Fault Diagnosis with Machine Learning Methods
Data Acquisition Report

Jiachang (Ernest) Xu
jiachang.xu@kcl.ac.uk 1819235
supervised by Dr. Hak-Keung Lam

March 15, 2019

1 Introduction

This report iterates the data sets that I acquired for my dissertation project, **Fault Diagnosis with Machine Learning Methods**. There are 3 sets of data: (1) the bearing data set from Case Western Reserve University (CWRU), (2) the bearing data set from Society for Machinery Failure Prevention Technology (MFPT), and (3) the bearing data set from National Aeronautics and Space Administration (NASA).

2 Description of Data

This section describes all 3 data sets in details (Table 1 to 5). Information includes (1) the attributes of data, (2) the format of data, (3) the number of instance, etc. This section cites information from each data set's corresponding website, whose URL is listed in Section 4.

Table 1: Summary of CWRU bearing data set

	Subset 1	Subset 2	Subset 3	Subset 4
Description	Normal baseline	12k drive-end bearing fault	48k drive-end bearing fault	Fan-end bearing fault
No. of files	4	60	52	45
No. of variables	3	4	3	4
Baseline accelerometer data collected	No	Yes	No	Yes
Drive-end accelerometer data collected	Yes	Yes	Yes	Yes
Fan-end accelerometer data collected	Yes	Yes	Yes	Yes
Motor speed collected	Yes	Yes	Yes	Yes
Sampling rate (samples/sec)	N/A	12,000	48,000	12,000
File format	.mat	.mat	.mat	.mat

Table 2: Variables of CWRU bearing data set

Name	Meaning	Data type
`experiment#_BA_time`	Time-series base accelerometer data	Column vector, double
`experiment#_DE_time`	Time-series drive-end accelerometer data	Column vector, double
`experiment#_FE_time`	Time-series fan-end accelerometer data	Column vector, double
`experiment#RPM`	Revolutions per minute	Scalar, double

Table 3: Summary of MFPT bearing data set

	Subset 1	Subset 2	Subset 3	Subset 4
Description	Baseline conditions	Outer race fault conditions with invariable loads	Outer race fault conditions with variable loads	Inner race fault conditions with variable loads
No. of files	3	3	7	7
Sampling rate (samples/sec)	97,656	97,656	48,828	48,828
Duration (sec)	6	6	3	3
Load (lb)	270	270	25, 50, 100, 150, 200, 250 and 300	0, 50, 100, 150, 200, 250 and 300
Input shaft rate (Hz)	25	25	25	25
File format	.mat	.mat	.mat	.mat

Table 4: Variables of MFPT bearing data set

Name	Meaning	Data type
`bearing.rate`	Input Shaft rate (Hz)	Scalar, integer
`bearing.load`	Load (lbs)	Scalar, integer
`bearing.gs`	Vibration amplitude (Gs)	Column vector, double
`bearing.sr`	Sampling rate (samples/sec)	Scalar, integer

Table 5: Summary of NASA bearing data set

	Subset 1	Subset 2	Subset 3
No. of files	2,156	984	4,448
No. of columns	8	4	4
Column(s) for bearing 1	1 and 2	1	1
Column(s) for bearing 2	3 and 4	2	2
Column(s) for bearing 3	5 and 6	3	3
Column(s) for bearing 4	7 and 8	4	4
File format	ASCII	ASCII	ASCII
Result	Inner race defect in bearing 3; roller element defect in bearing 4	Outer race failure in bearing 1	Outer race failure in bearing 3

3 Curation of Data

This project is scheduled to present an implementation written in MATLAB, with an option to migrate to Python. Therefore, the first task of the data cleaning process is to save all the data sets into .mat files. The CWRU and MFPT data sets have already satisfied this requirement; the NASA data set has an ASCII format, so a script is needed to process the raw data, and save into corresponding .mat files. The second task is to read the literature that uses these data sets, and to replicate their experiments. The schedule of these tasks are listed in Table 6

Table 6: Schedule of data cleaning process

Task #	Description	Deadline
1	Transform the NASA data set into .mat files	31st March, 2019
2	Read literature and replicate experiments	Ongoing

4 Source of Data

- CWRU bearing data:
  http://csegroups.case.edu/bearingdatacenter/pages/download-data-file
- MFPT bearing data:
  https://mfpt.org/fault-data-sets/
- NASA bearing data:
  https://ti.arc.nasa.gov/tech/dash/groups/pcoe/prognostic-data-repository