Applied Soft Computing 13 (2013) 3314–3323

Contents lists available at SciVerse ScienceDirect

Applied Soft Computing

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

Nonlinear speech coding model based on genetic programming

Xiaojun Wu [a,b,c,*], Zhanzhong Yang [c]

[a] Key Laboratory of Modern Teaching Technology (Shaanxi Normal University), Ministry of Education, Xi'an, Shaanxi 710062, China
[b] School of Computer Science, Shaanxi Normal University, Xi'an, Shaanxi 710062, China
[c] College of Automation of Northwestern University, Xi'an, Shaanxi 710072, China

ARTICLE INFO

Article history:
Received 16 August 2012
Received in revised form
21 November 2012
Accepted 6 February 2013
Available online 14 March 2013

Keywords:
Genetic programming
Nonlinear modeling
Speech coding

ABSTRACT

An improved genetic programming is proposed in this paper to construct the nonlinear models of speech signals, and the speech coding is further accomplished. After the preprocessing of the speech signals, the improved GP is used to construct the corresponding model of each speech frame. Then by analyzing these models, a normalized model that has generalization ability is obtained. And finally the process of speech coding is accomplished by the optimizing the parameters of the normalized model using an optimization algorithm. Experiments demonstrate that the feasibility of the improved GP in the modeling of speech signals, and show the superiority of the proposed method in speech coding based on the comparisons with the linear predictive coding.

1. Introduction

Speech coding is to transform the analog signals of speech into digital signals following some certain rules. The basic methods of speech coding are waveform coding [1], parametric coding [2] and hybrid coding [3]. Waveform coding makes effort to keep consistent with the original waveform, which has strong adaptability and high speech quality. However, the needed transmission bit rate is higher. Based on the analysis of the generation mechanism of speech, the parametric coding is accomplished by constructing the model of the generation mechanism under the principle that the decoded speech signals can be well understood. This method does not need to match the original waveform, which makes it have a lower transmission bit rate, but is more sensitive to the environmental noise and the synthesized speech has poor quality relatively. Formant coding and linear predictive coding are the typical approaches of parametric coding, of which the 'linear prediction' is a commonly used technology in speech processing, which has been successfully used in the applications of speech recognition [4], speech coding [5], etc.

Deep researches show that the speech signals are time series, which are time-varying and contain lots of nonlinear characteristics [6], and the linear prediction cannot meet the demand of modern speech processing. With the development of nonlinear theories, a few approaches, like neural network, have been widely used in speech processing, and the nonlinear research has become a hotpot in the domain of speech processing [7]. In this paper, the nonlinear models of speech signals are constructed based on the genetic programming.

Genetic programming (GP) [8] is a special optimization algorithm developed from genetic algorithm (GA). The hierarchical structure is used in GP and the solutions of different problems are boiled down to the corresponding computer programs with some given constraints. GP can accomplish the collateral optimization of the structure and the parameters of the model, which makes it extensively used in the modeling of nonlinear systems [9], data analysis [10], etc.

In this paper, an improved genetic programming is proposed to construct the nonlinear speech models based on the nonlinear characteristics of speech signals. By analyzing these models, a normalized model that has generalization ability is obtained. And finally, the speech coding is accomplished by optimizing the parameters of the normalized model using an optimization algorithm. The second part introduces some related works; the third part gives a general description of the proposed speech processing method; in part 4 and 5, the improved GP is proposed and the implementation of the speech coding is described particularly; Experiments is done in part 6 to demonstrate the method proposed in this paper.

2. Related works

2.1. Linear predictive coding

Linear predictive coding (LPC) is based on the assumption of all-pole model of the speech signals, whose parameters are estimated under the principle of the least-square error in the time domain.

* Corresponding author at: School of Computer Science, Shaanxi Normal University, Xi'an, Shaanxi 710062, China.
E-mail address: xjwu@snnu.edu.cn (X. Wu).

1568-4946/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.asoc.2013.02.008

LPC can preferably describe the spectrum of the speech and the characteristics of the vocal tract, and also can reduce the *kbps* of speech coding. The structure of LPC model is as follows,

$$s(n) = \sum_{i=1}^{p} a_i s(n-i) + Gu(n) \tag{1}$$

where G is the gain; $u(n)$ is the excitation, which is the unit pulse sequence when $s(n)$ is deterministic, otherwise is the white noise. $\hat{s}(n) = \sum_{i=1}^{p} a_i s(n-i)$ is called the 'linear predictor', which means $\hat{s}(n)$, the predicted value of $s(n)$, is the linear combination of P past values that are adjacent to $s(n)$. a_i is the linear prediction coefficient, and the difference between $s(n)$ and $\hat{s}(n)$ is called 'linear prediction error'.

By calculating the coefficients $|a_i|$, LPC is to minimize the linear prediction error under certain principle, such as the minimum mean square error (MMSE).

2.2. Nonlinearities in speech

The development of nonlinear theories, especially the technology of neural network, provides many effective solutions to the nonlinear systems. Neural network has gradually become the major approach to the nonlinear processing of speech signals, and has been applied in speech recognition [11], speech transmission[12], etc.

1994, Thyssen et al. [13] divided the speech signals into linear part and nonlinear part. The LPC was used to construct the model of the linear part. After certificating the nonlinearities in the residual signal, the volterra filter and the time delay neural network were used to construct the nonlinear models. And finally, the nonlinear predictor was proposed by combining the models of the two parts, which improved the coding precision. Matthew et al. [14] proposed a speech recognition system for local languages using the artificial neural network. The Mel-Frequency Cepstral Coefficients were extracted firstly to train the neural network, and in this process, a scaled conjugate gradient back-propagation algorithm was used to improve the rate of the convergence. The experimental Hausa language was well recognized by the proposed system. Nithin and Ganapati [15] proposed a novel low complexity nonlinear active noise control system. The nonlinear controller was composed of the adaptive Legendre neural network (LeNN), which was updated using a filtered-1 least mean square algorithm. And the computational complexity of the proposed scheme was further reduced by incorporating the partial update adaptive algorithm. Simulation study demonstrated that the new ANC had more strong ability to eliminate the noise compared with the traditional ANC schemes.

The technology of neural network, which can offer the structured models, has successfully been used in speech modeling. However, the structures of the models are alterable because of the weights in them. Thus, the neural network cannot give the fixed structured model like LPC does, which, to some extent, makes it go against the analysis of the speech signals.

2.3. Genetic programming

Compared with other algorithms while solving the nonlinear problems, GP could create nonlinear models which have the minimum error relatively. Besides, prior knowledge is not needed. And in the process of evolution, the structures of individuals can change proactively through the crossover and the mutation. The fitness values of new individuals are calculated by the fitness function, and the optimal individual is saved which makes the evolution goes along the established goal. Since the problem to be solved can be preferably described by the explicit structure of the model, GP has been used in lots of nonlinear modeling applications, especially in the prediction of time series.

Wagner et al. [16] divided the forecasting environment into the static environment and the dynamic environment. Since many studies, which only considered about the former, were unsuitable for the real-world time series, a Dynamic GP Forecasting (DyFor) model for nonstatic environment was proposed. A sliding window, whose size was natural adaptive, was brought into the process of evolution. Lee and Tong [17] separated the model of time series into the linear part and the nonlinear part. The autoregressive integrated moving average (ARIMA) was utilized to construct the linear model, while the GP was used to construct the nonlinear model on the basis of the residual signals. And then a hybrid forecasting model was proposed by combining these two models. Experiments certificated that the hybrid model was more outstanding than other forecasting models. Estévez et al. [18] proposed a voice activity detection (VAD) algorithm using GP. Logical operators, like AND, OR, NOT, GR (greater than), were introduced in the function set, while the terminators defined in the ITU-T G.729B VAD standard and a float constant, R, were brought into the terminal set. A process was finally created to judge that the given data was whether a speech or not. By comparing the experimental results, the error rate of the proposed approach is lower than the ITU-T G.729B VAD standard.

One important approach for speech coding is to construct a better model. GP is preferably used in nonlinear modeling, and the characteristics of the solving problems can be analyzed conveniently by the explicit structure of the model. So the basic thought of this paper is to construct nonlinear speech models using an improved GP algorithm and to accomplish the process of the speech coding further.

3. Proposed speech coding method

Discrete speech signals are nonlinear time series, and the samples are correlated with their neighbors. The traditional LPC model also indicates this phenomenon. Actually, the analysis of the speech signals shows that the largest correlation value exists between the adjacent samples. When the sample has a sampling rate of 8 kHz, the correlation value of the adjacent samples is larger than 0.85. Even there are 10 samples apart from one to another; the correlation value between them also has a value of 0.3. The higher sampling rate is, the larger correlation values are [19]. Since the existence of the correlation values increase the redundancy of the speech signals. One solution to compress the speech signals is to construct a model to make speech signals get rid of the correlation.

Because of the successful applications in nonlinear time series and speech processing, the GP algorithm is utilized in this paper to construct the nonlinear models of the speech signals. The short-term stability indicates that, despite the speech signals are time-varying, the changes of their characteristics are limited in a short time (10–30 ms). Thus, 'shot-term analysis' is the foundation for speech processing, which means to separate speech signals into frames before being processed.

In the evolutionary process of GP, the crossover and mutation operators are executed randomly, and a series of 'best' models are constructed finally. To find a normalized model that has generalization ability is the key in the modeling research of speech coding. And in this paper, the similarity of the 'best' models will been analyzed. By integrating different models, a normalized model with parameters is proposed. And the speech coding is implemented by the optimization of the parameters.The process of speech coding in this paper is shown in Fig. 1, which contains 5 aspects,