**Design/Build/Fly**

End of Semester Report

SP 2018 Dylan Tsai, dt439

**1 Abstract**

Throughout the past year, the propulsion team has struggled to relate theory to experimental results, and as a result, shifted towards a method closer to trial and error. To avoid such time consuming methods, machine learning algorithms can be applied to test data in order to identify the latent variables that cause experimental results to deviate from theory. Although the team does not have enough data yet, efforts from individuals on the team indicate that enough data will be collected soon. So, to begin developing machine learning models, data is scraped from eCalc for future reference as a source of what can be called ‘theory’. Then, basic data analysis and regressions are performed on this dataset in order to roughly estimate the performance of machine learning algorithms in the context of propulsion data. The results indicate that the discrepancy between eCalc’s estimates and the team’s test data is related to size. Also, regressions appear to be extremely flexible and effective.

[Propulsion, Data Acquisition, Data Analysis, Machine Learning, eCalc]

**2 Introduction**

This year, the propulsion team struggled greatly with finding propulsion systems that could supply enough power for the plane. The propulsion team sized propulsion systems primarily by starting with a goal thrust and a basic set of restrictions on weight and size, using intuition to guess what set of parts could fit those parameters, and testing systems until one worked. As a result, the propulsion team struggled to keep up with the quickly changing plane. Ultimately, the propulsion system used at competition was not tested thoroughly on the competition plane, and so was not nearly as dependable as the team could have hoped for.

While the plane’s aerodynamic changes can often be made quickly due to a considerable amount of reasonably accurate theory to pull from, the propulsion team has found that much of the already limited knowledge on electric propulsion systems is difficult to apply to our planes, which often push the propulsion systems beyond their normal limitations and are smaller than the planes considered in common published studies.

Several approaches to solving these problems have been considered and currently in development. Most of these approaches involve either accumulating more data by running a wider breadth of propulsion tests and constructing apparatus for wind tunnel tests or organizing. Others aim to summarize data through different organizational systems and automated database maintenance. These methods will provide the team with better tools to develop an intuition on how to size propulsion systems. The purpose of this paper is to begin developing basic machine learning tools for building models and performance estimates that can function independent of intuition and can eventually be integrated into the sizing estimates that the aerodynamicssize propulsion systems.

Since the propulsion team’s rush for data collection is only just beginning, the size of the team’s test data will not be large enough for accurate estimators to be produced until around the middle of next semester. So, this report focuses largely on data pulled from eCalc. Due to its size and basis in both experimental data and theory the eCalc data will be used in three ways: as a source of information from which propulsion theory can be extracted, as a dataset from which preliminary estimates can be made, and as a test set against which algorithms can be validated.

In order to accommodate the team’s propulsion data, the primary goals of the data analysis in this report are clarity and transparency, analysis with and without intuition, and forgiveness and stability and controls sub-teams use.

These machine learning algorithms may be able to help the team identify latent variables that must be recorded in the future. Furthermore, ability to be integrated into the team’s current sizing code will allow for greater cohesiveness between the propulsion team and the rest of the team. The most important feature of these machine learning algorithms, however, is their ability to uncover latent variables. This allows it to bridge the gap between theory and our results, which can allow the team to properly apply theory to determining how to of error and outliers.

**3 Procedure**

First, a program was written to scrape around one million sample points from eCalc. This was done by emulating a browser and iteratively inputting values into eCalc while recording the results. The inputted parameter values were, for the most part, random and restricted to a range of plausible values that the parameters might take on for the team. Because of the large number of variables and iterations, storage and computational efficiency were prioritized over code readability and maintainability.

Then, the data was compiled onto a single csv file for analysis. The data was then analyzed through four sections: data cleaning, K-Means analysis, observations on the effects of propeller pitch and diameter analysis, and a regression model for thrust. The procedure was documented on Jupyter Notebooks, for which the pdf can be found under the name Procedure.

**4 Results**

The K-Means analysis revealed that on average, only around 85% of the battery capacity is used, which could be a cause of the discrepancy between eCalc predictions and the team’s results. It also appears that systems with larger systems (that is, systems with higher current and voltage) have greater efficiency at maximum. This implies that downsizing propulsion systems has an exponential effect on output.

It was also found that at both high diameters as well as at high pitch ratios, efficiency approaches 0. This means that for a constant pitch, the optimal diameter is the one that balances the inefficiencies of low diameter and high diameter. Also, current decreases as pitch ratio increases. This indicates that with larger pitch ratios, a greater voltage (and thus a larger battery pack) may be needed; however, at larger diameters, efficiency and RPM decrease. Therefore, for a given pitch, there exists an optimization problem between diameter and battery size.

The relationship described above yields an algorithm for choosing diameter and battery size given a pitch and motor: starting from a large diameter, decrease diameter and increase battery size until a peak thrust is reached. The issue, then, is how to choose pitch. Torque and thrust appear to increase slightly with pitch; however, there also is a small peak between 4 and 5 inches. Torque and thrust both increase as pitch increases. Therefore, the optimal pitch is the highest pitch possible, bounded above by the pitch ratio relationships described earlier. This means that diameter is bounded by pitch and pitch is bounded by diameter; therefore, this problem is most easily solved by iterating through different pitches, finding the optimal diameter for each pitch, and choosing the best option out of these semi-optimal pitch-diameter pairings.

This algorithm requires machine learning on test data in order to find an accurate relationship between the control parameters (pitch, pitch ratio, diameter, and battery cells) and the outputs (thrust, torque, necessary current, etc.).

The weakness of this algorithm, however, is that it does not take into account the motor itself. With this algorithm, a Kv would be chosen given the found optimal point, and then a motor with that Kv would be selected; however, there exist important relationships that involve Kv, weight, and motor internal resistance. So, an improved algorithm would use a different variable as the primary variable to iterate through, use that variable to predict the tradeoffs between motor weight, Kv, and internal resistance, and then use the concepts discussed above to find the optimal points. This method requires an accurate estimate of some parameters using other parameters with unclear relationships, which can likely be accomplished through basic machine learning once enough data is collected.

As for predicting output using variables with somewhat clear relationships with the output, experimentation with the eCalc dataset proved promising. A simple regression for predicting thrust, using close to no intuition, nearly no data analysis by hand, and no feature engineering, resulted in an average relative error of at most 1%, even when using 500 data points to predict 100,000 data points or using 200 data points to predict 10,000 data points. Since no feature engineering was done, Lasso models prove to be extremely effective in this context.

An analysis of the error between eCalc predictions and test results indicated that the discrepancy between eCalc results and test results is related to the size of the propulsion system. This test does not determine whether the cause of this relationship is eCalc's predictions being less accurate on smaller scales or a systematic error in the team's tests on smaller systems; more data will need to be collected.

**5 Future Development**

The results of this test are primarily experimental, as the team does not have enough data for these algorithms to be verified. When more data is collected, the methods in this report should be applied to and adapted for the team’s test data. During that application, more attention will be required for engineering new features and especially identifying and dealing with outliers. Also, machine learning algorithms can estimate relationships between values that do not have a clearly defined relationship. For example, a database of motor attributes can be analyzed in order to estimate some motor attributes from other motor attributes. This would make the benefits and tradeoffs of different motors more clear and therefore aid in selecting propulsion systems, both by hand and programmatically.