"Environmental Engineering" 10th International Conference
Vilnius Gediminas Technical University
Lithuania, 27–28 April 2017
http://enviro.vgtu.lt

eISSN 2029-7092 / eISBN 978-609-476-044-0

Article ID: enviro.2017.055
DOI: https://doi.org/10.3846/enviro.2017.055

Calculation of the Probability Integral Indicator of the Level of Air Pollution

Tunakova Yulia Alekseevna[1], Novikova Svetlana Vladimirovna[2], Shagidullin Artur Rifgatovich[3]

[1,2]Kazan National Research Technical University named after A.N.Tupolev – KAI
10, K.Marx Str., Kazan, Republic of Tatarstan 420111, Russia
[3]Institute of Ecology and Subsurfase Tatarstan Academy of Sciences
28, Daursky Str., Kazan, Republic of Tatarstan 420087, Russia

E-mails: [1]juliaprof@mail.ru; [2]sweta72@bk.ru; [3]Artur.Shagidullin@tatar.ru

Abstract. This article describes the method for calculation of probability integral indicator of the level of air pollution. The calculation is based on information on easily identifiable meteorological parameters by using neural network technology. This is especially true in terms of space-time constraints for existing network stations for environmental monitoring system observations.

Keywords: fuzzy logical inference, membership functions, Mamdani, Takagi-Sugeno.

Conference topic: Environmental protection.

Introduction

The level of air pollution can be evaluated in two ways: quantitatively (in the form of specific values of the concentrations of toxic substances in the atmosphere) and qualitatively (in words of natural language). When the variables to calculate and research are set not quantitatively but qualitatively, we use Fuzzy logic in the form of linguistic variables. The variables expressed in words and not in numbers were first described by L. Zadeh (Zadeh 1965). Linguistic variables describe human reflection of the world as an imprecise (fuzzy). To analyse linguistic variables one of the basic concepts of mathematics has been extended – the concept of sets. For this it was introduced the definition of fuzzy set and developed a theory of fuzzy sets, which included a standard set as a special case.

Linguistic variables

In the standard set theory, there are several methods to define a set. One of them is definition using the characteristic function. Characteristic functions of set $A \in X$ is a function $\mu_A(x)$, whose values indicate whether $x \in X$ element of the set A:

$$\mu_A(x) = \begin{cases} 1, \text{ if } x \in A, \\ 0, \text{ if } x \notin A. \end{cases} \quad (1)$$

Fuzzy sets are the generalization of ordinary sets when we abandon the binary nature of this function, and we assume that it can take any value in interval [0,1].

The characteristic function of the fuzzy sets theory is called the membership function, and its value $\mu_A(x)$ – the degree of membership of element x to the fuzzy set A.

Different membership functions are used in various applications: piecewise linear, exponential (Poisson functions), Gaussian, and others.

To perform a calculation on fuzzy sets and fuzzy variables, we use fuzzy inference system. The foundation for the fuzzy inference is based on rules containing fuzzy statements in the form "IF-THEN" and the membership function for the corresponding linguistic variables. The result of fuzzy inference is a numeric value of the variable y * based on the specified numerical values x_k (k = 1..n).

Nowadays there exist two types of fuzzy inference systems: Takagi-Sugeno and Mamdani. The main difference between them is the Sugeno-system produces precise (quantitative) results as linear function value, and

Mamdani system produces qualitative result (imprecise fuzzy variable) (Klir, Yuan 1995). So if Sugeno system has fuzzy rules of the type:

$$\text{«IF } (x_1 \in A_{1j}) \text{ and } (x_2 \in A_{2j}) \text{ and ... } (x_n \in A_{nj}) \text{ THEN } y=b_{j0}+b_{j1}*x_1+b_{j2}*x_2+\dots b_{jn}*x_n\text{»}$$
$$j=1,2,\dots m; \quad (2)$$

here x_i – precise values of fuzzy variables, A_{ij} – fuzzy sets, b_{ji} – some numbers.
In Mamdani-system the same rules will be of the type:

$$\text{« IF } (x_1 \in A_{1j}) \text{ and } (x_2 \in A_{2j}) \text{ and ... } (x_n \in A_{nj}) \text{ THEN } y \in C_j\text{», } j=1,2,\dots m; \quad (3)$$

where C_j – fuzzy set of the output variable.

Method of generating Mamdani-system

If Sugeno-system is used as mechanism to calculate precise values, the algorithm Mamdani can be used for linguistic analysis of the acquired result. In this case there will be the problem of two identical systems creation: Sugeno and Mamdani, for the same input.

In this case:

1) The number of Mamdani-system input variables coincides with the number of Takagi-Sugeno input.
2) All the fuzzy terms and their membership functions of systems Takagi-Sugeno and Mamdani are identical.
3) All left parts of the rules of Takagi-Sugeno and Mamdani systems are the same.
4) The number of fuzzy sets of the output variable Mamdani and Takagi-Sugeno match. For each output variable of fuzzy terms we must define membership function.
5) The main problem lies in setting right parts of fuzzy inference rules in both systems. It is known that models of Mamdani type and Sugeno type will be exactly the same only when the conclusions of the rules are given as precise numbers. So if:
   – Fuzzy sets output variable in Mamdani model is given as singletons – fuzzy analogues of precise numbers. In this case, the degree of membership function for all the elements of the universal set equals to zero, except for one with a degree of membership being equal to one.
   – Conclusions of the rules in the knowledge base Sugeno are defined by linear functions in which all the coefficients of the input variables are equal to zero.

In the case where systems Takagi-Sugeno and Mamdani have arbitrary form (2)–(3), it is impossible to build two completely identical systems of these types. However, the two systems can be constructed, in a sense “close” to each other. Closeness is understood as the same exact response of the two systems if there appears defuzzification of Mamdani-system response.

Calculation Mamdani system parameters

If the Mamdani-system has already been built, then getting out of it the analogue Sugeno-system becomes quite simple. In some mathematical packages there are built-in functions for such a transformation. Eg, MathLab has function mam2sug (Sivanandam *et al.* 2007). However, inverse transformation is impossible, when the right parts of the system Sugeno are present, and we need to develop the right part of the output of a similar Mamdani system.

Therefore, there is a problem: to transform a linear polynomial right parts of the rules systems Takagi-Sugeno into membership functions right parts (the output variable) of Mamdani. The literature does not describe an algorithm for such a transformation. Therefore, the algorithm should be developed independently.

Generation of membership functions of Mamdani-systems analogous to the system of Sugeno could be realized on the basis of expert estimations but this method is labor-intensive and has a low level of adequacy. So it is advisable to develop an algorithm that will allow to build automatically Mamdani-system based on the existing Sugeno-system only on the basis of Sugeno characteristics.

We write down the value of the output variable for the inference Takagi-Sugeno systems with linear right parts of rules:

$$y = \frac{\sum_{j=1}^{M} w_j y_j}{\sum_{j=1}^{M} w_j}, \quad (4)$$

here: w_j – the weight (cogency) of the j-th rules when applied to the input systems the exact set $X = (x_1, x_2, \dots x_n)$; y_j –

right part of *j*-th rule (linear polynomial). In this case:

$$w_j = \prod_{i=1}^{N} \mu_i^j(x_i); \tag{5}$$

$$y_j = \sum_{i=1}^{N} t_i^j x_i + t_0^j, \tag{6}$$

$\mu_i^j(x_i)$ – membership functions of the input variables defined as Gaussians with parameters a_i^j and σ_i^j, *j* – number of inference rules, *i* – index of fuzzy sets in the rule:

$$\mu_i^j(x_i) = \exp\left[-\left(\frac{x_i - a_i^j}{\sigma_i^j}\right)^2\right]. \tag{7}$$

Then finally the output variable can be defined as:

$$y = \frac{\sum_{j=1}^{M}\left(\sum_{i=1}^{N} t_i^j x_i + t_0^j\right)\prod_{i=1}^{N}\exp\left[-\left(\frac{x_i - a_i^j}{\sigma_i^j}\right)^2\right]}{\sum_{j=1}^{M}\prod_{i=1}^{N}\exp\left[-\left(\frac{x_i - a_i^j}{\sigma_i^j}\right)^2\right]}. \tag{8}$$

If we use this way of defining the output variable algorithm Takagi-Sugeno is a universal approximator (Verbruggen, Babuska 1998; Xue *et al.* 1990).

It is natural to define system of Mamdani type fuzzy inference, which, on the one hand, would be possibly identical to the system. Takagi-Sugeno and on the other hand would be the universal approximator. The identity of the two systems is naturally understood in the "closeness" of output values for the same input parameters. It is worth considering that the system Mamdani unlike the Takagi-Sugeno gives a fuzzy result, which in the final stage should lead to definition (defazzifitsirovat). This means that the way of defuzzification will greatly affect the response of the system, and, consequently, on the degree of "closeness" to the Sugeno-system.

In 1992, L. Wang (Wang 1992) showed that Mamdani fuzzy system is universal approximator, if you use a standard set of rules of the form:

IF (x_1^j is A_1^j) and (x_2^j is A_2^j) and … (x_N^j is A_N^j) THEN (y^j is C^j)

with

1. Gaussian membership functions of the input variables:

$$\mu_{A_i}^j(x_i) = \exp\left[-\left(\frac{x_i - a_i^j}{\sigma_i^j}\right)^2\right]. \tag{9}$$

2. Gaussian membership functions of the output variable:

$$\mu_C^j(x_i) = \exp\left[-\left(\frac{x_i - c^j}{\sigma_C^j}\right)^2\right]. \tag{10}$$

3. composition in the form of a product:

$$[A_i(x_i) \text{ and } A_j(x_j)] = A_i(x_i)A_j(x_j).$$

4. implications in the form of Larsen:

$$[A_i(x_i) \text{ and } A_j(x_j)] \Rightarrow C_i(y) = A_i(x_i)A_j(x_j)C_i(y).$$

5. centroid method to bring clarity:

$$y = \frac{\sum_{j=1}^{M} c^j \prod_{i=1}^{N} \exp\left[-\left(\frac{x_i - a_i^j}{\sigma_i^j}\right)^2\right]}{\sum_{j=1}^{M}\prod_{i=1}^{N} \exp\left[-\left(\frac{x_i - a_i^j}{\sigma_i^j}\right)^2\right]}, \tag{11}$$

here c^j – centers of membership functions of output variable.

There are other ways of defining Mamdani systems with properties of universal approximator.

Thus if the system Takagi-Sugeno functions according to (8), and the system Mamdani – according to (11), both will be universal approximator. It is necessary to equate the formulas (8) and (11) to determine the characteristics of the future Mamdani-system, namely, notably centers of membership functions of output variable c^j (where the functions are defined by (10)),

$$\frac{\sum_{j=1}^{M}\left(\sum_{i=1}^{N} t_i^j x_i + t_0^j\right)\prod_{i=1}^{N} \exp\left[-\left(\frac{x_i - a_i^j}{\sigma_i^j}\right)^2\right]}{\sum_{j=1}^{M}\prod_{i=1}^{N} \exp\left[-\left(\frac{x_i - a_i^j}{\sigma_i^j}\right)^2\right]} = \frac{\sum_{j=1}^{M} c^j \prod_{i=1}^{N} \exp\left[-\left(\frac{x_i - a_i^j}{\sigma_i^j}\right)^2\right]}{\sum_{j=1}^{M}\prod_{i=1}^{N} \exp\left[-\left(\frac{x_i - a_i^j}{\sigma_i^j}\right)^2\right]}. \tag{12}$$

The denominators in both parts of the expression are equal and we use the equation to determine the parameters cj:

$$\sum_{j=1}^{M}\left(\sum_{i=1}^{N} t_i^j x_i + t_0^j\right)\prod_{i=1}^{N} \exp\left[-\left(\frac{x_i - a_i^j}{\sigma_i^j}\right)^2\right] = \sum_{j=1}^{M} c^j \prod_{i=1}^{N} \exp\left[-\left(\frac{x_i - a_i^j}{\sigma_i^j}\right)^2\right]. \tag{13}$$

Or separately for each parameter c^j:

$$\left(\sum_{i=1}^{N} t_i^j x_i + t_0^j\right)\prod_{i=1}^{N} \exp\left[-\left(\frac{x_i - a_i^j}{\sigma_i^j}\right)^2\right] = c^j \prod_{i=1}^{N} \exp\left[-\left(\frac{x_i - a_i^j}{\sigma_i^j}\right)^2\right]. \tag{14}$$

Equality (14) should be understood as equality of two functionals (the minimum distance between the functionals). To determine the parameter c^j from (11) we need to pass to the N-fold integral to both parts of the expression. Then the centers of membership function are found from the equation:

$$c^j = \frac{\int\ldots\int\left(\sum_{i=1}^{N} t_i^j x_i + t_0^j\right)\prod_{i=1}^{N} \exp\left[-\left(\frac{x_i - a_i^j}{\sigma_i^j}\right)^2\right] dx_1 dx_2 .. dx_N}{\int\ldots\int\prod_{i=1}^{N} \exp\left[-\left(\frac{x_i - a_i^j}{\sigma_i^j}\right)^2\right] dx_1 dx_2 .. dx_N} =$$

$$= \frac{\sum_{i=1}^{N} t_i^j \int\ldots\int x_i \prod_{i=1}^{N} \exp\left[-\left(\frac{x_i - a_i^j}{\sigma_i^j}\right)^2\right] dx_1 dx_2 .. dx_N}{\int\ldots\int\prod_{i=1}^{N} \exp\left[-\left(\frac{x_i - a_i^j}{\sigma_i^j}\right)^2\right] dx_1 dx_2 .. dx_N} + t_0^j. \tag{15}$$

For calculation we need to proceed to the definite integrals. To determine the limits of integration we use the rule of "three sigma". It can be interpreted as the values of a multiplicative functional defined by Gaussian beyond the distance of 3σ from the center of Gaussian, whose value is negligible in our context. Then the limits of integration for each variable x_i is defined as $[(a_i^j - 3\sigma_i^j), (a_i^j + 3\sigma_i^j)]$. Finally, we obtain a formula for determining the centers of Gaussians membership function of the output variable Mamdani-system:

$$c^j = \frac{\sum_{i=1}^{N} t_i^j \int_{a_1^j-3\sigma_1^j}^{a_1^j+3\sigma_1^j} \dots \int_{a_N^j-3\sigma_N^j}^{a_N^j+3\sigma_N^j} x_i \prod_{i=1}^{N} \exp\left[-\left(\frac{x_i-a_i^j}{\sigma_i^j}\right)^2\right] dx_1 dx_2 .. dx_N}{\int_{a_1^j-3\sigma_1^j}^{a_1^j+3\sigma_1^j} \dots \int_{a_N^j-3\sigma_N^j}^{a_N^j+3\sigma_N^j} \prod_{i=1}^{N} \exp\left[-\left(\frac{x_i-a_i^j}{\sigma_i^j}\right)^2\right] dx_1 dx_2 .. dx_N} + t_0^j. \quad (16)$$

We integrate the expression (16) by parts so thus we obtain a simplified formula for the calculation of the parameters c^j:

$$c^j = \sum_{i=1}^{N} t_i^j a_i^j + t_0^j. \quad (17)$$

Numerical experiments

Experiments on the use of expressions (16)–(17) to generate a functional form (11) show that it is equivalent to the functional (8). So for a simple system Takagi-Sugeno of the 1-st order with one input variable and one fuzzy set for the rule:

IF x∈A THEN y=3x+80,

where the membership function of the set A has the form:

$$\mu_A = \exp\left(-\frac{(x-10)^2}{4^2}\right).$$

Output variable is determined by functional:

$$y(x) = (3x+80) \times \exp\left(-\frac{(x-10)^2}{4^2}\right)..$$

The graph of this functional have the form (Fig. 1):

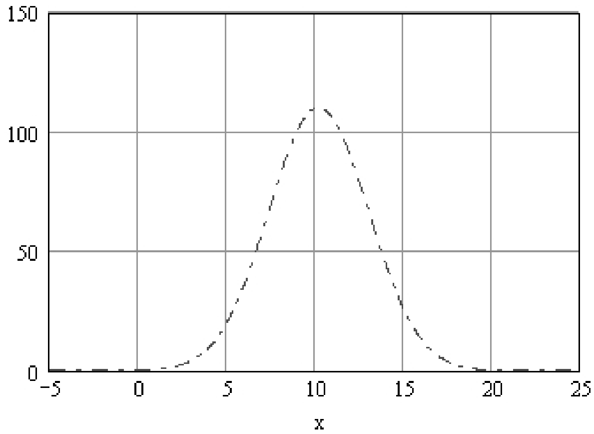

Fig. 1. Functional relationship between input and output variables Takagi-Sugeno system with one input variable

The center of the membership function calculated by the formula (17) of Mamdani-system identical to the given above resulted in: $c = 110$.

Constructed Mamdani-system with found parameters determines the output variable as (Fig. 2)

$$y(x) = 110 \times \exp\left(-\frac{(x-10)^2}{4^2}\right).$$

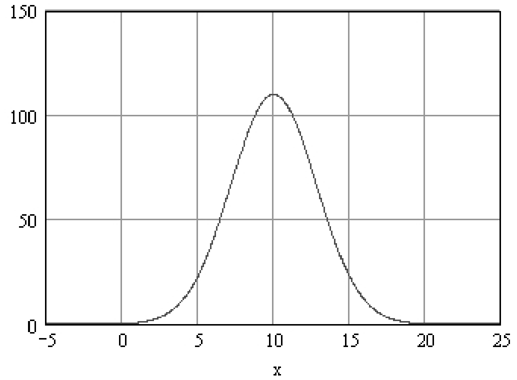

Fig. 2. The functional dependence of the input and output variables of the Mamdani-system with found center *c*

Comparison chart demonstrates their complete identity (Fig. 3)

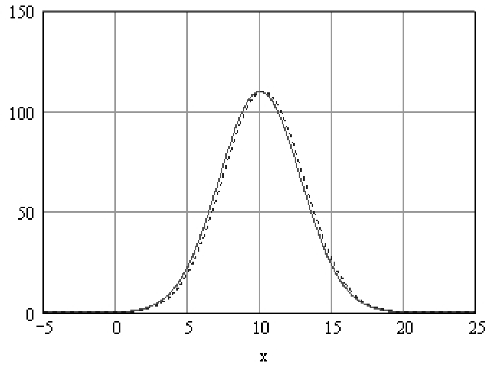

Fig. 3. Comparison of dependence charts of output variable from one input to the original system Takagi-Sugeno (dotted line), and built upon it Mamdani system (solid line)

Experiment with two input variables and two fuzzy sets in the Takagi-Sugeno system showed similar results. Characteristics of the original system Takagi-Sugeno:

– rules:

IF x∈A and y∈A THEN z=(3x+6y+7)

– membership functions of terms (sets):

$$\mu_A = \exp\left(-\frac{(x-7)^2}{2^2}\right); \ \ \mu_B = \exp\left(-\frac{(y-12)^2}{3^2}\right).$$

The calculated value of the Gaussian center of the output variables for Mamdani system:

$$c = 100.$$

Results of the comparison of output values in two systems shows Fig. 4.

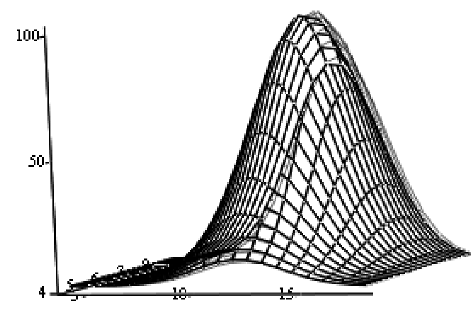

Fig. 4. Comparison of dependence charts of output variable with the two inputs in the original system Takagi-Sugeno (thin lines), and built upon it Mamdani system (thick lines)

Algorithm for obtaining a linguistic evaluation of atmospheric air

It is impossible to give definitive advice to determine the mixing σ^j of fuzz terms membership function of the output variable. Because to compare response of the Mamdani system to the response of Takagi-Sugeno, it is necessary to carry out th defuzzification of Mamdani-system response. In this process information about the dispersion might be completely lost. However, to prevent the possible data loss we can recommend to choose the dispersion so that the aggregate Gaussian membership function would fully cover the domain of the output variable $[a,b]$. To do this, use the following procedure:

1) Put in ascending order Gaussians centers ($c^1 < c^2 < ... < c^M$). Define limits of the output variable domain as: $c^0 = a$, $c^{M+1} = b$.
2) For all s from 1 to M:
   - Determine the distance from the center c^s to the previous ascending cluster center: $r_L = c^s - c^{s-1}$, and up to next higher cluster center: $r_R = c^{s+1} - c^s$.
   - Determine the dispersion for the Gaussian with center c^s as:

$$\sigma^s = \max(r_L, r_R).$$

Thus, to generate a fuzzy inference Mamdani type system based on existing system Takagi-Sugeno effectively you might apply for the following algorithm:

1) To set the number of input variables in the Mamdani system equal to the number of inputs in the Takagi-Sugeno system.
2) All the sets and their fuzzy membership functions of Takagi-Sugeno transfer to Mamdani system unchanged. It is necessary to have membership functions in the form of a Gaussian with arbitrary parameters of centers and dispersions.
3) Specify the number of fuzzy terms of the Mamdani output variable equal to the number of fuzzy sets of output variable Takagi-Sugeno. To set membership function for each fuzzy term of output variable as a Gaussian: $\mu^j(y) = \exp\left(-\frac{(y-c^j)^2}{\sigma^{j2}}\right)$
4) To evaluate the value of the Gaussians center c^j by the formula (17) proceeding from the parameters of Takagi-Sugeno system. To evaluate the values of Gaussians dispersion σ^j from the condition of uniform coverage determination under the proposed procedure. All the rules of inference Takagi-Sugeno are inference rules of the Mamdani system.

The proposed method of construction of identical systems of fuzzy inference in two types has been implemented in practice in the development of decision support systems for environmental safety. Practical use has proven its adequacy and efficiency.

Support

This article was supported by international study project Tempus NETCENG "New model of the third cycle in engineering education due to Bologna Process in BY, RU, UA".

References

Klir, G. J.; Yuan, B. 1995. *Fuzzy sets and fuzzy logic: theory and applications*. Prentice Hall PTR. 574 p.

Sivanandam, S. N.; Sumathi, S.; Deepa, S. N. 2007. *Introduction to fuzzy logic using MATLAB*. Springer-Verlag Berlin Heidelberg. 425 p. https://doi.org/10.1007/978-3-540-35781-0

Verbruggen, H. B.; Babuska, R. 1998. Constructing fuzzy models by product space clustering, in H. Helendorn, D. Driankov (Eds.). *Fuzzy model identification*. Berlin: Springer, 53–90.

Wang, L. X. 1992. Fuzzy systems are universal approximators, in *Proceedings of the IEEE International Conference on Fuzzy Systems,* San Diego, 1163–1169. https://doi.org/10.1109/FUZZY.1992.258721

Xue, Q.; Hu, Y.; Tompkins, W. 1990. Analysis of hidden units of back propagation model by SVD, in *Proceedings IJCNN*, Washington, 739–742.

Zadeh, L. A. 1965. Fuzzy sets, *Information and Control* 8(3): 338–353. https://doi.org/10.1016/S0019-9958(65)90241-X