2-2018

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
University of Texas at El Paso, vladik@utep.edu

Comments:
Technical Report: UTEP-CS-18-06

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

Abstract

Everyone knows Lotfi Zadeh as the Father of Fuzzy Logic. There have been – and will be – many papers on this important topic. What I want to emphasize in this paper 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.

1 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 [15]. 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.

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?

Problem with the traditional AI approach. Most people cannot answer the above 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. As a result, this control can be really bad.
- If we simply ignore expert rules, we also get a suboptimal control.

Also, what we want is imprecise. For example, for an elevator, we want a smooth ride, but it is difficult to describe this in precise terms.

What Zadeh proposed to solve this problem: main idea. 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.*

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 temperature 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.

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:

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

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

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

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*; see, e.g., [1, 5, 6, 9, 10, 14].

Let us explain all three stages of fuzzy logic technique.

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.

Similarly, 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. This way, we can find the degrees of $N(x)$, $N(u)$, $SP(x)$, etc.

Second stage. Based on the degrees obtained on the first stage, 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.* 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, and requirements that

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

Examples of "and"- and "or"-operations. All such operations are known.

We 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.

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 [1, 5, 6, 9, 10, 14]:

- fuzzy-controlled trains and elevators provide smooth ride;
- fuzzy rice cookers produce tasty rice; etc.

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.; see, e.g., [6, 7].
- Second, centroid defuzzification does not 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.

Fuzziness is ubiquitous. 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?

Three levels of applied mathematics. 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.
- *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 [13]; e.g.:

- quantum physics is perfectly described by Hilbert spaces;
- General Relativity is based on pseudo-Riemannian spaces.

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.

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!

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 counterexample 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.

2 Lotfi Zadeh: A Pioneer in Statistical Analysis

Formulation of the problem. 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:

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

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; see, e.g., [11, 12]. Which one should we select?

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?

Zadeh's idea. 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.

Relation to MYCIN. Zadeh's idea, in effect, formalizes the procedure successfully used for Stanford's MYCIN; see, e.g., [2]. 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.

3 Lotfi Zadeh: A Pioneer in Foundations of Mathematics

Main idea. From the logical viewpoint, the original fuzzy logic is simply $[0,1]$-valued logic. The main formulas for this logic were proposed by Lukasiewicz 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.

First example. How to extend data processing algorithm $y = f(x_1, \ldots, x_n)$ to fuzzy inputs $\mu_i(x_i)$? Main idea:

$$y \text{ is reasonable } \Leftrightarrow$$

$$\exists x_1 \ldots \exists x_n \,(x_1 \text{ is reasonable and} \ldots \text{and } x_n \text{ is reasonable, and}$$

$$y = f(x_1, \ldots, x_n)).$$

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 is this is *not* as arbitrary as it seems to some authors, this is a particular case of a general algorithm.

Second example. Another example is 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 $d(A) \ge d(B)$. Thus, we get the condition

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

This condition is equivalent to $|x - x'| \le K \cdot |f(x) - f(x')|$, i.e., to

$$|f(x) - f(x')| \le L \cdot |x - x'|,$$

for $L \stackrel{\rm def}{=} K^{-1}$. Thus, we get *Lipschitz condition*

Warning. Warning (emphasized by Elkan [3, 4, 8]): 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.

4 Lotfi Zadeh: A True Citizen of the World

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.
- He was passionate about US politics – and with his wife Fay, he spoke English (although they could communicate in many other languages).
- To many people, he was their own.

He was a true citizen of the world.

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.

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!

Acknowledgments

This work was supported in part by the National Science Foundation grant HRD-1242122 (Cyber-ShARE Center of Excellence).

The authors is thankful to all the participants of the Symposium on Fuzzy Logic and Fuzzy Sets (Berkeley, California, February 5, 2018) for valuable suggestions.

References

[1] R. Belohlavek, J. W. Dauben, and G. J. Klir, *Fuzzy Logic and Mathematics: A Historical Perspective*, Oxford University Press, New York, 2017.

[2] B. G. Buchanan and E. H. Shortliffe, *Rule Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project*, Addison-Wesley, Reading, Massachusetts, 1984.

[3] C. Elkan, "The paradoxical success of fuzzy logic", *Proceedings of th 1993 Annual Conference of the American Association for Artificial Intelligence AAAAI'93*, Washington, DC, July 11–15, 1993, pp. 698–703; reprinted in *IEEE Expert*, August 1994, pp. 3–8.

[4] C. Elkan et al, "Fuzzy Logic Symposium", *IEEE Expert*, August 1994.

[5] G. Klir and B. Yuan, *Fuzzy Sets and Fuzzy Logic*, Prentice Hall, Upper Saddle River, New Jersey, 1995.

[6] J. M. Mendel, *Uncertain Rule-Based Fuzzy Systems: Introduction and New Directions*, Springer, Cham, Switzerland, 2017.

[7] J. M. Mendel and D. Wu, *Perceptual Computing: Aiding People in Making Subjective Judgments*, IEEE Press and Wiley, New York, 2010.

[8] H. T. Nguyen, O. M. Kosheleva, and V. Kreinovich, "Is the success of fuzzy logic really paradoxical? Or: Towards the actual logic behind expert systems", *International Journal of Intelligent Systems*, 1996, Vol. 11, No. 5, pp. 295–326.

[9] H. T. Nguyen and E. A. Walker, *A First Course in Fuzzy Logic*, Chapman and Hall/CRC, Boca Raton, Florida, 2006.

[10] V. Novák, I. Perfilieva, and J. Močkoř, *Mathematical Principles of Fuzzy Logic*, Kluwer, Boston, Dordrecht, 1999.

[11] S. G. Rabinovich, *Measurement Errors and Uncertainty: Theory and Practice*, Springer Verlag, Berlin, 2005.

[12] D. J. Sheskin, *Handbook of Parametric and Nonparametric Statistical Procedures*, Chapman and Hall/CRC, Boca Raton, Florida, 2011.

[13] E. P. Wigner, "The unreasonable effectiveness of mathematics in the natural sciences.", *Communications on Pure and Applied Mathematics*, 1960, Vol. 13, pp. 1–14.

[14] L. A. Zadeh, "Fuzzy sets", *Information and Control*, 1965, Vol. 8, pp. 338–353.

[15] L. A. Zadeh and C. A. Desoer, *Linear System Theory: The State Space Approach*, Dover Pulc., New York, 2008.