Applied Mathematics and Computation 341 (2019) 204–214

Contents lists available at ScienceDirect

Applied Mathematics and Computation

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

Determination of endometrial carcinoma with gene expression based on optimized Elman neural network

Hongping Hu[a,*], Haiyan Wang[b], Yanping Bai[a], Maoxing Liu[a]

[a] School of Science, North University of China, Taiyuan, Shanxi 030051, PR China
[b] School of Mathematical and Natural Sciences, Arizona State University, USA

ARTICLE INFO

Keywords:
Elman neural network
Endometrial carcinoma
Gene expression
Grey wolf optimizer
Leave-one-out cross validation
Particular swarm optimization

ABSTRACT

Endometrial carcinoma is a life-threatening disease that causes serious damage to the women's health. This paper discusses classifications of 87 endometrial samples with gene expressions that are cancerous or cancer-free. Every sample has 5 indicators. For every indicator, the corresponding genes of the missing data are deleted and the signal noise ratios (SNRs) are calculated to filter the irrelevant genes. Then the obtained new samples use the principle component analysis to decrease the dimensions. Finally 10 random samples are selected to be the testing samples for classification. Thus the classification accuracy rate is given for every indicator. Based on cancer related to 5 indicators, the combination of the 5 indicators is used to classify to make new 87 endometrial samples as cancerous or cancer-free. We repeatedly process these new samples by deleting the missing data, filtering the irrelevant genes with SNRs, and decreasing the dimensions with PCA, an obtain the new data. The proposed method is that the particle swarm algorithm (PSO) and the grey wolf optimizer (GWO) is combined to optimize the parameters of Elman recurrent neural network (ERNN), written as PSOGWO-ERNN. The results show that PSOGWO-ERNN is superior to the single ERNN, ERNN optimized by PSO or GWO (PSO-ERNN or GWO-ERNN), and the classification accuracy rate of PSOGWO-ERNN reaches 88.8506%. The results also show that the neural networks optimized by some swarm intelligence algorithms are more useful for classification.

1. Introduction

In recent years, many researchers have based on the classifications and the predictions of the tumors based on gene expression which is convenient for precise diagnosis and classification of cancer. The researched data for gene expression characteristically include small samples, high dimensions, large noises and high redundancy. Therefore, it is essential to mine biological knowledge from gene expression and to select the genetic information in tumors for classifications and predictions.

Gene expression was firstly put forward in 1999 for cancer classification on acute myeloid leukemia (AML) and acute lymphoblastic leukemia (ALL), independent of previous biological knowledge [1]. Since then, the cancer classification based on gene expression has increasingly drawn the attention of researchers [2–4].

Many methods have been applied for classifications and predictions, such as support vector machines [5,6], independent component analysis [7], improvised interval value-based particle swarm optimization [8], interval valued classification [9],

* Corresponding author.
E-mail addresses: hhp92@163.com, huhongping@nuc.edu.cn (H. Hu).

https://doi.org/10.1016/j.amc.2018.09.005
0096-3003/

k-nearest neighbor [10], and classified neighbors algorithm [11]. Owing to the high dimensions and small samples of the researched data, gene selection can be viewed as a key stage for cancer classifications and predictions, such as ensemble gene selection by grouping [12], support vector data description [13], mutual information [6], interval valued classification [9].

Neural networks have also been applied for cancer classification and predication. For example, in [14], a support vector machine,a back propagation neural network and an S-Kohonen neural network were used to classify colon cancer patients with UICC II into 2 groups: relapse and no relapse, and the results showed that neural networks could be used for classification.

Endometrial carcinoma is the most common gynecologic malignancy. A thorough understanding of the epidemiology, pathophysiology, and management strategies for this cancer allows the obstetrician or gynecologist to identify women at increased risk, contributes to risk reduction, and facilitates early diagnosis [15]. Quantify cumulative incidence of, and risk factors for developing lymphedema following treatment for endometrial cancer and estimate absolute risk for individuals [16]. The disease's pathogenesis is closely related to the lifestyle and regional differences exist. The relative risks (RRs) of endometrial cancer associated with oral contraceptive use were estimated using logistic regression, stratified by study, age, parity, body-mass index, smoking, and use of menopausal hormone therapy [17]. Therefore, all these factors are must be considered in assessing whether a women suffers from endometrial cancer.

In this paper, classifications of 87 endometrial samples with 5 indicators based on gene expression that is cancerous or cancer-free are discussed. For every indicator,we delete the missing data, use the signal noise ratios(SNRs) calculated to filter the irrelevant genes, and apply principle component analysis to decrease the dimensions, and then 10 random samples are selected to be the testing samples for classification. The combination of 5 indicators are considered to classify 87 endometrial samples as cancerous or cancer-free. We use leave–one–out cross–validation (LOOCV) methods for classification. The results show that the proposed method, which is combined with the particle swarm algorithm (PSO) and the grey wolf optimizer (GWO) to optimize the parameters of an Elman recurrent neural network (ERNN), PSOGWO-ERNN, is superior to the single ERNN, ERNN optimized by PSO or GWO (PSO-ERNN or GWO-ERNN).

The structure of this paper is as follows. Section 2 is the basic theory of ERNN, PSO and GWO. Section 3 introduces the proposed method, PSOGWO. The experimental results and conclusions are shown in Section 4 and Section 5, respectively. Section 6 is the discussion in which the shortcomings and improvement are proposed.

2. Basic theory

2.1. Elman recurrent neural network

In 1990, Elman [18] proposed a simple recurrent neural network, now called the Elman recurrent neural network(ERNN), for solving the speech processing. After so many years of research and development, ERNN has time series and nonlinear prediction capabilities, faster convergence, and more accurate mapping ability. In ERNN, the outputs of the hidden layer are allowed to feed back onto themselves through a buffer layer, called the recurrent layer. This feedback allows ERNN to learn, recognize, and generate temporal patterns, as well as spatial patterns. Every hidden neuron is connected to only 1 recurrent layer neuron through a constant weight of value 1. Hence the recurrent layer virtually constitutes a copy of the state of the hidden layer 1 an instant before. The number of recurrent neurons is consequently the same as the number of hidden neurons. To sum up, the ERNN is composed of an input layer, a recurrent layer which provides state information, a hidden layer, and an output layer. Each layer contains 1 or more neurons which propagate information from 1 layer to another by computing a nonlinear function of their weighted sum of inputs.

The topological structure of ERNN is shown in Fig. 1, where the number of neurons in inputs layer is m, the number in the hidden layer is n, and the number in the output is r. Let k denote the kth iteration; $x_i^0(k)(i = 1, 2, \ldots, m)$ denote the ith input value in the input layer; $s_i^1(k)$ and x_i^1 denote the ith input value and ith output value in the hidden layer,respectively; s_i^2 and $c_i(k)$ denote the ith input value and the ith output value in the recurrent layer; s_i^3 and $y_i(k)$ denote the ith input value and the ith output value in the output layer,respectively; f_1 and f_2 denote the activate functions in the hidden layer and in the output layer,respectively; w_{ij}^0, w_{ij}^2 and w_{ij}^1 denote the connected weights in the hidden layer, in the recurrent layer, and in the output layer.

According to Fig. 1, the formula of ERNN is as follows.

In the input layer,

$$x_i^0(k) = x_i(k)(i = 1, 2, \ldots, m). \tag{2.1}$$

In the hidden layer,

$$\begin{cases} s_i^1(k) = \sum_{j=1}^{m} w_{ij}^0 x_j^0(k) + \sum_{j=1}^{m} w_{ij}^2 c_j(k), \\ x_i^1(k) = f_1(s_i^1(k) + b_i), \end{cases} \quad (i = 1, 2, \ldots, n) \tag{2.2}$$

In the recurrent layer,

$$\begin{cases} s_i^2(k) = x_i^1(k-1), \\ c_i(k) = s_i^2(k), \end{cases} \quad (i = 1, 2, \ldots, n) \tag{2.3}$$