B

Biomedical Signal Compression

Pedro de A. Berger
University of Brasilia, Brazil

Francisco A. de O. Nascimento
University of Brasilia, Brazil

Leonardo R. A. X. de Menezes
University of Brasilia, Brazil

Adson F. da Rocha
University of Brasilia, Brazil

Joao L. A. Carvalho
University of Southern California, USA

INTRODUCTION

Digitization of biomedical signals has been used in several areas. Some of these include ambulatory monitoring, phone line transmission, database storage, and several other applications in health and biomedical engineering. These applications have helped in diagnostics, patient care, and remote treatment. One example is the digital transmission of ECG signals, from the patient's house or ambulance to the hospital. This has been proven useful in cardiac diagnoses.

Biomedical signals need to be digitally stored or transmitted with a large number of samples per second, and with a great number of bits per sample, in order to assure the required fidelity of the waveform for visual inspection. Therefore, the use of signal compression techniques is fundamental for cost reduction and technical feasibility of storage and transmission of biomedical signals.

The purpose of any signal compression technique is the reduction of the amount of bits used to represent a signal. This must be accomplished while preserving the morphological characteristics of the waveform. In theory, signal compression is the process where the redundant information contained in the signal is detected and eliminated. Shannon (1948) defined redundancy as "the fraction of unnecessary information, and therefore repetitive in the sense that if it was missing, then the information would still be essentially complete, or it could at least be recovered."

Signal compression has been widely studied during the past decades, and several references discuss this subject (Gersho & Gray, 1992; Jayant & Noll, 1982; Sayood, 1996).

Signal compression techniques are commonly classified in two categories: lossless and lossy compression. Lossless compression means that the decoded signal is identical to the original one. In lossy compression, a controlled amount of distortion is allowed. Lossy signal compression techniques show higher compression gains than lossless ones.

BACKGROUND

Lossless Compression

Lossless signal compression techniques are less efficient with respect to compression gains. They may be used in combination with lossy compression techniques, especially in cases where the maximum allowed distortion has been reached, and additional compression is needed. Among several lossless compression techniques, we highlight entropy coding (Gersho & Gray, 1992), Run-Length, Huffman (Huffman, 1952), arithmetic coding (Witten, 1987), and delta coding (Ken, 1985).

Run-Length Coding

Data files frequently present sequentially repeated characters (a character run). For instance: text files use several spaces to separate sentences and paragraphs. Digital signals may contain the same value, or the same character representing that value in its data file, repeated many times sequentially. This indicates that the signal is not changing, as in the isoelectric segments of ECG signals, for example.

Figure 1 shows an example of Run-Length coding of a data set that contains runs of zeros. Each time the coder finds a zero in the entry data, two values are written in the output data. The first of these values is a zero indicating that the Run-Length codification started. The second value is the amount of zeros in the sequence. If the run of zeros in the input data set is in average larger than two, then the Run-Length coder will achieve data compression.

Huffman Coding

In Huffman coding, the data are represented as a set of variable length binary words. The lengths depend on the occurrence frequency of the symbols used for representing each signal value. Characters that are used often are represented with fewer bits, and those that are seldom used are represented with more bits.

Figure 2 shows an example of how a Huffman code is generated, given a data set X and its characteristic probability of occurrence—p(X). The character codes are generated by combining the bits of a tree with ramifications, adding their probabilities and restarting the process until only one character remains. This process generates a three with ramifications linked to bits 0 and 1. The codes for each character are taken in the inverse path of these ramifications. Note that initial character arrangement is not relevant. In this example, we chose to encode the upper ramifications with bit 0 and the lower ones with bit 1. However, the opposite representation could have been used as well. Any decision criteria may be used in ramifications with equal probabilities.

Huffman coding has the disadvantage of assigning an integer number of bits to each symbol. This is a suboptimal strategy, because the optimal number of bits per symbol depends on the information content, and is generally a rational number. Arithmetic coding is a more sophisticated compression technique, based on Huffman coding concepts. It results in compression gains closer to the theoretical limits. In this coding, character sequences are represented by individual codes depending on their probability of occurrence. Huffman and Arithmetic codes are often used in combination with Run-Length coding, for further compressing biomedical signals that were first compressed using orthogonal transforms.

Delta Coding

Delta coding refers to signal compression techniques that store a digital signal as the difference between successive samples. Figure 3 shows an example of how this is performed. The first sample in the encoded signal is equal to the first sample of the original signal. Subsequent encoded samples are equal to the difference between the current sample and the previous sample of the original signal.

Using this technique, the encoded signal has a smaller amplitude dynamic range than the original signal. Therefore, it takes fewer bits to store or transmit the encoded signal. Delta coding is a particular case of Differential Pulse Code Modulation (DPCM). DPCM is used in combination with Huffman coders in several biomedical signal compression algorithms.

Figure 1. Run-length coding example