Jurnal
Teknologi

Full Paper

AUTOMATIC DEVELOPMENT OF FUZZY MEMBERSHIP FUNCTIONS ON HEPATITIS PATIENTS DATA USING PARTICLE SWARM OPTIMIZATION (PSO)

Article history
Received
27 April 2015
Received in revised form
15 June 2015
Accepted
25 November 2015

Candra Dewi*, Ratna Putri P.S, Indriati

Informatics Department, Brawijaya University, Malang, Indonesia

*Corresponding author:
dewi_candra@ub.ac.id

Abstract

Information about the status of disease (prognosis) for patients with hepatitis is important to determine the type of action to stabilize and cure this disease. Among some system, fuzzy system is one of the methods that can be used to obtain this prognosis. In the fuzzification process, the determination of the exact range of membership function will influence the calculation of membership degree and of course will affect the final value of fuzzy system. This range and function can usually be formed using intuition or by using an algorithm. In this paper, Particle Swarm Optimization (PSO) algorithm is implemented to form the triangular membership functions in the case of patients with hepatitis. For testing process, this paper conducts four scenarios to find the best combination of PSO parameter values . Based on the testing it was found that the best parameters to form a membership function range for the hepatitis data is about 0.9, 0.1, 2, 2, 100, 500 for inertia max, inertia min, local ballast constant, global weight constant, the number of particles, and maximum iterations respectively.

Keywords: Hepatitis, fuzzy system, *membership* function, particle swarm optimization

1.0 INTRODUCTION

Hepatitis is a medical condition in which the liver becomes inflamed. This can be caused by many things, including viruses, bacteria, toxins as a result of liver injury, and by the body's own immune system [1]. Based on data from the World Health Organization (WHO), an estimated 3% of the world population has been infected with viral hepatitis. Beside, about approximately 170 million people have been chronically infected, and the number of Hepatitis patients grow 3-4 million people each year [2]. The prognosis is needed to determine the next treatment to stabilize and cure this disease. Prognosis is done by determining the status of the patient's condition based on the test results [3]. Thus, the status of the disease is very important to identify with the process of healing the patients with hepatitis.

The prognosis can be done using the computer instead of referring to the doctor. Some research on computer field has been done to diagnose this disease [4-6]. Among the methods used, the expert system is a method that performs inference based on the doctor's diagnosis. Recently, the usage of fuzzy to improve the performance of inference system is broadly used. Accordingly, determining the prognosis of some disease uses Fuzzy Inference System (FIS) has been done [7-9].

Fuzzy inference process the mapping of input to output with the fuzzy logic. This mapping will provide the basis for a decision or determination of the pattern. FIS could be used without hybrid it with another method [7]. But, to improve the performance, FIS is usually hybrid with some algorithm. This hybrid can be developed to pre-process the input data or to develop fuzzy rule [8-9].

Beside the data and rule base, the shape and boundaries of membership function are also a significant factor in determining the membership degrees of input parameter. However, there is often no reference or expert who can describe this limit. Therefore, the determination of membership function boundary would be very helpful if it can be done automatically. This process can be done using intuition or learning algorithm [9-14].

Particle Swarm Optimization algorithm (PSO) is a learning algorithm that is widely used. PSO is a stochastic search algorithm that uses a modeling problem through n - dimensional space to minimize or maximize the objective function of a problem [13, 15]. S study that compared the performance of PSO with clustering methods in data mining mentioned that in 500 running times, PSO has the best accuracy of 89.86 % with a mean accuracy of 82.66 % and the lowest accuracy is 73,55%. The Naïve Bayes method produces the best accuracy of 86.35%, with an average accuracy of 83.05 % and the lowest accuracy of 66.94 %. While the FDT (Fuzzy Decision Tree) method has the highest value of 78.15 % accuracy, average accuracy of 75.39 % and lowest accuracy of 61.49 % [6]. Therefore, this study used the PSO to seek the optimal limits of fuzzy membership functions of the hepatitis disease patient parameters.

2.0 EXPERIMENTAL

2.1 Membership Function

Fuzzy logic is used to overcome fuzziness phenomenon and is modeled through the truth value that is taken from the succession scale [16]. The fuzzy set can be characterized by using certain membership functions. This paper models a triangular membership function formed by three values, i.e. a, b and c [17, 18]. The term a and c in the triangular membership function describe the end point, while b describes the peak point. This function is represented using the following formula:

$$Triangular\,(x : a, b, c) = \begin{cases} 0, & x \le a \\ \dfrac{x-a}{b-a}, & a < x \le b \\ \dfrac{c-x}{c-b} & b < x \le c \\ 0 & x > c \end{cases} \qquad (1)$$

2.2 Particle Swarm Optimization

Particle Swarm Optimization (PSO) algorithm is a population based optimization technique to find the optimal solutions using a population of particles themselves. PSO is based on the idea that every particle crowd is a solution of the solution space [13]. The pseudo code of PSO algorithm is described in Figure **1**.

```
Start
for each particle i
   initialize the position and velocity
end for
while (not max iteration)
   for each particle i
      describe fitness value ωi
      if ωi is better than current pbest
           then pbest_fitness = ωi , pbest = sik
      if ωi is better than current gbest
           then gbest_fitness = ωi , gbest = sik
   end for
   for each particle i
      calculate vik
      update vik
   end for
   update ω
end while
End
```

Figure 1 Pseudo code of general PSO

The changing of velocity has been described in the pseudo code can be represented by the following equation [15]:

$$v_i^{k+1} = \omega v_i^k + c_1 rand_1 x(pbest_i - s_i^k) + c_2 rand_2 x(gbest - s_i^k) \qquad (2)$$

where,

v_{ik} : speed of agent i at iteration k
ω : weight function (inertia)
c_i : weighting factor
rand : random value between 0 and 1
s_{ik} : last position of agent i at iteration k
$pbest_i$: the best position of agent i
gbest : the best value of pbest

The value of inertia is updated at every iteration using the following equation [19]:

$$\omega = \omega_{\max} - \frac{\omega_{\max} - \omega_{\min}}{iter_{mx}} x \;\; iter \qquad (3)$$

where,

ωmax : initial value of inertia
ωmin : the last value inertia
itermax : maximum number of iterations
iter : number of the last iteration

The position of the agent will change at every iteration according to the velocity changing and can be calculated using the following equation [15]:

$$s_i^{k+1} = s_i^k + v_i^k \qquad (4)$$

where,

s_i^{k+1} : the position of the current agent
s_i^k : prior agent position
v_i^k : the current speed

2.3 Encoding Scheme of Particle

The process of encoding of particle is based on the parameter of input which is the hepatitis factors. The membership function of each linguistics variable uses triangular membership function. The system process six parameters, i.e. bilirubin, ALK Phosphate, AST, Albumin and ProTime of which each has three linguistics variable, and age parameter which has four linguistic variable. Therefore, the total linguistic function being processed is 19.

The triangular membership function is formed by three parameters, namely a, b, and c. Thus the total dimension of each particle is 57 dimensions obtained from 19 linguistic functions with each linguistic function have 3 parameters. The encoding scheme of particle for this paper is shown in Figure 2

a_1	b_1	c_1	a...	b...	c...	a_n	b_n	c_n
Linguistic 1			Linguistic ...			Linguistic n		

Figure 2 Encoding scheme of particle

2.4 Developing Membership Function

The process of developing the membership functions using PSO is shown in Fig. 3. The input of the process is maximum inertia (ω_{max}) and minimum inertia (ω_{min}), maximum iteration ($iter_{max}$), weight factor (c_1 , c_2), and the number of particles (particle_num). The fitness measurement is calculated based on the value of the membership degree of the particle positions (s). Initial gbest value (the best value of the particle of the whole dimension) use the membership functions are predefined and initial position value of the particle is randomly taken from the training data value. Each particle is designed in 3 dimensions, where each dimension (v) represents a boundary point of each membership function.

The results of modeling fuzzy membership function can be measured by the Mean Square Error (MSE). The calculation of MSE shown in Equation (5) is used as a fitness function for the implementation of the PSO algorithm [16]. The smaller of MSE value generated, the better of the results of experiments resulted.

$$MSE = \frac{\sum_{k=1}^{q}(y_k - \hat{y}_k)^2}{\sum_{k=1}^{q} y_k^2} \quad (5)$$

where,

y_k : target data value

$\hat{y}_k$: estimated data value

```
Start
Input : ωmax, ωmin,c1,c2, iter_max and particle_num
initialize s, v, spbest, msepbest, cost, r1, r2,
    gbest, msegbest
create random value of s0
v0=0, Gbest = initial_limit
while (msegbest > 0.01) or (i<= iter_max)
   For j=1 to particle_num
        calculate MSE
        check pbest
        create random value of r1, r2
        calculate ω, vi+1, si+1
        calculate gbest
   End for
End while
Ouput: new linguistic range
End.
```

Figure 3 The process of developing membership function using PSO

3.0 RESULTS AND DISCUSSION

The data used in this paper is a hepatitis patient data from the UCI machine learning repository. Total records are 80 without missing values. The data are divided into 50 training data and 30 testing data. Attributes used in this papers arenamely Age, bilirubin, ALK Phosphate, AST, Albumin, and Protime.

The test is done to obtain the best combination of PSO parameter values that is measured based on smallest MSE value. Testing is done with 4 scenarios. The first scenario is done by changing the parameter value of the maximum inertia (ω_{max}) and minimum inertia (ω_{min}). The second scenario is done by changing the values of constant local weights (c_1) and constant global weight (c_2). The third scenario is done by changing the value of particle number. The fourth scenario is done by changing the value of the maximum iteration. Each test was repeated 10 times and was recorded the value of the parameter that produces the smallest average error.

Table 1 Testing result of the combination of inertia

	Average MSE at ω_{min}		
ω_{max}	**0,1**	**0,2**	**0,3**
0,7	0,003619	0,004398	0,006481
0,8	0,002628	0,003738	0,004497
0,9	**0,002119**	0,002512	0,002905

The first scenario uses ω_{max}value of 0.7, 0.8, 0.9 and ω_{min} value of 0.1, 0.2, 0.3. This test is done at value of c1 = 2, c2 = 2, the maximum iterations = 100, and the number of particle = 10. The test result for the first scenario is shown in Table 1 where the greater the value of ω_{max} and the smaller the value of ω_{min} used produce increasingly smaller MSE value. The best results obtained at ω_{max} = 0.9 and ω_{min} = 0.1 with an average MSE is 0.002119.

The second scenario conducted with c1 of 0.5, 1, 2 and c2 of 0.5, 1, 2. Each experiment use ω_{max} = 0.9, and ω_{min} = 0.2, maximum iterations = 100, and the number of particles = 10. The testing results of this scenario are shown in

Table **2**.

By comparing the average MSE values at each testing value, it was found that the average value of MSE obtained at the value of c1 = 2, and c2 = 2. This table indicates that the best results obtained at balanced search of the local and global particles.

Table 2 Testing result for the combination of weight factors

	Average MSE at c_2		
c_1	**0,5**	**1**	**2**
0,5	0,005842	0,005455	0,004045
1	0,005581	0,005077	0,003647
2	0,003615	0,003132	0,002486

In the third scenario is used the number of particles between 10 to 200, ω_{max} = 0.9, ω_{min} = 0.1 , c1 = 2 , c2 = 2, and the maximum iteration = 100. The result of this scenario is shown in Table 3. This table shows that the experiment with the number of particles 200 produces the smallest MSE that is 0.002645. The table also shows that the decrease of MSE values on the number of particles 100, 150, and 200 is not too significant.

Table 3 Testing result for some particle number

Particle Number	MSE	Particle Number	MSE
10	0,004565	100	0,002830
20	0,004145	150	0,002697
50	0,003756	200	**0,002645**

In the fourth scenario uses the maximum iterations of 100, 200, and 500, ω_{max} = 0.9, ω_{min}= 0.1, c1 = 2, c2 = 2 and the number of particles = 100. The result of this scenario is shown in Table 4. This table shows that the smallest value of average MSE obtained at the number of iteration 700. However, a decrease in the value of MSE at iteration of 500, 600, and 700 is not too significant. This does not mean that the less iteration will give the worse result. The testing shows that the number of iteration has a great impact on the shifting value of the inertia. The greater the value of the maximum iteration, the shift value of inertia is also getting smaller.

Table 4 Testing result for some iteration number

Iteration Number	MSE
100	0,004771
200	0,003590
500	0,003016
600	0,002916
700	**0,002851**

As shown at (3), the maximum number of iterations associated with the particle velocity, in which the greater the maximum iteration will result the smaller inertia. Thus, these results suggest that the greater decline of inertia will result in a better MSE values and the resulting inertia value will go down faster. The amount of inertia will affect the speed of the particle. The magnitude of velocity at the beginning of the search will also affect the final result, because at high speeds these particles can reach the global optimum value freely. This is due to the condition that the search process will be more focused on the local search area at the smaller speed. The magnitude of the particle velocity that is move will be better if the particles move closer to the optimal value. However, when the particle moves away from the optimal value, then the value of the resulting MSE will also be greater.

Furthermore, the most optimal PSO parameter value is considered used in testing the limits of the membership functions. The values of the parameters used is 0.9, 0.1, 2, 2, 100, 500 respectively for maximum inertia, minimum inertia, weighting factor c1 and c2, the number of particles, and the maximum iteration. The results of these tests are shown in Table 5. The values of a, b and c indicate the end point and peak point of the membership function.

Table 5 The limits value of each membership function

Parameter	Linguistic	a	b	c
Bilirubin	Low	0,332	0,694	1,291
Bilirubin	Normal	0,385	0,929	1,660
Bilirubin	High	1,141	2,686	6,994
ALP	Low	34,313	61,734	64,288
ALP	Normal	45,569	95,807	157,626
ALP	High	103,788	137,850	256,502
Albumin	Low	1,249	2,566	3,768
Albumin	Normal	3,133	3,599	5,015
Albumin	High	4,706	5,323	6,970
SGOT	Low	10,542	12,521	21,222
SGOT	Normal	12,465	28,638	51,101
SGOT	High	36,211	55,588	358,174
ProTime	Low	13,453	18,286	24,169
ProTime	Normal	16,615	22,051	38,368
ProTime	High	22,613	52,588	131,779
Age	Child	6,473	14,935	22,943
Age	Adolescent	14,134	20,794	29,590
Age	Adult	17,317	35,090	54,372
Age	Old	44,917	61,610	72,766

4.0 CONCLUSION

This paper implements PSO algorithm to obtain the optimal end point and peak point of triangular membership function automatically for hepatitis patient data. The data is used for testing consist of 80 data without missing values. These data is obtained from the UCI machine learning repository. Based on the accuracy testing results that is measured using MSE, it can be concluded that the combination of PSO parameter that provide the best results are 0.9 and 0.1 for the maximum and minimum inertia, 2 for local and global weights, 200 for the number of particles, and 700 for maximum iterations. This optimal value of the PSO parameter can then be used to obtain the limits of the triangular membership function of the hepatitis disease data. The example of this limit value of each hepatitis parameter can be seen in Table 5. However, it needs further testing to observe the most optimal number of testing data due to the limitation of data in this paper. Furthermore, the use of the membership function of fuzzy inference system is needed to ensure whether it will offer the better accuracy or not.

Acknowledgement

I would like to thank to Program of Information Technology and Computer Science, Brawijaya University for the support to our research.

References

[1] B. Karlik. 2011. *Hepatitis Disease Diagnosis Using Backpropagation and the Naive Bayes Classifiers. Journal of Science and Technology*. 1 (1): 49-62.

[2] World Health Organization. 2002. *Hepatitis C*. http://www.who.int/csr/disease/hepatitis/Hepc.pdf. Accessed at 22 October 2013.

[3] N. Cutler. 2012. *Getting Real About Hepatitis C Prognosis*. http://www.hepatitiscentral.com/mt/archives/2012/07/getting_real_ab.html. Accessed 23 October 2013.

[4] E. Dogantekin, Akif D., Derya A. 2009. *Automatic Hepatitis Diagnosis System Based on Linear Discriminant Analysis and Adaptive Network Based on Fuzzy Inference System*. *Expert Systems with Applications*. 36 (2009): 11282–11286.

[5] M. Neshat, M. Sargolzaei, A. N. Toosi, A. Masoumi. 2012. *Hepatitis Disease Diagnosis Using Hybrid Case Based Reasoning and Particle Swarm Optimization*. International Scholarly Research Network. 2012. Article ID 609718, doi:10.5402/2012/609718.

[6] J. S Sartakhti, M. H. Zangooei, K. Mozafari. 2012. *Hepatitis Disease Diagnosis Using A Novel Hybrid Method Based On Support Vector Machine and Simulated Annealing (SVM-SSA)*. *Computer Methods and Programs in Biomedicin.* 2(2): 570–579.

[7] Debabrata P. K. M. Mandana, Sarbajit P., Debranjan S., Chandan C. 2012. *Fuzzy expert system approach for coronary artery disease screening using clinical parameters*. *Knowledge-Based Systems*. 36 (2012): 162–174.

[8] A. Keles, Ay. Keles, U. Yavuz. 2010. *Expert System Based on Neuro-Fuzzy Rules for Diagnosis Breast Cancer*. *Expert Systems with Applications.* 38 (2011): 5719–5726.

[9] U. Segundo, J.L. Cuadrado, L.A. Echevarria, T. A. Pérez, D. Buenestado, A. Iruetaguena, R. Barrena, J.M. Pikatza. 2014. *Automatic Construction of Fuzzy Inference Systems for Computerized clinical Guidelines and Protocols*. Applied Soft Computing 26 (2015): 257–269.

[10] C.C. Yang, N.K. Bose. 2006. Generating fuzzy membership function with self-organizing feature map. *Pattern Recognition Letters*. 27 (2006): 356–365.

[11] Byung-In Choi, F. Chung-Hoon Rhee. 2008. Interval type-2 fuzzy membership function generation methods for pattern recognition. *Information Sciences*. 179 (2009): 2102–2122.

[12] S. M. Bai, S. M. Chen. 2008. Automatically constructing grade membership functions of fuzzy rules for students' evaluation. *Expert Systems with Applications*. 35 (2008): 1408–1414.

[13] E.P. Kurniawan,. Hashim S.Z.M. 2010. *Fuzzy Membership Function Generation using particle Swarm Optimization*. *International Journal Open Problems Computation Math*. 3 (1): 27-41..

[14] A. Almeida,P. Orduña, E.Castillejo, D.López-de-Ipiña, M. Sacristán. 2013. *A method for automatic generation of fuzzy membership functions for mobile device's characteristics based on Google Trends. Computers in Human Behavior*. 29 (2013): 510–517.

[15] Y. Fukuyama. 2007. *Fundamentals of Particle Swarm Optimization Techniques in Modern Heuristic Optimization Techniques: Theory and Applications to Power Systems*. John Wiley & Sons, Inc, Hoboken, NJ, USA.

[16] V. Novak, I. Perfilieva, J. Mockor. 1999. Mathematical Principle of Fuzzy Logic. Kluwer Academic Publishers, London. http://www.academia.edu/5038898/Mathematical_Principles_of_Fuzzy_Logic.

[17] J. S. R. Jang, C.T. Sun, M. Eiji. 1997. *Neuro-Fuzzy and Soft Computing*. Prentice Hall, New Jersey, USA.

[18] S. Kusumadewi, H. Purnomo. 2004. *Aplikasi Logika Fuzzy Untuk Pendukung Keputusan*. Graha Ilmu, Yogyakarta.

[19] E. E. Omizegba, G.E. Adebayo. 2009. *Optimizing Fuzzy Membership Functions Using Particle Swarm Algorithm*. Proceedings of IEEE International Conference in Systems, Man, and Cybernetics, IEEE, San Antonio, TX, USA.

[20] B.Chen. 2013. *Fuzzy Inference System*. http://www.bindichen.co.uk/ post/AI/fuzzy-inference-system.html. Accessed 6 November 2013.