Why Sine Membership Functions

Sofia Holguin, Javier Viaña, Kelly Cohen, Anca Ralescu, and Vladik Kreinovich

Abstract In applications of fuzzy techniques to several practical problems – in particular, to the problem of predicting passenger flows in the airports – the most efficient membership function is a sine function; to be precise, a portion of a sine function between the two zeros. In this paper, we provide a theoretical explanation for this empirical success.

1 Formulation of the Problem

As all passengers know, passenger flow at the airports fluctuate widely hour by hour and day and 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; see, e.g., [3, 4, 5, 8, 12]. 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.

Sofia Holguin and Vladik Kreinovich
Department of Computer Science, University of Texas at El Paso
500 W. University, El Paso, TX 79968, USA
e-mail: seholguin2@miners.utep.edu, vladik@utep.edu

Javier Viaña and Kelly Cohen
Department of Aerospace Engineering & Engineering Mechanics
College of Engineering & Applied Science
University of Cincinnati, Cincinnati, OH 45219, USA,
e-mail: javier.viana.ai@gmail.com, cohenky@ucmail.uc.edu

Anca Ralescu
Department of Computer Science, College of Engineering & Applied Science
University of Cincinnati, Cincinnati, OH 45219, USA
e-mail: ralescal@ucmail.uc.edu

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 Lotfi Zadeh to use such knowledge – techniques of fuzzy logic (see, e.g., [1, 2, 6, 9, 10, 13]). 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*. 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 [11], one of the authors (JV) tried different families of membership functions to see which one would lead to a better prediction of the passenger flow. It turns out that the best results were obtained when he used the sine membership function (first introduced in [7]), i.e., the function of the form

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

A natural question is: how can we explain this empirical fact?

In this paper, we provide a possible explanation for this result.

2 Our Explanation

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 $\mu(x) = 0$.

Also, for most imprecise properties, if two values x_1 and x_2 are close, then the degrees $\mu(x_1)$ and $\mu(x_2)$ 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 different $\Delta x = x_2 - x_1$ is small, the difference $\mu(x + \Delta x) - \mu(x)$ between the values of the membership function for close values x and $x + \Delta x$ of the corresponding quantity should also be small – and the order of Δx. 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 $\mu(x+\Delta x)-\mu(x)\approx\mu'(x)\cdot\Delta x$, where, as usual, $\mu'(x)$ denotes the derivative.

Differentiable functions are continuous, so at the endpoints $\underline{x}$ and $\overline{x}$, we must have $\mu(\underline{x})=\mu(\overline{x})=0$.

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 $e_1,\ldots,x_n$ – 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. 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 $e_1(x)=1$, $e_2(x)=x$, $\ldots$, $e_n(x)=x^{n-1}$; 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, ... 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

$$\{c_1\cdot e_1(x)+\ldots+c_n\cdot e_n(x)\} \tag{1}$$

of all the functions of this type, where the functions $e_i(x)$ are fixed, and the coefficients $c_1,\ldots,c_n$ can take any real values.

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, i.e., that the original set of approximating functions and the set of approximating functions corresponding to the shifted basis should be the same:

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

Families that satisfy this property – i.e., that do no change (= are invariant) under shift are known as *shift-invariant.*

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 for which this set includes a non-zero function $f(x)$ for which $f(\underline{x}) = f(\overline{x}) = 0$.

When is a family shift-invariant. Scale invariance of the family (1) 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, since all the functions $e_1(x)$, $\ldots$, $e_n(x)$ belong to the family (1), 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

$$e_i(x+a) = c_{i1}(a)\cdot e_1(x)+\ldots+c_{in}(a)\cdot e_n(x). \tag{2}$$

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

Since all the functions $e_i(x)$ and $c_{ij}(a)$ are differentiable, we can differentiate both sides of the equality (1) 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:

$$e_i'(x) = C_{i1}\cdot e_1(x)+\ldots+C_{in}\cdot e_n(x), \tag{3}$$

where we denoted $C_{ij} \stackrel{\text{def}}{=} c_{ij}'(0)$.

Equations (3) for $i=1,\ldots,n$ form 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. Thus, each function $e_i(x)$ is equal to such a linear combination – and hence, all the functions from the family (1), which are themselves linear combinations of the functions $e_i(x)$, also have this form.

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 (1) 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 $x = \overline{x}$, and the function $C \cdot \exp(\lambda \cdot x)$ is not equal to 0 anywhere – unless, of course, $C = 0$, but in this 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.

In the first case, we conclude that a general function from the family (1) 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 $x = -C_1/C_2$, while 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 (1) has the form

$$f(x) = C_1 \cdot \exp(\lambda_1 \cdot x) + C_2 \cdot \exp(\lambda_2 \cdot x).$$

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

$$x = \frac{\ln(-C_1/C_2)}{\lambda_1 - \lambda_1}.$$

Thus, the function $f(x)$ is equal to 0 only at one point, while 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. 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, in this case, 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 (1) has the form

$$C_1 \cdot \exp((a + b \cdot \mathrm{i}) \cdot x) + C_2 \cdot \exp((a - b \cdot \mathrm{i})] \cdot x). \tag{4}$$

In general,

$$\exp((a + b \cdot \mathrm{i}) \cdot x) = \exp(a \cdot x) \cdot \exp(b \cdot \mathrm{i} \cdot x).$$

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

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

where $A \stackrel{\text{def}}{=} C_1 + C_2$ and $B \stackrel{\text{def}}{=} C_1 - C_2$. The trigonometric part can be equivalently described as $C \cdot \sin(b \cdot (x + \varphi))$ for some value φ, where $C \stackrel{\text{def}}{=} \sqrt{A^2 + B^2}$.

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

$$f(x) = C \cdot \exp(a \cdot x) \cdot \sin(b \cdot (x + \varphi). \tag{5}$$

For $a = 0$, we get exactly the desired form of the membership function – indeed, 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 – which may be useful in some applications.

Acknowledgments

This work was supported by:

- Fellowship for Postgraduate Studies in North America from “La Caixa” Banking Foundation, ID 100010434, grant LCF/BQ/AA19/11720045,
- Ohio State Excellence Scholarship & Recognition Grant, sponsored by the Hispanic Chamber of Commerce, Cincinnati, Ohio, USA,
- the Airport Cooperative Research Program Graduate Research Award, sponsored by the Federal Aviation Administration, administered by the Transportation Research Board and The National Academy of Sciences, and managed by the Virginia Space Grant Consortium.
- 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),
- the AT&T Fellowship in Information Technology,

• the program of the development of the Scientific-Educational Mathematical Center of Volga Federal District No. 075-02-2020-1478, and
• grant from the Hungarian National Research, Development and Innovation Office (NRDI).

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. G. Klir and B. Yuan, *Fuzzy Sets and Fuzzy Logic*, Prentice Hall, Upper Saddle River, New Jersey, 1995.
3. Ziyu Li, Jun Bi, and Zhiyin Li, "Passenger flow forecasting research for airport terminal based on SARIMA time series model", In: *IOP Conference Series. Earth and Environmental Science*, 2017, Vol. 100, Paper 012146.
4. Lijuan Liu and Rung-Ching Chen, "A novel passenger flow prediction model using deep learning methods", *Transportation Research. Part C, Emerging Technologies*, 2017, Vol. 84, pp. 74–91.
5. Xia Liu et al., "Prediction of passenger flow at Sanya Airport based on combined methods", in: *Data Science*, Springer, Singapore, 2017, pp. 729–740.
6. J. M. Mendel, *Uncertain Rule-Based Fuzzy Systems: Introduction and New Directions*, Springer, Cham, Switzerland, 2017.
7. K. Miyajima and A. Ralescu, "Spatial organization in 2D segmented images: representation and recognition of primitive spatial relations", *Fuzzy Sets and Systems*, 1994, Vol. 65, No. 2–3, pp. 225–236.
8. K. Munasingh and V. Adikariwattage, "Discrete event simulation method to model passenger processing at an international airport", *IEEE Proceedings of the 2020 Moratuwa Engineering Research Conference MERCon*, Moratuwa, Sri Lanka, July 28–30, 2020, pp. 401–406.
9. H. T. Nguyen, C. L. Walker, and E. A. Walker, *A First Course in Fuzzy Logic*, Chapman and Hall/CRC, Boca Raton, Florida, 2019.
10. V. Novák, I. Perfilieva, and J. Močkoř, *Mathematical Principles of Fuzzy Logic*, Kluwer, Boston, Dordrecht, 1999.
11. J. Viaña, *Explainable AI Process for Passenger Flow Prediction at the Security Checkpoint of the Airport*, United States Patent and Trademark Office. Approved US Provisional Patent. Application No. 63/232,782 filed on August 13, 2021.
12. Gang Xie, Shouyang Wang, and Kin Keung Lai, "Short-term forecasting of air passenger by using hybrid seasonal decomposition and least squares support vector regression approaches", *Journal of Air Transport Management*, 2014, Vol. 37, pp. 20–26.
13. L. A. Zadeh, "Fuzzy sets", *Information and Control*, 1965, Vol. 8, pp. 338–353.