Mining sensor data from complex systems

Vespier, U.

English Summary

Today, virtually everything, from natural phenomena to complex artificial and physical systems, can be measured and the resulting information collected, stored and analyzed in order to gain new insight. The adoption and deployment of measurement systems for all sorts of industrial, commercial and consumer applications has, in fact, paved the way to important opportunities for analyzing complex systems at a level of detail never experienced before.

In this thesis, I have shown how complex systems often exhibit diverse behavior at different temporal scales, and that data mining methods should be able to cope with the multiple resolutions (scales) at the same time in order to fully understand the data at hand and extract useful information from it.

Under these assumptions, I have designed and evaluated data mining and visualization methods for large time series data collected from complex physical systems by means of sensors. In particular, I have developed solutions to three fundamental problems: the detection of multi-scale patterns, the recognition of recurrent events, and the interactive visualization of massive time series data.

The methods and algorithms I have introduced combine concepts from data mining, signal processing, and information theory. I have shown how to combine different techniques in order to deal with many of the challenges present when analyzing real-world time series data, such as the presence of noisy measurements, the occurrence of spurious and anomalous events and, ultimately, the risk of over-fitting the data with models that would be hardly

general.

Although the research questions addressed in this thesis have a general applicability, I evaluated the proposed solutions on a real-world scenario provided by InfraWatch, a Structural Health Monitoring project centered around the management and analysis of data collected by a large sensor network deployed on a Dutch highway bridge.

The application of the methods I developed permitted the identification of the relevant scales of analysis in the InfraWatch data (and other datasets also), the detection of the different recurring motifs and the visualization of terabytes of time series data interactively.