Why Sine Membership Functions

Sofia Holguin[1], Javier Viaña[2], Kelly Cohen[2],
Anca Ralescu[3], and Vladik Kreinovich[1]

[1]University of Texas at El Paso, El Paso, Texas 79968
seholguin2@miners.utep.edu, vladik@utep.edu

[2]Department of Aerospace Engineering & Engineering Mechanics
[3]Department of Computer Science
University of Cincinnati, Cincinnati, OH 45219, USA,
[3]javier.viana.ai@gmail.com, cohenky@ucmail.uc.edu
[4]ralescal@ucmail.uc.edu

1. Formulation of the Problem

- As all passengers know, passenger flows at the airports fluctuate widely hour by hour and day by day.
- To avoid delays, it is important to predict the passenger flow as accurately as possible.
- Many methods have been applied to make such predictions, including machine learning.
- However, the resulting predictions are still far from perfect.
- Interestingly:
  - human professionals can often make better predictions that even the most complex models,
  - because they use their experience and their knowledge.
- It is therefore desirable to incorporate this knowledge into the predictions systems.

2. Formulation of the Problem (cont-d)

- This human expertise is usually formulated:
  - in terms of imprecise ("fuzzy") words from natural language
  - such as "high flow", "most probably", etc.
- To capture such knowledge, it is reasonable to use special techniques developed by Zadeh to use such knowledge; fuzzy techniques.
- In this technique:
  - for each imprecise statement like "the flow is high", and
  - for each specific value x of the corresponding quantity (e.g., of the passenger flow),
  - we ask the expert to indicate, on a scale from 0 to 1, the degree to which this statement is satisfied for this particular value x.
- In the ideal case, we elicit, from the expert, the degree $\mu(x)$ corresponding to each possible value x.
- The resulting function is known as the *membership function*.

3. Formulation of the Problem (cont-d)

- Of course, for each quantity, there are many possible values – thousands, millions, sometimes, infinitely many.
- We cannot ask thousands of questions to the experts.
- So, a reasonable idea is:
  - to select a few-parametric family of functions, as a few questions, and then
  - find the values of the corresponding parameters that best fit the answers.
- In a previous paper, one of the authors (JV) tried different families of membership functions: triangular, trapezoid, Gaussian, etc.

4. Formulation of the Problem (cont-d)

- It turns out that the best results were obtained when he used the sine membership function, i.e., the function of the form

  $\mu(x) = \sin(b\cdot(x+\varphi))$ when $-\dfrac{\pi}{2} \le b\cdot(x+\varphi) \le \dfrac{\pi}{2}$ and $\mu(x) = 0$ otherwise.

- A natural question is: how can we explain this empirical fact?
- In this talk, we provide a possible explanation for this result.

5. What we want from a membership function

- Usually, for each imprecise property P like "medium size":
  - if the value x of the corresponding quantity is too small,
  - this quantity is clearly not medium size.
- Similarly, if the value of the quantity is too large, this quantity is also clearly not medium size.
- Thus, for such properties:
  - the set of values x for which this property is to some extent satisfied – i.e., for which the corresponding degree is positive $\mu(x) > 0$
  - is bounded both from below and from above.
- In other words, there exists an interval $[\underline{x}, \overline{x}]$ such that for all the values x outside this interval, we have $\mu(x) = 0$.

6. What we want from a membership function (cont-d)

- Also, for most imprecise properties:
  - if two values x and y are close,
  - then the degrees $\mu(x)$ and $\mu(y)$ to which the given property is satisfied for these two values should also be close.
- For example:
  - if someone with height 180 cm is tall,
  - then someone whose height is close to 180 cm should be considered – to a large extent – tall.
- In other words:
  - when the difference $\Delta x = y - x$ is small,
  - the difference $\mu(x + \Delta x) - \mu(x)$ should also be small – and the order of Δx.

7. What we want from a membership function (cont-d)

- A natural precise formulation of this property is that the function $\mu(x)$ should be differentiable within the interval $(\underline{x}, \overline{x})$.
- Indeed, for differentiable functions, for small Δx, we have

$$\mu(x + \Delta x) - \mu(x) \approx \mu'(x) \cdot \Delta x.$$

- Here, as usual, $\mu'(x)$ denotes the derivative.
- Differentiable functions are continuous.
- So at the endpoints $\underline{x}$ and $\overline{x}$, we must have $\mu(\underline{x}) = \mu(\overline{x}) = 0$.

8. How to describe a class of membership functions

- It is well known that:
  - to describe general vectors v in an n-dimensional vector space (also known as linear space),
  - we can select a *basis* – i.e., n linearly independent vectors

$$e_1, \ldots, e_n,$$

  - and use the fact that every vector in this space can be represented as a linear combination of these vectors:

$$v = c_1 \cdot e_1 + \ldots + c_n \cdot e_n.$$

- Functions also form a linear space.
- We can define their sum as the function $f(x) + g(x)$ and the product of a function $f(x)$ and a number c as $c \cdot f(x)$.
- The only difference is that the resulting space is infinite-dimensional.

9. How to describe a class of membership functions (cont-d)

- Thus, to represent a general function, we can:
  - select a basis $e_1(x), \ldots, e_n(x), \ldots$ in the space of functions, and
  - represent every function as a linear combination of the basis functions:

$$f(x) = c_1 \cdot e_1(x) + \ldots + c_n \cdot e_n(x) + \ldots$$

- For example, we can take $e_1(x) = 1$, $e_2(x) = x$, $\ldots$, $e_n(x) = x^{n-1}$, $\ldots$
- This will correspond to Taylor series.
- Alternatively, we can take, as basis functions, sines and cosines – this corresponds to Fourier transform, etc.
- Since we are interested in differentiable functions, it makes sense to select differentiable functions $e_i(x)$.
- To describe a general function, we need to select the values of infinitely many parameters c_1, c_2, $\ldots$

10. How to describe a class of membership functions (cont-d)

- Of course, in a computer, at any given moment of time, we can only store finitely many values.
- So, to represent a function in a computer, we can only use finitely many parameters.
- In this case, approximating functions take the form

$$c_1 \cdot e_1(x) + \ldots + c_n \cdot e_n(x).$$

- In other words, we consider the following set of approximating functions – the set $\{c_1 \cdot e_1(x) + \ldots + c_n \cdot e_n(x)\}$, where:
  - the functions $e_i(x)$ are fixed, and
  - the coefficients $c_1, \ldots, c_n$ can take any real values.

11. Shift-invariance

- For many quantities like temperature or time, there is no fixed starting point;
  - if we select a different starting point – e.g., for time – which is a moments earlier,
  - then all numerical values x will be replaced by new values $x + a$.
- In mathematics, this transition $x \mapsto x + a$ is called a *shift*.
- After this shift, the original basis functions $e_i(x)$ will take the form

$$e_i(x + a).$$

- It is reasonable to require that the approximating family not change if we perform this shift:

$$\{c_1 \cdot e_1(x) + \ldots + c_n \cdot e_n(x)\} = \{c_1 \cdot e_1(x + a) + \ldots + c_n \cdot e_n(x + a)\}.$$

- Families that satisfy this property – i.e., that do not change (= are invariant) under shift are known as *shift-invariant*.

12. Let us consider the simplest possible family

- The larger the number of parameters n, the more complex (and thus, time-consuming) data processing; so:
  - to simplify and speed up computations,
  - it is reasonable to select a representation with the smallest possible number of parameters, i.e., with the smallest possible value n.
- So, we arrive at the following problem: find the smallest possible value n and differentiable functions $e_1(x), \ldots, e_n(x)$ for which:
  - the set of linear combinations is shift-invariant and
  - this set includes a non-zero function $f(x)$ for which

$$f(\underline{x}) = f(\overline{x}) = 0.$$

13. When is a family shift-invariant

- Scale invariance of the family means that:
  - if we take any function $f(x)$ from this family,
  - then, for every a, the shifted function $f(x+a)$ should also belong to this family.
- In particular, all the functions $e_1(x)$, ..., $e_n(x)$ belong to the family.
- So, the shifted functions $e_i(x+a)$ should also belong to this family.
- This means that for every i and for every a, there exist coefficients $c_{ij}(a)$ depending on i and a for which

$$e_i(x+a) = c_{i1}(a) \cdot e_1(x) + \ldots + c_{in}(a) \cdot e_n(x).$$

- We know that the functions $e_i(x)$ are differentiable.
- Let us show that the functions $c_{ij}(a)$ are differentiable too.
- Indeed, let us select any n different values $x_1, \ldots, x_n$.

14. When is a family shift-invariant (cont-d)

- Then to find n unknown values $c_{ij}(a)$, we get a system of n linear equations:

$$e_i(x_1 + a) = c_{i1}(a) \cdot e_1(x_1) + \ldots + c_{in}(a) \cdot e_n(x_1);$$

$$\ldots$$

$$e_i(x_n + a) = c_{i1}(a) \cdot e_1(x_1) + \ldots + c_{in}(a) \cdot e_n(x_n).$$

- It is known, from linear algebra, that:
  - the solution $c_{ij}(a)$ to this system is a linear combination of the left-hand sides,
  - namely, the product of the inverse matrix to the matrix $\|e_i(x_j)\|$ and the vector of the left-hand sides.
- Since the functions $e_i(x_j + a)$ are differentiable, their linear combinations $c_{ij}(a)$ are also differentiable.
- So all the functions $e_i(x)$ and $c_{ij}(a)$ are differentiable.

15. When is a family shift-invariant (cont-d)

- Thus, we can differentiate both sides of the above equality with respect to a, then we get the following:

$$e_i'(x+a) = c_{i1}'(a) \cdot e_1(x) + \ldots + c_{in}'(a) \cdot e_n(x).$$

- In particular, for $a = 0$, we get:

$$e_i'(x) = C_{i1} \cdot e_1(x) + \ldots + C_{in} \cdot e_n(x), \text{ where } C_{ij} \stackrel{\text{def}}{=} c_{ij}'(0).$$

- So, we get a system of linear differential equations with constant coefficients.
- It is known that a general solution to this system of equations is a linear combination of the terms $x^k \cdot \exp(\lambda \cdot x)$, where:
  - λ is an eigenvalue of the matrix $\|C_{ij}\|$ (which is, in general, a complex number), and
  - the integer k is smaller than the multiplicity of this eigenvalue.

16. When is a family shift-invariant (cont-d)

- Thus, each function $e_i(x)$ is equal to such a linear combination.
- Hence, all the functions from the family (1), which are themselves linear combinations of the functions $e_i(x)$, also have this form.

17. What is the smallest possible shift-invariant family

- In general, the smallest possible family corresponds to $n = 1$.
- In this case, according to the above description, all the functions from the family (has the form $C \cdot \exp(\lambda \cdot x)$.
- However, we want the desired function to be equal to 0 when $x = \underline{x}$ and when $x = \overline{x}$.
- But the function $C \cdot \exp(\lambda \cdot x)$ is not equal to 0 anywhere – unless, $C = 0$, in which case the function is just everywhere equal to 0.
- So, we cannot have $n = 1$.
- Let us therefore consider the next simplest case $n = 2$.
- In this case, we can have:
  - either one eigenvalue λ – in which case it must be a real number,
  - or two different eigenvalues $\lambda_1 \neq \lambda_2$ – in which case they can be either real or complex-valued.

18. What is the smallest possible shift-invariant family (cont-d)

- In the first case, we conclude that a general function from the family has the form $C_1 \cdot \exp(\lambda \cdot x) + C_2 \cdot x \cdot \exp(\lambda \cdot x)$, i.e., the form
$$(C_1 + C_2 \cdot x) \cdot \exp(\lambda \cdot x).$$
- In this case, this function is equal to 0 if $C_1 + C_2 \cdot x = 0$, i.e., only at one point $x = -C_1/C_2$.
- However, we want the desired membership function to be equal to 0 at two different points $\underline{x}$ and $\overline{x}$.
- Thus, this case is not possible.
- Let us now consider the case when we have two different *real* eigenvalues λ_1 and λ_2.
- In this case, a general function from the family has the form
$$f(x) = C_1 \cdot \exp(\lambda_1 \cdot x) + C_2 \cdot \exp(\lambda_2 \cdot x).$$

19. What is the smallest possible shift-invariant family (cont-d)

- For this function, the equation $f(x) = 0$ takes the form
$$C_1 \cdot \exp(\lambda_1 \cdot x) + C_2 \cdot \exp(\lambda_2 \cdot x) = 0.$$
- If we move the first term in the left to the right-hand side and divide both sides by $C_2 \cdot \exp(\lambda_1 \cdot x)$, we get $\exp((\lambda_2 - \lambda_1) \cdot x) = -C_1/C_2$.
- By taking logarithm of both sides and dividing the resulting equality by $\lambda_2 - \lambda_1$, we get $x = \dfrac{\ln(-C_1/C_2)}{\lambda_1 - \lambda_1}$.
- Thus, the function $f(x)$ is equal to 0 only at one point.
- However, we are looking for a function that is equal to 0 at two different points.
- So, for $n = 2$, the only remaining case is when both eigenvalues λ_1 and λ_2 are *complex* numbers, with non-zero imaginary parts.

20. What is the smallest possible shift-invariant family (cont-d)

- It is known that:
  - if $\lambda = a + b \cdot \mathrm{i}$, where $\mathrm{i} \stackrel{\text{def}}{=} \sqrt{-1}$, is an eigenvalue of a real-valued matrix,
  - then its complex conjugate $\lambda^* \stackrel{\text{def}}{=} a - b \cdot \mathrm{i}$ is also an eigenvalue of this matrix.
- Thus, the two eigenvalues are complex conjugates to each other, i.e.., $\lambda_1 = a + b \cdot \mathrm{i}$ and $\lambda_2 = a - b \cdot \mathrm{i}$ for some real numbers a and b.
- In this case, the general form of a function from the family has the form: $C_1 \cdot \exp((a + b \cdot \mathrm{i}) \cdot x) + C_2 \cdot \exp((a - b \cdot \mathrm{i}) \cdot x)$.
- In general, $\exp((a + b \cdot \mathrm{i}) \cdot x) = \exp(a \cdot x) \cdot \exp(b \cdot \mathrm{i} \cdot x)$.

21. What is the smallest possible shift-invariant family (cont-d)

- Since $\exp(\mathrm{i} \cdot z) = \cos(z) + \mathrm{i} \cdot \sin(z)$, we get

$$f(x) = \exp(a \cdot x) \cdot (A \cdot \cos(b \cdot x) + B \cdot \sin(b \cdot x)).$$

- Here $A \stackrel{\text{def}}{=} C_1 + C_2$ and $B \stackrel{\text{def}}{=} C_1 - C_2$.
- The trigonometric part can be equivalently described as

$$C \cdot \sin(b \cdot (x + \varphi)) \text{ for some value } \varphi, \text{ where } C \stackrel{\text{def}}{=} \sqrt{A^2 + B^2}.$$

22. Resulting explanation of the empirical success of sine membership functions

- So, we conclude all the functions from the corresponding family – including the desired membership function – have the form

$$f(x) = C \cdot \exp(a \cdot x) \cdot \sin(b \cdot (x + \varphi)).$$

- For $a = 0$, we get exactly the desired form of the membership function.
- In this case, the fact that the maximum value of the membership function should be equal to 1 implies that $C = 1$.
- In the general case, when the value a may be different from 0, we get a more general expression.
- This expression may be useful in some applications.

23. Conclusions

- In several practical problems – e.g., in predicting passenger flow at the airports:
  - fuzzy techniques are most effective
  - when the corresponding terms are interpreted as sine-shaped membership functions.
- In this talk, we provide a theoretical explanation for this empirical success.
- This explanation is based on the fact that for many quantities like time or temperature, there is no fixed starting point.
- For example, when predicting the passenger flow:
  - we can predict the actual flow
  - or – which users prefer – the difference between the forthcoming passenger flow and the average passenger flow at this season.

24. Conclusions (cont-d)

- We prove that:
  - of all the classes of membership functions which are invariant with respect to changing the starting point,
  - the simplest is the class of sine functions.
- This explains why the sine functions work so well in this problem.
- They work better than other tested classes of membership functions – classes that do not have the desired invariance.
- To be more precise, we come up with a slightly more general family of membership functions of which sine functions are a particular case.
- We hope that these more general membership functions may be useful in some applications of fuzzy techniques.

25. 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), and
  - HRD-1834620 and HRD-2034030 (CAHSI Includes).
- It was also supported by the AT&T Fellowship in Information Technology.
- 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.