Introducing Floating Singletons with an example application in clinic spirometry

Luis Argüelles Méndez[1]

European Centre for Soft Computing
Gonzalo Gutiérrez Quirós S/N, Mieres, Spain
[1]luis.arguelles@softcomputing.es

Abstract

Fuzzy-logic models where singletons are placed in the consequents of rules of the type if x is A_i and y is B_j then $z_i = s_i$ are used nowadays in many soft computing applications. In this work we introduce the concept of *floating singletons* where their precise location on the real axis is only known at run-time. At design time some intervals are established where this special singletons are allowed to float and later, the defuzzification of a complementary fuzzy model fixes their location. As an example, a complete model for detecting *Chronic Obstructive Pulmonary Disease* is introduced, taking into account values from spirometric tests and the effects of smoking in patients.

Keywords: Singletons, Defuzzification, Chronic Obstructive Pulmonary Disease, Spirometry.

1. Introduction

Fuzzy models composed by knowledge-based rules of the type if x is A_i and y is B_j then $z = C_k$ where A_i and B_j are fuzzy sets in the antecedent and C_k are fuzzy sets in the consequent are extensively used in fuzzy logic applications. When trapezoidal, triangular or Gaussian membership functions are chosen to represent C_k, the defuzzification method must be chosen with care in order to cover the complete universe of discourse where the C_k fuzzy sets are defined. More specifically, given a fuzzy partition $Z = \{C_1 \dots C_n\}$ defined on an interval $[a,b] \in R$ where the supports s_i of the respective fuzzy sets C_i, are completely included into $[a,b]$, defuzzification methods such as max-membership, centroid, weighted average, mean-max, etc. [1], always produce a defuzzified value $g \in R$ defined in the open interval (a,b), that is, g can never reach the extremes a or b that bound the universe of discourse where the consequents are defined. A quick workaround is to place the first and last fuzzy sets of the linguistic variable representing the consequent in such a way that their respective centre of gravity from their membership functions is projected respectively over the bounding extremes a and b. This approach exhibits some lack of elegance in the design of the model because the membership functions located at the extremes show an overflow with respect to the interval $[a,b]$.

Takagi-Sugeno fuzzy models [2] uses rules with the following form:

$$\text{if } x \text{ is } A_i \text{ and } y \text{ is } B_j \text{ then } z_i = q_i x + r_i y + s_i$$

where again, A_i and B_j are fuzzy sets in the antecedent, but this time q_i, r_i, and s_i are constants, thus resulting in a crisp value for z_i. A special case, where $q_i = r_i = 0$, is given by the following expression:

$$\text{if } x \text{ is } A_i \text{ and } y \text{ is } B_j \text{ then } z_i = s_i \quad (1)$$

defuzzification for the singletons s_i comes from the well-known expression:

$$\frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i} \quad (2)$$

where w_i are the weights associated to every singleton s_i resulting from evaluating the antecedents in the rules from the fuzzy model and x_i are the distances at which the singletons are located in $[a,b]$. Since every weight w_i projects itself exactly where the singletons s_i are defined, these models can easily reach the extremes a,b of any given universe of discourse by making $s_1 = a$, $s_n = b$

Aside the described ability to reach the extremes of the universe of discourse, this modelling strategy reduces computation time with respect to the required defuzzification processes in Mandami models and offer another set of desirable advantages: it works well with optimization and adaptive techniques, it too works smoothly with linear techniques such as PID control, and it has guaranteed continuity of the output surface [3]. For the sake of this exposition, it is convenient to remember that in a fuzzy model such as the described in these paragraphs, the rules are evaluated first (inferencing) and then a defuzzification procedure is applied for the singletons.

2. Floating singletons

As a first definition, a floating singleton, *sf*, is a special type of singleton that has the ability to move inside the universe of discourse defined in [*a,b*]. That is, the designer of the system does not know its exact position in advance and only at run-time the position of it is determined by means of a complementary fuzzy system, as we shall see later. A set of traditional singletons s_i defines a set of $I_i = i\text{-}1$ internal intervals in [*a,b*] as can be observed in figures #1a and #1b:
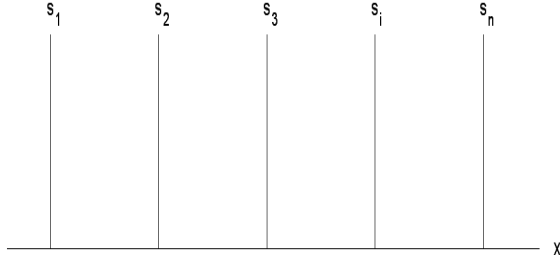

Figure 1a: A traditional set of singletons, s_i

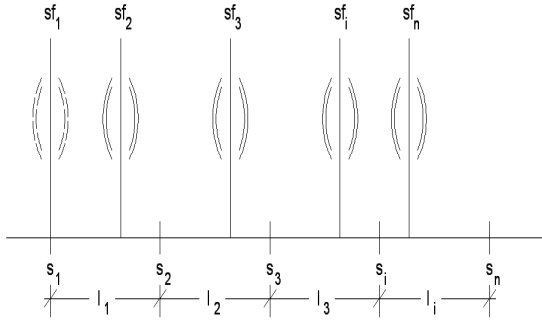

Figure 1b: A set of floating singletons by-left, sf_i. Note the position of traditional singletons s_i on the horizontal axis.

We say that sf_i is a floating singleton by-left if :

$$s_{i-1} \le sf_i \le s_i \Rightarrow sf_i \in I_{i+1}$$

with $sf_1 = s_1$. Conversely, sf_i is a floating singleton by-right if:

$$s_i \le sf_i \le s_{i+1} \Rightarrow sf_i \in I_i$$

with $sf_n = s_n$ (see figures #1a and #1b). Now we can expose a more restricted definition: a floating singleton, sf_i, is a special type of singleton that has the ability to move inside its assigned interval I_i.

The floating nature of floating singletons is explained by means of the existence of another knowledge-based model composed by fuzzy rules of the type:

$$\text{if x } is\ M_g\ and\ y\ is\ N_h\ then\ z_i = t_{i,j} \quad (3)$$

where $t_{i,j} \in [s_i, s_{i+1}]$ are a set of singletons $t_{i,j}$, expressed as consequents. Here we must remark that every singleton $t_{i,j}$ is defined on the interval I_i. Graphically:
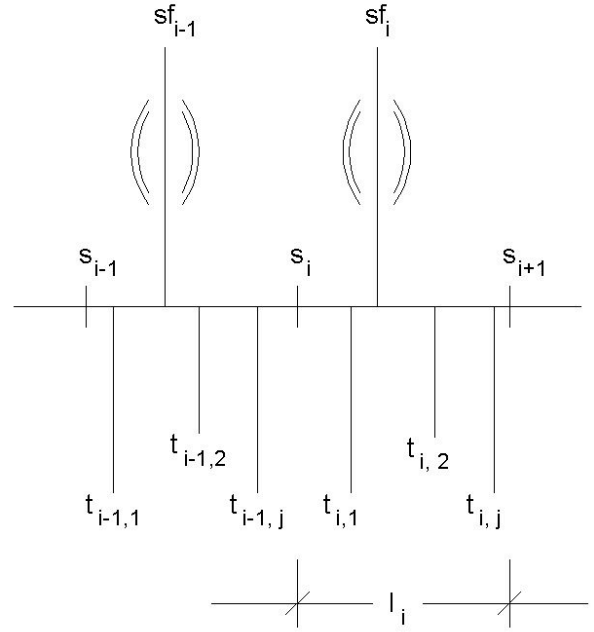

Figure 2: Graphic interpretation of floating singletons sf_i, and their relationship with standard singletons $t_{i,j}$

now, the position of every floating singleton sf_i comes from the defuzzification of the singletons t_i, expressed by:

$$\frac{\sum_{i=1}^{n} h_i d_i}{\sum_{i=1}^{n} h_i} \quad (4)$$

where h_i are the weights associated to every t_i while d_i are the distances at which the singletons t_i are located inside the interval I_i. As can be seen, in this arrangement of models, inferences are made first for rules given by (3), then defuzzification of t_i are made for "fixing" the position of the floating singletons and then, after making the inferences for the rules given by (1), the final defuzzification process is made.

The term "floating distance" is defined as the shift towards left or right of a floating singleton with respect to a singleton of reference.

$$fd = sf_i - s_i \text{ (by-right)}$$
$$fd = s_i - sf_i \text{ (by-left)}$$

For normal floating singletons, "floatability" is defined as the maximum floating space available for a floating singleton, as can be observed in figure 3:

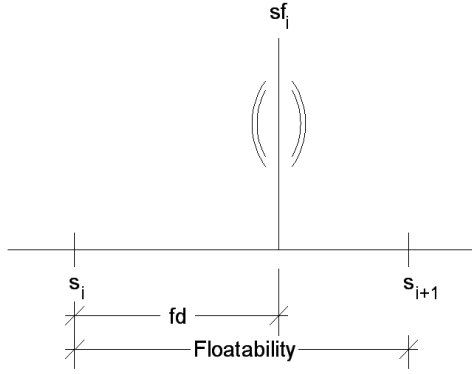$$F = s_{i+1} - s_i \text{ and hence:}$$
$$fd \le F$$

Figure 3: Graphic interpretation of floatability of sf_i with respect to its assigned interval

Finally, a set of partially floating singletons Spf is defined as a set of singletons containing both traditional singletons and floating singletons:

$$\forall Spf \; \exists \; sf_i \, , \; 1 < i < n \, / \, sf_i \in (s_1, s_n) \quad (5)$$
$$\forall Spf \; \exists \; s_i \, , \; 1 < i < n \, / \, s_i \in (s_1, s_n) \quad (6)$$

In this definition we exclude floating singletons placed in the extremes ($sf_1 = s_1$, $sf_n = s_n$), hence the open interval used.

3. Spirometry

Spirometry is a clinical test where a device (spirometer) records the volume of air exhaled by a patient and plots it as a function of time, producing a curve that shows the lung function, specifically the volume and flow of air that can be exhaled. The patient inhales maximally and then exhales as rapidly and completely as possible. This technique is the most common of the pulmonary function tests and aside other conditions, is a suitable clinical test for detecting Chronic Obstructive Pulmonary Disease (COPD). As an example, this disease appears usually with time in coal mine workers due to coal dust exposure [4], so it is a common test in order to check their health status every six or twelve months because spirometry is the only test able to detect COPD years before the shortness of breath develops [5]. A typical spirometric curve can be observed in figure 4. The most important numerical parameters obtained by means of a spirometry are the following ones:

- FVC: Forced Vital Capacity: It is the whole volume of air that can be exhaled by a patient after full inspiration in the test.

- FEV1: Forced Expiratory Volume in one second: It is the volume of air that can be exhaled in the first second of a spirometric test. Both FVC and FEV1 are measured in litres.

- TI: Tiffenau Index: It is a relationship between FEV1 and FVC values given by the following expression:

$$TI = 100 \cdot \frac{FEV1}{FVC} \quad (7)$$

From the height, x, in metres and age, y, in years of an individual, theoretical values of FEV1 and FVC for men, in litres, can be obtained using the following expressions [6]:

Theoretical FVC = 5,76x – 0,026y – 4,34 (8)
Theoretical FEV1 = 4,3x – 0,026y – 2,49 (9)

Comparing the experimental FEV1 and FVC values obtained from a spirometry with the theoretical values, lung obstruction is detected applying the following algorithm: If (Observed FVC > 80% theoretical FVC) and (Observed FEV1 > 80% theoretical FEV1) and TI > 80% then spirometry is normal, else the individual suffers an obstructive pulmonary disease (similar algorithms are given, as example, in [7]. However, the expressions (8) and (9) show two regression lines from a general population, but physicians know that the requirements for detecting COPD in workers exposed to ambient dust must be more strict than the simple application of the aforementioned algorithm, and test results must be evaluated in light of the patient's clinic history, physical examination and pertinent laboratory findings. Moreover, medical practice shows that smoking is a strong factor that worsens COPD. In the following section we introduce a fuzzy-logic based model that gathers the experience of physicians and also has into account the smoking factor. This model uses floating singletons.
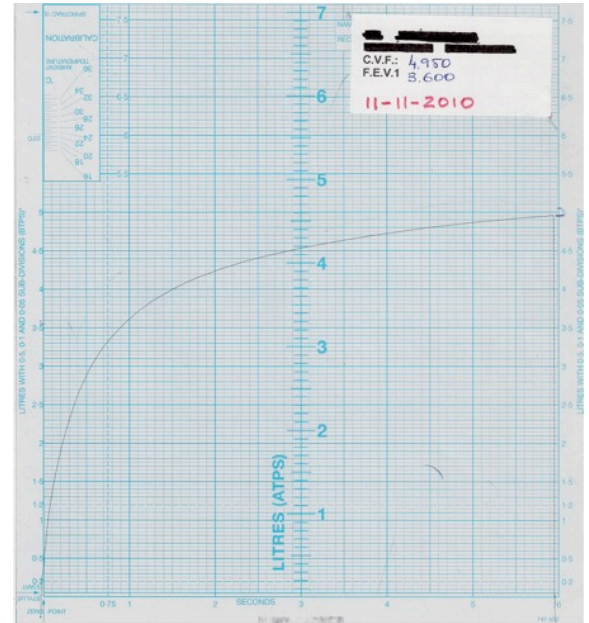

Figure #4: A typical spirometry. The horizontal axis represents time in seconds, while the vertical axis represents litres of exhaled air. This spirometry show moderate COPD (see test)

4. Fuzzy model

First, we introduce two new magnitudes, *f1*, *f2*, that have into account both the test results from a spiromet-

ric test and the theoretical values shown in expressions (8) and (9):

$$f1=100(\text{FVC spirometry})/(\text{Theoretical FVC}) \quad (10)$$
$$f2=100(\text{FEV1 spirometry})/(\text{Theoretical FEV1}) \quad (11)$$

Then, we build two linguistic variables, Pct-FVC and Pct-FEV1 using triangles and trapezoids as membership functions for representing $f1$ and $f2$ percentages:

- Pct-FVC = {Severe(20,30,50), Moderate(35,50,70), Slight(50,70,85), Normal(70,85,100,120), Excellent(100,120,150,150)}

- Pct-FEV1 = {Severe(20,30,50), Moderate(35,50,70), Slight(50,70,85), Normal(70,85,100,120), Excellent(100,120,150,150)}

Our goal is to obtain an index z between 0 and 100 for representing the obstruction degree for an individual. We use the following floating singletons for modelling the consequents:

- Obstruction = {Null($sf1$[0,15]), Appreciable($sf2$[15,65]), Important($sf3$[65,85]), Very Important($sf4$[85-100]), Extreme($sf5$[100])} (12)

As can be observed, these floating singletons are expressed by intervals I_i. Note also that the last floating singleton is fixed at [100], that is, we use in this case floating singletons by-right.

The rules for the model are of the type:

If $f1$ is Pct-FVC and $f2$ is Pct-FEV1 then z = Obstruction

In order to model smoking, we use two numerical magnitudes: Years of smoking, y, and number of cigarettes/day, c. From here, we introduce two linguistic variables to handle these magnitudes:

- Smoking-years = {Few(0,0,2,5), Quite-a-while(2,5,10), Many(5,10,75,75)}

- Cigarettes-day = {Few(0,0,10), Moderate(0,10,15,20), Many(15,20,100,100)}

and then a set of rules with the following form:

If y is Smoking-years and c is Cigarettes-day then s = t
(13)

where s is the pursued value for being used in the Obstruction intervals I_i that represent the floating singletons sf_1 ... sf_5 shown in (12). This value is obtained by inferencing and defuzzifying the system of rules shown in (13). Since in this model we have five intervals, I_1 ...

I_5, we need five sets of rules, so expression (13) becomes:

If y is Smoking-years and c is Cigarettes-day then $s_i = t_i$
(14)

where t_i are the following singletons:

$$t_i = \{\text{left}_i, \text{medium}_i, \text{right}_i\}$$

For every i index we have the following values associated to them:

- t_1 = {left(0), medium(8), right(15)}
- t_2 = {left(15), medium(40), right(65)}
- t_3 = {left(65), medium(75), right(85)}
- t_4 = {left(85), medium(92.5), right(100)}
- t_5 = {left(100), medium(100), right(100)}

As a practical example, let's take an individual 40 years old, 1,80 metres height, smoker from 20 years ago that smokes about 15 cigarettes/day with spirometric results FVC = 4.82 litres and FEV1 = 4.15 litres. From expressions (8) and (9) the *Theoretical FVC* and *Theoretical FEV1* values are 4.99 and 4.21, respectively, obtaining a Tiffenau Index of 86.10. f1(Pct.-FVC) is 96.63 and f2(Pct-FEV1) is 98.57. These figures, using the conventional, sharp, algorithm informs about an individual without obstructive pulmonary disease. However, smoking causes obstruction that is *invisible* to traditional methods. From the input data, and after making inferences in rules of the type shown in (14) the t_i singletons get the following values:

- $t1$ = 15
- $t2$ = 65
- $t3$ = 85
- $t4$ = 100
- $t5$ = 100

Translating these values to the fuzzy singletons shown in (12) and after inferencing and defuzzification, we obtain an obstruction index z = 15 that tells the physician that the individual, while having a normal ability to blow air, definitively suffers from obstructive pulmonary disease. Table 1 shows different t_i singletons obtained from changing smoking years and number of cigarettes/day in the model, leaving the rest of input data unchanged. This table shows how the singletons *float* in their respective intervals.

years	cig/day	t_1	t_2	t_3	t_4
0	0	0.00	15.00	65.00	85.00
3	5	1.33	19.17	66.67	86.25
15	5	11.50	52.50	80.00	96.25
5	12	8.00	40.00	75.00	92.50
10	10	15.00	65.00	85.00	100.00

Table 1: Floatability of singletons t_i inside their intervals I_i from different input data. Note that t_5 is constant t_5 = 100

5. Conclusions

Fuzzy logic models where singletons are used as consequents in the body of knowledge-based rules have been a success for building applications in the field of soft computing since its inception. Aside its intrinsic elegance, efficiency and its relative easiness for being implemented [8], it reduces computation time with respect to the required defuzzification processes in Mandami models, among other advantages.

In this work we have introduced *floating singletons.* They can be seen as a special type of singletons able to move inside a certain set of intervals defined by the designer that cover the complete output universe of discourse. Since the exact location of every floating singleton is unknown at design-time (only the floating intervals are known), a complementary fuzzy model is used in order to define their positions at run-time. In this way, and when used, for example, in fuzzy control applications, floating singletons stop for an instant in every control loop and are able to get different values for next outputs over time, hence the name "floating". An external observer could see how these special singletons float over timer inside their assigned floating intervals. This strategy offers more flexibility and adds another tool for modelling soft computing applications.

As an example, we have described in this paper a model for determining, in an index z valued from 0 to 100, the degree of Chronic Obstructive Pulmonary Disease (COPD) present in a patient after performing a clinical spirometric test. Through the use of floating singletons, the model considers not only the forced vital capacity (FVC) and the forced expiratory volume in one-second (FEV1) values from spirometry, but also smoking, a negative factor that, despite worsening COPD, many times remains invisible in spirometric tests. In this way, the model shows a better agreement with medical experience than traditional algorithms.

References

[1] Ross, T., "Fuzzy-Logic with Engineering Applications", Wiley, 2010.

[2] Takagi, T., Sugeno, M., "Fuzzy Identification of Systems and Its Applications to Modelling and Control", IEEE Transactions on systems, man and cybernetics, Vol. SMC-15, No.1, 1985.

[3] Klinkhachorn, P. "Fuzzy Rule based models", Lane department of Computer Science and Electrical Engineering. West Virginia University, 2004.

[4] Santo-Tomas, LH., "Emphysema and chronic obstructive pulmonary disease in coal miners", Curr Opin Pulm Med., 2011.

[5] Hyatt, R., Scanlon, P., Nakamura, M. "Interpretation of Pulmonary Function Tests: A Practical Guide", Lippincott Williams & Wilkins, 2009.

[6] Morris, JF. Koski A., Johnson, LC., "Spirometric Standards for Healthy, non-Smoking Adults", American Rev. of Respiratory Diseases, 1971; 103:57-67.

[7] Qaseem, A., et al. "Diagnosis and Management of Stable Chronic Obstructive Pulmonary Disease: A Clinical Practice Guideline Update from the American College of Physicians, American College of Chest Physicians, American Thoracic Society, and European Respiratory Society", Annals of Internal Medicine, 2011.

[8] Argüelles, L. "Diffucalc: A simple computational tool for Fuzzy-Logic based modelling", ESTYLF, 2008.