Instructions: Airfoil Data Digitization

I. Overview

The goal of this project is to develop a machine learning model (i.e. "Large Airfoil Model") that is capable of predicting airfoil pressure distribution given its geometry and flow conditions. What differentiates this work compared to existing literature is that we seek to use *experimental* measurements as the training data. Consequently, in order to develop this model, we need to establish a database of experimental airfoil data by data mining for experimental reports and digitizing its data into CSV files. This guide aims to give you a tutorial on how to do this efficiently. I will share with you the Git repository (https://github.com/hwlee924/GP-Aero) where you will be able to add whatever data you digitized.

II. Setting up Repository Access

In order to set up your access to the Git repository:

1) Download and install the latest version of Git (https://git-scm.com/downloads).
2) In terminal, confirm the installation was done properly by entering the *git* command.
3) Clone the repository using the command *git clone https://github.com/hwlee924/GP-Aero.git*
4) Make sure your git is synced using the command *git config –global user.email "you@example.com"* and go through the necessary authentification process.
5) This should let you commit your changes to the repository.

III. Data Mining

There are largely three ways that I have been using to obtain experimental airfoil pressure data: 1) Google Scholar, 2) Aerospace Research Central, and 3) NASA Technical Reports Server (NTRS). While any contributions to the data mining is welcome, I ask you to prioritize experimental data where **the chord Reynolds number is order of $\times 10^6$ and a Mach number range of 0.0 to 0.8**. **If you find a good paper that's worth digitizing, please upload it to *GP-Aero/Airfoil Database/Sources***.

A. Google Scholar

In Google Scholar, you can essentially "google" existing papers. Use appropriate key words such as "airfoil pressure distribution", "wind tunnel investigation", etc. to find a wide variety of existing papers from different sources. Most of these will involve graphical plots of pressure data.

B. Aerospace Research Central

Aerospace Research Central is hosted by AIAA and will give you papers that they have hosted in AIAA conferences and journals. If you have a GT account and GT VPN, you should be able to access this without an issue. The website is more focused than Google Scholar but has a good variety of all aerospace-related literature. As with Google Scholar, most of these will involve graphical plots of pressure data.

C. NASA Technical Reports Server

NTRS specifically hosts NASA's technical reports. While these documents are often very old, they also provide the most extensive coverage of a variety well-performed airfoil experiments. Most of the data has a combination of tabulated and graphical results. This lets you verify your digitization relatively easily. However, due to the age of many documents, the quality of the scanned PDFs may be poor.

IV. Data Digitization

The digitized data should be **uploaded to GP-Aero/Airfoil Database/Digitized data**. Each airfoil will get its own folder with the folder name being the name of the airfoil (e.g. *GP-Aero/Airfoil Database/Digitized data/NACA0012*).

A. Data Format

Each airfoil folder should contain 2 types of CSV files: the coordinate files and the pressure data files.

1. Coordinate Files

The coordinate files will define the geometry of the airfoil, with the first column being the normalized chordwise location (x/c) and the second column being the normalized thickness (y/c) as seen in Figure 2. The order of the coordinates should start from the upper surface of the trailing edge ($x/c = 1.0$) to the leading edge ($x/c = 0.0$) and end at the lower surface of the trailing edge ($x/c = 1.0$). The coordinate files should be named: *<airfoilname>_coordinates.csv*.

x/c	y/c
0.996018	0.001732
0.980268	0.001524
0.973987	0.001844
0.961987	0.00263
0.948393	0.005271
0.941174	0.005986
0.929174	0.007647
0.912299	0.009818
0.899174	0.011611
0.891674	0.012343
0.879486	0.014745
0.863549	0.016705
0.854174	0.017856
0.843861	0.019148
0.830736	0.02112
0.81058	0.022978
0.795111	0.024714
0.779173	0.026907
0.762767	0.028852
0.747298	0.030631
0.740501	0.031352
0.728548	0.032697
0.714954	0.034081
0.70761	0.034767
0.695735	0.036023
0.680735	0.037564
0.671922	0.03827
0.661047	0.039198
0.628234	0.041645
0.611828	0.042915
0.594953	0.044414
0.578546	0.045535
0.561671	0.046308

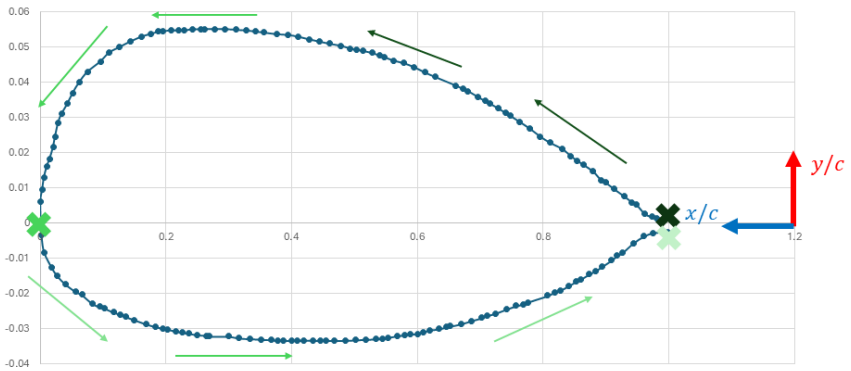

Fig. 1 Visualization of coordinates file

2. Pressure Data

Each pressure data CSV file consists of the pressure measurements along the entire wing chord at different Mach numbers at a given angle of attack. The file should following the naming scheme of *<airfoilname>_A<angle of attack> _A.csv*. This should be at a single Reynolds number. Please keep track of what Reynolds number the data is in a separate text file and upload that together as well.

	Mach number																
x/c	0.31	0.41	0.51	0.56	0.6	0.63	0.66	0.68	0.7	0.73	0.76	0.78	0.81	0.84	0.87	0.9	0.92
1	0.08	0.12	0.13	0.12	0.12	0.1	0.12	0.14	0.12	0.14	0.12	0.14	0.12	0.04	-0.08	-0.21	-0.24
0.947	0.01	-0.01	0	0	0	-0.02	0.01	0.01	-0.01	0	-0.01	0.02	0.04	-0.06	-0.18	-0.3	-0.35
0.902	-0.03	-0.06	-0.05	-0.05	-0.05	-0.06	-0.05	-0.05	-0.07	-0.06	-0.07	-0.04	-0.03	-0.17	-0.27	-0.38	-0.45
0.852	-0.12	-0.16	-0.14	-0.14	-0.15	-0.16	-0.14	-0.14	-0.17	-0.16	-0.17	-0.13	-0.15	-0.3	-0.37	-0.49	-0.56
0.802	-0.22	-0.26	-0.26	-0.26	-0.27	-0.3	-0.27	-0.28	-0.31	-0.29	-0.32	-0.27	-0.29	-0.43	-0.51	-0.64	-0.62
0.752	-0.25	-0.3	-0.3	-0.31	-0.33	-0.36	-0.34	-0.36	-0.38	-0.37	-0.39	-0.32	-0.55	-0.71	-0.7	-0.69	-0.69
0.701	-0.28	-0.32	-0.33	-0.33	-0.36	-0.39	-0.38	-0.4	-0.44	-0.43	-0.45	-0.45	-0.81	-0.77	-0.72	-0.71	-0.68
0.651	-0.3	-0.35	-0.36	-0.37	-0.4	-0.43	-0.41	-0.44	-0.48	-0.48	-0.54	-0.7	-0.84	-0.81	-0.78	-0.75	-0.67
0.6	-0.33	-0.38	-0.39	-0.4	-0.43	-0.46	-0.45	-0.48	-0.52	-0.54	-0.61	-0.78	-0.9	-0.88	-0.8	-0.73	-0.65
0.551	-0.35	-0.4	-0.4	-0.41	-0.45	-0.48	-0.46	-0.5	-0.54	-0.56	-0.64	-0.81	-0.88	-0.82	-0.74	-0.68	-0.6
0.5	-0.36	-0.42	-0.42	-0.43	-0.47	-0.51	-0.49	-0.53	-0.57	-0.6	-0.68	-0.84	-0.85	-0.78	-0.7	-0.63	-0.55
0.45	-0.37	-0.42	-0.43	-0.44	-0.48	-0.51	-0.5	-0.54	-0.58	-0.61	-0.7	-0.8	-0.79	-0.72	-0.64	-0.57	-0.5
0.4	-0.37	-0.42	-0.42	-0.44	-0.48	-0.51	-0.5	-0.54	-0.58	-0.61	-0.7	-0.77	-0.74	-0.67	-0.59	-0.52	-0.45
0.352	-0.38	-0.42	-0.42	-0.44	-0.48	-0.51	-0.5	-0.54	-0.58	-0.6	-0.68	-0.73	-0.69	-0.62	-0.54	-0.47	-0.39
0.298	-0.38	-0.42	-0.42	-0.43	-0.47	-0.51	-0.49	-0.53	-0.57	-0.59	-0.66	-0.68	-0.63	-0.56	-0.48	-0.41	-0.33
0.251	-0.37	-0.41	-0.41	-0.43	-0.46	-0.49	-0.48	-0.51	-0.55	-0.57	-0.62	-0.61	-0.56	-0.48	-0.4	-0.33	-0.26
0.2	-0.33	-0.37	-0.37	-0.39	-0.42	-0.45	-0.44	-0.46	-0.48	-0.49	-0.52	-0.5	-0.45	-0.37	-0.3	-0.22	-0.15
0.15	-0.31	-0.36	-0.36	-0.36	-0.39	-0.42	-0.41	-0.43	-0.45	-0.46	-0.47	-0.46	-0.42	-0.33	-0.25	-0.18	-0.1
0.101	-0.3	-0.35	-0.34	-0.34	-0.37	-0.4	-0.38	-0.4	-0.42	-0.42	-0.43	-0.41	-0.36	-0.26	-0.17	-0.09	-0.02
0.075	-0.26	-0.31	-0.3	-0.3	-0.32	-0.34	-0.33	-0.34	-0.35	-0.34	-0.34	-0.31	-0.24	-0.15	-0.06	0.01	0.08
0.051	-0.21	-0.24	-0.23	-0.23	-0.25	-0.26	-0.24	-0.24	-0.25	-0.23	-0.23	-0.19	-0.13	-0.03	0.06	0.14	0.2
0.025	-0.23	-0.26	-0.26	-0.25	-0.26	-0.27	-0.26	-0.24	-0.25	-0.23	-0.22	-0.17	-0.1	0.02	0.13	0.22	0.28
0.013	-0.15	-0.17	-0.15	-0.14	-0.13	-0.14	-0.12	-0.1	-0.09	-0.06	-0.04	0.02	0.09	0.22	0.32	0.41	0.48

Fig. 2 Visualization of pressure data file

Will add more details when time permitting

B. Tabulated Data

If the source material presents the experimental data in a tabulated format, I recommend that you use the online Optical Character Recognition (OCR) tool ExtractTables (https://extracttable.com), which I found to yield the best results. I have paid for the Web-Pro version which allows us to digitize as many files as we want. Unlock the pro version by clicking on the Web-Pro icon on the top right of the website and entering the API key: yEahkEhRKQ5oYCz87FfrAcZXvTTCfLkG8tzreTQ6. Let me know if we run out of credits so that I can recharge it.

The workflow of using ExtractTables is as follows:

1) Take a screenshot of the table that you want to digitize.
2) In ExtractTables, upload the screenshot.
3) When the conversion is complete, copy the data to clipboard and paste it into the csv file in Excel.
4) Make sure the digitization was done accurately. Fix erroneous results as necessary.
5) Rearrange the data into the data format described earlier.

C. Graphical Data

If the source material presents the experimental data in a graphical format (plot), I recommend that you use the online tool WebPlotDigitizer (https://automeris.io/WebPlotDigitizer.html).

The workflow of using WebPlotDigitizer is as follows:

1) Click *Launch Online*
2) Take a screenshot of the graph that you want to digitize.
3) In WebPlotDigitizer, upload the screenshot (*Files* → *Load Image(s)*).
4) Calibrate the X- and Y-axes. Click on the minimum and maximum values of each axes and click *Complete!*. Then, enter their numerical values. Tick the box for *Assume axes are perfectly aligned with image coordinates (skip rotation correction)*.
5) Click on the plot points to digitize the data. Make sure you click the points in the order of: Trailing edge, upper surface → Leading edge → Trailing edge, lower surface.
6) When finished, click *View Data* on the left and *Copy to clipboard*. Paste the result into the csv file in Excel.
7) Make sure the digitization was done accurately. Fix erroneous results as necessary.
8) Rearrange the data into the data format described earlier.

D. list_airfoils.xlsx

Once you fully digitize the results for a single airfoil from a source document, please update the list_airfoils.xlsx and upload to the repository.

Will add more details if necessary

E. Notes

1) If you cannot digitize a certain data point due to poor quality, either do your best to estimate or omit the point. If you make an educated guess, make sure you make a note of it somewhere (e.g. in a text file or in the airfoil list file) and upload.