論 文 の 内 容 の 要 旨

論文題目　Cluster Analysis of Audio Data
with Varying Amounts of Metadata
（メタデータを含む音響データのクラスタ分析）

氏　　名　ルイ笠原　純ユネス

The present thesis proposes methods for defect detection based on cluster analysis of audio data with varying amounts of metadata.

In Chapter 1 is presented the notion of defect detection, which is a growing concern in modern societies. Audio-based inspection methods are a class of non-destructive testing methods that have a long history in the field. One example of such methods is the hammering test for defect detection in concrete structures such as tunnels. Being mainly conducted by humans, their automation is desirable from the perspective of both efficiency and objectivity.

Previous works on this topic employed supervised learning methods which rely heavily on both the quality and quantity of available training data. Such characteristic of supervised learning methods is ill-suited for detect detection. This is because training data needs to be adequate for the considered dataset, meaning that data needs to be collected and labeled for each specific inspection target, a tedious and time-consuming task. Moreover, supervised learning methods

require training data to be representative as a condition for performance. While non-defect is the normal state and therefore can be sampled or predicted, defects are by nature unpredictable and can come in a virtually infinite number of configurations, making the process of gathering samples for each type of defect near impossible. Unsupervised learning approaches, which do not require training data, therefore appear as more suited for defect detection. However, this characteristic of unsupervised learning approaches is also their drawback: they require human guidance in the design phase. The critical aspect of the design process of unsupervised learning systems is the choice of feature vector. Feature vectors define which characteristics of the data are emphasized and greatly impact the final result.

Additional data is always beneficial to machine learning approaches. Obtaining training data cannot be considered for most cases of defect detection, as explained previously. Gathering more data samples is also a tedious process. Other, easily accessible, sources of additional data, are required. Two such easily accessible additional data are considered and taken advantage of in the scope of the present thesis: sensor-provided metadata and human-provided metadata.

Sensor-provided metadata refers to additional data generated by sensors along the main type of data. For audio-based inspection methods, spatial information is such sensor-provided metadata. Sensor-provided metadata can be automatically collected and therefore can be gathered in large amounts. However, they vary greatly in nature with the main data and therefore present an apparent incompatibility.

Human-provided metadata refers to weak supervision, i.e., hints, provided by a human expert. Compared to sensor-provided metadata, human-provided metadata bear much denser information. However, since those are provided by a human, only scarce amounts can be obtained.

In Chapter 2 are first presented the different unsupervised feature learning methods, which aim at building appropriate feature vectors without training data. Then, a survey on cluster analysis with spatial constraints, where data of different nature are combined, are introduced. Finally, a survey on semi-supervised metric learning methods, where hints provided by a human are used

to build an adequate feature space, is conducted.
From here onwards, the main contributions of this thesis are detailed.

In Chapter 3 a novel method for unsupervised defect detection using audio-based inspection methods without any metadata is proposed. This method is motivated by the fact that humans are able to conduct audio-based inspection and therefore their auditory perception is appropriate for this task. Audio samples are first transformed into Fourier spectrum and normalized before Mel-Frequency Cepstrum Coefficients (MFCC) feature vectors, a hand-crafted feature vector mimicking the human auditory perception, are built. Then K-Means clustering is used to separate the dataset between defect and non-defect audio samples. Experiments are conducted for audio-based inspection of concrete structures using concrete test blocks and shows that the proposed method outperforms popular unsupervised feature learning methods.

In Chapter 4 is proposed a new method that combines audio analysis with spatial metadata, a type of sensor-metadata, for defect detection using audio-based inspection. Along with the audio data, each sample of the dataset is provided with spatial information. As in the previous chapter, MFCC is used as feature vectors. Then, a modified Fuzzy C-Means with a spatial estimator is introduced. The spatial estimator builds estimates of the fuzzy coefficients of a sample based on its spatial neighbors. Furthermore, a process to automatically find the appropriate number of clusters is presented. Experiments conducted both in laboratory conditions and in the field shows the proposed method effectively improving defect detection performance by successful incorporation of spatial metadata in the analysis.

In Chapter 5 the availability of a human expert, able to provide weak supervision, i.e., human-provided metadata, is considered. More specifically, a semi-supervised metric learning clustering approach is considered for defect detection using audio-based inspection methods. In those settings, the fewer the amount of weak supervision, the lesser the burden on the human expert. However, reducing the amount of weak supervision negatively affects the performance of semi-supervised metric learning clustering methods. Therefore, a novel weak supervision augmentation scheme based on overclustering, an initial clustering of the dataset with a number of clusters in excess, is proposed in order to increase the weak supervision and therefore allow semi-supervised metric learning clustering approaches to have better performance with less amounts of supervision, effectively alleviating the burden on the human expert. Experiments are conducted on general UCI datasets as well as on real audio-based inspection data, both in laboratory and field conditions. Results shows that the proposed augmentation scheme allows semi-supervised metric learning methods to be competitive with considerably less amounts of weak supervision.

In Chapter 6, recommended use-cases for each of the proposed methods of this thesis based on the available metadata are presented. Furthermore, a comparison of the proposed methods in the present thesis is conducted on selected settings to provide an insight on the relative performances of the proposed methods.

Finally, in Chapter 7, the present thesis is summarized and potential future work is briefly explained.