JOURNAL OF PURE AND APPLIED MICROBIOLOGY, September 2013. Vol. 7(3), p. 1989-1994

Research on Water Microorganism Recognition Based on BP Neural Network Algorithm

Xinwu Li

Electronic Business Department, Jiangxi University of Finance and Economics,
Nanchang - 3300 13, P. R. China.

(Received: 18 October 2012; accepted: 21 December 2012)

Technology of water microorganism recognition plays a key role in water quality detection and fields related. The paper presents a new water microorganism recognition algorithm based on BP neural network classifier technology and genetic algorithm. First, the structure of the presented BP neural network algorithm is analyzed and designed based on the analysis of the requirement of water microorganism recognition system. Second, in order to speed up algorithm calculation and simplify algorithm structure, genetic algorithm is used to improve ordinary BP neural network algorithm, and some specific measures are taken, and the calculation procedures of improved algorithm are redesigned. Finally, the realization and experiment results show that, compared with some methods which have relatively high accuracy, the algorithm can improve water microorganism recognition accuracy, decrease the calculation time needed greatly and can satisfy the engineering requirement in water microorganism recognition.

Key words: Water microorganism recognition, Classifier technology, BP neural network algorithm, Genetic algorithm.

Due to the intensified human activities, especially the acceleration of industrialization and globalization, water pollution is increasingly highlighted. Once the water resources are polluted, the water will be rich in nutrition; chemical elements like nitrogen, phosphorus, and etc. in the water will dramatically increase, providing a favorable condition for the mass reproduction of microorganism. Thus, the immoderate reproduction of microorganism will worsen the water body, even producing large quantity of microorganism and chemical elements harmful to human body. Therefore, technology for the recognition and detection of water microorganism appears particularly important. However, as it is difficult to recognize the water microorganism currently, traditional microscopic artificial recognition requires excellent professionals, which is also expensive, even having large personal errors, it is a simplest and efficient technology to automatically recognize microorganism by making use of information technology. Hence, technology for the recognition and detection of microorganism based on artificial intelligence technology is one of the research hotspots in the industry at present[1].

Literature Review

Currently, technologies for the recognition and detection of microorganism at home and abroad include the following major methods. Aerobic plate count (APC) is a method to detect the total number of bacteria generally used at home and abroad, which is to dilute the food to be detected into 2-3 kinds of appropriate dilution under sterile conditions, culture the same for 48 hours in the incubator of 36°C, and carry out the counting of colony count artificially after taking out. The method is accurate in counting, simple in operating, easy in mastering, thus widely used till now. However, the method needs long detection

---

* To whom all correspondence should be addressed.

time, high detection cost, and high detection environment requirement[2]; ATP luminous detection is to determine whether there is microbial infection in the sample food and the quantity of microorganism through the detection of the content of ATP substances in sample container. The method is easy and convenient in operating, but the detection results may have certain errors sometimes as it is unable to directly detect the food[3]; Impedance method is an electrochemical technology used for the rapid detection of bacteria. Generally, it takes less than 16 hours to detect the total number of bacteria by using the method; however, the detection time and accuracy of the method are related to the quantity of bacteria in the food sample[4]; Microorganism recognition based on artificial intelligence technology is mainly the classifier technology based on feature recognition. The method has such advantages as high automatic degree, less manual intervention, high recognition precision, and etc. But different classifiers may have certain technical defects, for example, neural network classifier is easy to fall into local optimization in the case of specific computation, thus leading to slow rate of convergence and long computing time[5].

The paper improve BP neural network with integration of genetic algorithm to overcome the question of slow convergence speed of BP neural network and presents some improvements of genetic algorithm. In so doing a new algorithm for water microorganism recognition is advanced and try to speed up algorithm convergence and simplify algorithm structure.

Design of Water Microorganism Classifier

Structure Design of water microorganism classifier

Flow chart of recognition classifier of water microorganism designed in the thesis is as shown in Picture 1, the steps of which are: ‘$ Collection of image information; a$impurities processing on images, i.e. images preprocessing ; b$ adopting image segmentation technique to separate images and extract the targets to be recognized; what is used in the system is watershed algorithm; making use of feature extraction technique to carry out feature extraction and optimization selection on feature vectors, thus feature vectors can be transferred to classifier to be recognized. Based on the self-learning and easy realization of programming of BP neural network, this thesis adopts BP neural network classifier. The designed the structure of water microorganism classifier see figure 1.

The design of BP neural network classifier shall be combined with projects. In this system, as what we need to solve is nonlinear problems in the closed interval, which can be solved with a hidden layer, three-layer BP neural network is hereby adopted. After extracting plenty of feature information of five kinds of microorganism, make use of conventional reservation method to train and detect classifier, where two thirds of the data are used to train the network, and one third of the data are used to detect the network, so as to detect the recognition rate.

The system is to recognize vorticella, epistylis plicatilis, colpidium, rotifer and nematode according to the characteristics of water quality. This paper selects 10 characteristics in total, so 10 input neurons of network are required. Due to the feature of S-type function in BP neural network, the characteristic values shall be normalized, and the normalized values shall be limited within [0.1,0.9]; so equation 1 is adopted to carry out the normalization[6].

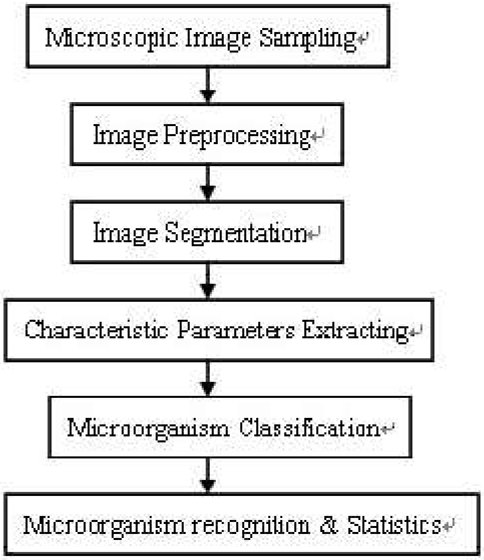

Fig. 1. The designed the structure of water microorganism classifier

$$x' = \frac{0.8(x - x_{min})}{(x_{max} - x_{min})} + 0.1 \qquad ...(1)$$

As the system here is mainly to recognize five kinds of microorganism, the number of neurons of network output layer is designed as 5. Binary five-digit is used to indicate the output. (10000), (01000), (00100), (00010) and (00001) represent vorticella, epistylis plicatilis, colpidium, rotifer and nematode respectively. Thus the mathematical corresponding relations of BP neural network are established. Input a ten-digit vector and output a five-digit binary number through function corresponding relations of network, so as to distinguish the classification of five kinds of microorganism. Therefore, we have designed the rough input-output form of BP network classifier required by the recognition system of water microorganisms by and large[7].

Improvement of BP Neural network Algorithm

An ANN (Artificial Neural Networks) algorithm can be described by the connecting method of finite parameters such as neuron, network layers, neuron number of each layer and neuron, weight of each connection and transfer function. So we can encode an ANN algorithm and realize the learning process of neural network with genetic algorithm.

Parameter Setting

Input population size P, network layers (not including input layer), neuron number of each layer. Genetic algorithm has excellent robustness towards the setting of these parameters; changing these parameters won't exert great impact on obtained results[8].

Initialization and Evaluation

Randomly generate initial population $P=(x_1,x_2,...,x_n)$, any $x_i \in P$ being a neural network weight, which is comprised of a weight vector and a threshold vector, weight vector being n -dimensional real vector, being the number of all the connection weights, threshold vector also being -dimensional real vector (not including neuron of input layer). Each network weight is equal to a chromosome; there are such chromosomes, i.e. population size. The neurons are numbered from the bottom to the top, from the left to the right (including input neuron).

Table 1. The part characteristic values used in the paper

S. No	Area	Circumference	Rectangularity	length-width ratio	intensity	morphology ratio
1	0.1073	0.3940	0.1304	0.2089	0.9000	0.2257
2	0.1011	0.3647	0.2150	0.3472	0.8103	0.2695
3	0.1126	0.2057	0.1304	0.2089	0.8375	0.1490
4	0.1000	0.3227	0.3695	0.2299	0.7422	0.1709
5	0.1023	0.4254	0.3583	0.1359	0.7223	0.1586
6	0.1600	0.1918	0.6253	0.2732	0.3080	0.2347
7	0.3326	0.1777	0.6393	0.6029	0.1916	0.7345
8	0.1171	0.1380	0.6293	0.4391	0.3131	0.5322

Table 2. The Average Recognition Performance of Different Microorganism

	Vorticella	Epistylis	Plicatilis	Colpidium	Rotifer	Nematode
Accuracy Rate	90.15%	96.89%	94.72%	92.24%	85.78	89.73%
Time Consuming(S)	16	19	11	17	15	14

Table 3. The Recognition Performance of Different Algorithms

Algorithm	Algorithm in This Paper	Ordinary BP Neural Network Algorithm	ordinary Genetic algorithm
Accuracy Rate	92.94%	81.69%	61.33%
Time Consuming(S)	16	631	24

According to corresponding neural network of randomly generated weight vector and threshold vector, as for the given input set and output set, calculate the global error of each neural network, as genetic algorithm can only evolve towards the direction of increasing fitness. So the fitness function can be formed according to equation 2 and equation 3, among which is the adaptive value of the i^{th} individual, being the number of chromosome, being the number of nodes of output layer, being the number of learning samples, being the output value of the x^{th} node while inputting the p^{th} training sample, being the anticipated output value[9].

$$f_i = 1 / E_i \qquad \text{...(2)}$$

$$E_i = \sum_{p=1}^{m}\sum_{k=1}^{n}(v_{pk} - T_{pk}) \qquad \text{...(3)}$$

Selection Operator

This thesis adopts the mass selection operator combining spinning roulette wheel strategy with optimal retention strategy. Selecting process takes the spinning roulette wheel as basis, which is a kind of playback random sampling method. All the selections are to select good individual according to individual fitness from current population in the light of certain criterion to enter the next generation population, the basic ideal of which is that the selective probability of each individual is equal to the ratio of its fitness to the individual fitness among the entire population. The higher the individual fitness is, the greater the possibility to be selected is, and the greater the probability to enter next generation is. However, due to random operation, the selection error of this method is relatively big, sometimes even making the individual with high fitness be selected. In order to improve the convergence of genetic algorithm, this thesis adopts optimal retention strategy, selecting individual with the largest fitness as seeded player, directly retaining to the next generation. Substitute the worst individual in the population with the optimal individual recorded by the preceding generation while forming new population every time, so as to preventing the individual with optimal fitness in current population from being destroyed.

Crossover and Mutation Operator

Improved Adaptive Crossover Probability and Mutation Probability

In the parameters of genetic algorithm, the selection of crossover probability P_c and mutation probability P_m is the key to influence the behavior and performance of genetic algorithm, exerting a direct impact on the convergence of algorithm. In the simple genetic algorithm, as the values of P_c and P_m are constant, it is not efficient enough to solve multivariable complication optimization problems, having the problems of prematurity or misconvergence. Srinivas and etc. put forward adaptive genetic algorithm, AGA, the basic idea of which is that the individual with fitness higher than average fitness in the population adopts the smaller crossover probability P_c and mutation probability P_m, aiming at retaining individual with favorable structure so as not to be destroyed and to enter the next generation; as for individual with fitness lower than average fitness, using higher crossover probability and mutation probability to facilitate the elimination of such individual. Although this method is improved compared with simple genetic algorithm, there are still some problems. For example, while the fitness is close to the largest fitness, the crossover probability and mutation probability are; while equal to the largest fitness, the crossover probability and mutation probability are zero, which makes AGA undesirable in the early stage of evolution. As in the population of early stage of evolution, more optimal individuals are in an unchangeable state, and the favorable individual at this time is not always the globally optimal solution, which is easy to make the evolution tend to be locally converged. Hence, this thesis, based on this, adopts improved adaptive algorithm, making the individual crossover probability and mutation probability of largest fitness in the population be not zero, as shown in equation 4 and equation 5, f_{avg} in which represents the average fitness of population of each population; f_{max} represents the largest fitness in the population; f' represents larger fitness of two individuals to be crossed over; f represents the fitness of individual to be mutated in the population. p_{c1}, p_{c2}, p_{m1} and p_{m2} are design parameters, which are 0.9, 0.6, 0.1, 0.001 respectively[9].

$$P_c = \left\{ \begin{array}{ll} p_{c1} - \dfrac{(p_{c1} - p_{c2})(f' - f_{\max})}{f_{\max} - f_{avg}}, & f' \geq f_{avg} \\ p_{c1} & f \leq f_{avg} \end{array} \right] \quad ...(4)$$

$$P_m = \left\{ \begin{array}{ll} p_{m1} - \dfrac{(p_{m1} - p_{m2})(f' - f_{\max})}{f_{\max} - f_{avg}}, & f' \geq f_{avg} \\ p_{m1} & f \leq f_{avg} \end{array} \right] \quad ...(5)$$

Improved AGA not only keeps the adaptive advantage of AGA but also conquers the shortage of slow evolution of population in the early stage, having favorable optimization function.

Crossover Operator

First, in the population, according to the crossover probability P_c obtained in, randomly select certain quantity of chromosomes as parents, and randomly select a breakpoint, exchanging the gene strand on the right (or top) of the breakpoints of parents, generating new filial generation; finally, substitute the paternal chromosome with filial generation chromosome, generating new population.

Mutation

Similar to the selection of paternal generation in crossover process, as for each selected chromosome to be mutated, in order to get better mutation, multiple mutation is permitted. While mutating, first randomly generate a vector with the same dimension as each weight and threshold of chromosome, and add to the selected vector to be mutated. As to the result of each mutation, restore neural network and carry out performance evaluation. If the descendant is better than paternal generation, the mutation of paternal generation shall be ended; otherwise, carry out next mutation on paternal generation, until finding out descendant better than paternal generation.

Immigration Operator

It is found through the test that in the search process of genetic algorithm, the individual with highest fitness in the population at present is possible to participate in crossover and mutation calculation, just with small probability; on the contrary, the lower the fitness of the individual is, the larger the probability to be selected to participate in crossover and mutation is, but the generated individual fitness is very low, and the global search performance on algorithm is not obviously increased. Therefore, this thesis introduces immigration operator which is a good method to avoid prematurity. In the immigration process, it can only accelerate the elimination of bad individual, but also increase the diversity of solution, further meeting the evolutionary mechanism of creatures. Immigration operator eliminates the worst individual with certain elimination rate (generally 15%~20%) in the evolutionary process of each generation, and generates part of excellent immigrants to supplement the population. Excellent immigrants here are mutated and generated through the multiple crossovers on those individuals to be eliminated. Thus, not only fully retain the good gene genetic pattern of paternal generation but also guarantee the diversity of population, improving the optimization searching performance of GA.

End of Operation

If the network error meets the requirement or reaches certain evolution generations, the evolution shall be stopped and the evolution result shall be outputted; otherwise, turn to Step (3).

Calculation Procedures of Improved Algorithm

Initial Population, including the population size and the initialization of each weight (generate according to the method for neural network to generate initial weight), and encode it; a. Calculate the selection probability of each individual and sort them; b. Select good individual to enter next generation population according to spinning roulette wheel selection strategy; c. In the new generation population, select adaptive individual to carry out crossover and mutation operation according to adaptive crossover probability and mutation probability to generate new individual; d. Insert the new individual into the population and calculate the fitness of new individual; e. Immigration operator operation. Judge whether there is “prematurity phenomenon”, if there is, immigration strategy shall be adopted and turn to step a; f If the satisfactory individual is found, it shall be ended; otherwise, turn to a.

After reaching required performance indicator, decode the optimal individual in final population, then the optimized network connection weight can be obtained.

Experimental Results and Analysis

Data Collection and Process

This paper adopts the water in East Lake in Nanchang, Jiangxi, China as detection sample; due to limited space, here only list such six normalized characteristic values as area, circumference, rectangularity, length-width ratio, intensity and morphology ratio of colpidium, all the above characteristic values can be seen in Table 1.

Experimental Results

As for the performance of the presented algorithm, this paper also realizes the application of the improved algorithm, the ordinary BP neural network[3] and ordinary genetic algorithm[9], recognition performance of different algorithms is shown in table 2 and table 3. Table 2shows the recognition accuracy rates of different water microorganism are shown. Table 3 shows the avarage the recognition accuracy rates and calculation time needed of the three algorithms. And the calculation platform as follows: hardware is Dell Poweredge R710, in which processor is E5506, memory 2G, hard disk 160G; software platform is Windows XP operating system, C programming language environment.

CONCLUSION

The recognition of microorganisms in water is a complicated and multi-factor system problem, the study on which has certain difficulty. So, this paper, on the consideration of actual characteristics of microorganisms in water, designs a classifier for microorganisms recognition in water based on imporved BP neural network, and put forward some improvements bansed on genetic algorithm to speed up the calculation of the BP neural network algorithm according to the recognition requirement of multi-factor complicated system. Test results indicate the engineering practicability of the algorithm presented in this paper .Hence, algorithm in this paper has universal applicable value. In the next study, we shall pay attention to the combination of generality with individuality of the algorithm, as well as the robustness of recognition methods of microorganisms in water.

REFERENCES

1. Doloca S P. Feature Selection for Texture Analysis Using Genetic Algorithm .*International J Comp Math.* 2010; **74**: 274-292.
2. Handels S. Y. Feature Selection for Optimized Skin Tumor Recognition Using Genetic Ajogrithms, *Artificial Intelligence in Medicin,* 2011; **18**(4): 283-297.
3. Dinstein G.G. and Moral S.T.Textural features for image classification. *IEEE Trans System Man Cybern,* 2012; **76**(4): 296-309.
4. Patricia M, Victor H. Genetic Optimization of Neural Networks for Person Recognition Based on the Iris. *TELKOMNIKA Indonesian Journal of Electrical Engineering.* 2012; **10**(2): 309-320.
5. Wang X Q. Study on Genetic Algorithm Optimization for Support Vector Machine in Network Intrusion Detection. *Advances in Information Sciences and Service Sciences.* 2012; **4**(2): 282-288.
6. Yan X S,Wu Q H,Zhang C. An Improved Genetic Algorithm and Its Application. *TELKOMNIKA Indonesian Journal of Electrical Engineering.* 2012; **10**(5): 1081-1086.
7. Zhang D M, Lu X L. Minimizing Network Coding Resources with Improved Genetic Algorithm. *International Journal of Digital Content Technology and its Applications.* 2013; **7**(3): 464-471.
8. Li X H, Yu A. Aquaculture Monitoring System Design Based on Genetic Algorithm. *International Journal of Advancements in Computing Technology.* 2013; **5**(3): 779-788.
10. Ding S F,Jia W K. An Improved Genetic Algorithm Based on Factor Analysis. *Journal of Convergence Information Technology.* 2010; **5**(4): 103-108.