Algorithm Engineering Group
Prof. Dr. Tobias Friedrich
Bachelor Project 2018/19

Lossy Compression of Time Series Data

Background. Time series data is data derived from consecutive measurements over time. It appears in all domains of applied science and engineering. Examples include industrial applications (status and parameters of machines), mathematical finance (interest and exchange rates), climate research (weather forecast and earthquake prediction) and digital health (glucose and insulin levels of diabetes patients). Algorithms for processing such data face the additional challenge of exploiting the linear structure of the data rather than treating the time component as a mere additional data value.

Challenge. Storing time series data can quickly overload any available storage. Even dedicated time series databases (e.g., InfluxDB) are challenged by storing high-frequency data that come into the system at frequencies higher than 10 kHz. It is therefore crucial to compress the data in a meaningful sense. A lossless data compression algorithm can only reduce the space requirement to Shannon's information-theoretic lower bound. The random noise introduced by physical measurements fundamentally limits the achievable compression rate. For an effective compression of high-frequency data, we are therefore forced to apply lossy compression schemes.

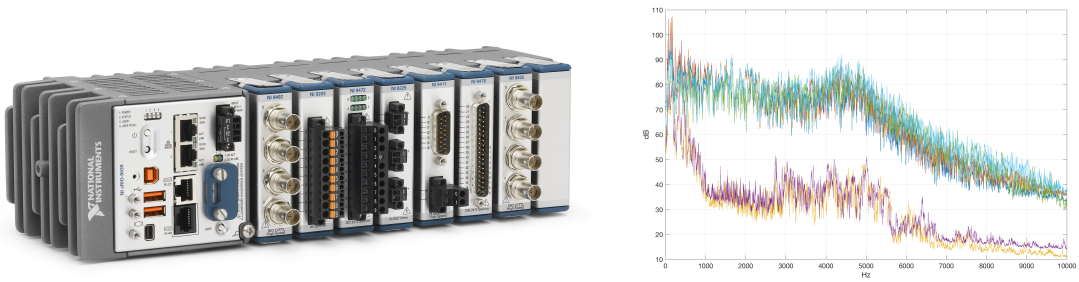

Figure 1: A process automation controller by National Instruments and a sample industrial time series.

Vision. We want to evaluate existing algorithms and develop new algorithms for lossy compression schemes for high-frequency time series data from different domains. The key question is what compression can be achieved without losing too much information while still being able to analyze the data and using it for modeling and statistical predictions. The answer will depend on the intended data analytics. The scientific method will be two-fold: (1) an empirical study of different methods on real-world data and (2) a thorough analysis of the theoretical limitations.

Industry partner contribution. Our project partner is a young IoT company based in Berlin with a scientific background in physics and mechanical engineering. They supply the team with high-dimensional industrial data sets spanning a period of several years. The data has been obtained from physical measurements of large industrial machines. For understanding the time series data, the partner company provides details of the particular industrial machines and the underlying thermodynamics. During the project, they will support the team with the required background in mechanical engineering and industrial data analytics.

Our contribution. The research group will work closely with you on the project and provide algorithmic and statistical expertise. We will teach some fundamentals of digital signal processing, like discrete Fourier transform, sampling, kernel and window functions. For studying the fundamental limitations of lossy time series compression, we will provide theoretical models for time series data. As in previous bachelor and master projects of our group, we plan to write a joint scientific publication about our research findings in the end.

Your contribution. In your work as a team, you will develop and customize a bundle of algorithms for handling time series data in an agile manner. For exploring industrial-size time series datasets, you will create the necessary code infrastructure and efficient algorithms for applying statistical models. You start from scratch and work with modern programming environments. Besides the team experience in an industrial context, you will learn about efficient data processing and the practical use of probability theory and statistics. You are not expected to become an expert in mechanical engineering or physical measurements.

Supervisors and project partners.

HPI

Prof. Dr. Tobias Friedrich
Dr. Timo Kötzing
Nikolay Dichev

INDUSTRIAL ANALYTICS

Dr. Richard Büssow
Dr. Eike Hylla
Markus Schildhauer