Structural health monitoring meets data mining

Miao, S.

Chapter 8

Conclusion

8.1 Conclusion

In this thesis, we have applied data mining techniques to the SHM domain, more precisely, to a highway bridge under normal in-service conditions. Long-term changes in the system can be analysed through a bridge's dynamic response, variables of which are known as modal parameters: natural frequencies, damping ratios and mode shapes. In reality, modal parameters are not only sensitive to structural damage and degradation, but also to varying operational and environmental conditions, such as traffic, sunshine, wind and most importantly, temperature. Understanding the performance of the bridge, and investigating the influences of operational and environmental variables, are two tasks we have discussed in this thesis.

We explore the nature of the bridge through measurements collected with a sensor network installed on the bridge, which consists of three types of sensor: strain, vibration and temperature sensors. The signals collected by strain sensors are sensitive to traffic loadings (including normal traffic events and traffic jams), as well as temperature changes and various types of noise. Individual vehicles are represented as peaks (with various durations and amplitudes), and traffic jams are represented as sharp baseline jumps, which last much longer than normal traffic peaks. Temperature changes cause a gradual baseline drift, which is a low-frequency effect. Finally, noise is any high-frequency component, which appears as small fluctuations in strain signals. The vibration signals are sensitive to noise and normal traffic events, but not sensitive to low-frequency components, such as environmental changes and traffic jams. Temperature is a low-frequency signal, which is just sensitive to local temperature changes, with some level of delay to environmental temperature, in the order of several hours.

In the end, all sensors are attached to the same bridge, and respond to dynamics occurring on the bridge, so there must be some kinds of dependencies among the various sensors. To look into the dependency between each sensor type pair, we employ datasets of different scales, and analyse them in both the time and the frequency domains. In the time domain,

- The dependency between strain and temperature sensors is strong at a large scale, but is weak at a small scale; temperature is not affected by traffic loadings on the bridge.
- The dependency between strain and vibration sensors is weak at both large and small scales, because the former are not only sensitive to traffic events, but also to temperature, and the latter are just sensitive to traffic events; what's more, the responses of these two types of sensors to traffic events are different.
- The dependency between vibration and temperature sensor types is fairly weak at both big and small scales.

In the frequency domain,

- The dependency between the strain and temperature sensor types is weak at a small scale. At a large scale, due to the daily and seasonal fluctuations, the low components of strain spectrum correlates with that of temperature spectrum.
- When the bridge is excited by traffic events, some spectral components in the strain and vibration spectra are highly correlated.

- the dependency between the vibration and temperature sensor types is the same as that in the time domain.

Having a big picture about dependencies among sensor types in both the time and the frequency domains at various scales, we analysed the dependency between the strain sensor type and the temperature sensor type in the time frequency at a big scale, and employed an exponential decay model to overcome delays caused by concrete properties. We analysed the dependency between the strain sensor type and the vibration sensor type in the frequency domain at a small scale, and applied bandpass filters to the spectra. We observed that the dependency between vibration and temperature sensors is fairly weak in both the time and the frequency domains at any scales, so we first conducted modal analysis on vibration signals, and then associate natural frequencies with temperature.

In the sensor network, there are 145 sensors of different properties (sensor type, location, orientation), and the level of correlation between different sensors appears to depend on these properties. To further look into the dependencies in detail, we employed Subgroup Discovery techniques to analyse correlations of all sensor pairs of different types, and obtained a number of interesting patterns (rules).

As mentioned above, temperature has a strong influence on strain signals. To separate the influence of temperature from other effects, we proposed a baseline correction method (the *most-crossing* method), which is based on the probability density function. Within a sliding window, we assume the baseline is a constant value, and divide data points into two categories: noise and peaks. The PDF of the noise category is different from that of peaks. We take the peak value of the former PDF as baseline, and model adjacent sliding windows with linear interpolation. The most-crossing method is not only capable of catching baseline drift in our strain signals, but also works well on other kinds of datasets.

In the entire thesis, traffic events play a crucial rule. We propose two supervised methods to identify traffic events: one of which is a classification method based on video labels, as described in Section 7.2.2. The other one is a predefined pattern detection method based on template, landmarks and constraints, as proposed in

Chapter 6. The former method is precise, but requires a lot of effort to manually label traffic events from video; the latter method is based on prior knowledge and the MDL principle, which is fast, and of satisfactory precision. In Chapter 7, we employed the first method to illustrate the procedure of data selection and modal parameter extraction. To look into the influence of environmental factors, we needed to process a huge amount of datasets, so the second method is chosen for traffic event identification for this purpose.

There are a number of modal analysis methods to extract modal parameters, such as the PP method and the SSI method. We employed the SSI method to extract natural frequencies, mode shapes and damping ratios, and verified that the natural frequencies obtained with the SSI method are correlated well with those obtained with the PP method. Because the PP method is simple and of acceptable accuracy, we employed it to extract natural frequencies from datasets selected from more than two years' measurements, and associated them with temperature. Generally speaking, natural frequencies decrease when temperature increases. We looked into the relationship between temperature and natural frequencies mode by mode, and found that high-frequency modes are more sensitive to temperature than low-frequency modes. We also analysed the influence of vehicle mass on natural frequencies, and found that natural frequencies are less sensitive to traffic mass.

8.2 Discussion

As mentioned in Chapter 1, this thesis focuses on Part 2 and Part 3 of the SHM process. In this section, we will discuss some topics related to Part 1: Operational Evaluation. The first topic we want to discuss is whether the sensor network is the best solution for structural health monitoring. To answer this question, we should compare it with some other solutions. In the literature, there is some research utilising radar systems, which are flexible and capable of catching the (absolute) displacements accurately. However, the radar systems are more suitable for short-term measurements than long-term tasks. In long-term

SHM, we are usually interested in more factors related to the bridge than just displacements. The sensor network composed of multiple sensor types is more informative, which is usually preferred for long-term measurements.

The second topic relates to the optimal number and distribution of sensors within a sensor network. In a sensor network, if we employ too few sensors, the measurements are not enough to catch the performance of the bridge; however, if we employ too many sensors, there will be a lot of redundant measurements, which increases the burden of data-storage. Another factor related to the number of sensors is the distribution of sensors. Given a bridge, not every location on it is of the same importance. There are some points that are sensitive to loads, while some points are not. Understanding the structural properties of the bridge helps us determine the distribution of sensors, and then figure out the optimal number of sensors.

In our sensor network, there is considerable redundancy, and the sensor distribution can also be improved. As introduced in Chapter 3, the sensors in the sensor network are distributed along three cross-sections of the last half part of a single span. The number of sensors within each cross-section is also different, one of which covers 78.6% of all sensors. The improved sensor network should at least cover the whole span (to improve modal analysis), and the sensor number within each cross-section should be approximately equal.

8.3 Future work

In the future, we try to improve our work in the following directions:

Minimal sensor network We will build a minimal sensor network model. To monitor the health of a bridge, it is necessary to employ a number of sensors. However, it is not true that the more sensors there are, the better for an SHM system. To figure out the optimal sensor number (minimal sensor network), we have explored the dependencies among multiple sensor types in Chapter 4. In the future, we will also take the dependencies within the same sensor type into account.

Automatic baseline correction method In Chapter 5, we have proposed a baseline correction method, the most-crossing method. One of the important parameters in the method is the size of the sliding window. In the method, we empirically select a parameter as the size of the sliding window. In the future, we will introduce the MDL principle to the most-crossing method to select the optimal window size. We suppose that the optimal window size will lead to the minimal MDL score.

Multiple-scale pattern detection In Chapter 5, we focus on catching the trend (baseline) hidden in the time series. The baseline can be viewed as a large-scale pattern. In Chapter 6, we pay more attention to extract the predefined patterns, which are usually of small scale. In the future, we will develop a method to identify patterns of multiple scales, by combining the most-crossing method with the predefined pattern detection method. We suppose that the MDL scores of patterns of the same scale follow similar density distributions, and the critical points between adjacent MDL-score distributions can be taken as thresholds of multi-scale patterns.

More accurate modal analysis results In Chapter 7, we employed a number of datasets for modal analysis. The experimental results indicate that the temperature has a clear influence on natural frequencies, and the influence of mass on natural frequencies is obscure. To look into the environmental influence on modal parameters in more detail, we will employ more datasets for modal analysis, and instead of just considering the environmental influence on natural frequencies, we will also take mode shapes and damping ratios into account.