Diversity of Maintenance Logs and Delay

Bryan Anderson

Cobec Consulting

March 26, 2018

Abstract

The FAA records all maintenance activity occurring on NAS assets. For work that cannot be automated, technicians detail what actions they performed to resolve the maintenance event. This unstructured data provides useful, but difficult data to utilize. With the use of NLP techniques, a simple measure of diversity of log words is correlated with delay incurred at airports.

1 Introduction

Text data provides a wealth of information that is generally difficult to use. With the advent of new technologies and techniques, analysis of this data type is becoming easier to perform. This field of analysis is known as Natural Language Processing (NLP). NLP is concerned with processing human text into a machine usable format for analysis. An example of this style of work can be found in the Remote Monitoring Logging System (RMLS). RMLS is the Federal Aviation Administration's (FAA) logging system for recording maintenance work performed on facilities in the National Airspace System (NAS). RMLS provides a rich database of unstructured text. Reading every log file, however, is not pragmatic solution in understanding the data. Using simple NLP techniques, this paper explores the correlation of the diversity of words used in logs with delay experienced at an airport.

2 Motivation

The motivation for this report comes from the capital investment process in the FAA. The FAA has mature process of acquiring capital. This process requires at various points in time specific analyses to be conducted. A brief overview of the FAA acquisition process and two data systems are given in the subsequent subsections.

2.1 FAA Capital Investment Plan

The FAA has been primarily responsible for the safety of civil aviation and fostering air commerce since it's inception [3]. The FAA has over the years evolved into a complex system of automation, organization entities, and management processes to support the growth of U.S. aviation. An example of a management process is the Capital Investment Plan (CIP). The CIP identifies and describes the capital investments required to sustain and modernize the infrastructure, system, and services, and procedures required for the safe and efficient operation of the NAS [1] [5]. This report helps FAA plan for the next 5 years and support funding requests from the President's budget.

The CIP requires detailed and thorough analysis of the potential costs associated with the capital investments being proposed. The analysis process in the FAA is known as the Acquisition

[1] National Airspace System

Management System (AMS). The AMS establishes policy and guidance for all aspects of life cycle acquisition management for the FAA [6]. The AMS can be loosely viewed as significant points in time in which certain analyses are needed during the review of the investment (see Figure 1). The AMS artifacts are then reviewed by the Joint Resource Council (JRC) for approving investment [5]. An example of an AMS artifact is a life cycle cost estimate (LCCE).

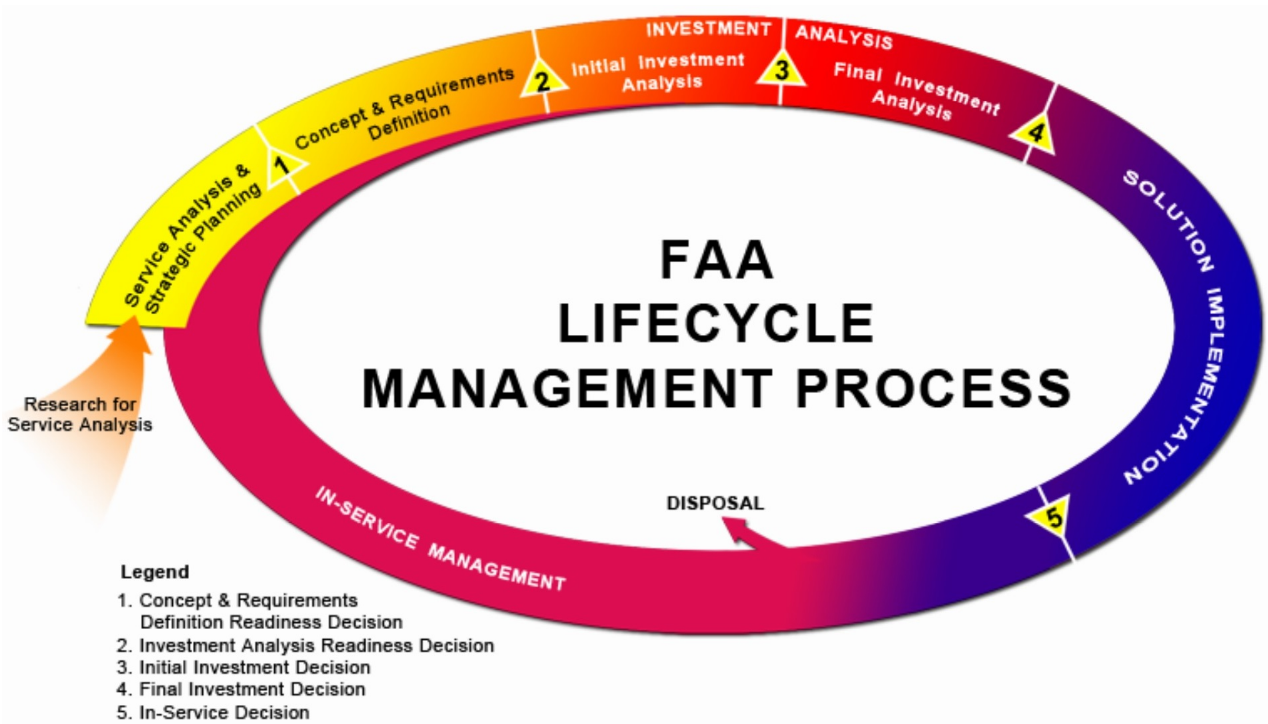

Figure 1: FAA AMS Process

The significant points in the AMS process are known as decision points. Critical decision points in the AMS for LCCEs are the Initial Analyses Readiness Decision (IARD), Initial Investment Decision (IID), and the Final Investment Decision (FID). Each of theses decision points require increasing quality of LCCEs. Various methods exists to create high quality LCCEs. Some estimation methods include by analogy, parametric, or by engineering. Data for constructing these estimates can come from many sources. One of these sources is text data. New technology and methods like NLP is opening up new possibilities for analysis. This paper starts to explore the potential use of NLP techniques in FAA cost analysis.

2.2 Data Description

Data analyzed in the report came from the RMLS and Bureau of Labor Statistics (BTS). The data was retrieved from two different data sources. RMLS data was retrieved from an internal Oracle database. The BTS data was retrieved from the public BTS T100 database. Both data sets needed to be cleaned before analysis could be done. The following two subsections describe both data sets in more detail.

2.3 RMLS

The RMLS data consists of maintenance activities occurring in the NAS. RMLS provides the logging, monitoring, and controlling functionality for the official record of assets ([4], [1]). The system is logically bifurcated into the National Logging Network (NLN) and National Remote Maintenance Monitoring (RMM) Network (NRN). The NLN sub-system contains the applications used by the on-site technicians when maintenance is being performed on NAS assets. The NRN is the sub-system in RMLS that provides the automation of mechanical reporting and the IT infrastructure for system to function. The data used in this paper is primarily generated in the NLN.

The NLN sub-systems manages and records the maintenance activity that can't be automated. Three applications comprise the NLN:

- Event Manager (EM)
- Simplified Automated Logging (SAL)
- Peabody

The EM monitors an asset remotely and when status thresholds are breached; it manages the ticketing and dispatching of technicians to the assets [4]. These status thresholds are set by the Peabody application [4]. The technician activities needed to resolve that breach are manually entered via the SAL. The SAL application creates a log record and captures all the relevant information to describe the maintenance. Essential data included in every logging record is the date-time, code category, technician remarks, and technician initials [2]. Often, when a technician is creating a log record, the detailed required to document the work activity is too much for the

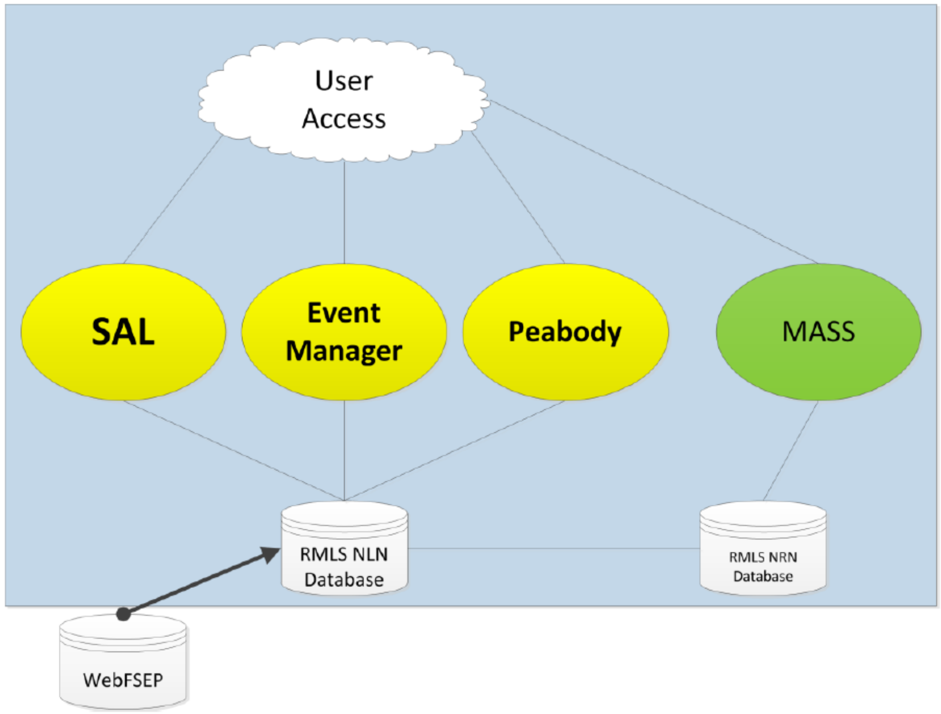

Figure 2: NLN Application View

default log remarks field. Thus, SAL also has the functionality for the technician to creating additional log comment records that allow for longer, more detailed remarks (i.e. there is a one-to-many relationship between log records and log comments).

The detail remarks by technicians is valuable information that can be leveraged in analysis. The technician is essentially a complex measuring sensor the has for understanding the status of the NAS. Technicians in the field can process and record multiple factors in the environment that are related to the maintenance activity. This complexity, however, comes with a cost of passing this information onto other applications for further use. Examples of this include understanding the nuances in local vernacular, misspelling of words, unstructured nature of text,and translating concepts into machine code. This paper doesn't try to solve all these issues, it takes an exploratory assessment to understand some of these issues in the RMLS.

2.4 Bureau of Labor Statistics Airline On Time Performance

RMLS information is correlated with the Bureau of Transportation Statistics (BTS) Airline On-Time Performance. This data contains the on-time arrival for major, non-stop domestic flights [8].

Some examples in this source are the arrival/departure times, airline, airport, and several others. The data of interested for this paper was delay incurred by the flight. BTS categories delay by weather, security, arrival, and departure delay. By incorporating the BTS delay dataset, effects of the log data can be labeled. It was assumed that log records and delay occurred in the same day. With the label identified, the RMLS text data can be measured against a separate data source.

3 Methodology

Data gathered needed to be transformed before any meaningful result could be found. The following subsections describe the process of manipulating the data in a suitable format. This includes a overview how the RMLS text data was parsed, what features were extracted, and how it was joined with the BTS delay data.

3.1 Parsing RMLS Text

As detailed in subsection 2.3, data from the log records is difficult to utilize. Manually entered in text data is non-structured by nature. Or there is no standard format. The data needs to be parsed into a useable format before any analysis can be done. A common model to use is the bag-of-words model to represent the text data. The bag-of-words is a set ,"bag", of all the words (or some feature) that appear in the text documents. Then for each document the count of occurrence (or some measure) of that feature is given. This paper used a single word as the feature to extract from the documents. A document in this paper was defined as the text associated with an airport's logs record for a given day. Thus, each log summary for the day and any log comments were joined together to form a single document for that airport.

To create the bag-of-words for the airports, words were extracted from the documents associated with the given airport. Identifying words that appear in the document is a challenge too. Since the text is unstructured, there isn't a common parser to pull out each unique word. The general procedure was to break down documents into sentences. Then for each sentence, decompose into words.

After words were identified, features were parsed to lower case and removing non-value words. Converting to lower case (a.k.a. case-folding) is a common heuristic to reduce the number of features into a more tractable model. An example is “Runway” and “runway” are two words that appear in a document. They generally are referring to the same object and thus should be considered the same. Next, a second parsing technique is applied to the words was removing non-value words. Non-value words are words that appear but generally have little content value. An example of a non-value word is a stop-word. Stop-words are “I”, “the”, “is”, and several others. This words are common support words in helping the reader structure sentences. They are removed since they do not actually provide information on the activity being performed. A third example of non-value words is punctuation characters. In a detailed log comment the technician can delimit their comment by creating visual sections with repeating several symbols. Once again, these “words” add little context to the actual maintenance activity being performed. With relevant features extracted from the documents, the bag-of-words for each airport is constructed. The set of all the features parsed from the documents comprised of the bag for each airport.

With each airport’s bag-of-words identified, the Term-Frequency Inverse Document-Frequency (TFIDF) was calculated on each airport-day’s document to measure the features. The TFIDF is another common NLP technique for weighting features not only on the number of times counted in the documents, but also the frequency they appear in documents. An example, of this could be when technicians are referring to a single asset but use different acronyms to identify it. A hypothetical situation is two technicians are assigned to fix runway 26. The first technician could refer to it as “runway 26” while the second refers to it as “rnwy 26”. Performing word counts will miss this important concept that runway 26 is having multiple maintenance activities being performed on it. With the TFIDF measure, “26” will be assigned a higher weight than before to account for this conceptual case. With each feature weighted, the average of the weights was calculated to represented the diversity of the language at the airport during that day. Higher the metric, the less diverse the language is. This proxy provides a simple and quick metric to measure each airport’s log day.

With text measure for each airport, it was joined with the BTS delay data to label our feature. The BTS data, however, needed to manipulated before being joined. The BTS data had individual flight data. Since the maintenance activity is performed at singular airport, the flights needed to be mapped to one airport. Flights were aggregated by the destination airport since arrival delay was assumed to be harder to mitigate against random maintenance. The BTS was aggregated by

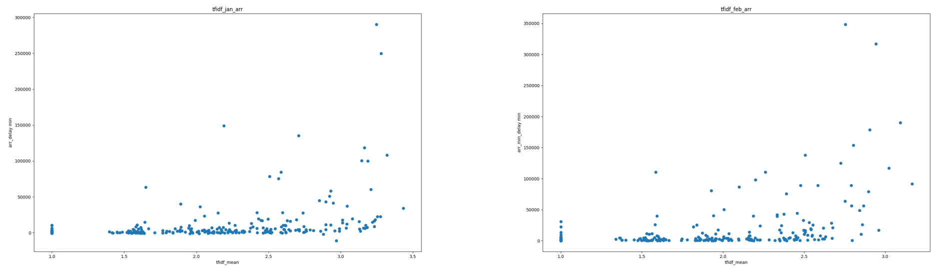

Figure 3: Log Days Average Feature Weight: January & February

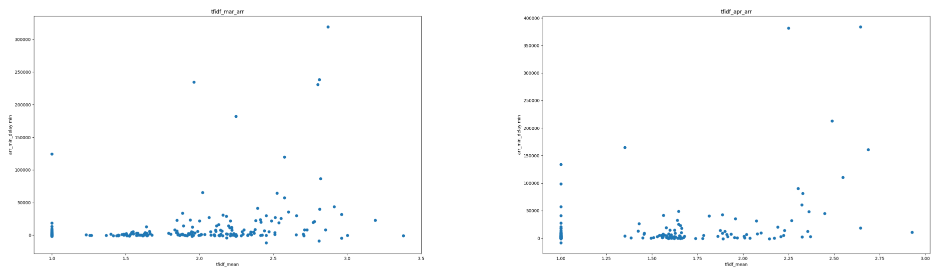

Figure 4: Log Days Average Feature Weight: May & April

airport and flight date. The arrival delay was summed to get the labels to correlate with the RMLS data.

The algorithm process above was implemented in Python using various third party libraries. The sentence and word tokenization was done with the popular NLP python nltk package. Removal of non-value characters such as stop-words and punctuation was also done with ntlk. The TFIDF measure was calculated with scipy.

4 Experiments

The methodology was applied to actual data for testing. Data was pulled from the RMLS and BTS systems for January 2017 to December 2017. The data was grouped by months to understand the effect of time. For each month the algorithm discussed above was ran to obtain the results.

The results shows a potential relationship between the diversity of language and delay incurred. Generally for each month, lower the average TFIDF correlates to lower delay experienced at the airport. Higher the average TFIDF, the airport would generally experience higher delay. There

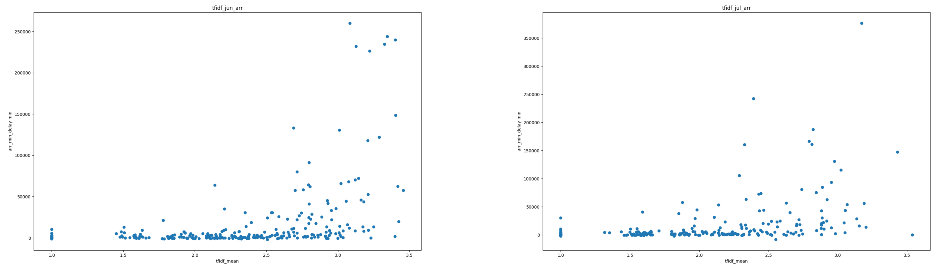

Figure 5: Log Days Average Feature Weight: June & July

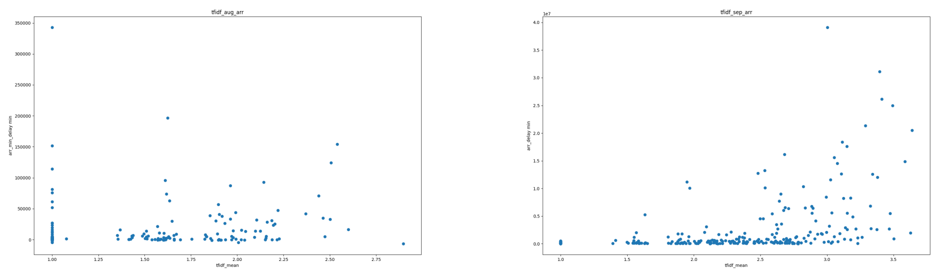

Figure 6: Log Days Average Feature Weight: September & August

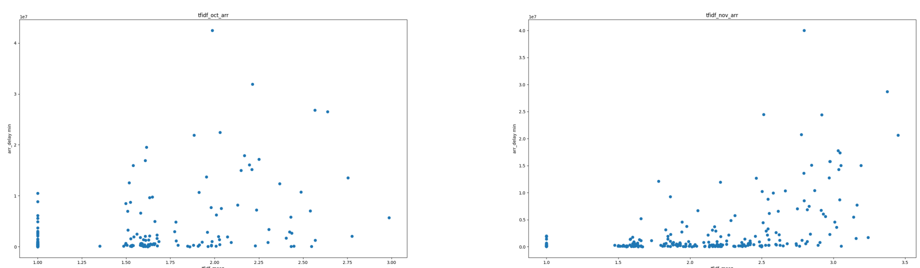

Figure 7: Log Days Average Feature Weight: October & November

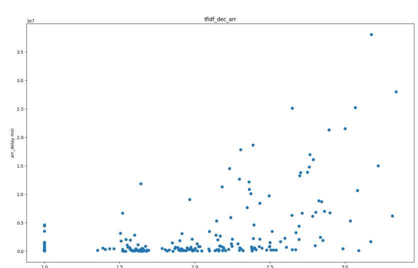

Figure 8: Log Days Average Feature Weight: December

also appears to be a acceleration of growth of delay as the TFIDF increases. These, however, could be signalling that certain words are delay drivers. A higher average TFIDF means the words, on average, appeared more frequently than lower average TFIDF. Another interesting result is the clustering of data points around certain average TFIDFs. For example, for a few exceptions, on the x-axis there seems to be a defined cluster of data points around the 1.5 mark, then a gap before the next clustering of data points. It is intriguing why there is a dense cluster at that average TFIDF mark and with virtually zero delay.

5 Conclusion

This paper attempted to perform simple NLP techniques to quantify diversity of log records in terms of delay. Parsing and joining RMLS and BTS delay data, a empirical relationship starts to take form. The results seem to generally agree on a higher average Airport-Day TFIDF measures correlates to a higher delay for a given day. This could indicate that repeated words are being used to describe logging events. Identifying those words would be interesting for the FAA. Understanding what words drives delay may give us understanding on what assets need to focused on when LCCEs or cost relationships.

The work from the paper also opens the possibility for several other areas to investigate. First, is the role of NLP techniques in analyzing RMLS data. In the FAA capital investments decisions are also based on the benefit they provide to public. One of the difficulties in estimating this benefit is finding a reliable monetizing function of the benefit. If this relationship can be empirically defined, those benefits can be captured and factored in the decision on what investments to prioritize. Second, the question of how do various facility's logging mean TFIDF compare to each other should be answered. This could help create a another comparison metric when costing by analogy. And a last suggestion for future investigation is understanding the ontology of the maintenance community. There are common conventions that are being used in the logging data across facilities. Understanding these terms could shed light onto cost drivers of maintenance work.

References

[1] Operations (AJW). Remote monitoring and logging system (rmls) maintenance handbook. Technical Report 17D, Federal Aviation Agency, 2016.

[2] Technical Operations Services (AJW-17). *Remote Monitoring and Logging System (RMLS) Simplified Automated Logging (SAL): User's Manual*, ti 6030.6b edition, 10 2017. Techincal Issuance.

[3] FAA. Birth of federal aviation agency. https://www.faa.gov/about/history/brief_history. Date Accessed: 2018-03-16.

[4] FAA. *NAS-MD-797E*, v1.11 edition, 11 2017. Remote Monitoring and Logging System (RMLS): RMLS System Design Document.

[5] FAA. National airspace system capital investment plan fy2018-2022. Technical report, Federal Aviation Administration, 2017.

[6] FAA. Acquisition management policy. Technical report, Federal Aviation Administration, 2018.

[7] FAA IPA. Cost analysis: Analogous cost estimating. https://www.ipa.faa.gov/Tasks.cfm?PageName=Analogous Cost Estimating. Date Accessed: 2018-03-19.

[8] Bureau of Transportation. Airline on-time performance. https://www.transtats.bts.gov/Tables.asp?DB_ID=120DB_Name=Airline Date Accessed: 2018-03-25.

[9] Phaltane K. Ahmad S. Ranjan N., Mundad K. A survey on techniques in nlp. *International Journal of Computer Applications*, 132:6–9, Janurary 2016.