$.
- Since $A\,\&\,(B\,\&\,C)$ means the same as $(A\,\&\,B)\,\&\,C$, the "and"-operation must be associative.
- There are also natural requirements of monotonicity, continuity,

$$f_{\&}(1,1) = 1,\ f_{\&}(0,0) = f_{\&}(0,1) = f_{\&}(1,0) = 0, \ldots$$

- All such operations are known.

12. Examples of "'And"- and "Or"-Operations

- W may want to also require that $A \,\&\, A$ means the same as A: $f_\&(a,a) = a$.
- In this case, we get $f_\&(a,b) = \min(a,b)$.
- This is one of the most widely used "and"-operations.
- Others include $f_\&(a,b) = a \cdot b$, etc.
- Similar properties hold for "or"-operations $f_\vee(a,b)$ (a.k.a. t-conorms).
- For example, if we require that $A \vee A$ means the same as A, we get $f_\vee(a,b) = \max(a,b)$.
- Others include $f_\vee(a,b) = a + b - a \cdot b$, etc.

13. Third (Final) Stage and Resulting Success Stories

- By applying "and"- and "'or"-operations, we get, for each u, the degree $R(x,u)$ to which u is reasonable.
- Now, we need to select a single control value $\overline{u}$.
- It is reasonable to use Least Squares, with $R(x,u)$ as weights: $\int R(x,u)\cdot(u-\overline{u})^2\,du \to \min$.
- The resulting formula is known as *centroid defuzzification*:

$$\overline{u} = \frac{\int R(x,u)\cdot u\,du}{\int R(x,u)\,du}.$$

- This technique has led to many successes:
  - fuzzy-controlled trains and elevators provide smooth ride;
  - fuzzy rice cookers produce tasty rice; etc.

14. This Was A Simplified Description

- The above description only contains the main ideas, real-life applications are more complex.
- First, just like experts cannot say with what force their press the brakes, they cannot tell whet exactly is their degree of confidence.
- An expert can say 7 or 8 on a scale of 0 to 10, but cannot distinguish between 70/100 and 71/100.
- Thus, a more adequate description of expert's confidence is not a number but an interval of possible values.
- An expert may also say how confident she is about each degree – so we have a type-2 fuzzy degree.
- This leads to control which is closer to expert's – and thus, better: smoother, more stable, etc.

15. This Was A Simplified Description (cont-d)

- Second, centroid defuzzification doe snot always work.
- For example, if we want to avoid an obstacle in front, we can steer to the left or to the right.
- The situation is completely symmetric, thus the defuzzified value is symmetric.
- So it leads us straight into the obstacle.
- Thus, we need to only select control values for which degree of confidence exceeds some threshold.
- Third, we also often have additional constraints – which could also be fuzzy.
- Finally, we often want not just to follow expert, but to optimize – thus further improving their advice.
- Optimization under fuzzy uncertainty can also be handled by fuzzy logic techniques.

16. Why This Is Great

- Many of our important notions are “fuzzy”.
- No one is absolutely good or bad – it is a matter of degree.
- It is difficult to find *the* cause of an event – usually, many factors have different degrees of causality.
- How can we describe this fuzziness?
- There are three levels of applied mathematics.
- *Level 1:* most researchers are well familiar with one formalism and use it.
- Statisticians use statistics, others use differential equations. etc.

17. Why This Is Great (cont-d)

- *Level 2:* some researchers have mastered several mathematical techniques.
- These researchers select, for each practical problem, the most appropriate of these techniques.
- However, existing techniques are often not perfectly adequate for a practical problem.
- *Level 3:* a researcher designs a new formalism, specially for the given application.
- Philosophers like to cite Nobelist Eugene Wigner who wrote about unexplainable efficiency of mathematics:
  - Quantum physics is perfectly described by Hilbert spaces.
  - General Relativity is based on pseudo-Riemannian spaces.

18. Why This Is Great (cont-d)

- However:
  - Hilbert spaces were invented by John von Neumann explicitly to describe quantum physics.
  - Pseudo-Riemannian spaces were invented by A. Einstein explicitly to describe curved space-time.
- Zadeh's ideas follow the same pattern.
- Before Zadeh, researchers described human uncertainty by known math: e.g., probabilities.
- This covered some cases well, some not so well.
- Zadeh came up with a new technique specifically designed for describing non-probabilistic uncertainty.
- As a result, he got many successful applications.

19. Comment on Simplicity

- From the mathematical viewpoint, his main ideas were simple.
- This makes it even better: if we can get good empirical results by using simpler techniques, good!

20. Misunderstandings

- Zadeh's AI ideas were often misunderstood.
- Some folks falsely believed that in fuzzy logic, $d(A \,\&\, B)$ is uniquely determined by $d(A)$ and $d(B)$.
- They thought that a simple counterexamples to this Straw-man belief can prove that fuzzy logic is wrong.
- Some falsely believed that Zadeh recommended min and max only.
- In reality, in his very first fuzzy paper he introduced other operations as well.
- Some believed that Zadeh wanted to replace probabilities with fuzzy logic.
- In reality, he always emphasized the need to have 100 flowers bloom.

21. Lotfi Zadeh: A Pioneer in Statistical Analysis

- In many practical situations:
  - we know the probabilities $p_1, \ldots, p_n$ of individual events $E_1, \ldots, E_n$, and
  - we would like to know the probabilities of different propositional combinations, such as $E_1 \,\&\, E_2$.
- To describe all such probabilities, it is sufficient to find the probabilities of all "and"-combinations

$$E_{i_1} \,\&\, \ldots \,\&\, E_{i_m}.$$

- If the events are independent, the answer is easy:

$$p(E_{i_1} \,\&\, \ldots \,\&\, E_{i_m}) = p(E_{i_1}) \cdot \ldots p(E_{i_m}).$$

22. Statistical Analysis (cont-d)

- However, often:
  - we know that the events are not independent,
  - but we do not have enough data to find out the exact dependence.
- Traditional statistical approach was to assume some prior joint distribution.
- The problem is that different prior distributions lead to different answers.
- In statistical analysis, we usually select the easiest-to-process distribution.
- However, real life is often complex – so why should we select the simplest method?

23. Statistical Analysis (cont-d)

- Zadeh's revolutionary idea was to select an appropriate "and"-operation for:
  - converting probabilities $a = p(A)$ and $b = p(B)$
  - into an estimate $f_{\&}(a, b)$ for $p(A \,\&\, B)$.
- A natural requirements that estimates for $A \,\&\, B$ and $B \,\&\, A$ should be the same lead to commutativity

$$f_{\&}(a, b) = f_{\&}(b, a).$$

- The requirement that estimates for $A \& (B \,\&\, C)$ and $(A \,\&\, B) \,\&\, C$ coincide lead to associativity.
- The corresponding "and"-operation should be experimentally determined.

24. Relation to MYCIN

- This idea, in effect, formalizes the procedure successfully used for Stanford's MYCIN.
- This was the world's first successful expert system – designed for diagnosing rare blood diseases.
- Interestingly, MYCIN's authors first thought that their "and"-operation describes general human reasoning.
- However, when they tried to apply it to geophysics, they realized that we need a different $f_\&(a, b)$.
- This makes sense – in geophysics:
  - we start digging for oil if there is a good chance of success,
  - even if further tests could clarify the situations.
- In contrast, in medicine, we do not recommend a surgery unless we have made all possible tests.

25. Lotfi Zadeh: A Pioneer in Foundations of Mathematics

- From the logical viewpoint, the original fuzzy logic is simply $[0,1]$-valued logic.
- The main formulas for this logic were proposed by Lukaciewicz in the 1920s.
- Zadeh succeeded in transforming this abstract theory into a successful practical tool.
- He also came up with an idea of how to generalize all mathematical notions into fuzzy, e.g.:
  - replace & (and $\forall$ – infinite &) with min, and
  - replace $\vee$ (and $\exists$ – infinite $\vee$) – with max.
- Example: how to extend data processing algorithm $y = f(x_1, \ldots, x_n)$ to fuzzy inputs $\mu_i(x_i)$?

26. Foundations of Mathematics (cont-d)

- Main idea: y is reasonable $\Leftrightarrow$

  $\exists x_1 \ldots \exists x_n\,(x_1$ is reasonable and ... and x_n is reasonable,

  $$\text{and } y = f(x_1, \ldots, x_n)).$$

- The above transformation leads to

$$\mu(y) = \max_{x_1,\ldots,x_n:\ f(x_1,\ldots,x_n)=y} \min(\mu_1(x_1), \ldots, \mu_n(x_n)).$$

- This is known as *Zadeh's extension principle.*
- Instead of min, we can use other 'and'-operations.
- Important point: this is *not* as arbitrary as it seems to some authors.
- This is a particular case of a general algorithm.

27. Foundations of Mathematics: 2nd Example

- Another example – intuitive continuity: if x and x' are close, then $y = f(x)$ and $y' = f(x')$ should be close.
- Let $\mu_{\rm in}(x' - x)$ describe closeness of inputs.
- Closeness of outputs may be described in a different scale: $\mu_{\rm out}(y - y') = \mu_{\rm in}(K \cdot (y - y'))$.
- Implication $A \to B$ can be understood as $d(A) \geq d(B)$.
- Thus, we get the condition

$$\mu_{\rm in}(x - x') \geq \mu_{\rm out}(f(x) - f(x')) = \mu_{\rm in}(K \cdot ((x) - f(x')).$$

- This condition is equiv. to $|x - x'| \leq K \cdot |f(x) - f(x')|$, i.e., to $|f(x) - f(x')| \leq L \cdot |x - x'|$, for $L \stackrel{\rm def}{=} K^{-1}$.
- Thus, we get *Lipschitz condition*

28. Foundations of Mathematics: Warning

- Warning (emphasized by Elkan): we need to use the original logical formulation of the property.
- Indeed, e.g., $A \vee \neg A$ is not always true:

$$\max(d(A), d(\neg A)) = \max(d(A), 1 - d(A)) \not\equiv 1.$$

- Thus, a classically equivalent logical formula can lead to a different translation.

29. Lotfi Zadeh: A True Citizen of the World

- Who was he?
- For Azeris, Zadeh is a national hero who passionately cared about the country of his birth.
- To Iranians, he is a great Iranian who knew a lot and cared a lot about their country.
- People of Russia knew him as passionate and well-informed about Russian events.
- With his wife Fay, he spoke English – although they could communicate in many other languages.
- And he was passionate about US politics.
- To many people, he was their own.
- He was a true citizen of the world.

30. A True Citizen of the World (cont-d)

- He was *not* coldly above struggle, no way.
- He was passionate about everyone, his heart bled about all the injustices of the world – as if they were his own.
- And he wa passionately happy about the successes of everyone – as if they were their own.
- He was the true embodiment of Apostle Paul's famous statement:

  "There is no Jew and no Greek, we are all one".

- This was his attitude to nations, this was his attitude to people.

31. A Good Person

- Zadeh's"take everything as a compliment" life stance helped him remain calm, cheerful – and successful.
- He promoted his fuzzy ideas – but never at the expense of others.
- Vice versa, he always emphasized the need to combine them with others – probabilistic, neural, etc.
- He inspired:
  - a combination of different AI directions – fuzzy, neural, etc.
  - into a single soft computing direction, with successful conferences, journals, and applications.

The world needs more people like Lotfi Zadeh!