October 2021

ACOUSTICS PROJECT REPORT

HP INC

Acoustics Project Report

1 Introduction

This project focuses on diagnosis of problems in a printer from captured acoustic signals. Previous researchers already successfully designed a feature generator to encode the acoustic signals into 6 dimensional features. And they also trained classifiers on artificial data to predict if a printer is in a good or bad status. The final target is to make a diagnosis based on the acoustic data, such as pre-alert. This will help engineers to fix the problems of the printers before customers making service calls.

2 Dataset Introduction

The input data comes from the feature generator mentioned previously. They are automatically uploaded to the cloud by printers. And the data is saved in excel files. In the current stage of this project, we do not directly work on the raw acoustic data but extracted features. For each printer, the excel files recorded the time tags, platform names, printer serial numbers and 6 extracted features for acoustic signals. And there are multiple records for each printer every day. Some dates are skipped. The missing data for the skipped dates can cause a certain inconvenience in data processing. So, we did a causal data interpolation first to estimate the status of printers on the missing dates.

Another excel file records the failure data for different printers for various reasons including regular maintenance, device noises, printing quality, paper-feed jams, and other unclear issues. In this file, many details in the failures of printers are recorded, which includes event id, event open time, event close time, platform name, printer serial number, symptoms, symptom details and so on. We merge the data from this file together with the previous excel file and plot them together to look for potentials to file pre-alerts for the incoming symptoms in printers.

Figure 1: Pipeline

3 Methodology

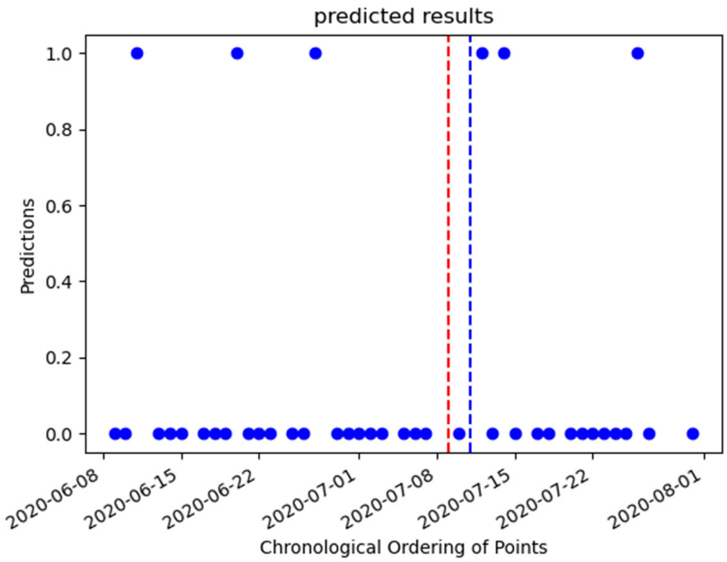
The complete workflow is shown in Figure 1. In this pipeline, we complete the causal interpolation and pre-alert analysis steps. For previous steps, features are extracted from acoustic data using a detector, which was designed by a previous researcher. There are 6 features in total: strong-tone frequency, strong-tone relative amplitude, strong-tone absolute amplitude, peak width, modulation frequency and modulation absolute amplitude. Next, a classifier, which is pretrained on the augmented acoustic data, is used by us to classify the real acoustic data. It takes a 6-dimensional feature as input and outputs a single label (normal or abnormal). The label indicates the working status of the printer at that specific time. From this, we can get a temporal 1D data sequence, which can be used to investigate the relations between acoustic data and printer failure. An example of such a 1D data sequence is shown in Figure 2a. The vertical broken lines represent the start and end of a service call.

3.1 Data Interpolation

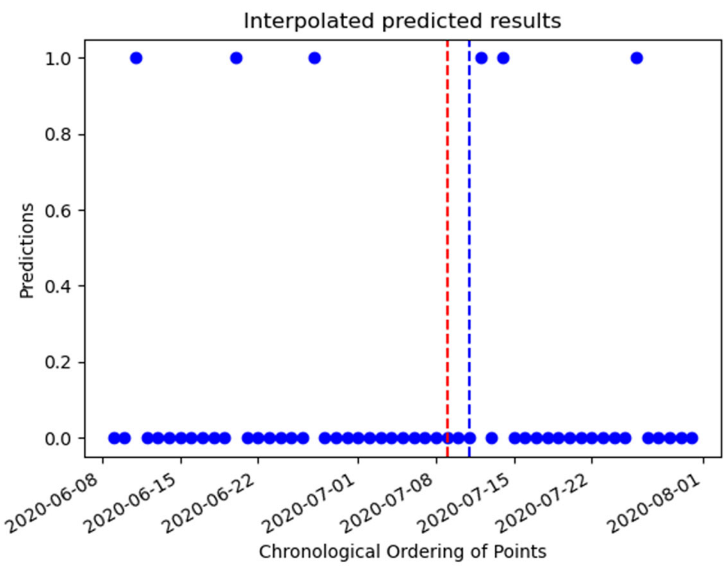
Before we start, we found that the acoustic data does not cover every day. For some days, the data points are missing. This will bring some inconvenience to the pre-alert analysis later. So, we decide to do an interpolation to get the complete data for each day.

To get an effective data interpolation, we try several different interpolation methods. At first, we use a single non-causal averaging window with length of 30 days. It interpolates well but causes certain local inconsistencies. A printer will switch between good and bad status within 3 to 4 days. Intuitively, if the printer fails on day one and day three, it will be very likely to fail on day two also. But if the averaging window is too large, it will cause status of printer in day two to be normal. So, we change to use causal filters with varied length. The lengths we pick including 3, 14, 21 and 30. We found that 14 is the most effective one. It smooths the curve and prevents it from being too flat. And it is not helpful for pre-alerting if the filter is non-causal. So, our filters are causal filters. In Figure 2, we show an example of two plots before and after the interpolation.

2



(a) Original predicted results.



(b) Interpolated predicted results.

Figure 2: Original and interpolated binary classification results. The red vertical line represents start of a service call while the blue vertical line represents the end of a service call. The prediction ’1’ corresponds to ’abnormal’ and ’0’ corresponds to ’normal’.

3.2 Causal Integration and Slope Feature

We plot out the predicted temporal results for normal and abnormal status of printers. Based on these one-dimensional data, we investigate the relation between temporal records and potential future failures of printers.

We apply causal averaging filters of varied lengths to the predicted temporal results. Some peaks in the filtered results are correlated with the failure records. Different failure records are marked with broken vertical lines in different colors. Red marks represent regular maintenance. Blue marks represent service calls for noise. Green marks represent service calls for printing quality. Magenta marks represent service calls for paper feed jam. And black marks represent other non-specified issues. An example is shown in shown in Figure 3.

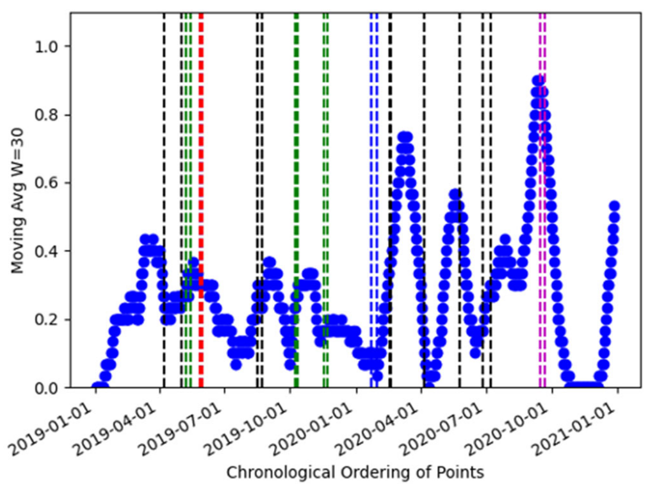
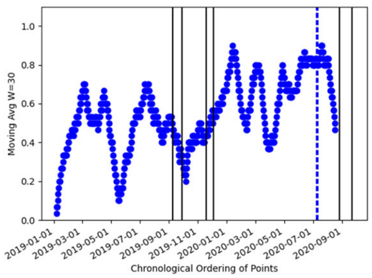
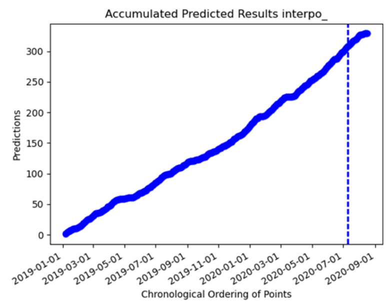


Figure 3: Causal integration over classification results. In this plot, different colors of vertical lines represent different symptoms. Red represents regular maintenance. Blue represents noise. Green represents printing quality. Magenta represents paper feed jam. Black represents other non-specified issues.

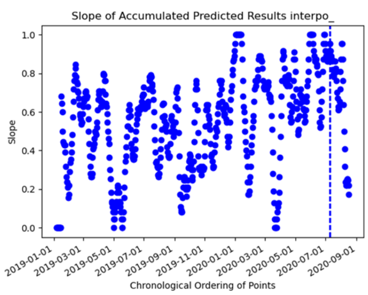
The slope feature is calculated based on the casual integration. We accumulate the integration first. Then, over a fixed temporal length, we fit the accumulation curve with a line. The slope of the line is plotted. We found that peaks in the slopes are correlated with time for the service calls. Figure 4 shows an example to reveal this relationship. The peaks on plots of the slope features occur at the time when a service call happens (vertical blue broken lines).



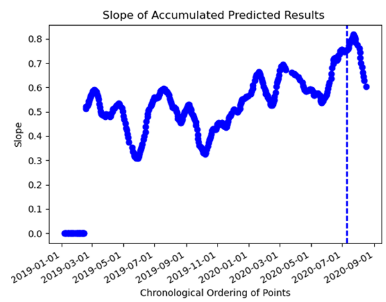
(a) Original predicted results.



(b) Causal integration of predicted results.

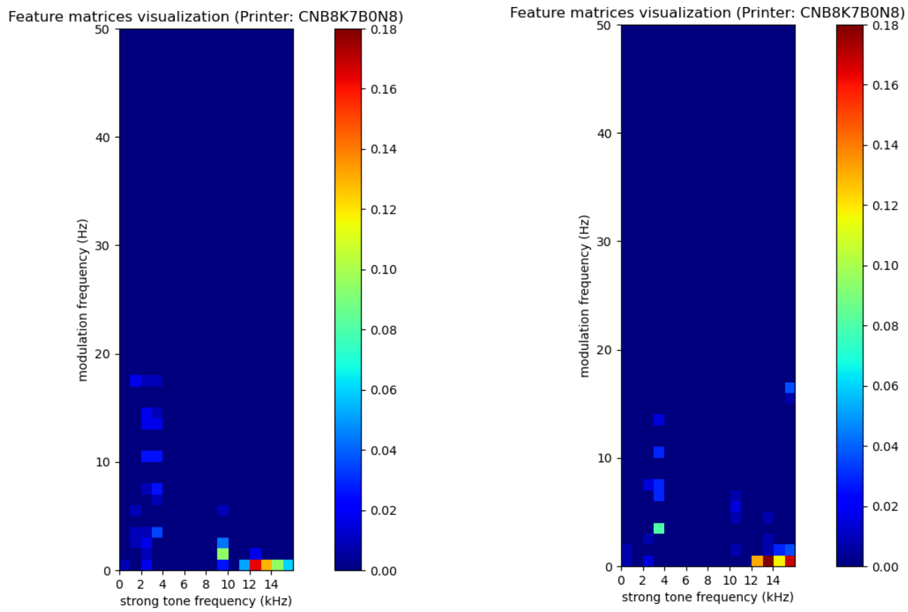


(c) Slope feature of causal integration (window length 10).



(d) Slope feature of causal integration (window length 30).

Figure 4: Original predicted results, its accumulation, and the slope features in different temporal window length (10 and 30 days). For the window length 30 (d), peaks on plots of the slope features occur at the time when a service call happens (vertical blue broken lines).



(a) Histogram plot before a service call. (b) Histogram plot after a service call.

Figure 5: Histogram plots of features before and after a service call. The vertical axis is 1 Hz per pixel while the horizontal axis is 1 kHz per pixel. We can see that the acoustic features are greatly reduced around 3 kHz and 10 kHz after a service call.

3.3 Visualization of data

At this stage of current project, we process the features extracted from acoustic data. We do not process the recorded acoustic data directly. To better understand the data, we plot a temporal histogram of the features. We pick strong-tone frequency and modulation frequency from the 6 features. We accumulate them in time to show happening frequency of instances. An example histogram plot is shown in Figure 5. In this plot, the vertical axis represents modulation frequency and is 1 Hz per pixel. The horizontal axis represents strong-tone frequency and is 1 kHz per pixel. By comparing the two plots, we can see that the acoustic features around 3 kHz and 10 kHz are greatly reduced after a service call. The visualization results will help us to focus on the most informative frequency regions.

4 Conclusions and Future work

In this project, we make a few contributions:

1. Apply classifiers pretrained on artificial data to analyze real data.

2. Apply causal filters to interpolate for the missing dates.
3. Reveal the correlation between calculated features and service calls.
4. Calculate the slope features of accumulated integration of classification results for various temporal lengths and show the potential to do pre-alerting.

Here are some possible future extensions of this project:

1. Continue to analyze the relation between calculated features and service calls.
2. Design a system that can automatically predict the service calls in the future and alert before failures.

Disclosed by Qiulin Chen and Jan P. Allebach, Purdue University and Katy Ferguson and Mark Q. Shaw, HP Inc.