Demo Abstract: Acoustic Signal Processing for Anomaly Detection in Machine Room Environments

Bong Jun Ko, Jorge Ortiz, Theodoros Salonidis, Maroun Touma, Dinesh Verma,
Shiqiang Wang, Xiping Wang, David Wood *
IBM T.J. Watson Research Center, Yorktown Heights, NY, United States
{bongjun_ko, jjortiz, tsaloni, touma, dverma, wangshiq, xiping, dawood}@us.ibm.com

ABSTRACT

We present a system that uses acoustic signals to monitor equipment in commercial buildings, such as in machine rooms with HVAC system components. The system uses an ensemble of machine learning classifiers to effectively label signals as either "normal" or "abnormal". We collect audio clips from mobile devices in a machine room and an elevator shaft in the main building of IBM Research. We use these to learn the spectrum of normal sound signatures and identify abnormal sounds that fall outside this range. Abnormal sounds detected by the system are presented to the end user for anomaly confirmation. We also integrate a work-order system to automatically issue a repair work-order if the sound is abnormal.

CCS Concepts

•Information systems → Information systems applications; •Computing methodologies → Machine learning;

Keywords

Audio signals, building monitoring, machine learning

1. INTRODUCTION

Commercial buildings have multiple machine rooms with a variety of motors, HVAC equipment, fans and generators. Currently, system health checking and maintenance occurs when either a complaint is made or several months have passed since the last check. In the latter case, it is common for the inspector or walk around the machine room and listen for abnormal sounds. These two approaches can cause small problems to eventually become big and costly, since faults are only found after the fault has been in effect for some time. For instance, the cost of replacing a damaged motor or fan can be around $50K dollars.

*Authors listed alphabetically by last name.

BuildSys '16 November 16-17, 2016, Palo Alto, CA, USA

ACM ISBN 978-1-4503-4264-3/16/11.

DOI: http://dx.doi.org/10.1145/2993422.2996401

Mobile audio sensing is a very active area of research [1, 2] and has been applied for a variety of applications. We choose mobile audio sensing due to the low cost and versatility of mobile phones. Phones can easily connect to a backend server via WiFi or cellular networks, support onboard processing, can easily record sound, and can be easily deployed and re-used. In this demonstration, we present a system architecture and live demonstration of an acoustic anomaly detection system using mobile phones and ambient sounds. Specifically, we present a software framework that analyzes sound clips previously collected from mobile phones in an elevator shaft and machine room setting to create a model of normal behavior. We demonstrate how a new sound clip is uploaded to our system and how anomalies are identified, displayed, and reported.

2. SYSTEM ARCHITECTURE

The system architecture is shown in Fig. 1. It collects audio signals from a mobile device for both classifier training and real-time classification. Some of the collected audio clips are labeled, i.e., a human has identified the correct description of this clip, such as whether the machine is operating in normal or abnormal condition. These labeled data are used for training machine learning models that are comprised of feature extractors and classifiers.

We train multiple classifiers on different features during the learning process. These classifiers are then combined by the model combiner building block using ensemble learning methodologies. The reason for using ensemble learning is due to the observation that a mix of multiple classifiers may perform better than a single classifier. Moreover, this framework allows us to reuse classifiers trained on separate domains into a new domain. For example, we may reuse classifiers trained with signals from two different machine rooms in a third machine room, where the domain descriptor can take some human input to describe how similar the sounds in each of the two rooms is with the third room. This enables rapid configuration of the system into new environments, even though there may not exist a sufficient amount of training data that are specifically collected in the new environment.

The combined model becomes a classifier engine which is used for real-time classification of audio signals. The classifier engine can be different for different domains, such as rooms that have different types of machines inside. The classification result is referred to as an event, which is analyzed by the root cause analysis and policy-based action generation building blocks, to determine whether an action should

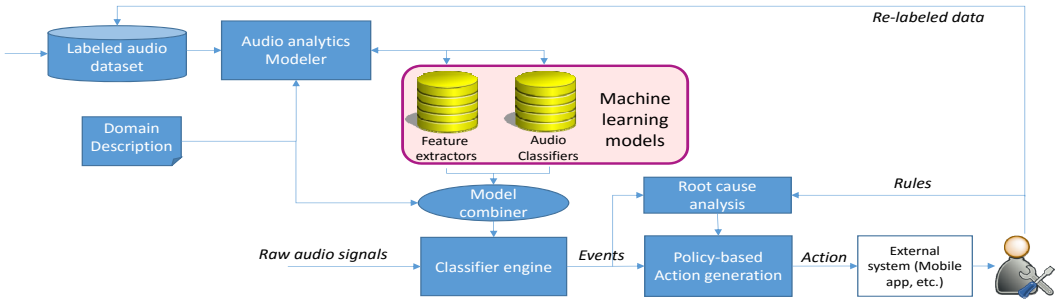

Figure 1: System architecture

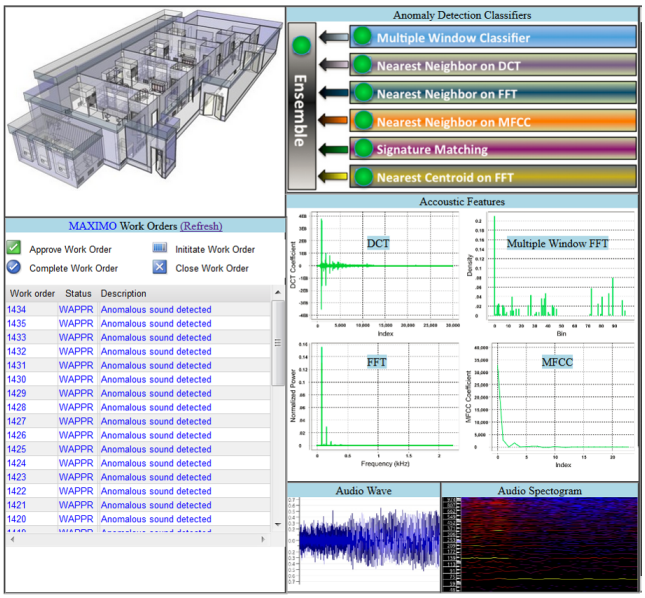

Figure 2: Demonstration view

be generated. An action can be in the form of a work-order to a service personnel for maintaining the machine. The action will be sent to an external management system, where we use IBM Maximo in our current implementation. There is an interface for the user to provide feedback, such as re-labeling sounds and re-defining action generation rules etc. The system's accuracy is increasingly enhanced by receiving user feedback over time.

3. DEMONSTRATION ENVIRONMENT

The demonstration interface on the server side is shown in Fig. 2, while there is another interface on the mobile app for collecting and submitting sound data which is not shown here. The current implementation of the system contains four feature extractors, namely discrete cosine transform (DCT), multiple window fast Fourier transform (FFT), single window FFT, and mel-frequency cepstral coefficients (MFCC), all applied onto the sound signal. The classifier implementation includes nearest neighbor, which stores the features of labeled sound clips and classifies the new sound clip into the label of the feature with smallest Euclidian distance. Mathematically, it can be expressed as follows:

$$l = L(\arg\min_i \|\mathbf{y} - \mathbf{x}_i\|^2) \quad (1)$$

where $\mathbf{y}$ is the feature vector of the new sound clip to be classified, $\mathbf{x}_i$ is the ith feature vector with known labels, the function $L(i)$ maps index i to its label, l is the classification result (i.e., the label) of the new sound clip. Other classifiers, such as nearest centroid on each label, signature matching by counting the number of frequency components with high amplitude, and a bag-of-words model based on multiple window FFT are also implemented.

The system will be demonstrated in an anomaly detection scenario where sound clips are classified as normal or abnormal. Initially, the classifiers are trained for outlier detection with normal sounds only. Abnormal sounds can be used in training once they are detected and confirmed by the user. The user also has the option of tagging new labels for sound clips to train the classifiers for more detailed classification. The demonstration shows this online training and classification procedure.

The top-right of Fig. 2 shows the outputs of different classifiers with different feature extractors, where only some of the possible combinations are shown. In the figure, all classifiers classify the sound as normal (shown by the green bulb), but this is not always the case. An ensemble based on simple weighted voting, where weights are equal to the training accuracy of individual classifiers, is applied to combine the different outputs. When anomaly is detected by the ensemble, a Maximo work order is generated, as shown in the bottom-left of Fig. 2.

Acknowledgments

The authors would like to thank Mandis Beigi and Joshua Rosenkranz who also contributed to the development of the demonstration system.

4. REFERENCES

[1] N. D. Lane, P. Georgiev, and L. Qendro. Deepear: Robust smartphone audio sensing in unconstrained acoustic environments using deep learning. In *Proc. of ACM UbiComp '15*, pages 283–294, New York, NY, USA, 2015. ACM.

[2] E. C. Larson, T. J. Lee, S. Liu, M. Rosenfeld, and S. N. Patel. Accurate and privacy preserving cough sensing using a low-cost microphone. In *Proc. of ACM UbiComp '11*, pages 375–384, New York, NY, USA, 2011. ACM.