Rongjuan Yang*, Chuxue Wu, Yuhui Ge, Cheng Fang, Decai Yang

Cancer Trends Predication Based on an Intelligent Method

Abstract: Cancer incidence is a complicated dynamic system affected by many factors. Predicting cancer trends can help evaluate the progress in cancer control while reinforcing prevention activities. To forecast cancer trends, one mathematical model to analyze the development of cancer incidence is proposed. Based on the analysis of this model, one new intelligent method for predicting cancer trends has been proposed. For this new intelligent method, one new evolutionary neural network, combined with the modified back propagation algorithm and the black hole algorithm, was applied. Finally, long term monitoring of data on cancer trends from Canada and China were used to verify the good performance of this new intelligent method.

Keywords: cancer incidence, cancer trends, predication, evolutionary neural network, intelligent method

1 Introduction

Nowadays, cancer has become a leading cause of the death throughout the world. In China, it is the leading cause of death in urban areas and the second most common cause in rural areas for many years [1]. Many important factors, such as the ageing population, severe environmental pollution, uncontrolled chronic infection, and adoption of westernized lifestyles, cancer in China is expected to become more and more severe [2]. Moreover, in Canada, at the beginning of 2007, nearly 750 000 Canadians had a diagnosis of cancer more than in the previous 10 years [3]. Therefore, cancer has become the enemy of human health. To control the cancer incidence, it is very important to predicate the cancer trends based on the long time monitoring of cancer data.

***Corresponding author: Rongjuan Yang**, Hospital of Wuhan Polytechnic University, Wuhan, Hubei, 430023 P. R. China, E-mail: xiao12yang@163.com

Chuxue Wu, School of Optical-Electrical and Computer Engineering, University of Shanghai for Science & Technology, Shanghai, 200093 P. R. China

Yuhui Ge, Business School, University of Shanghai for Science & Technology, Shanghai, 200093 P. R. China

Cheng Fang, College of Health Science & Nursing, Wuhan Polytechnic University, Wuhan, Hubei, 430023 P. R. China

Decai Yang, Hubei University of Chinese Medicine, Wuhan, Hubei, 430065 P. R. China

To predict cancer trends, some basic laws hidden in the historical data of cancer incidence such as some mathematic functions, must be recognized and analyzed [4-5]. Because the artificial neural network (ANN) can draw its law from very complicated data, and can approximate very complicated function, then the artificial neural network can be used to model the non-linear dynamic system [6]. Generally, modeling the non-linear dynamic system is to model the time series of this system. The ANN can solve this problem very well [7]. Therefore, the ANN can be used to predicate cancer trends. To improve the ANN, in this study, one new intelligent method using one new evolutionary neural network (ENN) is applied. At last, some long time monitoring of cancer trends data from Canada and China were used to verify the good performance of this new intelligent method.

2 Mathematic Model of Cancer Incidence

As the cancer incidence system is a complicated and dynamic system [4-5], it can be described mathematically as follows.

The evolvement of a dynamic system can be analyzed using some differential equations [8]. In this study, the cancer incidence is supposed to be described by one q -th order ordinary differential equation, as.

$$\frac{d^p x}{dx^p} = f(x, x', \dots, x^{(p-1)}) \quad (1)$$

If the solution for above equation is one state variable of this system, then the time series of this variable can be used to model the dynamic system. In this study, the follow third order ordinary differential equation is applied.

$$\frac{d^2 x}{dt^2} = f(x, \frac{dx}{dt}) \quad (2)$$

If a time series $x(j)$ $j = 1 \sim n$ with step h , is the solution of the equation (2), then the numerical solution of the equation (2) can be used to model the system, which is,

$$\frac{d^2 x}{dt^2} \Big|_j = \frac{x(j+1) - 2x(j) + x(j-1)}{h^2} + o(h^2) \quad (3)$$

$$\frac{dx}{dt} \Big|_j = \frac{x(j) - x(j-1)}{h} + o(h) \quad (4)$$

Thus, the time series can be described as,

$$x(j+1) = h^2 f[x(j), \frac{x(j) - x(j-1)}{h}] + 2x(j) - x(j-1) + o(h^3) \quad (5)$$

The function (5) can be generalized to p -th order by retrieving its high order minim. Thus, the time series can be described as,

$$x(j+1) = F[x(j), x(j-1), h] \quad (6)$$

Therefore, we can see that, the above function can be used to describe the dynamic evolvement model of cancer incidence. And the late state of this system can be described by some of its prophase states.

In the traditional method, the function relationship F is supposed as a linear function. But in the cancer incidence system, the F should be a very complicated non-linear function. Thus, the above function is always very complicated and its description cannot be confirmed. To this problem, the traditional method cannot solve it. For the ANN, the information is stored in linkage of network hidden, and cannot be described with one function. Moreover, the neural network has the adaptivity for a dynamic process [9]. Thus, the ANN can be used to approximate the function relationship F of equation (6) very well.

3 Evolutionary Neural Network for Cancer Trends Predication

From the above analysis, the number of input neuron for ANN can be confirmed by the dynamic system, but the cancer incidence system described by n -th order differential equation is an uncertain system. Therefore, the lingering time step of this time series for the ANN cannot be obtained. However, it is a very important parameter to construct ANN samples, as it can seriously affect the predication results [9]. Nowadays, there is no good method to determine this parameter. Normally, in general applications, it is determined by the experiment.

Therefore, to model the cancer incidence system by the ANN, the most important work is to determine the construction of the ANN. Because the selection of the construction of the ANN is a very typical optimization step, the global optimization-evolutionary algorithm can solve this problem very well [10]. And this kind of ANN is called ENN. Nowadays, the ENN has been used in time serials data forecasting field [11].

In this study, the neural numbers of hidden layer and input layer are determined by the evolutionary algorithm, and one new ENN is proposed. In this new ENN, the construction of ANN is determined by the black hole algorithm [12] and the weights are determined by a modified back propagation algorithm. In this study, the three layers ANN is used. Therefore, in this ENN, there are four parameters to be evolved by the black hole algorithm, which are the numbers of the input neuron and the hidden layer neuron, iterating step η and inertia parameter α . Moreover, in this study, the indirect coding method and full-linking network are used.

3.1 Black Hole Algorithm

One new nature-inspired metaheuristic algorithm inspired from the black hole phenomenon, is the black hole algorithm (BHA) which was firstly proposed by

Hatamlou [10] in 2013. Studies show that, for BHA, the controlling parameters are very few and the structure is very simple. However, the performance of the BHA is very good. Thus, nowadays, the BHA has been used in engineering problems [13]. In this ENN, to obtain the good ENN, the BHA is used.

The detailed process for the BHA is as following.

1. *Creation the initial stars*: Generally, in the search range of the problem, the initial stars (candidate solutions) are created randomly. In this algorithm, one star is its representation. In other words, for the search range of a n-dimension space, the star is a n-dimension vector.
2. *Fitness value computation*: The fitness value is to evaluate the star. Moreover, the progression of the algorithm is controlled by the fitness value, which is its criteria. In this method, the transformation of the objective function for one star is its fitness function.
3. *Black hole selection*: The star whose fitness value is the best is selected as the black hole.

The stars that surround the black hole will be absorbed by the black hole. Thus, when the first black hole is initialized, all the stars that start to move towards the black hole and the stars around it will be absorbed.

4. *Stars movement*: Because the black hole has the ability to absorb the stars that surround it, thus, all stars must move towards the black hole and change their locations. This process is as follows,

$$x_i(t+1) = x_i(t) + rand \times (x_{BH} - x_i(t)). \quad (7)$$

where n is the stars number; $rand$ is a random number whose interval is $[0, 1]$; x_{BH} is the location of the black hole; $x_i(t)$ and $x_i(t+1)$ are the locations of the i -th star at iterations t and $(t+1)$ respectively; $i = 1, 2, 3, \dots, n$.

5. *Stars replacement*: For BHA, to select the more optimal star, the probability to cross over the event horizon for the black hole by the moving stars is applied. The radius of the event horizon is calculated as follow,

$$R = \frac{f_{BH}}{\sum_{i=1}^n f_i}. \quad (8)$$

where f_{BH} is the fitness values of the black hole and f_i is the fitness value of the i -th star.

6. *Update black hole*: In the moving process of stars, when a star reaches one location whose cost is the lowest cost, that is, the fitness value of this star is better than that of the black hole. Thus, this star will replace the black hole. This process is called update black hole.

After all the stars have been moved, the next iteration will begin.

7. *Termination condition*: In this study, when the number of iterations performed without replacing the black hole reaches five, the algorithm process will stop. In other words, if the black hole is not replaced for five iterations, the algorithm will stop. Moreover, to avoid infinite iteration of the algorithm process, in this study, the maximum number of iterations must be given.

Based on above analysis, the basic flow chart for the BHA can be determined, shown in Fig. 1.

Figure 1: Flow chart of BHA

3.2 Evolutionary Neural Network

The details of ENN are as follows.

- 1) Firstly, the search ranges for the four optimization parameters, which are numbers for the input and layer neurons, iterating step η and inertia parameter α , must be given. Moreover, the controlling parameters for ENN, including maximum number of iterations, star number in one population, output neuron number, criteria for the iterating stop and iterating error of the modified back propagation algorithm, must be given too.
- 2) In this ENN, two random numbers for neuron of the input and hidden layers can determine the network construction. And two random values for parameters η and α can determine one modified back propagation algorithm. Using this network construction and this modified back propagation algorithm, one initial star can be created.
- 3) The fitness value for one star is computed by the following steps.
 - a) According to the number of input neuron, the whole time series of cancer trends data is divided to construct the learning samples for the ENN.
 - b) To improve the performance of ENN, the learning samples are divided into two parts, which are training samples and testing samples. The training samples are used to determine the non-linear mapping of the ANN, and the testing samples are used to test the generalization of the ANN.
 - c) The initial linking weights for one star are generated randomly.
 - d) For the modified back propagation algorithm, the iterating step is taken as $j = 1$.
 - e) For one individual, its square error $E(j)$ trained by testing samples is computed, and taken as the minimum error, $\min E = E(j)$. And then, the algorithm process returns to step (3).
 - f) For one individual, if the square error trained by training samples is less than the iterating error criteria for modified back propagation algorithm, its fitness value is still to be $\min E$. And then, the process returns to step (3) too.
 - g) Using the modified back propagation algorithm, the linking weights for the individual stars are adjusted.
 - h) The iterating step increased by one, that is, $j = j + 1$, then, the process returns to step e.
 - i) If the iterating step reaches the given iterating stop criteria for modified back propagation algorithm, the fitness value of the individual star equals to $\min E$. And then, the process returns to step (3) too.
- 4) The black hole is selected.
- 5) All the individual stars in the population are moved.
- 6) The event horizon radius of the individual stars are computed. And the sucked stars are replaced.
- 7) Using the fitness computation method shown in step (3), the fitness values for all individual stars are computed.

- 8) The black hole is selected again.
- 9) The number of iterations increases 1. If the number of iterations reaches its stop criteria, the algorithm process will stop. Therefore, the black hole at this time is the best ANN for prediction the time series of cancer trends data.

Using the best searched ANN for predicting the time series of cancer trends data, the real time series of cancer trends data are applied, and the forecasting data can be obtained.

4 Real Example Application

Using the obtained ENN, some long term monitoring of cancer data from Canada and China have been used to predict cancer trends.

Firstly, the data of four main cancers in Canada, which are prostate cancer for males, lung cancer for males, colorectal cancer for males, and breast cancer for females, were applied. In this study, 21 sets of data from 1970-1990 for the age-standardized incidence rate were used [3], which are shown in Figures 2-5. The detailed description of long term monitoring of cancer data can be found in the reference [3].

Figure 2: Data of prostate cancer for males

Figure 3: Data of lung cancer for males

Figure 4: Data of colorectal cancer for males

Figure 5: Data of breast cancer for females

In this study, to construct the ENN model, the former 19 sets of data were used. Moreover, to verify the predication, the last 2 data sets were applied. In the 19 data sets, the former 14 sets of data were selected to establish the training samples, and the last 5 data sets were selected to establish the testing samples. For four data sets, the controlling parameters of ENN were as follows,

The maximum number of iterations was 200. The star number for one population was 50. The output neuron was 1. All the criteria of the iterating stop and iterating error for the modified back propagation algorithm were $10e-5$.

After computation, the best construction of ANN for four data sets can be obtained, which are 6-11-1 for the data set of prostate cancer, 7-15-1 for the data set of lung cancer in males, 6-13-1 for the data set of colorectal cancer in males, and 8-12-1 for the data set of breast cancer in females. Using those ANNs, the predication results for four data sets of cancer trends are shown in Fig. 2-5.

From the above four figures, it can be found that, the predication results of new intelligent method are comparable with the real monitoring data, thus, the cancer trends can be predicted by the new intelligent method very well.

Finally, the data from two main types of cancers in China, which are lung cancer for males and breast cancer for females, were applied. In this study, 12 sets of data from 2000-2011 for the incidence rate were used [14], which are shown in Figures 6 and 7. The detailed description of the long term monitoring of cancer data can be found in the reference [14].

Figure 6: Data of lung cancer for males

Figure 7: Data of breast cancer for females

In this study, to construct the ENN model, the former 9 data sets were used. Moreover, to verify the predication, the last 2 data sets were used. In the 9 data sets, the former 7 data sets were selected to establish training samples, and the last 2 data sets were selected to establish testing samples. For two data sets, the controlling parameters of ENN were as follows,

The maximum number of iterations was 200. The star number for one population was 50. The output neuron was 1. All the criteria of the iterating stop and iterating error for the modified back propagation algorithm were $10e-5$.

After computation, the best construction of ANN for two data sets were obtained, which are 4-15-1 for data set of lung cancer for males and 4-12-1 for data set of breast cancer for females. Using those ANNs, the predication results for two cancer trends data set are shown in Fig. 6 and 7.

From the above two figures, the predication results of the new intelligent method correlate with the real monitoring data well, thus, the cancer trends can be predicted by the new intelligent method very well.

5 Conclusions

Predication of cancer trends can help to evaluate the progress in cancer control while reinforcing prevention activities. In this study, the change law of cancer incidence is analyzed by the mathematical method, and one new intelligent method to predict the cancer incidence trends based on new ENN is proposed. Finally, using six long term monitoring of cancer trends data from Canada and China, this new intelligent method was verified. The predication results show that, the new intelligent algorithm can be applied to predict the real cancer incidence trends very well.

References

- [1] Chen Z., The Third National Retrospective Sampling Survey on Causes of Death [M], Beijing: Peking Union Medical College Press, 2008: 5-30.
- [2] Goss P.E., Strasser-Weippl K., Lee-Bychkovsky B.L., Fan L., Li J.J., Chavarri-Guerra Y., et al, Challenges to effective cancer control in China, India, and Russia [J], *Lancet Oncol.*, 2014, 15: 489-538.
- [3] Canadian Cancer Society's Steering Committee on Cancer Statistics. Canadian Cancer Statistics 2011. Toronto (ON): Canadian Cancer Society, 2011: 25-54.
- [4] Zeng H.M., Zheng R.S., Guo Y.M., Zhang S.W., Zou X.N., Wang N., et al, Cancer survival in China, 2003–2005: A population-based study [J], *Int. J. Cancer*, 2015, 136(8): 1921-1930.
- [5] Kachuri L., De P., Ellison L.F., Semenciw R., Cancer incidence, mortality and survival trends in Canada, 1970–2007 [J], *Chronic Dis. Inj. Can.*, 2013, 33: 69-80.
- [6] Luo F.L., Unbehauen R., *Applied Neural Networks for Signal Processing* [M], New York: Cambridge University Press, 1997: 121-130.

- [7] Gao W., Modeling Methods Analysis for Non-linear Time Series [J], J. Tsinghua Univer., 2000, 37: 6-10.
- [8] Beltrami E.D., Mathematics for Dynamic Modeling [M], New York: Academic Press, 1998: 115-139.
- [9] Jacek Z.M., Introduction to Artificial Neural Systems [M], St. Paul: West Publishing Company, 1992: 57-60.
- [10] Jin F., Neural Computing Intelligence Foundation-Principles and Method [M], Chengdu: Xi'nan Jiaotong University Press, 2000: 158-165.
- [11] Gao W., New Neural Network Based on Ant Colony Algorithm for Financial Data Forecasting [C], New York: IEEE Press, 2008: 1437-1440.
- [12] Hatamlou A., Black hole: a new heuristic optimization approach for data clustering [J], Inform. Sciences, 2013, 222: 175-184.
- [13] Gao W., Ge M.M., Chen D.L., Wang X., Back Analysis for Rock Model Surrounding Underground Roadways in Coal Mine Based on Black Hole Algorithm [J], Eng. Comput., 2016, 32(4): 675-689.
- [14] Chen W.Q., Zheng R.S., Baade P.D., Zhang S.W., Zeng H.W., Bray F., et al, Cancer Statistics in China, 2015 [J], CA: Cancer J. Clin. 2016, 66(2): 115-132.