

Abstract

The relentless advancement in mobile communications and the Internet has contributed to a rapid increase in the amount of digital data created and replicated worldwide, which is estimated to double every three years. In this context, data compression algorithms, which allow for the reduction of the number of bits needed to represent digital data, have become increasingly relevant. In this work, we focus on the compression of multichannel signals with irregular sampling rates and with data gaps. We consider state-of-the-art algorithms, which were designed to compress gapless signals with regular sampling, adapt them to operate with signals with irregular sampling rates and data gaps, and then evaluate their performance experimentally, through the compression of signals obtained from real-world datasets. Both the original and our adapted algorithms work in a near-lossless fashion, guaranteeing a bounded per-sample absolute error between the decompressed and the original signals. This includes the important *lossless compression* case, which corresponds to an error bound of zero. The algorithms compress signals by exploiting correlation among signal samples taken at close times (temporal correlation), and, in some cases, among samples from various signals (spatial correlation). For most algorithms we design and implement two variants: a masking (M) variant, which first encodes the position of all the gaps, and then proceeds to encode the data values separately, and a non-masking (NM) variant, which encodes the gaps and the data values together. For each algorithm, we compare the compression performance of both variants: our experimental results suggest that variant M is more robust and performs better in general. Every implemented algorithm variant depends on a window size parameter, which defines the size of the blocks into which the data are partitioned for encoding. We analyze the sensitivity of variant M of each algorithm to this parameter: for each dataset, we compress each data file, and compare the results obtained when using a window size optimized for said specific file, against the results obtained when using a window size optimized for the whole dataset. Our experimental results indicate that the difference in compression performance is generally rather small. The last part of our experimental analysis consists in comparing the compression performance of our adapted algorithms, among each other, and with the general-purpose compression algorithm `gzip`. Following previous experimental results, we only consider variant M of each algorithm, and we always use the optimal window size for the whole dataset. Our experimental results reveal that none of the algorithm variants obtains the best compression performance in every scenario, which means that the optimal selection of a variant depends on the characteristics of the data to be compressed, and the error threshold that is allowed. In some cases, even a general-purpose compression algorithm such as `gzip` outperforms the specific algorithm variants. Nevertheless, we extract some general conclusions from our analysis: for large error thresholds, variant M of algorithm APCA achieves the best compression results, while variant M of algorithm PCA (and, in some cases of lossless compression, algorithm `gzip`) are preferred for lower threshold scenarios.

Keywords— multichannel signal compression, near-lossless compression, irregular sampling rate, data gaps.