Optimal Structure of Pipelined Recurrent Neural Network in Modeling of MPEG Video

J. Ahmadi Shokouh1, H. Keshavarz2
1-Department of Electronic, University of Sistan and Baluchistan
Email: shokouh@hamoon.usb.ac.ir
2-Department of Electronic University of Sistan and Baluchistan
Email: keshavarz@hamoon.usb.ac.ir

Abstract—This paper investigates optimal structure of Pipelined Recurrent Neural Network (PRNN) for adaptive traffic prediction of MPEG video signal via dynamic ATM networks. The traffic signal of each picture type (I, P, and B) of MPEG video is characterized by a nonlinear autoregressive moving average (NARMA) process. Since those modules of PRNN can be performed simultaneously in a pipelined parallelism fashion, this would lead to a significant improvement in the total computational efficiency of PRNN. In order to further improve the convergence performance of the adaptive algorithm for PRNN, a learning-rate annealing schedule is proposed to accelerate the adaptive learning process. The measure that is used for optimum structure of PRNN i.e. number of neurons in each module, number of modules is the one-step forward prediction gain. Results are shown to be promising and practically feasible in obtaining the best structure for a adaptive predictor of Noise.

Keywords — MPEG, Video modeling, NARMA, PRNN, RTRL algorithm, One-step forward prediction gain

1. INTRODUCTION
In B-ISDN, a major part of the traffic will be produced by multimedia sources like teleconferencing terminals and video-on-demand servers, because all the video applications are high-bandwidth-required, even the huge link capacity provided by optical fibers can be quickly saturated by few video streams, in order to transmit them efficiently, the compression process is necessary. There are many video coding algorithm such as MPEG, QuickTime, Indeo, and so on. But MPEG is the most important, because MPEG is a factual international standard and almost all the video terminals support it. There are several phases of MPEG, they are MPEG-I, MPEG-II etc. MPEG-II is a compatible extension of MPEG-I, it supports interlaced video formats and a number of other advanced features, but the main difference which respect to video transmission on ATM is that MPEG-II allows for scalability, this is useful in application areas such as video communication, video on ATM, HDTV with embedded TV etc. Scalability enables different decoders to construct different versions of the same video source by using subset of total encoded bit stream. It includes spatial scalability, temporal scalability, SNR scalability, and datapartitioning. Using this technology, the video data stream can include a base layer stream, which contains the most important video data, and one or more enhancement layer, which can be used to improve the quality of the video sequence. In this paper, we focus on data streams of MPEG-I type. Most of encoders will use this scheme and in case of MPEG-II encoded stream, the statistical properties of the most important information will be almost identical to this type of stream.

On ATM networks, MPEG streams can be transmitted in two ways, CBR or VBR, but only the latter can fully take the advantage of statistical multiplexing, the fundamental character of ATM. So, it is expected that VBR video service will make up a significant portion of the traffic on ATM networks. For the VBR traffic, properly traffic control mechanism is very important to maintain the required QOS, at the same time, modeling of VBR video traffic is the basis of VBR traffic control mechanism, it has gained the interest of so many ATM researchers an a
large variety of papers about VBR model can be found in the current teletraffic literature. These modeling approaches can be divided into different classes such as Markov chains, Autoregressive processes, Self-similar models and so on. Moreover, an alternative method based on neural networks has been proposed to further improve the prediction accuracy of the traffic statistics of the VBR video. Several examples of using neural network approaches for on-line traffic prediction of VBR video signals can be found in the literature [17], [18]. Neves et al. [17], [18] applied the multilayered perceptron neural networks with backpropagation training to the traffic prediction in the flow control and traffic enforcement mechanism for ATM networks, respectively. Recently, Chong et al. used the high-order pi–sigma neural network for the bandwidth prediction of VBR video over an ATM network. All of them have shown that satisfactory traffic prediction accuracy can be achieved by those neural networks. However, those neural networks suffer from drawbacks of slow convergence and unpredictable solutions during learning. To overcome this difficulty, an alternative architecture to the traffic prediction of MPEG video with the flexibility to adapt to a changing ATM network environment is based on recurrent neural networks (RNN’s). An RNN is well suited for the adaptive prediction of a nonstationary time series [1]. Several algorithms have been proposed for the training of the RNN’s. The most widely known algorithm is the real-time recurrent learning (RTRL) algorithm, proposed by Williams and Zipser [2], that can be used to update the synaptic weights of the RNN in real time. According to the theory of prediction [3], the minimum mean-squared error noise predictor is the conditional mean which can be expressed in terms of the functional expression of the NARMA process. However, the explicit functional expression of the NARMA model is actually unknown. Conner et al. [4] have introduced a recurrent neural network (RNN) implementation to approximate the NARMA-based conditional mean predictor, and have shown the superior accuracy of its prediction. In Section II, we will show the source characteristic for MPEG video signals. In Section III, a pipelined recurrent neural network (PRNN), proposed by Haykin and Li [5], is introduced to implement the NARMA-based optimal traffic predictor with low complexity. Furthermore, in section IV, we would like to apply the learning-rate annealing schedules [6] to the PRNN in order to improve its convergence performance further. Finally, in Section V, we present an experimental study of the PRNN applied to the prediction of MPEG video data stream [7] using of real data for optimal structure of this predictor. Of course we optimized forgetting factor in [8].

2. SOURCE CHARACTERISTICS AND TRAFFIC PREDICTION MODEL OF MPEG VIDEO SIGNALS

The MPEG video international standard specifies the coded representation of the video data. The video source coding is based on motion-compensated hybrid DCT coding which employs two basic techniques: motion compensation for the reduction of temporal redundancy and DCT transform compression for the education of spatial redundancy. In order to achieve the highly efficient compression and to meet the conflicting requirements of random access, the input video is divided into units of group-of-pictures (GOP’s) consisting of an intra (I) picture, coded without reference to other pictures an arrangement of predictive (P) pictures, coded with reference to previous (I or P) pictures, and bidirectionally predictive (B) pictures, coded with reference to an immediate previous (I or P) picture, as well as an immediate future (P or I) picture. The I picture at the beginning of a GOP serves as a basic entry point to facilitate random seek or channel switching, and also provides coding robustness to transmission error, but is coded with only moderate compression to reduce the spatial redundancies. P pictures are coded more efficiently using motion-compensated prediction from a past I or P picture, and are generally used as a reference for further prediction.

B pictures provide the highest degree of compression, but require both past and future reference pictures for motion compensation. It should be mentioned that B pictures are never used as references for prediction. A GOP is defined by its length N(GOP) which is the distance between I pictures. An example of a GOP in MPEG is presented in Fig. 1, where I, P, and B denote picture encoding in the intrapicture mode, predicted mode, and bidirectionally predicted mode, respectively.

![Figure 1. Example of MPEG GOP.](image-url)
The macroblocks of 16x16 pixels are the basic coding units for the MPEG algorithm. Each macroblock is divided into four blocks, where each block contains 8x8 pixels. The main extension from monochrome video to color is the addition of two 8x8 chrominance blocks to the macroblock. A row of macroblocks that makes up a horizontal strip in the image is called a slice, and a number of slices are combined to form a picture. The coding mode of each macroblock within a specific picture depends on its picture type. For I pictures, a discrete cosine transform (DCT) is performed on each block. The resulting two-dimensional block of DCT coefficients is quantized and scanned in a zig-zag order to convert it into a one-dimensional string of quantized DCT coefficients. Runlength coding is used for the quantized coefficient data. The predicted pictures (P and B) use motion-compensated prediction of the contents of the macroblock based on past or future reference pictures. This prediction is subtracted from the actual data in the current macroblock to form an error signal. The prediction error is coded like the intracoded macroblocks. None of the analytical models available today can adequately represent VBR MPEG video traffic. Here, we choose a 10000 frames film equivalents to approximately 10 minutes. It is coded by an MPEG compression technique called PVRG. In this research, we use GOP by case: IBBPBPPBBPBB.

3. PIPELINED RECURRENT NEURAL NETWORKS (PRNN)

The PRNN shown in Fig. 2 is composed of \( q \) identical modules, each of which is designed as a fully connected recurrent network with \( N \) neurons.

![Figure 2. Structure of PRNN with \( q \) modules](image)

Fig. 3 shows the detailed structure of module \( i \) with \( N \) neurons and \( p \) external inputs.

![Figure 3. Module \( i \) with \( N \) neurons](image)

The updated value of the synaptic weight matrix \( W \) is computed using the RTRL algorithm [2]. Haykin and Li [5] have demonstrated that the PRNN is able to provide satisfactory accuracy of the nonlinear adaptive prediction of a non-stationary signal and time series process. An important feature of the PRNN is its high computational efficiency.
Specifically, the total computational requirement of processing a single sample on a PRNN is $O(qN^4)$ arithmetic operations. Prediction error for module is given by:

\[ \hat{e}_i(n) = s_i(n+1) - \hat{s}_i(n+1) \]  

(1)

Thus, an overall cost function for the PRNN is defined by:

\[ \varepsilon(n) = \sum_{i=1}^{q} \lambda^{i-1} e_i^2(n) \]  

(2)

Where $\lambda$ is an exponential forgetting factor that lies in the range of with the RTRL algorithm.

4. Real Time Recurrent Learning Algorithm (RTRL)

For the case of a particular weight $w_{kl}$ its incremental change $\Delta w_{kl}$ made at time $n$ according to the method of steepest descent is given by:

\[ \Delta w_{kl}(n) = -\eta \frac{\partial \varepsilon(n)}{\partial w_{kl}} \]  

(3)

Where $\eta$ is the learning-rate parameter.

The most well-known annealing schedule is the search-then-converge schedule, defined by:

\[ \eta(n) = \frac{\eta_0}{1 + n/\tau} \]  

(4)

Where $\tau$ denotes the search time constant.

5. AN OPTIMAL STRUCTURE FOR PRNN ON MPEG VIDEO

In this section we present the results of applying the statistical method described in section III to several MPEG-II encoded video sequences. Each video consists of 10000 frames-equivalents to approximately 10 minutes. The frame size traces were extracted from MPEG-2 sequences, after begins encoded with an adapted version of Stanford PVRG-MPEG-2 encoder (Avi2MPG), that have been produced by Brent Beyeler [25], this software can produce statistic file include number of bytes in each frame after encoding. This number is used for output time series of MPEG video source. All of samples are NTSC standard in 30 frame/s A standard MPEG encoder tries to reduce the spatial and temporal redundancy using several modes of compression resulting in the generation of three different types of compressed frames: I-frames (I), P-frames (P) and B-frames (B). Typically I-frames are larger than P-frames, while B-frames have the lowest size. The MPEG coding technique, arrange the compressed frames using a deterministic sequence called Group of Pictures (GOP). The pattern used to encode the MPEG sequences used in this work follows an “IBBPBBPBBPBB” GOP pattern. Three categories of videos constitute this sample, respectively:

1. news/talk (low motion)
2. movie (medium motion)
3. sport-soccer (full motion)

In figures 4-6 you see GOP size stream for each one of video stream.

![Figure 4. news/talk (low motion)](image-url)
And for checking performance of this predictor in optimum conditions, we used one-step forward prediction gain defined:

\[ R_p = 10 \log_{10} \left( \frac{\hat{\sigma}_s^2}{\hat{\sigma}_e^2} \right) \, dB \quad (5) \]

Where \( \hat{\sigma}_s^2 \) denotes the estimated variance of the video stream data, and \( \hat{\sigma}_e^2 \) denotes the estimated variance of the error signal. The most important parameters in PRNN structure are number of neurons in each module and number of modules. For obtaining optimum structure for PRNN we computed \( R_p \) versus above parameters. Simulation results have been shown in Fig. 7-12.
Figure 7. $R_p$ versus number of modules
Low Motion (Talk/News)

Figure 8. $R_p$ versus number of modules
Medium Motion (Film)

Figure 9. $R_p$ versus number of modules
Full Motion (Soccer)

Figure 10. $R_p$ versus number of neurons
in each module, Low Motion (Talk/News)

Figure 11. $R_p$ versus number of neurons
in each module, Medium Motion (Film)

Figure 12. $R_p$ versus number of neurons
in each module, Full Motion (Soccer)
6. Conclusion

This paper has presented a new optimal prediction model based on a PRNN which is capable of predicting the MPEG video signal by using a modification of the RTRL algorithm. The modified learning algorithm with learning-rate annealing schedules is well suited for the PRNN to learn highly dynamic situations such as the status of video stream. In this paper we have shown a technique for optimal design of PRNN by one-step forward prediction gain. Simulation results have shown that the optimum values for number of neurons in each module and number of modules, i.e. essential parameters of PRNN. This verifies the effectiveness of the best approximation capability of the PRNN as a predictor for modeling of MPEG video.

7. References

