A Fuzzy-based Algorithm for Auditors to Detect Elements of Fraud in Settled Insurance Claims - An Illustrative Example

Jean Felipe Oehrwald

Januar 2017

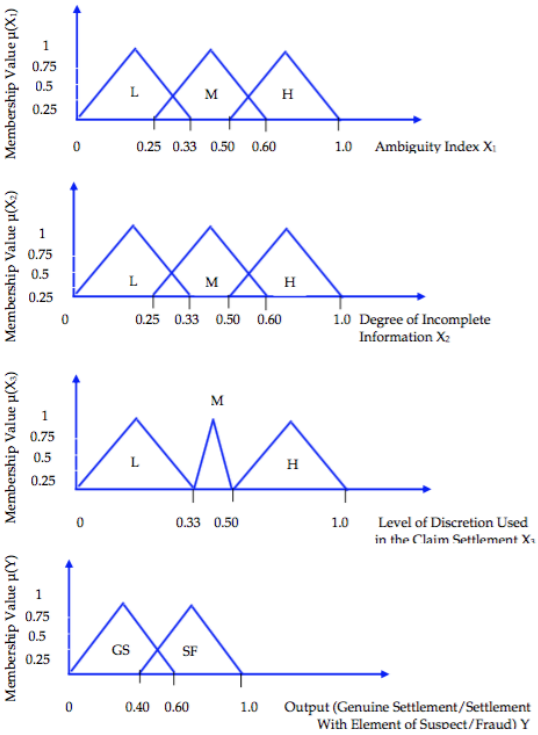
Abstract

This articel is based on the worked from Jagdish Pathak, Navneet Vidyarthi and Scott L. Summers.

Step One

Sample rule base for the Fuzzy Logic Based Expert System and the Membership Function.

Rule No.	INPUTS			OUTPUT
	X_1	X_2	X_3	Y
1	Low	Low	Low	Genuine Settlement (GS)
2	Low	Low	Medium	Genuine Settlement (GS)
3	Low	Low	High	Settlement with Fraud Element (SF)
4	Low	Medium	Low	Genuine Settlement (GS)
5	Low	Medium	Medium	Genuine Settlement (GS)
6	Low	Medium	High	Settlement with Fraud Element (SF)
7	Low	High	Low	Genuine Settlement (GS)
8	Low	High	Medium	Settlement with Fraud Element (SF)
9	Low	High	High	Settlement with Fraud Element (SF)
10	Medium	Low	Low	Genuine Settlement (GS)
11	Medium	Low	Medium	Genuine Settlement (GS)
12	Medium	Low	High	Settlement with Fraud Element (SF)
13	Medium	Medium	Low	Genuine Settlement (GS)
14	Medium	Medium	Medium	Genuine Settlement (GS)
15	Medium	Medium	High	Settlement with Fraud Element (SF)
16	Medium	High	Low	Genuine Settlement (GS)
17	Medium	High	Medium	Settlement with Fraud Element (SF)
18	Medium	High	High	Settlement with Fraud Element (SF)
19	High	Low	Low	Genuine Settlement (GS)
20	High	Low	Medium	Genuine Settlement (GS)
21	High	Low	High	Settlement with Fraud Element (SF)
22	High	Medium	Low	Genuine Settlement (GS)
23	High	Medium	Medium	Settlement with Fraud Element (SF)
24	High	Medium	High	Settlement with Fraud Element (SF)
25	High	High	Low	Genuine Settlement (GS)
26	High	High	Medium	Settlement with Fraud Element (SF)
27	High	High	High	Settlement with Fraud Element (SF)



Step Two

Step 1: Evaluate the authenticity of claim settlement: Determine the ambiguity index X_1, degree of incomplete information of the claims X_2, level of discretion used by the claim settlers X_3.

Lets say that the input values are: $X_1 = 0.27$, $X_2 = 0.55$ and $X_2 = 0.40$.

Step Three

Fuzzify the values of inputs: Through the use of membership functions defined for each fuzzy set for each linguistic variable (as in figure 1), determine the degree of membership of a value in each fuzzy set. Each of these three ambiguity indices have been divided into three fuzzy sets (LOW - L, MEDIUM - M and HIGH - H). The equations for computing memberships are:

$$\mu\left(X_i\right)_L = max\left\{min\left(\frac{X_i - a_L^{X_i}}{b_L^{X_i} - a_L^{X_i}}, \frac{c_L^{X_i} - X_i}{c_L^{X_i} - b_L^{X_i}}\right), 0\right\}$$

$$\mu\left(X_i\right)_M = max\left\{min\left(\frac{X_i - a_M^{X_i}}{b_M^{X_i} - a_M^{X_i}}, \frac{c_M^{X_i} - X_i}{c_M^{X_i} - b_M^{X_i}}\right), 0\right\}$$

$$\mu\left(X_i\right)_H = max\left\{min\left(\frac{X_i - a_H^{X_i}}{b_H^{X_i} - a_H^{X_i}}, \frac{c_H^{X_i} - X_i}{c_H^{X_i} - b_H^{X_i}}\right), 0\right\}$$

where (a, b, c) are the vertices of the triangular membership function and L, M and H represents the fuzzy set low, medium, and high respectively.

For our input we receive the following result:

$$\mu\left(X_1\right)_L = max\left\{min\left(\frac{X_1 - a_L^{X_1}}{b_L^{X_L} - a_L^{X_1}}, \frac{c_L^{X_1} - X_1}{c_L^{X_1} - b_L^{X_1}}\right), 0\right\} = 0,86$$

$$\mu\left(X_1\right)_M = max\left\{min\left(\frac{X_1 - a_M^{X_1}}{b_M^{X_M} - a_M^{X_1}}, \frac{c_M^{X_1} - X_1}{c_M^{X_1} - b_M^{X_1}}\right), 0\right\} = 0,11$$

$$\mu\left(X_3\right)_M = max\left\{min\left(\frac{X_2 - a_M^{X_2}}{b_M^{X_M} - a_M^{X_2}}, \frac{c_M^{X_2} - X_2}{c_M^{X_2} - b_M^{X_2}}\right), 0\right\} = 0,29$$

$$\mu\left(X_2\right)_H = max\left\{min\left(\frac{X_2 - a_H^{X_2}}{b_H^{X_H} - a_H^{X_2}}, \frac{c_H^{X_2} - X_2}{c_H^{X_2} - b_H^{X_2}}\right), 0\right\} = 0,20$$

$$\mu\left(X_2\right)_M = max\left\{min\left(\frac{X_3 - a_M^{X_3}}{b_M^{X_M} - a_M^{X_3}}, \frac{c_M^{X_3} - X_3}{c_M^{X_3} - b_M^{X_3}}\right), 0\right\} = 0,64$$

$\mu\left(X_1\right)_H = \mu\left(X_2\right)_L = \mu\left(X_3\right)_L = \mu\left(X_3\right)_H = 0$
where:

For X_1 :

$\mathrm{a}_L^{X_1}, b_L^{X_1}, c_L^{X_1} = (0, 0.165, 0.33)$

$\mathrm{a}_M^{X_1}, b_M^{X_1}, c_M^{X_1} = 0.25, 0.425, 0.60)$

$\mathrm{a}_H^{X_1}, b_H^{X_1}, c_H^{X_1} = 0.50, 0.75, 1)$

For X_2 :

$\mathrm{a}_L^{X_2}, b_L^{X_2}, c_L^{X_2} = 0, 0.165, 0.33)$

$\mathrm{a}_M^{X_2}, b_M^{X_2}, c_M^{X_2} = 0.25, 0.425, 0.60)$

$\mathrm{a}_H^{X_2}, b_H^{X_2}, c_H^{X_2} = 0.50, 0.75, 1)$

For X_3 :

$\mathrm{a}_L^{X_3}, b_L^{X_3}, c_L^{X_3} = 0, 0.165, 0.33)$

$\mathrm{a}_M^{X_3}, b_M^{X_3}, c_M^{X_3} = 0.33, 0.44, 0.55)$

$\mathrm{a}_H^{X_3}, b_H^{X_3}, c_H^{X_3} = 0.50, 0.75, 1)$

Example

I will give a more detailled explaination for:

$\mu\left(X_1\right)_M = 0,11$
$\mu\left(0,27\right)_M = max\left\{min\left(\frac{0,27-0,25}{0,425-0,25}, \frac{0,60-0,27}{0,60-0,425}\right), 0\right\} = 0,11$

The placeholder (a, b, c) are the vertices of the triangular membership function and L, M and H represents the fuzzy set low, medium, and high respectively.

Step Four

Fire the rule bases that correspond to these inputs: Every fuzzy logic based expert system uses fuzzy IF-THEN rules. A fuzzy IF-THEN rule is of the form IF X_1 =A_1 and X_2 = A_2 ... and X_n = A_n THEN Y = B,where X_i and Y are linguistic variables and A_i and B are linguistic terms. There are two outputs: GENUINE SETTLEMENT - GS, CASES SETTLED WITH FRAUD ELEMENT OR ELEMENT OF SUSPECT– CF. The set "GENUINE SETTLEMENT" refers to the insurance claims which are genuinely settled without any room for ambiguity or suspect whereas "CASES SETTLED WITH FRAUD ELEMENT OR ELEMENT OF SUSPECT– CF" refers to the claims that have been settled but contain suspicious elements that need to be substantively audited.

Step Five

Based on the value of the fuzzy membership function values for the example under consideration, the following rules apply:

Rule 5: If X_1 is LOW and X_2 is MEDIUM and X_3 is MEDIUM THEN Y is GENUINE SETTLEMENT

Rule8:If X_1 is LOW and X_2 is HIGH and X_3 is MEDIUM then Y is SETTLEMENT WITH ELEMENT OF FRAUD (SF).

Rule 14: If X_1 is MEDIUM and X_2 is MEDIUM and X_3 is MEDIUM then Y GENUINE SETTLEMENT (GS).

Rule 17: If X_1 is MEDIUM and X_2 is HIGH and X_3 is MEDIUM then Y is SETTLEMENT WITH ELEMENT OF FRAUD (SF).

Step Six

Execute the Inference Engine: We use the "root sum squares" (RSS) method to combine the effects of all applicable rules, scale the functions at their respective magnitudes, and compute the "fuzzy" centroid of the composite area. This method is more complicated mathematically than other methods, but was selected for this example since it seemed to give the best weighted influence to all firing rules.

The respective output membership function strengths (range: 0-1) from the possible rules (R1-R27) are:

"Genuine Settlement"

$$\sqrt{\sum \left(\mu_{R_1}\right)^2}$$

GS=1,2,4,7,8,10,11,13,16,17,19,20,22,25

$$\sqrt{(0.29)^2 + (0.11)^2} = 0.31$$

"Settlement with element of fraud/suspect"

$$\sqrt{\sum \left(\mu_{R_1}\right)^2}$$

SF=3,5,6,9,12,14,15,18,21,23,24,26,27

$$\sqrt{(0.20)^2 + (0.11)^2} = 0.23$$

Step Seven

The defuzzification of the data into crisp output is accomplished by combining the results of the inference process and then computing the "fuzzy centroid" of

the area. The weighted strengths of each output member function are multiplied by their respective output membership function center points and summed.