Fuzzy System Model For Insurance Industry

Karolina Tatarczyk[a]

[a]Faculty of Applied Mathematics, Silesian University of Technology), Kaszubska 23, 44-100 Gliwice, Poland

Abstract
The insurance industry has to perform an accident risk assessment when insurance is purchased by drivers. Unfortunately, but this type of assessment can be quite often problematic because it depends on many parameters. The most common criteria are the age of the driver, the engine power of the car and its year of manufacture. This paper presents the Takagi-Sugeno system model to assess risk for driver insurance companies. The proposed model presents a fuzzy approach to using these three parameters. The results were presented and discussed due to the pros and cons of the proposed solution and practical use.

Keywords
Takagi-Sugeno system, fuzzy logic, insurance problem

1. Introduction

Fuzzy logic belongs to one of the most popular artificial intelligence techniques next to artificial neural networks. Recent years have allowed for great development, not only theoretical but also practical, as evidenced by numerous studies and implementations. This is particularly evident in the area of research on uncertain rule-based fuzzy models and hesitant fuzzy data [1, 2]. Also, many generalizations of this type of logic is analyzed for a much broader practical application. A special case is the analysis of generalized orthopair fuzzy sets in [3]. Again the authors of [4] compare different types of fuzzy logic systems used for control problems. All of these theoretical and comparative studies are very important but practical use has a big impact on today's industry.

The main application of the fuzzy approach and the artificial intelligence, in general, is making decisions about the information which is given to the system [5, 6]. In [7], the authors presented a detection and classification solution based on neural networks and fuzzy controllers that decide about the clustering of images. An interesting solution was described in [8], where the idea of smartphone-based intelligent indoor positioning was described in detail. Fuzzy time series forecasting is a less popular use of these solutions [9] or public support for insurgency and terrorism [10]. Another possible application is a multiword search over encrypted data [11] and medicine. In medicine, the artificial intelligence use it is a very important solution that can be useful for a doctor for analyzing some patient' data [12, 13, 14, 15].

A deep learning also finds possible use in security, which is described in a blockchain approach for analyzing weights in neural systems [16].

In this paper, we propose a system model for risk analysis in companies insuring drivers using three parameters. The described solution is based on a fuzzy system using *If-Then* rules and was tested for sample data and discussed compared to the pros and cons of this solution in practice.

2. Problem analysis

Fuzzy systems find practical use in many areas of life. By using fuzzy systems we can define the membership to the selected set. In comparison to the classic system, where the only values are 0 and 1, the fuzzy system value of the argument can be a number between 0 and 1. Moreover, one argument can belong to many sets at various levels. Fuzzy set A is a set of elements, which partly belongs to it.

$$A = \{(X, \mu_A(x)) \, : \, x \in X\}, \tag{1}$$

where

$$\mu_A \, : \, X \longrightarrow [0, 1]. \tag{2}$$

and X is a space with complete number of elements, $X = \{x_1, x_2, \dots, x_n\}$

$$A = \frac{\mu_A(x_1)}{x_1} + \dots + \frac{\mu_A(x_n)}{x_n} = \sum_{i=1}^{n} \frac{\mu_A(x_i)}{x_i}. \tag{3}$$

Linguistic variables are also used in fuzzy systems. Linguistic variables are the statements of natural language, which are descriptions of fuzzy sets defined on a specific space. Fuzzy systems may be used in many

IVUS 2020: Information Society and University Studies, 23 April 2020, KTU Santaka Valley, Kaunas, Lithuania
karotat400@student.polsl.pl (K. Tatarczyk)

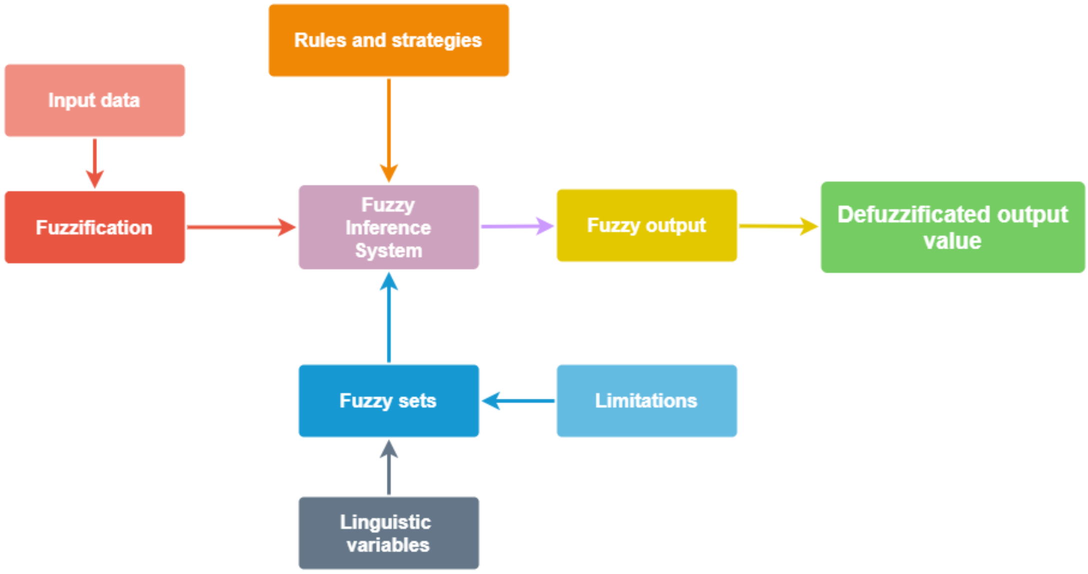


Figure 1: Flow chart of the fuzzy logic system.

fields. In this paper, we focus on using fuzzy systems in the car insurance industry, especially in the case of calculating the risk.

2.1. Fuzzy system

As an example in car insurance, the linguistic variables that are used are describing the age of a driver, the power of the engine and the car's production date. For each of them, there are three different labels describing the state. For the age: "young", "middle", "old". For the engine's power: "low", "medium", "high". For the car's production date: "old", "middle", "new". Charts 3 are presenting membership functions for the age of the driver, power of engine and car's production date respectively.

The overall membership function is presented as follows

$$\mu_{low/high} = \begin{cases} 0 \quad \text{if} \quad x < a \\ (x-a)/(b-a) \quad \text{if} \quad x \in [a,b] \\ 1 \quad \text{if} \quad x \in [b,c] \\ (d-x)/(d-c) \quad \text{if} \quad x \in [c,d] \\ 0 \quad \text{if} \quad x > d \end{cases} . \tag{4}$$

$$\mu_{medium} = \begin{cases} 0 \quad \text{if} \quad x < a \\ (x-a)/(b-a) \quad \text{if} \quad x \in [a,b] \\ (c-x)/(c-b) \quad \text{if} \quad x \in [b,c] \\ 0 \quad \text{if} \quad x > c \end{cases} . \tag{5}$$

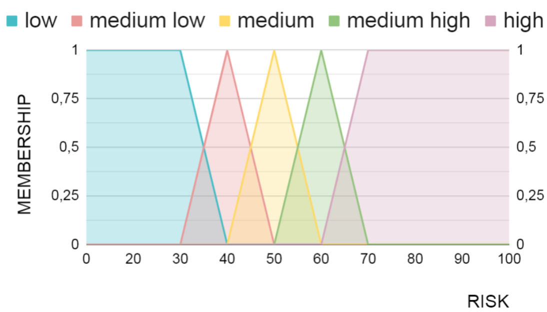


Figure 2: Chart of a membership function.

To count the membership each set's parameters a, b, c, d ($a \le b \le c \le d$) has to be replaced by individual numbers typical to each of the specific membership functions. For example to count the membership of argument x (relevant to the age) to fuzzy set "middle", the parameters a, b, c have to be replaced by $20, 40, 60$ respectively. The next step to cost out the risk is building **IF-THEN rules**.

If x_1 *in* A_1 **and** x_2 *in* A_2 **and** ... **then** y

where y is a variable of the consequence whose value is inferred.

Taking into consideration circumstances mentioned above, we get undermentioned rules:

1. If the driver is young and the power of the en-

```
Data: age of driver, power of engine, car's
      production date
Result: risk
age;
power;
year;
Form the membership function for every
  linguistic variables;
w[27];
weight[27];
Check the membership μ of each data to
  appropriate membership functions ;
i := 0;
j := 0;
k := 0;
for i < 3 do
    for j < 3 do
        for k<3 do
            w[l] =
             μ_{A_i}(age) · μ_{B_j}(power) · μ_{C_k}(year);
        end
    end
end
for i< 27 do
    Assign value of result for each rules;
end
numerator := 0;
for i<27 do
    numerator+ = w[i] · weight[i];
end
denominator := 0;
for i<27 do
    denominator+ = w[i];
end
risk := numerator · 100/denominator;
Display amount of risk;
```

Algorithm 1: Fuzzy system in insurance.

gine is low and the car is new, then the risk is low.

2. If the driver is young and the power of the engine is low and the car is middle-age, then the risk is medium-low.
3. If the driver is young and the power of the engine is low and the car is old then the risk is medium.
4. If the driver is young and the power of the engine is medium and the car is new, then the risk is medium-low.
5. If the driver is young and the power of the engine is medium and the car is middle-age, then the risk is medium.
6. If the driver is young and the power of the engine is medium and the car is old, then the risk is medium-high.
7. If the driver is young and the power of the engine is high and the car is new, then the risk is medium.
8. If the driver is young and the power of the engine is high and the car is middle-age, then the risk is medium-high.
9. If the driver is young and the power of the engine is high and the car is old, then the risk is high.
10. If the driver is in middle age and power of the engine is low and the car is new, then the risk is low.
11. If the driver is in middle age and power of the engine is low and the car is middle-age, then the risk is medium-low.
12. If the driver is in middle age and power of the engine is low and the car is old, then the risk is medium.
13. If the driver is in middle age and power of the engine is medium and the car is new, then the risk is low.
14. If the driver is in middle age and power of the engine is medium and the car is middle-age, then the risk is medium-low.
15. If the driver is in middle age and power of the engine is medium and the car is old, then the risk is medium-high.
16. If the driver is in middle age and power of the engine is high and the car is new, then the risk is medium-low.
17. If the driver is in middle age and power of the engine is high and the car is middle-age, then the risk is medium.
18. If the driver is in middle age and power of the engine is high and the car is old, then the risk is medium-high.
19. If the driver is old and the power of the engine is low and the car is new, then the risk is medium.
20. If the driver is old and the power of the engine is low and the car is middle-age, then the risk is medium-high.
21. If the driver is old and the power of the engine is low and the car is old, then the risk is high.
22. If the driver is old and the power of the engine is medium and the car is new, then the risk is medium.

Table 1
Tables with examples of linguistic variables values and their affiliation to each set.

Driver's age	young	middle	old
20	1	0	0
21	0.95	0.05	0
25	0.75	0.25	0
30	0.5	0.5	0
35	0.25	0.75	0
40	0	1	0
44	0	0.8	0.2
50	0	0.5	0.5
80	0	0	1
Engine's power	low	medium	high
50	1	0	0
75	0.9	0.1	0
80	0.8	0.2	0
100	0.4	0.6	0
110	0.2	0.8	0
145	0	0.5	0.5
160	0	1	0
180	0	0.2	0.8
190	0	0	1
Year of production	old	medium	new
1998	1	0	0
2001	0.875	0.125	0
2004	0.5	0.5	0
2006	0.25	0.75	0
2008	0	1	0
2009	0	0.857	0.143
2010	0	0.714	0.286
2012	0	0.429	0.571
2017	0	0	1

23. If the driver is old and power of the engine is medium and the car is middle-age, then the risk is medium-high.
24. If the driver is old and the power of the engine is medium and the car is old, then the risk is high.
25. If the driver is old and the power of the engine is high and the car is new, then the risk is medium-high.
26. If the driver is old and the power of the engine is high and the car is middle-age, then the risk is high.
27. If the driver is old and the power of the engine is high and then a car is old, the risk is high.

The results of the implication can be presented as a membership function (see Fig. 2).

By having rules and values, we can go to the next step - **inference**. Here for each of rules, we use a formula

$$\mu_R(x, y, z) = \mu_A(x) \cdot \mu_B(y) \cdot \mu_C(z). \tag{6}$$

The result μ_R means the activation level of each conclusion.

There are many different methods of **defuzzification** available. In this paper, assigning weight δ_i for every conclusion holds by using the center of area (COA). The general visualization of the proposed system is presented in Fig. 1.

In conclusion to get the final result, all data have to be apllied to the formula:

$$y = \frac{\sum_{i=1}^{n} \delta_i \cdot \mu_R(x_i, y_i, z_i)}{\sum_{i=1}^{n} \mu_R(x_i, y_i, z_i)}. \tag{7}$$

The result of that formula is a percentage of risk. Fuzzy system algorithm is presented in Alg. 1.

3. Experiments

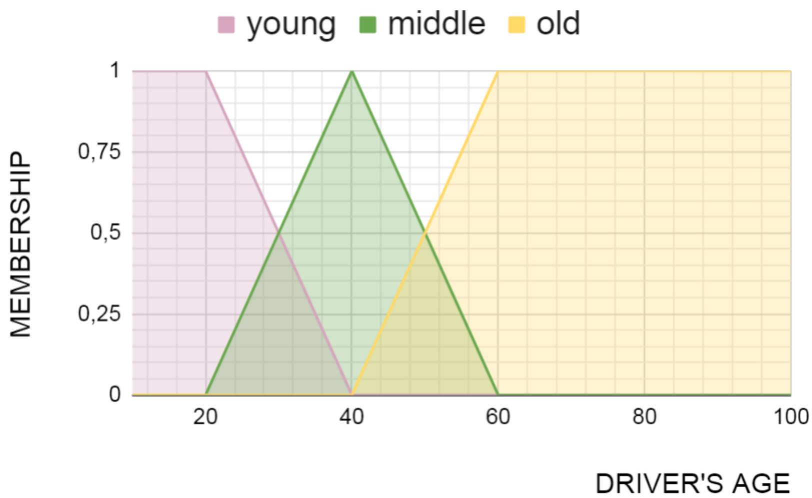
As part of checking the propriety of the proposed solution, for the selected data (see Tab. 1) we described average steps described in Tab. 2 . The table shows only results for 8 rules, but we should note that the implemented application has to check each of the rules respectively and count each of the elements. The sample data shows information that was used for analyzing the proposed system. A presented solution was implemented in C# language.

In Tab. 2, a sample output result was presented. In these experiments, we analyze the results output for a drive aged 25 and a six-year car with a capacity of 100. It is easy to notice that the more information about the driver, the better and more accurate are the results. The reason for that is the defuzzification – we have more specific elements of summation in the last step (called defuzzification). It allows saying that fuzzy systems work more effectively when we have more linguistic variables. However, more variables contribute to modeling more rules and analyzing relationships between data.

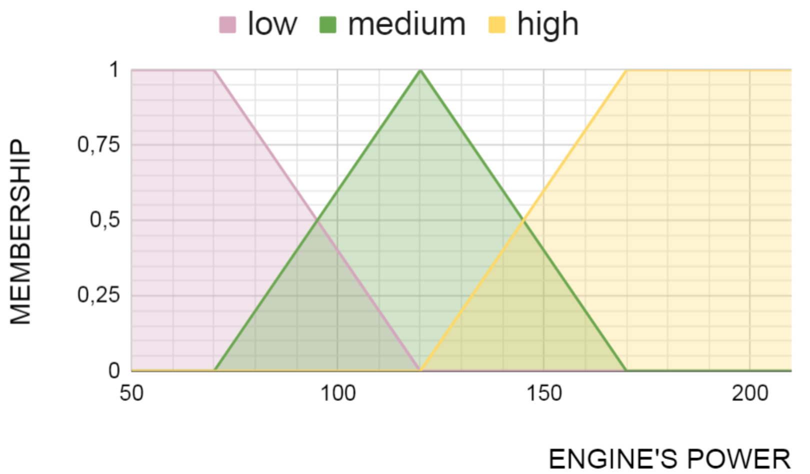
Obtained results show the possibilities of using the Takagi-Sugeno system for analyzing risk for insurance companies. In conducted experiments, we analyzed the ability to model fuzzy rules and their multitude (see Tab. 3). During modeling the technique, it was also noticed that the biggest problem with this approach is rule modeling. During the simulation, it was noticed that it is a flexible system in terms of use, which was reflected in the possibilities of introducing new linguistic variables as well as modifying values.

4. Conclusion

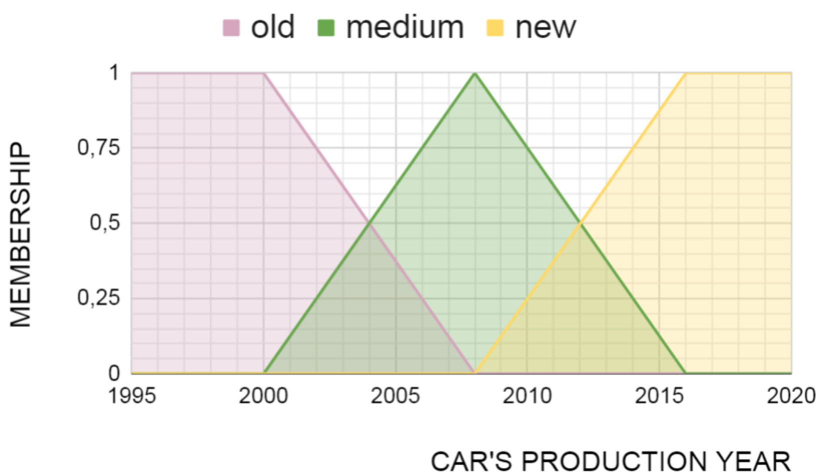
In this paper, we proposed a solution for calculating the risk of causing a car accident and directly the possibility of insurance. Proposed rules and used system based on *If-Then* rules shows a large dependence on the number of variables. The more variables, the more



(a)



(b)



(c)

Figure 3: Charts of membership functions for each of liguistic variables.

Table 2
Table of selected rules and the conclusion, when drivers are 21, the engine's power is 100 and the year of production is 2014 (other rules will give us zero results).

Number of rule i	$\mu_{age}(25)$	$\mu_{power}(100)$	$\mu_{year}(2014)$	μ_R	result of rule	$\mu_R \cdot \delta_i$
1.	0.75	0.4	0.875	0.2625	low	0.0525
2.	0.75	0.4	0.125	0.0375	medium-low	0.015
4.	0.75	0.6	0.875	0.39375	medium-low	0.1575
5.	0.75	0.6	0.125	0.05625	medium	0.028125
10.	0.25	0.4	0.875	0.0875	low	0.0175
11.	0.25	0.4	0.125	0.0125	medium-low	0.005
13.	0.25	0.6	0.875	0.13125	low	0.02625
14.	0.25	0.6	0.125	0.01875	medium-low	0.0075
SUM			1			**0.309375**

Table 3
Table showing the dependence between the number of linguistic variables and amount of rules, where n-amount of labels describing each of linguistic variable (assuming that the amount is the same for all linguistic variables).

Amount of linguistic variables	Amount of rules
2	n^2
3	n^3
4	n^4
$\vdots$	$\vdots$
k	k^n

accurate the results. However, it is worth noting that the proposed system shows fast calculations and a small amount of computing power required for its operation. These are undoubted advantages, although rule modeling is a time-consuming activity that needs to be analyzed without using self-adaptive methods.

Based on the obtained results, this approach is very promising. In future works, we plan to analyze and model a solution that can generate rules automatically.

References

[1] J. M. Mendel, Uncertain rule-based fuzzy systems, in: Introduction and new directions, Springer, 2017, p. 684.

[2] L. Aliahmadipour, V. Torra, E. Eslami, On hesitant fuzzy clustering and clustering of hesitant fuzzy data, in: Fuzzy sets, rough sets, multisets and clustering, Springer, 2017, pp. 157–168.

[3] R. R. Yager, Generalized orthopair fuzzy sets, IEEE Transactions on Fuzzy Systems 25 (2016) 1222–1230.

[4] O. Castillo, L. Amador-Angulo, J. R. Castro, M. Garcia-Valdez, A comparative study of type-1 fuzzy logic systems, interval type-2 fuzzy logic systems and generalized type-2 fuzzy logic systems in control problems, Information Sciences 354 (2016) 257–274.

[5] S. Guo, H. Zhao, Fuzzy best-worst multi-criteria decision-making method and its applications, Knowledge-Based Systems 121 (2017) 23–31.

[6] G. Capizzi, G. Lo Sciuto, C. Napoli, E. Tramontana, An advanced neural network based solution to enforce dispatch continuity in smart grids, Applied Soft Computing Journal 62 (2018) 768–775.

[7] G. Capizzi, C. Napoli, S. Russo, M. Woźniak, Lessening stress and anxiety-related behaviors by means of ai-driven drones for aromatherapy, in: CEUR Workshop Proceedings, volume 2594, 2020, pp. 7–12.

[8] F. Orujov, R. Maskeliūnas, R. Damaševičius, W. Wei, Y. Li, Smartphone based intelligent indoor positioning using fuzzy logic, Future Generation Computer Systems 89 (2018) 335–348.

[9] S.-H. Cheng, S.-M. Chen, W.-S. Jian, Fuzzy time series forecasting based on fuzzy logical relationships and similarity measures, Information Sciences 327 (2016) 272–287.

[10] O. A. Osoba, B. Kosko, Fuzzy cognitive maps of public support for insurgency and terrorism, The Journal of Defense Modeling and Simulation 14 (2017) 17–32.

[11] Z. Fu, X. Wu, C. Guan, X. Sun, K. Ren, Toward efficient multi-keyword fuzzy search over encrypted outsourced data with accuracy improvement, IEEE Transactions on Information Forensics and Security 11 (2016) 2706–2716.

[12] L. C. Jain, A. Kandel, H.-N. L. Teodorescu, Fuzzy and neuro-fuzzy systems in medicine, CRC Press, 2017.

[13] D. Połap, Analysis of skin marks through the use of intelligent things, IEEE Access 7 (2019) 149355–149363.

[14] F. Beritelli, G. Capizzi, G. Lo Sciuto, C. Napoli, M. Woźniak, A novel training method to preserve generalization of rbpnn classifiers applied to ecg signals diagnosis, Neural Networks 108 (2018) 331–338.

[15] G. Capizzi, G. Lo Sciuto, C. Napoli, D. Polap, M. Wozniak, Small lung nodules detection based on fuzzy-logic and probabilistic neural network with bioinspired reinforcement learning, IEEE Transactions on Fuzzy Systems 28 (2020) 1178–1189.

[16] A. Winnicka, K. Kęsik, Idea of using blockchain technique for choosing the best configuration of weights in neural networks, Algorithms 12 (2019) 163.