Expert Systems with Applications 37 (2010) 2066–2071

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

A hybrid immune-estimation distribution of algorithm for mining thyroid gland data

Wei-Wen Chang [a,*], Wei-Chang Yeh [a], Pei-Chiao Huang [b]

[a] Department of Industrial Engineering and Engineering Management, National Tsing Hua University, 101, Section 2, Kuang-Fu Road, Hsinchu, Taiwan, ROC
[b] Department of Industrial Engineering and Management, National Chiao Tung University, 1001 University Road, Hsinchu 300, Taiwan, ROC

ARTICLE INFO

Keywords:
Estimation of distribution algorithms
Immune algorithms
Genetic algorithms
Classification rules

ABSTRACT

In this paper we combine the main concepts of estimation of distribution algorithms (EDAs) and immune algorithms (IAs) to be a hybrid algorithm called immune-estimation of distribution algorithms (IEDA). Both EDAs and IAs are extended from genetic algorithms (GAs). EDAs eliminate the genetic operation including crossover and mutation from the GAs and places more emphasis on the relation between gene loci. It adopts the distribution of selected individuals in search space and models the probability distributions to generate the next population. However, the primary gap of EDAs is lock of diversity between individuals. Hence, we introduce the IAs that is a new branch in computational intelligence. The main concepts of IAs are suppression and hypermutation that make the individuals be more diversity. Moreover, the primary gap of IAs is to pay no attention to the relation between individuals. Therefore, we combine the main concepts of two algorithms to improve the gaps each other. The classification risk of data mining is applied by this paper and compares the results between IEDA and general GAs in the experiments. We adopt the thyroid gland data set from UCI databases. Based on the obtained results, our research absolute is better than general GAs including accuracy, type I error and type II error. The results show not only the excellence of accuracy but also the robustness of the proposed algorithm. In this paper we have got high quality results which can be used as reference for hospital decision making and research workers.

1. Introduction

In this paper, we attempt to propose the hybrid algorithm that considers the advantage of estimation of distribution algorithms (EDAs) and immune algorithms (IAs). The EDAs was first introduced by Larranaga and Lozano (2001). It is a search method that eliminates crossover and mutation from the genetic algorithms (GAs) and places more emphasis on the relation between gene loci. More precisely, it generates the next generation based on probability distribution of N superior population samples. In this way, the probability distribution estimated at each generation is progressively converted into a probability distribution that generates more superior individuals (Chen & Zhao, 2008). In addition, many combinatorial optimization algorithms have no mechanism for capturing the relationships among the variables of the problem (Inza, Larranaga, & Sierra, 2001). The EDAs considers the interactions between individuals that are performed by probability distribution, hence, this is main improvement from general GAs.

IAs emerged in the 1990s as a new branch in computational intelligence. The biological immune system is a complex adaptive system that has evolved in vertebrates to protect them from invading pathogens (Dipankar, 2006). The operative mechanisms of immune system are very efficient from a computational standpoint. The immune system mostly consists of the immune cells that most are lymphocytes. We can summarize the two main concepts of IAs; first, the immune response to secondary encounters could be considerably enhanced by storing some high affinity antibody producing cells from the first infection (memory cells), so as to form a large initial clone for subsequent encounters. Second, the process of hypermutation that means the mutation processes in lymphocytes. Random changes (mutations) take place in the variable region genes of antibody molecules. These random changes are mutational events and cause structurally different cells (Engin & Döyen, 2004). According to the above, we known the major ideas of two algorithms were the representation of probability model in EDAs and the mechanism of physiological immune systems in IAs. In addition, the main drawback in IAs is not to consider the interaction of variables that causes the phenomenon of local optimal. Therefore, we combined the two major ideas to be a hybrid algorithm called immune-estimation of distribution algorithms

* Corresponding author.
E-mail address: changwerwen@gmail.com (W.-W. Chang).

0957-4174/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2009.06.100

(IEDA) that was proven much effective and efficient in the experiments.

With the rapid growth of databases, data mining has become an increasingly important approach for data analysis (Yeh, Chang, & Chung, 2008). The operations research community has contributed significantly to this field, especially through the formulation and solution of numerous data mining problems such as optimization problems. Several operations research applications have also be addressed using data mining methods. One of the important tasks in data mining is classification (Olafsson, Li, & Wu, 2008). In classification, there is a target variable which is partitioned into predefined groups or classes. The classification system takes labeled data instances and generates a model that determines the target variable of new data instances. The discovered knowledge is usually represented in the form of if – then prediction rules, which have the advantage of being a high level, symbolic knowledge representation, contributing to the comprehensibility of the discovered knowledge (Mohamadi, Habibi, Abadeh, & Saadi, 2008).

The health care related data mining is one of the most rewarding and challenging area of application in data mining and knowledge discovery. The challenges are due to the data sets which are large, complex, heterogeneous, hierarchical, time series and of varying quality. The available healthcare data sets are fragmented and distributed in nature, thereby making the process of data integration is a highly challenging task (Delen & Patil, 2006). Moreover, data classification method has been applied in problems of medicine, social science, management, and engineering (Ryu, Chandrasekaran, & Jacob, 2007). In this paper, we adopt IEDA to discover the classification rules that the thyroid gland data from UCI database is applied. We will compare the effectiveness between IEDA and general GAs in experiments.

2. Literature review

2.1. Estimation of distribution algorithms

The EDAs emerged as a generalization of genetic algorithms (GAs), for the purpose of overcoming the two main problems: poor performance in certain deceptive problems and the difficulty of mathematically modeling a huge number of algorithm variants (Sun, Zhang, & Tsang, 2005). The primary concept was to extracts directly the global statistical information about the search space from the search so far and builds a probability model of promising solutions. New solutions are sampled from the model thus built (Gonzalez, Lozano, & Larranaga, 2002). Hence, the representation of probability model was a crucial process in EDAs. An appropriate probability model could ensure the effectiveness and efficient of algorithm. However, it was difficult and complicated to build an appropriate probability model. From now, many references had proposed various methodologies to build an appropriate probability model. The concept of EDAs was first introduced by Muhlenbein and Paas (1996) in 1996 and was later termed by Larranaga and Lozano (2001) in 2001. EDAs derive optimal solutions by developing probability models of each population. Compared with GAs, EDAs do not require genetic operations such as crossover and mutation to estimate the next generation. Instead, EDAs rely on selected individuals to model the joint probability distribution that can reflect the important feature of EDAs—precise description of the association between variables. General evolutional algorithms are not equipped with any mechanism that can correctly capture the global statistics in the previous search as a basis for future searches. EDAs develop probability distributions and use sampling and estimation to generate new generation solutions and improve the above drawbacks. Besides, the absence of crossover and mutation operators in EDAs can avoid the problem of prematurity and is thus an important contribution.

2.2. Immune algorithms

IAs also extend from GAs and use ideas gleaned from immunology to develop intelligent systems capable of learning and adapting, and have been widely applied to the various areas (Zuo & Fan, 2006). IAs is evolutionary algorithms based on physiological immune systems that have mechanisms to enable them to eliminate foreign substances. The mechanisms work by first recognizing foreign substances known as antigens. The immune systems then generate a set of antibodies to eliminate the antigens. These antibodies interact with the antigens to produce different results. The mechanisms are able to recognize which antibodies are better at eliminating the antigens and produce more variations of those antibodies in the next generation of antibodies (Alisantoso, Khoo, & Jiang, 2003). Fig. 1 shows the flowchart of a typical immune network algorithm (Timmis, 2007).

Therefore, in this paper we combine the main viewpoint of two algorithms to be a hybrid algorithm. This algorithm not only considers the interaction between individuals but also maintain the diversity between individuals.

3. The procedure of IEDA

We described the detailed steps of IEDA in the following:

Step 0: Set the three number T_L, T_C and T_U where $0 < T_L < T_C < T_U < 1$.
Step 1: Generate an initial solution randomly labeled by P_1.
Step 2: Evaluate the fitness of individuals and arranged in an order.

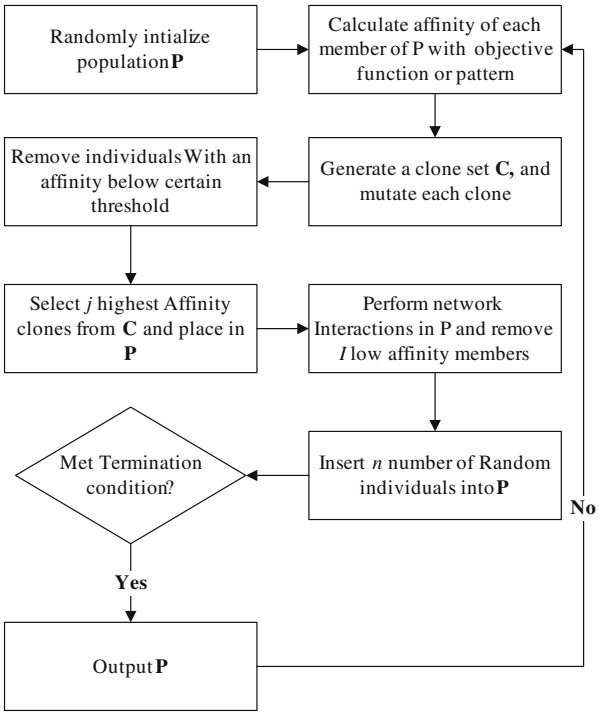

Fig. 1. The flowchart of a typical immune network algorithm (Timmis, 2007).

Step 3: Select individuals of the best $\delta\%$; M-inds and put them into the memory pool. We adopt each dimension of individuals in memory pool to build a probability model; $M_i - \text{Prob}_j$ where $i = 1,\ldots,m$, $j = 1,\ldots,n$.
Step 4: Select individuals of the worst $\delta\%$; S-inds and put them into the suppress pool. We adopt each dimension of individuals in suppress pool to build a reverse probability model; $S_i - \text{Prob}_j$ where $i = 1,\ldots,m$, $j = 1\ldots,n$.
Step 5: Generate a random number; ζ where $0 < \zeta \leqslant 1$.
Step 6: If $0 < \zeta \leqslant T_L$ then kept the present variable.
Step 7: If $T_L < \zeta \leqslant T_C$ then generate a variable via $M_i - \text{Prob}_j$.
Step 8: If $T_C < \zeta \leqslant T_U$ then generate a variable via $S_i - \text{Prob}_j$.
Step 9: If $T_U < \zeta \leqslant 1$ then generate a random variable; ξ where the domain of ξ depended on the problem.
Step 10: Weather the criterion of stopping is reached or not? If not then goto Step 2.
Step 11: End algorithm.

In the step 4, we not only achieve the effect of suppression but also maintain the diversity between individuals. Since, we still keep the worse individuals and build the reverse probability model. The different combination of variables may create better individuals, therefore in proposed IEDA does not eliminate worse individuals to maintain the diversity between individuals. We adopt the reciprocal of original probability and recalculate new probability of each dimension. Afterward we build the new probability model called reverse probability model. The example of Table 1 shows this process in detail. We assume the domain of dimension j to be 1, 2 and 3. At first, we sum up the number of each variable, and then we derive the original probabilities; 0.5, 0.333 and 0.167, respectively. Second, we transform the original probabilities into the type of reciprocal; 2, 3 and 6, respectively. Finally, we use these reciprocals to build the new probability model called reverse probability model. Hence, this process not only improves the diversity between individuals but also considers the mechanism of suppression. Fig. 2 shows the process of IEDA in detail.

4. IEDA for mining thyroid data set

4.1. Introduction of data set

This paper adopts the thyroid gland data from UCI database. Number of instances is 215 and includes five features and one class that is showed by Table 2. The data set contains 150 to be Normal (class = 1), 70 Hyper (class = 2), and the reminders is Hypo (class = 3).

4.2. Data preprocessing

We rearrange the configuration of data set according to the order. Since the original type of data set is floating point in part. The way of transformation is to convert the original data set to be integers according to order. For example, the feature of f1 that the first value is "65", and we convert "65" to be "1" and so on. It is convenient to be executed by IEDA. Table 3 shows parts of new data set via transformation.

4.3. Encoding

This paper adopts the Yeh et al.'s (2008) approach that the method of encoding is showed by Fig. 3. We define the feasible

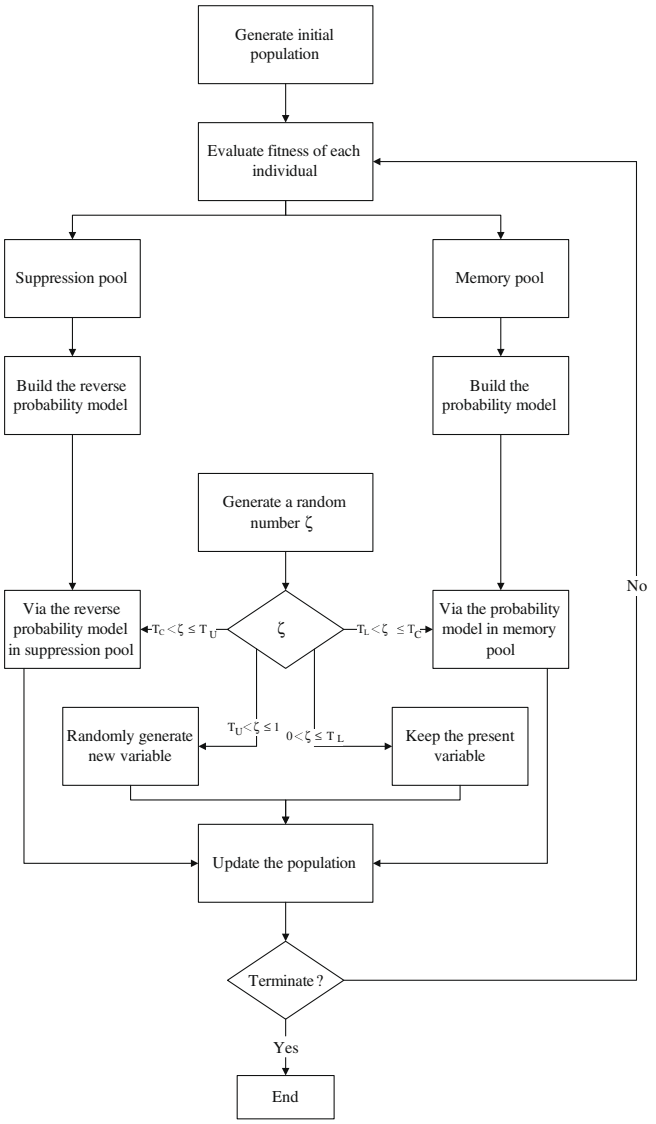

Fig. 2. The flowchart of IEDA.

Table 1
The process of reversing probability distribution.

	Dimension j		
Individual 1	2		
Individual 2	3		
Individual 3	2		
Individual 4	1		
Individual 5	1		
Individual 6	1		
Variable	1	2	3
Count	3	2	1
Original probability	0.5	0.333	0.167
Reciprocal	2	3	6
New probability	0.182	0.273	0.545

Table 2
The features of thyroid gland dataset.

Feature name	Domain	Simplified form
T3-resin uptake test	65–144	f_1
Total serum thyroxin	0.5–25.3	f_2
Total serum triiodothyronine	0.2–10	f_3
Basal thyroid-stimulating hormone	0.1–56.4	f_4
Maximal absolute difference of TSH value	−0.7 to 56.3	f_5
Class	1,2,3	Y
Class 1: Normal, class 2: Hyper, class 3: Hypo		

Table 3
Parts of results via transformation.

	Original dataset					Transformed dataset				
1	f_1	f_2	f_3	f_4	f_5	f_1	f_2	f_3	f_4	f_5
2	65	25.3	5.8	1.3	0.2	1	88	47	12	8
3	65	18.2	10	1.3	0.1	1	101	42	12	9
4	67	23.3	7.4	1.8	−0.6	2	98	45	17	2
5	68	14.7	7.8	0.6	−0.2	3	76	46	5	5
6	76	25.3	4.5	1.2	−0.1	4	101	37	11	6
7	79	19	5.5	0.9	0.3	5	90	41	8	10
8	80	23	10	0.9	−0.1	6	97	47	8	6
9	84	21.5	2.7	1.1	−0.6	7	89	36	10	4
10	84	18.5	4.4	1.1	−0.3	7	94	26	10	2
Domain	65–144	0.5–25.3	0.2–10	0.1–56.4	−0.7 to 56.3	1–54	1–101	1–47	1–47	1–85

Number of Feature Variable	Variable 1	> or = or <	Threshold	...	Variable k	> or = or <	Threshold

Fig. 3. The form of encoding.

Table 4
The TP, TN, FP and FN rate parameters.

Actual state	Predicted patient state	
	Classified as "true" (Positive)	Classified as "false" (Negative)
Class is "true" (Positive)	TP	FN
Class is "false" (Negative)	FP	TN

solution like that is a $1^{*}(3m+1)$ array called an "individual" and the grids called "dimensions". The first dimension represents the amounts of features that are chosen by IEDA. The second dimension represents the variable 1, third dimension that uses "1" represents ">", "2" represents "=" and "3" represents "<" and forth dimension represents threshold.

4.4. Fitness function

We can calculate the accuracy that represents the fitness value of each individual. The equation of accuracy, sensitivity and specificity are showed by the following. In terms of relative reference, we define the TP, TN, FP and FN rate parameters to show in the Table 4. The calculation of accuracy is the amount of the "class<> 1" to be select correctly plus the amount of the "class = 1" to be not select that divide the amount of data (Yeh et al., 2008).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Sensitivity = \frac{TP}{TP + FN}$$

$$Specificity = \frac{TN}{TN + FP}$$

4.5. The process of IEDA

We divide the raw data set into training data set and testing data set. The number of training data set is 142, and the remainders are testing data set. We adopt the method of 10-fold-validation (Chen & Hsu, 2006) to perform the robustness and reliability of algorithm. The process of mining thyroid gland data set by IEDA shows in Fig. 4.

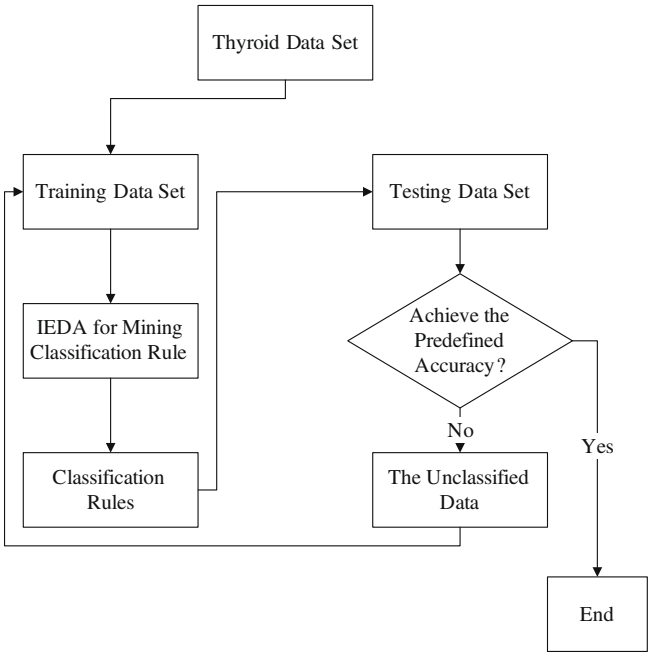

Fig. 4. The process of mining thyroid gland data set by IEDA.

5. Experiment

In our experiment, we want to classify between "Normal" and "Hyper" and between "Normal" and "Hypo". Hence, we divide the two parts that the purpose of first part is to find rules to classify between "Normal" and "Hyper", and the second part is to find rules to classify between "Normal" and "Hypo". Table 5 shows the results of first part that classify between "Normal" and "Hyper". We find the best and average accuracy of classification by IEDA to be better than by GAs. The best and average accuracy by IEDA are 0.9839 and 0.96775, respectively. Besides, the best and average accuracy by GAs are 0.9516 and 0.90805, respectively. In the second part, we want to find the classification rules between "Normal" and "Hypo". Similarly, both the best and average accuracy

Table 5
The results of first part.

	Rule	Accuracy of training	Accuracy of testing	Type I error	Type II error
The results of IEDA					
1	$f_2 > 69$ and $f_5 > 3$	0.8943	0.9839	0	1
2	$f_2 > 59$ and $f_5 < 21$	0.9675	0.9355	4	0
3	$f_2 > 55$, $f_4 < 21$ and $f_5 < 15$	0.9675	0.9839	1	0
4	$f_2 > 77$, $f_3 < 47$ and $f_4 < 19$	0.9106	0.9516	0	3
5	$f_2 > 68$ and $f_5 < 20$	0.9675	0.9839	0	1
6	$f_2 > 75$ and $f_5 > 3$	0.9106	0.9516	0	3
7	$f_2 > 70$ and $f_5 < 24$	0.9593	0.9839	0	1
8	$f_2 > 76$ and $f_3 < 48$	0.9268	0.9516	0	3
9	$f_2 > 70$ and $f_3 > 15$	0.9675	0.9677	0	2
10	$f_2 > 68$ and $f_5 < 20$	0.9675	0.9839	0	1
Average		0.94391	0.96775	0.5	1.5
Std.		0.0298	0.018639		
The results of GAs					
1	$f_2 > 87$ and $f_5 > 1$	0.9187	0.8387	0	10
2	$f_3 > 27$ and $f_5 > 3$	0.935	0.8871	1	6
3	$f_2 > 38$, $f3 < 4_5$ and $f_5 < 11$	0.9187	0.8548	7	2
4	$f_3 > 26$	0.935	0.9032	1	5
5	$f_3 > 27$ and $f_5 > 3$	0.935	0.8871	1	6
6	$f_2 > 75$ and $f_5 > 3$	0.9106	0.9516	0	3
7	$f_2 > 77$, $f_3 < 47$ and $f_4 < 19$	0.9106	0.9516	0	3
8	$f_2 > 74$ and $f_3 < 44$	0.9024	0.9516	0	3
9	$f_3 > 26$	0.935	0.9032	1	5
10	$f_2 > 75$ and $f_5 > 3$	0.9106	0.9516	0	3
Average		0.92116	0.90805	1.1	4.6
Std.		0.012757	0.042369		

Table 6
The results of second part.

	Rule	Accuracy of training	Accuracy of testing	Type I error	Type II error
The results of IEDA					
1	$f_2 < 22$	0.9667	0.9833	0	1
2	$f_2 < 21$ and $f_5 < 83$	0.95	0.95	2	1
3	$f_3 > 0$ and $f_5 > 63$	0.9417	0.9333	3	1
4	$f_4 > 25$	0.9667	0.9667	2	0
5	$f_2 < 20$ and $f_4 < 45$	0.9333	0.9667	2	0
6	$f_2 < 22$ and $f_3 < 28$	0.9667	0.9833	0	1
7	$f_2 < 28$ and $f_5 > 45$	0.9667	0.9833	1	0
8	$f_2 < 63$ and $f_4 > 24$	0.9667	0.95	2	1
9	$f_1 > 13$ and $f_4 > 25$	0.9667	0.9667	2	0
10	$f_2 < 22$	0.9667	0.9833	0	1
Average		0.94391	0.96775	1.5	0.5
Std.		0.0298	0.018639		
The results of GAs					
1	$f_2 < 46$, $f_3 < 9$ and $f_5 < 83$	0.8917	0.9167	5	0
2	$f_3 < 11$, $f_4 > 29$ and $f_5 < 85$	0.8917	0.9333	4	0
3	$f_4 < 47$ and $f_5 > 69$	0.9083	0.9167	5	0
4	$f_2 < 21$ and $f_5 < 84$	0.9583	0.95	2	1
5	$f_2 > 2$ and $f_5 > 61$	0.9167	0.9167	4	1
6	$f_2 < 63$ and $f_4 > 24$	0.9667	0.95	2	1
7	$f_3 < 13$, $f_4 < 48$ and $f_5 > 55$	0.925	0.9333	4	0
8	$f_3 < 21$, $f_4 > 20$ and $f_5 > 26$	0.9583	0.9333	2	2
9	$f_2 < 16$, $f_3 < 28$ and $f_4 < 46$	0.9083	0.9333	4	0
10	$f_1 > 18$, $f_3 > 5$ and $f_4 > 23$	0.9333	0.9	5	1
Average		0.92116	0.90805	4.6	1.1
Std.		0.012757	0.042369		

Table 7
The comparisons of IEDA and GAs.

	Algorithm	Best rule	Best accuracy	Type I error	Type II error
Part 1	*IEDA*	$f_2 > 55$, $f_4 < 21$ *and* $f_5 < 15$	*0.9839*	*1*	*0*
	GAs	$f_2 > 75$ and $f_5 > 3$	0.9516	0	3
Part 2	*IEDA*	$f_2 < 28$ *and* $f_5 > 45$	*0.9833*	*1*	*0*
	GAs	$f_2 < 63$ and $f_4 > 24$	0.95	2	1

Table 8
The type I and type II error of IEDA and GAs.

Actual class		Classified class	
		I(Normal)	II(Hyper)
Part 1			
I(Normal)	*IEDA*	*49(98.00%)*	*1(2.00%)*
	GAs	50(100.00%)	0(0.00%)
II(Hyper)	*IEDA*	*0(0%)*	*12(100.00%)*
	GAs	3(0.25%)	9(0.75%)
Part 2			
I(Normal)	*IEDA*	*49(98.00%)*	*1(2.00%)*
	GAs	49(98.00%)	1(2.00%)
II(Hypo)	*IEDA*	*0(0%)*	*10(100.00%)*
	GAs	2(0.20%)	8(0.80%)

Table 9
The average type I and type II error of IEDA and GAs.

	Average type I error (%)	Average type II error (%)
Part 1		
IEDA	1	12.5
GAs	2.2	38.3
Part 2		
IEDA	3	5
GAs	9.2	11

of classification by IEDA are better than by GAs. The details of result are showed in Table 6. Table 7 lists the best results in IEDA and GAs, respectively. In addition, Table 8 shows the best type I error and type II error of IEDA are better than GAs. Table 9 shows the average type I error and type II error of IEDA are also better than GAs.

6. Conclusion

Data mining is the search for valuable information in large volumes of data (Xiong, Kim, Baek, Rhee, & Kim, 2005). In this paper, we propose the hybrid algorithm that combines the immune algorithm and estimation distribution of algorithm called immune-estimation distribution of algorithm; IEDA and successfully applied to the classification risk of UCI thyroid gland data set. We compare the results between IEDA and traditional GAs. Based on the obtained results, our research absolute is better than GAs including accuracy, type I error and type II error. The results show not only the excellence of accuracy but also the robustness of the proposed algorithm. In this paper we have got high quality results which can be used as reference for hospital decision making and research workers. In future research, we will improve the effectiveness and efficiency of IEDA and make it apply more domains. Hence, we will draw the concept of estimation distribution of algorithm for continuous problems to our IEDA, and consider the conditional probability in establishing probability distribution of the IEDA. It will improve the IEDA to be more effective.

References

Alisantoso, D., Khoo, A L. P., & Jiang, A P. Y. (2003). An immune algorithm approach to the scheduling of a flexible PCB flow shop. *International Journal of Advanced Manufacturing Technology, 22*(11–12), 819–827.

Chen, T.-C., & Hsu, T.-C. (2006). A GAs based approach for mining breast cancer pattern. *Expert Systems with Applications, 30*(4), 674–681.

Chen, Y., & Zhao, Y. (2008). A novel ensemble of classifiers for microarray data classification. *Applied Soft Computing, 8*(4), 1664–1669.

Delen, D., & Patil, N. (2006). Knowledge extraction from prostate cancer data. In *Proceedings of the 39th annual Hawaii international conference on system sciences* (Vol. 5, pp. 92b–92b).

Dipankar, D. (2006). Advances in artificial immune systems. *IEEE Computational Intelligence Magazine, 1*(4), 40–49.

Engin, O., & Döyen, A. (2004). A new approach to solve hybrid flow shop scheduling problems by artificial immune system. *Future Generation Computer Systems, 20*(6), 1083–1095.

Gonzalez, C., Lozano, J. A., & Larranaga, P. (2002). Mathematical modelling of UMDAc algorithm with tournament selection. Behaviour on linear and quadratic functions. *International Journal of Approximate Reasoning, 31*(3), 313–340.

Inza, I., Larranaga, P., & Sierra, B. (2001). Feature subset selection by Bayesian networks: A comparison with genetic and sequential algorithms. *International Journal of Approximate Reasoning, 27*(2), 143–164.

Larranaga, P., & Lozano, J. A. (2001). *Estimation of distribution algorithms: A new tool for evolutionary computation*. Kluwer Academic Publishers.

Larranaga, P., & Lozano, J. A. (2001). *Estimation of distribution algorithms a new tool for evolutionary computation*. Dordrecht: Kluwer Academic Publishers.

Mohamadi, H., Habibi, J., Abadeh, M. S., & Saadi, H. (2008). Data mining with a simulated annealing based fuzzy classification system. *Pattern Recognition, 41*(5), 1824–1833.

Muhlenbein, H., & Paas G. (1996). From recombination of genes to the estimation of distributions. I. Binary parameters. In H.-M. Voigt, W. Ebeling, I. Rechenberg, & H.-P. Schwefel (Eds.), *Parallel problem solving from nature, PPSN-IV* (pp. 178–187).

Olafsson, S., Li, X., & Wu, S. (2008). Operations research and data mining. *European Journal of Operational Research, 187*(3), 1429–1448.

Ryu, Y. U., Chandrasekaran, R., & Jacob, Varghese S. (2007). Breast cancer prediction using the isotonic separation technique. *European Journal of Operational Research, 181*(2), 842–854.

Sun, J., Zhang, Q., & Tsang, E. P. K. (2005). DE/EDA: A new evolutionary algorithm for global optimization. *Information Sciences, 169*(3–4), 249–262.

Timmis, Jon (2007). Artificial immune systems—Today and tomorrow. *Natural Computing, 6*, 1–18.

Xiong, X., Kim, Y., Baek, Y., Rhee, D. W., & Kim, S.-H. (2005). Analysis of breast cancer using data mining and statistical techniques. In *Proceedings of the sixth international conference on software engineering, artificial intelligence, networking and parallel/distributed computing and first ACIS international workshop on self-assembling wireless networks* (pp. 82– 87).

Yeh, W.-C., Chang, W.-W., & Chung, Y.-Y. (2008). A new hybrid approach for mining breast cancer pattern using discrete particle swarm optimization and statistical method. *Expert Systems with Applications*. Available online 21 October 2008.

Zuo, X. Q., & Fan, Y. S. (2006). A chaos search immune algorithm with its application to neuro-fuzzy controller design. *Chaos, Solitons & Fractals, 30*(1), 94–109.