Symmetry approach explains the computational advantage of intuitionistic fuzzy sets over their interval-valued reformulation

Olga Kosheleva[1] and Vladik Kreinovich[2]

Departments of [1]Teacher Education and [2]Computer Science
University of Texas at El Paso,
500 W. University, El Paso, Texas 79968, USA
olgak@utep.edu, vladik@utep.edu

1. Why symmetries?

- In this paper, we will use symmetry approach – one of the main tools of modern physics.
- Since many people in the computing community are not well familiar with this approach, let us briefly explain its main ideas.
- One of the main objectives of science and engineering is:
  - to predict future events – and
  - to make sure that these future events are as beneficial to the humankind as possible.
- In general, what is the reason why we can make predictions in the first place?
- For example, why are we sure that if we drop a pen (or any other object), it will start falling down with the acceleration of 9.81 m/sec^2?

2. Why symmetries (cont-d)

- Because the same experiment was performed:
  - at different locations on the Earth,
  - with the experimenters facing different directions.
- In all these cases, we observed the same phenomenon.
- In other words, we can move to a different location, we can rotate – and this phenomenon will not change.
- In geometric terms, this phenomenon is symmetric with respect to shifts and rotations.
- Such situations are ubiquitous.
- In all these situations, we have some transformations that do not change the corresponding physical process.
- In physics, such transformations are known as *symmetries*.

3. Why symmetries (cont-d)

- In ordinary speech, we most apply this word to invariance under geometric transformations such as rotations or reflections in a plane.
- However, in physics, this term is applied to any transformations, not necessarily geometric.
- For example, electromagnetic interactions will not change if we replace all positive charges with negative ones and vice versa.
- In physics terms, this is symmetry.
- The idea of symmetry can be (and is) applied to even more complex situations than these.
- For example, how do we know that Ohm's law will hold when students do experiments in our lab?
- Because this law was followed on many different locations on Earth – and even on many locations outside the Earth surface.

4. Symmetries are one of the main tools in modern physics

- Physicists fully understood the importance of symmetries around the 1960s, and this understanding revolutionized physics.
- Even the way physical theories are proposed has changed.
- In the past, starting from Newton, new physical theories were proposed in terms of differential equations.
- Nowadays, starting with quarks, new theories are described by listing the corresponding symmetries.
- And differential equations can be then derived from these symmetries.
- This is not only about new theories.
- It turned out that most fundamental physical theories can be reformulated in terms of the corresponding symmetries.

5. Symmetries in modern physics (cont-d)

- Examples include Maxwell's electrodynamics, quantum physics, special relativity, general relativity, etc.
- These symmetries uniquely determine the corresponding differential equations.

6. Symmetries have been useful in computing as well

- The same symmetries ideas have been applied to many application areas beyond physics.
- For example, it turned out that many empirically successful selection of intelligent techniques can be explained by symmetries:
  - selection of membership functions and "and"- and "or"-operations in fuzzy logic,
  - selecting activation functions in neural networks,
  - seclecting appropriate functions in genetic algorithms and simulated annealing, etc.
- Many empirically successful features of deep learning can also be explained by symmetries.

7. Symmetries and optimality

- In the previous section, we mentioned that symmetries can help explain the empirically successful algorithmic (and other) choices.
- Let us explain how exactly the symmetries approach can help find optimal solutions to practical problems.
- To explain this, we first need to describe, in precise terms, what we mean by transformations and what we mean by optimality.

8. What we mean by symmetries

- If transformations g and g' both preserve the corresponding phenomenon, then:
  - if we apply them one after another – i.e., if we form what is known as a *composition* of two transformation,
  - then this composition also preserves the phenomenon.
- For example, suppose that:
  - we first go 10 meters forward – which preserves the above-described pen-dropping behavior,]
  - and then go 10 meters to the right – which also preserves this behavior.
- Then the resulting diagonal shift also preserves this behavior.
- Similarly, if we perform the inverse transformation g^{-1} – e.g., go back 10 meters – the behavior with not change.

9. What we mean by symmetries (cont-d)

- Thus, the set of phenomenon-preserving transformations must be closed under composition and under taking the inverse.
- Such sets of transformations are known as *transformation groups.*

10. What we mean by optimal

- There may be different optimality criteria, but they are all based on the possibility to compare two alternatives a and b.
- Sometimes, we can conclude that a is better; we will denote it by $a > b$.
- Sometimes, we can conclude that b is better, i.e., that $b > a$.
- Sometimes, we can conclude that a and b are, from the decision maker's viewpoint, of equal value: we will denote it by $a \sim b$.
- And sometimes, we cannot meaningfully compare the two alternatives.
- In other words, to describe what is optimal, we need to have two binary relations $>$ and $\sim$.
- Of course, these relations must be consistent.
- For example, if a is better than b and b is better than c, then a should be better than c.

11. What we mean by optimal (cont-d)

- In these terms, an alternative a is optimal:
  - if it is either better than or of the same quality as all other alternatives b,
  - i.e., if for every b, we have either $a > b$ or $a \sim b$.
- What if there are several different optimal alternatives which are equally good with respect to the given optimality criterion?
- This means that we can now use some other criterion to optimize something else.
- For example, if we are looking for the most healthy diet, and we came up with several equally healthy diets.
- Then a natural idea is to select the one which is the tastiest.
- If we have several algorithms that lead to the same optimal accuracy, we can select the one which is, e.g., the fastest.

12. What we mean by optimal (cont-d)

- In general, if an optimality criterion does not lead to a unique optimal alternative, this means that this criterion is not final.
- We can use this non-uniqueness to optimize something else.
- So, it makes sense to consider *final* optimality criteria, for which there is exactly one optimal alternative.

13. The optimality criterion should be invariant

- Suppose that there is a transformation that preserves the corresponding phenomenon.
- Then which alternative is better should not change if we apply this transformation.
- For example, the laws of electrodynamics do not change under spatial shift.
- So, we expect that:
  - the relative quality of two electric devices will remain the same
  - if we simply move both devices the same distance in the same direction.

14. Definitions

- Now we are ready to formulate all this in precise terms, and to finally explain how symmetries lead to optimality.
- Let A be a set. We will call its elements *alternatives*.
- By an *optimality criterion* on the set A, we mean a pair of relations $(<, \sim)$ for which the following conditions are satisfied for all a, b, c:
  - if $a > b$ and $b > c$, then $a > c$;
  - if $a > b$ and $b \sim c$, then $a > c$;
  - if $a \sim b$ and $b > c$, then $a > c$;
  - if $a \sim b$ and $b \sim c$, then $a \sim c$;
  - if $a \sim b$, then $b \sim a$;
  - we always have $a \sim a$;
  - we cannot have $a > b$ and $a \sim b$.

15. Definitions (cont-d)

- We say that an alternative a_{opt} is *optimal* if for every $a \in A$, we have either $a_{\text{opt}} > a$ or $a_{\text{opt}} \sim a$.
- We say that the optimality criterion is *final* if for this criterion, there is exactly one optimal alternative.
- Let G be a transformation group.
- We say that the optimality criterion is G-*invariant* if for all $a, b \in A$ and $g \in G$, we following conditions hold:
  - if $a > b$, then $g(a) > g(b)$; and
  - if $a \sim b$, then $g(a) \sim g(b)$.

16. Result

- *Proposition.*
  - Suppose that an alternative a_{opt} is optimal with respect to some final G-invariant optimality criterion.
  - Then this alternative is also G-invariant, i.e., $g(a_{\text{opt}}) = a_{\text{opt}}$ for all $g \in G$.
- Thus, to find the optimal alternative, we only need to consider G-invariant (*symmetric*) alternatives.
- And if there is only one such invariant alternative, then this is exactly the optimal one.

17. Proof of the Proposition

- Let us show that for every g, we indeed have $g(a_{\text{opt}}) = a_{\text{opt}}$.
- Indeed, by definition of optimality, for every alternative a, we have either $a_{\text{opt}} > a$ or $a_{\text{opt}} \sim a$.
- In particular, this is true for each alternative $g^{-1}(a)$: for every a, we have either $a_{\text{opt}} > g^{-1}(a)$ or $a_{\text{opt}} \sim g^{-1}(a)$.
- Since the optimality criterion is G-invariant, for all a, we have either $g(a_{\text{opt}}) > g(g^{-1}(a)) = a$ or $g(a_{\text{opt}}) \sim g(g^{-1}(a)) = a$.
- Thus, for every a, have either $g(a_{\text{opt}}) > a$ or $g(a_{\text{opt}}) \sim a$.
- By definition of the optimal alternative, this means that the alternative $g(a_{\text{opt}})$ is optimal.
- But since our optimality criterion is final, this means that it has only optimal alternative.
- Thus indeed $g(a_{\text{opt}}) = a_{\text{opt}}$. The statement is proven.

18. What we do in this paper

- In this paper, we apply the symmetry approach to intuitionistic and interval-values fuzzy logics.
- So, let us briefly recall these two ideas.
- First, let us briefly recall traditional fuzzy logic.
- In the early 1960s, Lotfi Zadeh was one of the leading specialists in control.
- He noticed that in many cases:
  - skilled human controllers control systems – like chemical plants – more effectively
  - than the supposedly optimal automatic controllers.
- Of course, one cannot be better than the optimal, so this simply meant that the supposedly optimal controllers were not really optimal.
- To be more precise, they were optimal with respect to the model used in their design.

19. What we do in this paper (cont-d)

- However, this model was not always an adequate description of the actual physical systems.
- In other words, skilled controllers have some additional knowledge that was not incorporated into the existing models.
- Skilled controllers were willing to share this knowledge.
- However, unfortunately they could only describe it by using imprecise ("fuzzy") words from natural language such as "small".
- To describe this knowledge, Zadeh designed special techniques – that he called *fuzzy*.

20. What we do in this paper (cont-d)

- Specifically, to describe a word like "small", he suggested to ask the experts to mark:
  - for each possible value x of the corresponding quantity,
  - the degree $m(x)$ – on the scale from 0 to 1 – to which the expert believes that this value satisfies the given property
  - e.g., to which x is small.
- This technique has indeed led to many practical successes.
- In practice, we also use:
  - logical combinations of statements,
  - i.e., complex statement obtained by applying logical connectives like "and", "or", and "not".

21. What we do in this paper (cont-d)

- To estimate our degree of confidence in such statements, we need to extend the usual logical operations:
  - from the set $\{0, 1\} = \{$"false", "true"$\}$
  - to the whole interval $[0, 1]$.
- Such extensions are known as, correspondingly:
  - "and"-operations $f_{\&}(a, b)$ (also called *t-norms*),
  - "or"-operations $f_{\vee}(a, b)$ (also called *t-conorms*), and
  - negation operations $f_{\neg}(a)$.
- The simplest such operations are $f_{\&}(a, b) = \min(a, b)$, $f_{\vee}(a, b) = \max(a, b)$, and $f_{\neg}(a) = 1 - a$.

22. Limitation of the traditional fuzzy techniques

- One of the limitations of the traditional fuzzy technique is that it cannot distinguish between two completely different situation:
  - when we know nothing about a statement, and
  - when we have a large number of arguments for the statement and an equal number of arguments against it.
- In both cases, we have equal number of arguments for and against.
- So it makes sense to describe each of these situations:
  - by a number which is equidistant from both 1 ("definitely yes") and 0 ("definitely no"),
  - i.e., by the number 0.5.
- To overcome this limitation, two different techniques were developed: intuitionistic fuzzy logic and interval-valued fuzzy logic.

23. Intuitionistic fuzzy logic

- In the above situation, a natural idea is to explicitly take into account how many arguments we have for and again this statement.
- In the first case, we have no arguments for each side, so we can say that $m_+(x) = m_-(x) = 0$, while
- In the second case, we have many arguments for both side, so we can take $m_+(x) = m_-(x) = 0.5$.
- In general, instead of a single value $m(x)$, we have two values:
  - the value $m_+(x)$ describes to what extend the expert is confident that the given statement is true, and
  - the value $m_-(x)$ describes to what extend the expert is confident that the given statement is false.
- This idea – first proposed by Krassimir Atanassov – is known as *intuitionistic fuzzy* approach.

24. Intuitionistic fuzzy logic (cont-d)

- Its name came from intuitonistic logic, in which some statements are neither true nor false.
- This makes sense, e.g., in constructive interpretation, where:
  - $\forall x A(x)$ would mean that we have an algorithm that proves $A(x)$ for each x, and
  - its negation would mean that we can algorithmically construct an object x for which this property does not hold.
- Such algorithms do not always exist.
- Logical operations can be naturally extended to this case:

$$f_{\&}((a_+, a_-), (a_+, a_-)) = (\min(a_+, b_+), \max(a_-, b_-)),$$

$$f_{\vee}((a_+, a_-), (a_+, a_-)) = (\max(a_+, b_+), \min(a_-, b_-)), \text{ and}$$

$$f_{\neg}((a_+, a_-)) = (a_-, a_+).$$

25. An alternative approach: interval-valued fuzzy logic

- Another approach to overcoming the above-described limitation of the traditional fuzzy logic is to take into account that:
  - in the above two situations,
  - we may eventually end up with different degrees:
- in the first case, after we gain more information about the statement, we may end up in any degree of confidence:
  - from 0 when all future evidence will be against the statement,
  - to 1, when all future evidence will be for this statement.
- So, the set of all possible values of the future expert's degree of confidence is the whole interval $[0, 1]$.
- In the second case, only the value 0.5 is possible – which corresponds to the degenerate interval $[0.5, 0.5]$.
- In general, we can have all possible intervals $[\underline{a}, \overline{a}] \subseteq [0, 1]$.

26. Interval-valued fuzzy logic (cont-d)

- This idea was proposed by Zadeh himself and then actively developed by Jerry Mendel and his group.
- Logical operations can be naturally extended to such case.
- For each operation $f(a, b)$, we can take:

$$f([\underline{a}, \overline{a}], [\underline{b}, \overline{b}]) \stackrel{\text{def}}{=} \{f(a, b) : a \in [\underline{a}, \overline{a}], b \in [\underline{b}, \overline{b}]\}.$$

- As a result, we get the following operations:

$$f_{\&}([\underline{a}, \overline{a}], [\underline{b}, \overline{b}]) = [\min(\underline{a}, \underline{b}), \min(\overline{a}, \overline{b})];$$

$$f_{\vee}([\underline{a}, \overline{a}], [\underline{b}, \overline{b}]) = [\max(\underline{a}, \underline{b}), \max(\overline{a}, \overline{b})];$$

$$f_{\neg}([\underline{a}, \overline{a}]) = [1 - \overline{a}, 1 - \underline{a}].$$

27. Interestingly, from the purely mathematical viewpoint, these two approaches are equivalent

- Let us start with the well-known fact:
  - while the ideas behind these two approaches are different,
  - from the purely mathematical viewpoint, these two approaches are equivalent.
  - in the sense that we can transform them into one another while preserving all three logical operations.
- This can be attained when we:
  - assign, to each intuitionistic degree (a_+, a_-), the interval $[a_+, 1 - a_-]$ and, vice versa,
  - assign, to each interval $[\underline{a}, \overline{a}]$, the intuitionistic degree $(\underline{a}, 1 - \overline{a})$.

28. However, from the computational viewpoint, the intuitionistic approach is more efficient

- Indeed, in the intuitionsitic approach, all we need is min and max.
- In the interval-valued approach, we also need the subtraction $1 - a$.
- To perform a substraction of an n-bit number a, we need to flip all n bits.
- So, we need n bit operations, which for 64-bit words means 64 bit operations.
- In contrast, to compute min or max, we only need to decide:
  - which of the two numbers to use.
  - i.e., which of the two numbers is larger.
- In half of the cases, the first (largest) bits are different, so we get our answer already after 1 bit comparison.

29. Intuitionistic approach is more efficient (cont-d)

- In situations in which their first bits are the same:
  - in half of these cases – i.e., in 1/4 of overall cases.
  - the second bits are different, so we get the answer after 2 bit operations, etc.
- In general, with probability 2^{-k}, we need k bit operations.
- So the overall expected number of bit operations is equal to

$$s = 2^{-1} \cdot 1 + 2^{-2} \cdot 2 + \ldots + 2^{-k} \cdot k + \ldots$$

- To compute this sum, let us divide all the terms by 2, then we get:

$$\frac{s}{2} = 2^{-2} \cdot 1 + 2^{-3} \cdot 2 + \ldots + 2^{-k} \cdot (k-1) + 2^{-(k+1)} \cdot k + \ldots$$

- If we subtract the second sum from the first one, for each k, we get the term $2^{-k} \cdot k - 2^{-k} \cdot (k-1) = 2^{-k}$.

30. Intuitionistic approach is more efficient (cont-d)

- So the overall difference takes the following form:

$$\frac{s}{2} = 2^{-1} + 2^{-2} + \ldots + 2^{-k} + \ldots$$

- The right-hand side is a well-known geometric progression whose sum is 1. So, $s/2 = 1$ and $s = 2$.
- Thus, each min or max operation requires, on average, 2 bit operations.
- Thus, it is 32 times fewer than 64 bit operations needed to compute $1 - a$.
- Thus, intuitionistic fuzzy logic is indeed computationally much more efficient than its interval-valued form.

31. How can we explain it in terms of symmetries?

- When we do not known whether the statement S is true or its negation $\neg S$ is true, we can view both S or $\neg S$ as the main statement.
- In other words, we have a natural symmetry: we can swap a statement and its negation.
- Intuitionistic fuzzy logic is invariant with respect to this swap.
- All we have to do to perform this operation is to swap the two degrees, i.e., to go from (a_+, a_-) to (a_-, a_+).
- No arithmetic operations are needed.
- In contrast, in the interval-valued approach, a similar swap requires two negations.
- Our general result says that for any invariant optimality criterion, the optimal alternative is also invariant.

32. How can we explain it in terms of symmetries (cont-d)

- Clearly:
  - comparing computational complexity – i.e., average running time – of the logical operations
  - should not depend on whether we consider the statement or its negation as basic.
- So, this criterion is invariant with respect to the swap – and the only swap-invariant representation is the intuitionstic fuzzy one.
- So, symmetries explain why the intuionistic fuzzy techniques are computationally more efficient than their interval-values reformulation.

33. Acknowledgments

This work was supported in part:

- by the US 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),
  - EAR-2225395 (Center for Collective Impact in Earthquake Science C-CIES),
- by the AT&T Fellowship in Information Technology, and
- by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) Focus Program SPP 100+ 2388, Grant Nr. 501624329,