INSTITUTE FOR DEFENSE ANALYSES

A Feasibility Study on the Use of Artificial Intelligence for Defense Acquisition Program Review, Volume I: Main Report

Brian Q. Rieksts, Project Leader
Kristen M. Guerrera

July 2020
Approved for public release;
distribution is unlimited.
IDA Paper P-13239
Log: H 20-000222

INSTITUTE FOR DEFENSE ANALYSES
4850 Mark Center Drive
Alexandria, Virginia 22311-1882

About This Publication

This work was conducted by the IDA Systems and Analyses Center under contract HQ0034-14-D-0001, Project AB-7-4624, "Artificial Intelligence (AI) Applied to Program Data," for the Office of the Secretary of Defense for Acquisition (ASD(A)). The views, opinions, and findings should not be construed as representing the official position of either the Department of Defense or the sponsoring organization.

Acknowledgments

Thank you to Patricia F. Bronson, Matthew S. Goldberg, and Arun S. Maiya for performing a technical review of this document.

For More Information

Brian Q. Rieksts, Project Leader
brieksts@ida.org, (703) 575-1404

David E. Hunter, Director, Cost Analysis and Research Division (CARD)
dhunter@ida.org, (703) 575-4686

Copyright Notice

INSTITUTE FOR DEFENSE ANALYSES

IDA Paper P-13239

A Feasibility Study on the Use of Artificial Intelligence for Defense Acquisition Program Review, Volume I: Main Report

Brian Q. Rieksts, Project Leader
Kristen M. Guerrera

Executive Summary

The Office of the Assistant Secretary of Defense for Acquisition (ASD(A)) tasked the Institute for Defense Analyses (IDA) to explore the feasibility of applying artificial intelligence (AI) and related tools to structured and unstructured data sources. The goal is to streamline processes, enhance predictive power, and better and more efficiently inform senior decision-makers about trends and possible outcomes in acquisition.

Current Department of Defense (DoD) policies and statutes require Major Defense Acquisition Programs (MDAPs) to report cost, schedule, performance, and financial information so that DoD can better monitor performance. The assessment process has been criticized as being too slow and lacking predictive value. However, recent advances in AI have led to improved efficiency in a broad range of applications. These efficiencies arose from both improving predictions of outcomes and assisting people in their analysis of large volumes of information. In this study, IDA explores the feasibility of applying emerging AI techniques to acquisition program assessments.

Approach

Our study focuses on exploring how AI can help acquisition analysts be more effective and efficient, as a complex process like acquisition may be well suited for human-machine teaming. We take a broad approach in this project and use examples to demonstrate which areas might be most promising for further investigation. To accomplish our goal, we interview staff from ASD(A), survey relevant examples of AI in the private sector, and identify potential applications of text analytics in the acquisition process.

Options

We identify several options, outlined in the following list, where AI and text analytics can help analysts assess acquisition programs. We note that while no single person can review all available information about all acquisition programs, a machine can consume large volumes of data in a short period of time. Our research explores enabling analysts to leverage the capabilities of a machine. We also examine unconventional data sources, which program reviewers may not have the time, ability, or access to examine, in order to exploit these data for insights. We examine some repetitive, manual processes that have the potential to be streamlined through the use of AI.

- The most promising option we identify is to use text analytics to prioritize information for analyst review. For example, we explore 367 documents from

Congressional testimony that mention “F-35,” amounting to more than 75,000 pages. We use sentiment analysis and topic models to identify pages that discuss problems with the testing and cost of the F-35 aircraft. This approach could be adapted to many data sources.

- Another promising option is examining patterns of issues across functional areas and time. We consider an example from the Defense Acquisition Executive Summary (DAES) narrative assessments for the WIN-T Increment III program.
- We explore estimating outcomes of programs directly with structured and unstructured data. We demonstrate this concept by predicting whether a program ever had a Nunn-McCurdy breach or was cancelled. We find the sentiment of narratives to have more relative influence in the model than the green, yellow, and red ratings.
- We examine the trends in sentiment and frequency of tweets for several defense acquisition programs. Awareness of the recent public conversation about defense acquisition could aid leadership during an engagement with external stakeholders
- We examine the possibility of using program office emails as an unconventional data source for assessing acquisition programs. We find examples where industries have examined employee emails to gain insights about their companies.
- We explore using AI on unstructured data. We find examples of companies in private industry using text analytics to extract metadata from unstructured data. The U.S. Air Force also sponsored a feasibility study to extract metadata from contracts.
- We consider how identifying differences in documents could be useful both in comparing data sources and comparing the same data source across time. We examine Congressional testimony and DAES assessments for the KC-46A aircraft to identify common and dissimilar information between these data sources.
- We discuss using artificial intelligence to estimate whether a document is ready for approval or whether it needs further work. This AI application is used for legal review in the private sector.

Observations

Our assessments over the course of this study led us to the following observations on the value of AI in assessing acquisition programs:

- Acquisition experts are necessary to work with an AI tool given the complexity and diversity of acquisition programs.
- The number of outcomes on DoD acquisition programs is limited, but a large volume of unstructured data exists on each program.
- Significant effort would be required to transition to a production tool given the barriers to sharing data, approvals required for software installation on DoD computers, and processes required to automate real-time data feeds for a tool.
- Data required by a contract, resulting from an audit or other mechanism that encourages high data quality, could provide good input for text analytics.
- An ideal tool would help an analyst identify and explain an issue so that decision-makers could further explore the issue before deciding on appropriate action.

Path Forward

The next step in applying AI to acquisition data is to demonstrate for a particular use case the improved efficiency of using AI tools. We could examine the results with and without text analytics tools and note any improvement—either in the time required to perform an assessment or in the quality the assessment itself. The text analytics algorithms could also be customized for identifying acquisition-specific information.

One specific use case is identifying challenges in a program that require management attention. These challenges could include the supply of labor, such as welders; weight growth in an aircraft; or issues like those with the KC-46A boom. We can analyze a particular use case to determine:

- Whether analysts who use text analytics to prioritize documents can review information more quickly
- Whether analysts who use text analytics to prioritize documents achieve improved results compared to analysts who perform reviews manually
- Whether analysts tracking a program over time can save time by using text analytics to focus reviews only on those documents that have changed since the last month
- Whether issues that ultimately required management attention could have been identified sooner

Contents

1. Task Description

Current Department of Defense (DoD) policies and statutes require Major Defense Acquisition Programs (MDAPs) to report information, including cost, schedule, performance, and financial data, to help DoD monitor program performance. In addition, DoD collects various unstructured program data in the form of charts, spreadsheets, and narrative assessments across various dimensions and echelons. These unwieldy data and information, which are stored in various databases, require many man-hours to retrieve and evaluate for any one program. The assessment process has been criticized as being too slow and lacking predictive value.

Recent advances in artificial intelligence (AI) have led to improved efficiency in a broad range of applications, through both improving predictions of outcomes directly and assisting people in their analysis of large volumes of information. For example, AI techniques have been used in private industry to assess performance, increase efficiencies, and reduce costs. Currently, an opportunity exists to apply these emerging AI techniques to DoD's methods for assessing acquisition programs.

The Office of the Assistant Secretary of Defense for Acquisition (ASD(A)) tasked the Institute for Defense Analyses (IDA) to explore the feasibility of applying AI and related tools to structured and unstructured data sources to streamline processes, enhance predictive power, and better and more efficiently inform senior decision-makers about trends and possible outcomes. We examine the myriad data sources available as part of the defense acquisition process as well as non-traditional data related to acquisition, such as Congressional testimony and social media. We also explore the potential use of various advanced analytic techniques and, where possible, apply them to the available data sources to prototype possible applications. This feasibility study offers an exploratory view of a broad range of applications and provides initial examples to demonstrate how the AI techniques might be implemented. Providing well-refined models was out of scope and not a part of this feasibility study.

This paper first outlines our approach, including the methodology for identifying potential applications of AI to acquisition program review. Next, we review each of the potential AI applications in greater detail, including a description of the idea, a discussion of previous similar applications, a review of the results from any initial analyses, and possible next steps or additional information needed to advance the idea to realize its full potential for defense acquisition.

2. Approach

This chapter discusses the project approach, including the methodology applied, data sources used, and relevant acquisition background for context in understanding the various applications.

A. Methodology

First, before considering options for using AI to assist with program review, we interviewed staff from ASD(A) to understand their individual roles in the process. Staff members described their responsibilities and how they currently perform their jobs. The goal of these discussions was to determine whether some aspects of their jobs could be streamlined with AI and to identify the information they would like to analyze if given more time. We also met with staff who have roles in acquisition data management as well as those using analytics within the military services.

Second, we examined AI activities in the private sector where unstructured data has been analyzed for related purposes. The goal of this literature review was to identify concepts used to solve problems in the private sector that could be transferred to the DoD review process for acquisition programs.

Finally, we identified potential options for using AI to assist with the acquisition program review process. These options, which focused on text analytics, included further exploring data sources in a systematic way, streamlining processes, and analyzing large amounts of unstructured data. In addition, we developed prototypes of the more promising options.

For this study, we explored using the following AI analytical tools:

Topic model. An algorithm that *scans* a set of documents, *examines* how words and phrases co-occur in them, and automatically *learns* groups or clusters of words that best characterize those documents. For example, in a newspaper, words like "stock" and "sales" may be grouped into a single topic by appearing together in the business section, while "baseball" and "pitching" could be another topic from the sports section. The latent Dirichlet allocation (LDA) model is one algorithm to fit a topic model. An output of the LDA model is an estimate of the probabilities that a document is associated with each topic. One way to visualize a topic is through a word cloud. A word cloud is a graphical representation of words from documents with the font size or color scaled to the prevalence of the words within the documents.

Sentiment analysis. A text analytics technique to interpret, classify, and quantify emotions (positive, negative, and neutral) within text data. A variety of natural language processing (NLP) algorithms have been developed to estimate sentiment by accounting for negators, amplifiers, and other aspects of language.

Supervised machine learning. An AI tool that helps estimate an output from a set of input variables based on learning from a training set of input-output pairs. Many supervised machine-learning algorithms exist. In particular, our study uses gradient boosting, which is a tree-based machine-learning algorithm that performs well on tabular data. It is a prediction model in the form of an ensemble of weak prediction models.

B. Acquisition Background

To understand the current acquisition process, we engaged with OUSD(A&S) staff who perform Defense Acquisition Executive Summary (DAES) assessments, analyze middle-tier acquisition programs, develop scorecards for acquisition portfolios, manage Earned Value (EV) data, work on legislative and Congressional issues, produce Selected Acquisition Reports (SARs), and liaise with the Government Accountability Office (GAO) for their weapons systems report.

Within the Acquisition Analytics and Policy (AAP) office, several analysts are responsible for producing the DAES assessments for the MDAPs. The MDAPs are partitioned into groups, and each analyst is responsible for several MDAPs. Each MDAP is assessed every 3 months, with assessments staggered so that about one-third of the MDAPs are assessed each month. Analysts are scheduled to examine information for the programs that they are assessing. This information includes Program Assessment Reports (PARs); Earned Value Management-Central Repository (EVM-CR) data, including cost and schedule assessments as well as EV Format 5 data; information from Defense Acquisition Management Information Retrieval (DAMIR), GAO, and other Congressional reports; and information from the program offices. These analysts also produce assessment scorecards. For more information on the DAES process, please consult the following sources:

- DoD document from December 2012, “Defense Acquisition Executive Summary Assessment Guidance”[1]
- RAND report from September 2017, “Issues with Access to Acquisition Data and Information in the United States Department of Defense”[2]

---

[1] Office of the Under Secretary of Defense for Acquisition, Technology, and Logistics, “Defense Acquisition Executive Summary Assessment Guidance,” December 13, 2012.

[2] Megan McKernan et al., “Issues with Access to Acquisition Data and Information in the Department of Defense,” 2017, https://www.rand.org/pubs/research_reports/RR1534.html, accessed June 22, 2020.

- Gwozdz and Kodzwa, 2019, “The DAES Process”[3]

Some analysts in the Acquisition Approaches and Management (AAM) office are working to develop the new middle-tier acquisition (MTA) process. The MTA process consists of both rapid fielding and rapid prototyping programs. As of fall 2019, OUSD(A&S) was still developing rapid fielding processes, and OUSD(R&E) was developing rapid prototyping processes. OUSD(A&S) does not plan to collect much text data, but instead will collect mostly structured numerical or categorical data. Because the MTA process is still very new we did not explore these options as prototypes for AI applications.

Another office with potential interest in applied AI for acquisition is the Legislative and Congressional Oversight (LCO) group within the Office of the Chief of Staff, OUSD(A&S). They were interested in mining public information, particularly Twitter feeds, to stay informed about the issues of the day for particular members of Congress, and possibly the general public. This information could be especially useful when preparing for meetings between senior DoD leadership and members of Congress. The LCO also thought it might be useful to examine trends over time for particular acquisition programs.

---

[3] Lawrence T. Gwozdz and Paul M. Kodzwa, “The DAES Process,” *Defense Acquisition Magazine Blog*, January 30, 2019, https://www.dau.edu/library/defense-atl/blog/The-DAES-Process-, accessed May 22, 2020.

3. Potential Applications

In this feasibility study, we identified and explored the following potential applications of AI to acquisition program review:

- Finding useful documents for a human to review
- Identifying patterns across functional areas
- Predicting program outcomes
- Gaining insights from social media and news articles
- Examining program office email data
- Extracting metadata from text documents
- Analyzing differences among documents
- Estimating whether a document is ready for approval

These potential applications arose from key observations about how a computer could be used to assist an acquisition analyst. We note, for example, that while a single person is likely unable to consume all information about a program, a machine can process a large set of information. Thus, a machine could be used to systematically analyze existing and unconventional data sources, such as emails, news articles, and social media. Additionally, some currently time-consuming, repetitive, manual processes could benefit from computer automation through the use of AI.

A. Find Useful Documents to Review

1. Description

In our discussions with staff involved in the acquisition process, they indicated that to be most useful an AI tool needs to produce explainable results. Human-machine teaming is one approach to achieve this objective. Teaming allows text analytics to identify useful information for analysts to review. The analysts can then explain the insights they derive from this information. Although the algorithms may at times incorrectly estimate the usefulness of some information, this approach is more efficient than selecting a document at random, particularly if the source is something the analysts would not have time to consider otherwise.

Sentiment analysis could be used to prioritize documents to be examined for potential acquisition issues. Documents with more negative sentiment are more likely to identify acquisition problems.

Topic modeling is another AI tool to analyze documents. Topic modeling allows analysts to identify and focus on information related to specific topics. Combining this approach with sentiment analysis can identify documents describing issues for a particular topic.

Supervised machine learning can also be used to find documents of interest. In this approach, each program is categorized according to outcome (e.g., canceled yes/no, Nunn-McCurdy breach yes/no). Documents for each program are partitioned into smaller segments of text, and each segment is labeled according to the outcome of its program. A supervised machine-learning model could be fit to these data to estimate the probability that a segment of text is associated with a program having a negative outcome. These estimates could be used as a queuing tool for a large set of information, allowing analysts to first review the text segments most likely to be associated with a program having an issue. An alternative for modeling an overall outcome for a program is to label individual pieces of information as "useful" or "not useful" and then use supervised machine learning to predict whether information is useful. This method requires effort to create a dataset of labeled information.

Transfer learning, unlike supervised learning, is one approach to fit models with smaller samples of labeled information. Transfer learning leverages knowledge gained from a related problem. That is, transfer learning is a machine-learning technique to transfer knowledge from one problem with a labeled training set to a different but related problem.

2. Previous Research

We reviewed and examined relevant applications in both the commercial world and the military.

Commercial Applications

Companies like Daegis and TERIS use AI and machine learning to facilitate the process of document review for legal analysis. They use computers to learn which documents might be most useful to the reviewers, eliminating the need for a human to read a very large corpus of documents.

A company called Datascope Analytics was used by at least one private company to identify internal subject matter experts using social network analysis, research reports, and internal directories. The algorithm recommended potential collaborators based on each expert's set of skills and previous research and experience. The approach successfully matched researchers for collaboration.

Investment firms, like Blackrock, use AI to assist in market research and determine which industries or specific companies might be good investments. They take the language of public conference calls and perform NLP analysis to uncover top insights. For example, they might identify a company for potential investments that numerous executives are discussing in the same period. They can also use the topics from various documents to identify firms that are doing well or poorly.

Military Applications

The Air Force contracted with Decisive Analytics Corporation to analyze the cost, schedule, and performance risks of their programs as well as to provide an early warning of issues. The vendor demonstrated an example of identifying an issue by assessing the sentiment of various sentences from program documentation.

Maiya et al. (2013) provides a case study on searching for information pertaining to a critical military technology within a document collection. In the case study, 25 of almost 30,000 documents pertained to this technology. A machine-learning classifier correctly identified 18 of the 25 documents within seconds, while text analytics enabled correct identification of the remaining 7 documents within 30 minutes. In comparison, analysts spent 14 hours identifying the documents manually, and in addition to identifying the 25 documents, incorrectly identified 5 additional documents that were not related to the technology.

3. Examples

Some OUSD(A&S) staff indicated an interest in reviewing Congressional testimony about acquisition programs. However, Congressional testimony covers a broad range of topics, and it would be onerous to read all testimony to find just the portions regarding the acquisition program of interest. To address this issue, we created a prototype model using open source tools. We then used the F-35 as a case study to identify useful information from Congressional testimony. We downloaded 1,295 documents on Congressional testimony from the House Armed Services Committee and Senate Armed Services Committee websites.[4] Next, we filtered these documents only to those that mention the F-35, reducing the number of documents to 367 with about 75,000 pages in total. Figure 1 shows the trend of the number of mentions of the F-35. The number of mentions increased

---

[4] These data were downloaded from https://www.govinfo.gov/app/collection/chrg in August 2019. Documents were collected with the earliest document in the 106th Congress that convened on January 3, 1999. The website indicates that most Congressional hearings are published 2 months to 2 years after they are held, so not all data were available at the time of the download. We also observed fewer documents in the earlier years. In examining trends, we considered data only from the 110th Congress that convened on January 3, 2007, through the 114th Congress that ended on January 3, 2017. Individual committee websites may have more up-to-date information that could be useful in future analyses.

in the 111th Congress and the 112th Congress, although the program did have a Nunn-McCurdy breach during the 111th Congress. An analyst could further explore testimony to understand changes over time in the number of mentions of a program.

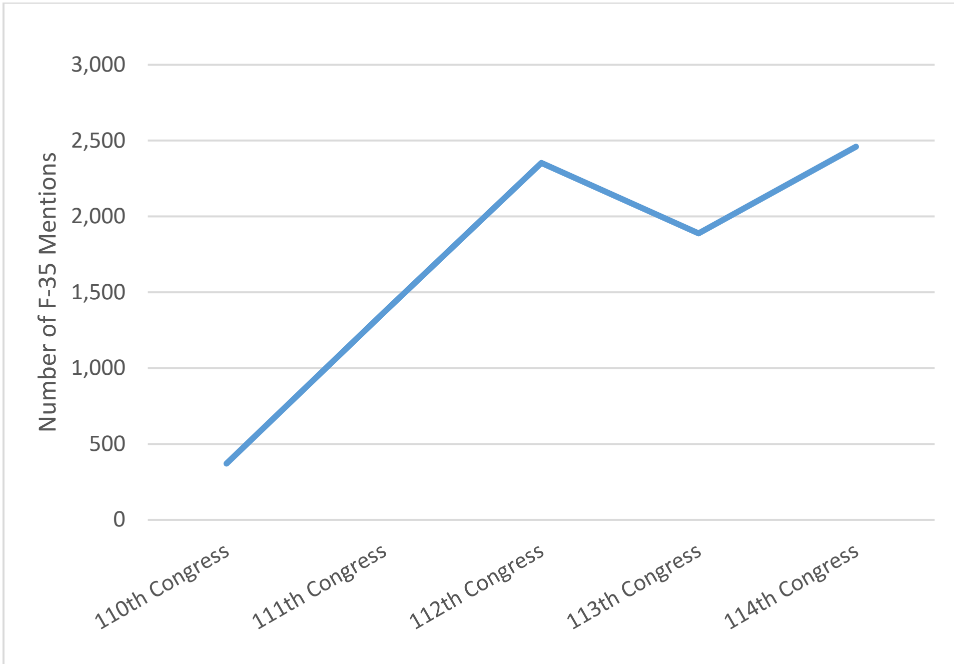

Figure 1. Number of F-35 Mentions across Congresses

We filtered almost 75,000 pages from documents mentioning F-35 to identify only those pages mentioning F-35, resulting in about 3,600 pages. To better understand the content of these pages, we fit a topic model with 9 topics using the LDA model. Figure 2 shows the word clouds corresponding to these topics.

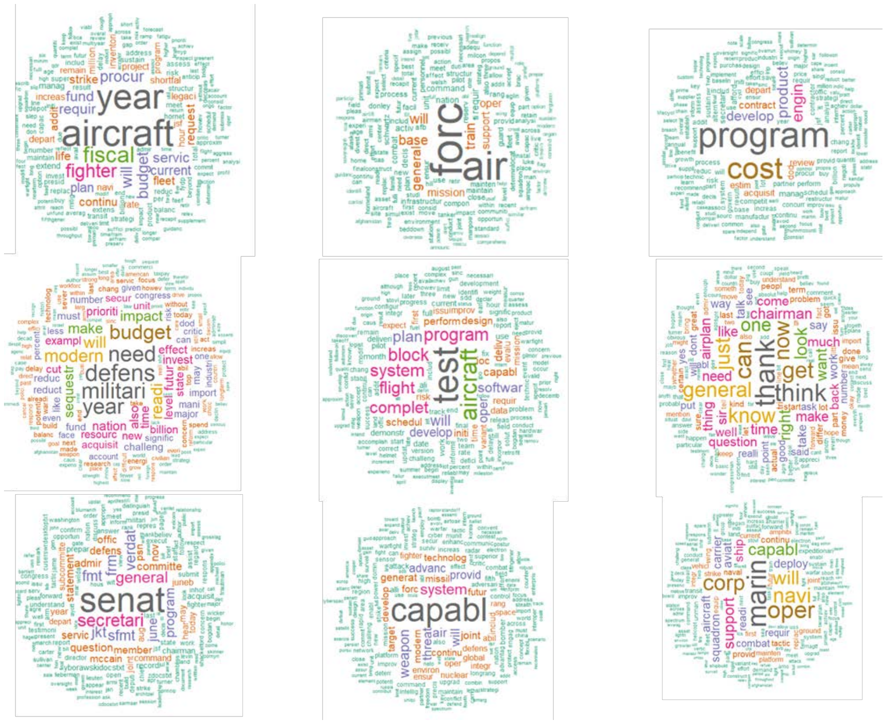

Figure 2. Word Cloud for Topics in Congressional Testimony Mentioning F-35

We can associate pages that mention F-35 to the topics in Figure 2. With this association of pages and topics, we can further explore the documents associated with a topic of interest. The topic in the upper right corner of Figure 2 corresponds to costs. In the center, we have a topic about testing. Nearly 400 pages of testimony mention F-35 and are associated with the topic about testing.

We sorted these pages about testing according to sentiment. When we consider the five pages with the most negative sentiment, each of them discusses program risks or problems. The following example describes an incident where the F-35 engine caught on fire during a training event in June 2014. We have underlined some of the words identified as being associated with negative sentiment in the sentiment algorithm's dictionary.

Mr. TURNER: What concerns do you have regarding the June 2014 engine failure?

Dr. GILMORE: The program has done a good job of explaining what happened, but we do not fully understand why this happened. We do not understand what occurred differently than expected during the relatively restricted flight maneuvers on the flight prior to the engine failure that set up the failure event. The effect on engine performance and reliability with the modified stators is not well understood. We understand the engine contractor has acknowledged limitations in their models, particularly associated with axial movement within the engine, and have updated those models as a result of the engine failure. It is not clear what exactly the models did not correctly predict in the original design and use of the aircraft that turned out to be incorrect, or how the modeling was improved. We should *also determine through additional analysis if the containment of the damage is sufficient, given the nature of the failure. In this case, the pilot was able to stop the aircraft and safely get away from the burning airframe. However, inflight failures of this kind should be examined to determine if the uncontained damage is tolerable from an aircraft vulnerability perspective.*[5]

Conversely, if we consider the five pages with the most positive sentiment associated with the testing topic, each page has an optimistic outlook. The following example describes improvements in availability. We underline several words associated with positive sentiment in the sentiment algorithm's dictionary.

JOINT PREPARED STATEMENT BY THE HONORABLE FRANK KENDALL AND LT. GEN. CHRISTOPHER C. BOGDAN: … Aircraft availability rates continue to be a focus area for the program and in this area various program initiatives are now showing a positive trend. A disciplined Reliability & Maintainability program, improved maintenance procedures and manuals, continued improvement in the Autonomic Logistics Information System (ALIS), better forecasting of spares requirements, improved repair turnaround times from suppliers, and incorporation of aircraft design improvements have resulted in gains in mission capability rates and aircraft availability rates. Today, across the fleet, we are seeing 55 to 60 percent availability rates with units performing at 63 percent.[6]

Similar results are obtained when analyzing the cost topic. Although the cost example does have a page that, while correctly categorized as a cost issue with negative sentiment, is not about the F-35. The discussion is about V-22 cost issues and mentions F-35 only at the end of the page as the topic shifts. Over time the models could be refined to improve the accuracy of finding desired topics, but analysts are still necessary to evaluate whether information is useful. Overall, tools like this can help analysts be more efficient in finding the information they need.

4. Next Steps

We assess finding useful information to review as the most promising option in our study. This option leverages the expertise of acquisition analysts, who would still need to analyze and explain the issues. Text analytics identifies likely useful information and could improve analyst productivity.

The next step to implement this option is to engage with people working in the acquisition program assessment process and have them use the output from a prototype

5 House Hearing, 114th Congress - [H.A.S.C. No. 114-35], "Update on the F-35 Joint Strike Fighter Program and the Fiscal Year 2016 Budget Request," October 14, 2015.

6 Senate Hearing, 114th Congress - [S. Hrg. 114-658] Part 1, "Department of Defense Authorization for Appropriations for Fiscal Year 2017 and the Future Years Defense Program," February–April 2016.

that finds the most useful information. Acquisition experts need to be involved in an iterative process to determine whether the information is useful and how to make the algorithms more beneficial.

In addition, we should precisely define the set of documents that are analyzed by the prototype model. The initial model should be simple. An iterative process with acquisition experts could be used to make incremental improvements and customize the tool to acquisition.

Two potential data sources would be the Performance Assessment Report (PAR) and the EV Format 5 data. The PAR contains narrative information from the Defense Contract Management Agency (DCMA) and stoplight ratings. Preliminary examination suggests these types of data could be useful in finding problems across programs.

The EV Format 5 report contains narrative descriptions from the contractor about variances that are required as part of the contract. Personnel from OUSD(A&S) indicated that these data are used to manage individual programs. Personnel also suggested that a systematic analysis of these data could produce useful insights. Freeman (2013) uses EV Format 5 data to predict estimated cost at completion and provides an overview of previous text analytic research with Format 5 data. In our preliminary exploration of the data, we observed that EV Format 5 data are not as consistently organized as PARs. Therefore, a systematic analysis across programs would require more effort to collect and prepare EV Format 5 data.

B. Identify Correlations across Data in Different Functional Areas

1. Description

Acquisition program review involves several functional areas such as cost, schedule, testing, and performance. A subset of analysts in DoD might look at the cost of a program, while the engineers might understand the technical functionality of the program, while the evaluators examine test results. Problems in one functional area might begin to spill over into another functional area and still not be widely understood until someone with a view across all functional areas can identify the problem. Text analytics may provide insights by simultaneously exploring unstructured information across functional areas, thereby finding potential issues that no single person could otherwise identify.

One approach to analyzing these data is to use topic modeling algorithms to identify topics across functional areas and over time. From a historical standpoint, these topics may show patterns of how major issues develop within a program. For example, a topic may appear in one functional area and then occur later in other functional areas. Another potential use of these topic models is to identify the topics within a program that are spanning functional areas at a particular time.

2. Previous Research

One example of previous research using topic models in this way is in the medical field. A company called Tempus uses AI to extract data and insights from multiple types of medical reports in order to look holistically at a single patient for clinical decision support. For example, perhaps a particular patient has seen several doctors over time and has undergone several tests (imaging, biopsies, etc.) for each doctor. However, the doctors cannot view the other test results. A key insight, conclusion, or diagnosis might be evident if a single person had synthesized all the test results. Tempus can take data that might traditionally be in various data silos and bring the data together to provide increased insights. They use optical character recognition and then apply NLP to tease out insights.

3. Examples

We consider the text assessments in the DAES process for a prototype analysis to identify issues across functional areas. The DAES assessments are consistently provided over time for MDAPs by functional area, thus affording us a single data source. Although these data are convenient for a prototype analysis, a more comprehensive model could consider additional data sources. For our analysis, we fit a topic model for DAES assessments over time and across functional areas for a program. For a topic corresponding to a problem within a program, we illustrate the pattern of the issue appearing in DAES assessments. The details of this example are shown in Appendix A, which is marked "For Official Use Only" and issued separately.

4. Next Steps

In practice, identifying correlations across data in different functional areas requires a decision of either using this approach as a forensic tool to better understand patterns of issues, or using it to identify issues as they start to span multiple functional areas. After choosing the objective, the next step is to identify more patterns in issues across functional areas. This identification might be aided by expanding the data sources from just the DAES to include data sources with more details on issues within programs. We provide one example of this in Appendix A, although a challenge in the prototype analysis was finding topics that described a specific issue instead of more generic topics that were not easily mapped to an issue. Models with further refined stop words could help mitigate this issue. Another potential challenge is that the language about a single issue may change over time. Developing this option may require more work and is higher risk than other options, but analysis across functional areas and with multiple data sources could provide valuable new insights.

C. Predict Program Outcomes

1. Description

Another application is to use machine learning to predict program outcomes from structured and unstructured data. A major challenge of this approach is defining program outcomes. Several characteristics about a program can be measured, but program outcomes can be difficult to define, especially during execution. Programs that ultimately provide useful capabilities to the warfighter may encounter problems during their acquisition. A long-term outcome may be whether a program is terminated before production. Cost growth and schedule slip are other examples of quantifiable program outcomes. Associating outcomes with particular periods within a program can also be difficult. For example, a program in development may not acknowledge any cost growth at a particular time, but issues that ultimately led to cost growth could be discussed in program documentation created during that time.

The limited number of programs available for predictive modeling also presents a challenge to predicting outcomes. At any time, DoD and the services have roughly 100 MDAPs that each span several years. We do have large volumes of information about each program, however, so one approach to predicting outcomes is to use different assessments or other documents for a single program as multiple data points. Predictive modeling could also predict problems that arise during the acquisition of a program, but a dataset with these issues labeled would need to be created before the model could be developed.

Another approach to modeling outcomes is to develop a two-stage model to predict outcomes from text data in assessments. The first stage is to estimate sentiment from assessments. The second stage is to predict outcomes directly from sentiment.

Another potential approach is to directly model outcomes using supervised machine learning. There are many ways to implement this approach. Unstructured text data could be used directly to predict an outcome with methods such as gradient boosting or deep learning. These unstructured data could also be supplemented with structured numerical data like earned value data. Topic models could also be fit to unstructured data, which could be used as features in a model. For example, Miller (2012) used topic modeling to create features from EV Format 5 data to predict the future estimated cost at completion to be used in statistical process control.

A prediction of an outcome alone may not be useful to decision-makers if there is no explanation about why an outcome is predicted by the model. However, subject matter experts could examine the input data that the model uses to evaluate whether an issue warrants further attention. This approach is similar to identifying useful information but adds an attempt to predict an outcome.

2. Previous Research

A problem related to predicting program outcomes for MDAPs is predicting the outcome of investments in the private sector. The use of machine learning is significant and growing for companies that analyze the stock market, including forecasting stock prices.[7] Although financial institutions do not disclose the details of their algorithms as they seek to gain a competitive advantage, several academic papers have explored using machine learning and sentiment analysis to analyze the stock market. Shah, Isah, and Zulkernine (2019) present a review and taxonomy of the different analytical approaches used in the literature to predict stock prices.

OUSD(A&S) is also developing predictive models. In particular, models have been developed to predict from the text the DAES stoplight assessment of red, yellow, or green. A goal of these predictions is to understand where the model forecasts an assessment that differs from the actual rating. For example, the model predicts a red assessment, but the rating is yellow. These differences could be further investigated. Hassler and Clark (2020) from Rotunda Solutions developed a neural network model to predict these ratings. Also, Kloke et al. (2019) from Ayasdi predicted these ratings using a topological modeling approach.

3. Examples

We used DAES assessments to create a prototype for predicting outcomes based on whether a program ever had a Nunn-McCurdy breach or cancellation. We labeled each assessment by an office in a functional area as 1 if at any point in the life of the program it had a Nunn-McCurdy breach or was cancelled. Conversely, we labeled these assessments as 0 if the program never had a Nunn-McCurdy breach and was not prematurely cancelled.

The dataset for this prototype contained 125 complete, canceled, and active programs with DAES assessments. We randomly partitioned these programs into 2 datasets, one with 63 programs in a training set and the other with 62 programs in a test set, for a combined total of almost 80,000 assessments. About 46 percent of these assessments corresponded to programs that at some point had a Nunn-McCurdy breach or a cancellation.

We used a machine-learning algorithm called gradient boosting with the open source XGBOOST package to make predictions about the outcome that a program had a Nunn-McCurdy breach or a cancellation. For each of the narrative assessments, we estimated sentiment offline with multiple methods. Each assessment also had stoplight ratings of red, yellow, or green as outlined in the DAES assessments. We used both estimates of sentiment and the stoplight ratings as input features to predict the outcome. The relative influence of

---

7 M. Thomas, "How AI Trading Technology is Making Stock Market Investors Smarter," April 10, 2020, https://builtin.com/artificial-intelligence/ai-trading-stock-market-tech, accessed June 24, 2020.

these variables in the gradient boosting model to predict the outcome is shown in Figure 3. The model included both sentiment of narrative assessments and stoplight ratings. However, we can see that much of the relative influence, or predictive value, is from the sentiment estimates of the narrative assessments. This result suggests that the narrative assessments contained useful information about this outcome beyond that in the stoplight ratings.

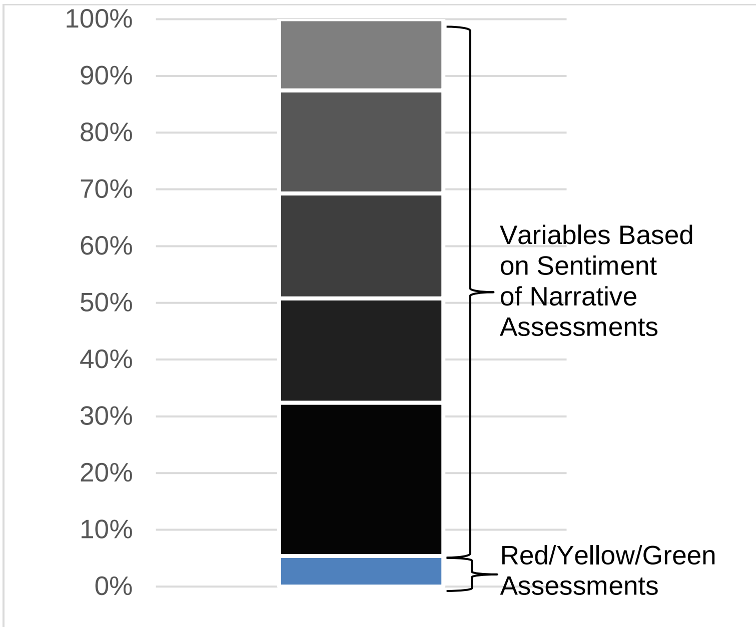

Figure 3. Relative Influence of Variables in Gradient Boosting Model for Outcomes

On average, we would expect more issues with programs that at some point had a Nunn-McCurdy reach or cancellation. However, we do not expect this model always to distinguish the overall program outcome from individual narrative assessments and stoplight ratings. These observations reflect a particular point in time for a particular functional area. The reason a program had a Nunn-McCurdy breach or cancellation may have nothing to do with a particular assessment. Similarly, programs that never had a Nunn-McCurdy breach or cancellation still experienced problems. If we consider the extremes, of the 5,000 assessments in the test set that the model predicted were most likely to have a Nunn-McCurdy breach or cancellation, 65 percent of them did have such an outcome. Of the 5,000 assessments in the test set that the model predicted were least likely to have this outcome, 46 percent had the outcome. That is, the 5,000 assessments that the model predicted would most likely have a Nunn-McCurdy breach or cancellation are about 40 percent more likely to have it than the 5,000 assessments that the model predicted were least likely to have this outcome. If we expand the test set to the approximately 40,000 observations, the performance of the model overall is weak with an area under the curve

(AUC) for the test set of 0.55.[8] This example demonstrates the type of predictive model that could be developed. The accuracy of predictions could be improved with other types of data and different modeling approaches.

4. Next Steps

Choosing the appropriate outcome is the first key to successfully using estimates of outcomes to assist in program assessment. Some outcomes, like problems that need management attention, may require people to create labeled training and test data. Other outcomes, like cost growth, can be directly measured. After identifying the appropriate outcome to study, the next step is to develop a predictive model for this outcome. This effort may require a model customized to the language used in acquisition. For example, the text from narratives could be used to create learned word embeddings. That is, the text could be transformed into a vector of continuous numbers through an intermediate supervised learning task. Similarly, categorical variables could be converted into learned entity embeddings. A predictive model for the overall outcome could include both structured data and learned embeddings. Analysts who develop the predictive models would need to work closely with subject matter experts to understand how to adjust the model to make output more useful and explainable.

D. Gain Insights from News Articles and Social Media

1. Description

One type of data source that could be further exploited is public information in news articles or on social media. We discussed the value of these data sources with government staff who assess acquisition programs. Some staff indicated that this public information is a lagging indicator compared to internal data, but they did describe how this type of information is useful when leadership in OUSD(A&S) engages with external stakeholders such as the Congress. OUSD(A&S) staff suggested that it would be useful to see the trends in popularity and sentiment of news articles or social media about acquisition.

2. Previous Research

This application of AI amounts to monitoring defense programs using social media, news stories, or other internet traffic. This approach is very similar to how the private sector handles brand monitoring, which refers to a company having an interest in ensuring an online presence that aligns with their brand, generally to maintain a positive perception. They may hire a company to monitor and track any brand-related developments so that they can make the most of positive mentions and, perhaps more importantly, respond

---

8 AUC = 0.5 for a random prediction and AUC = 1 for a perfect prediction.

quickly to problems. Companies like Critical Mention and Synthesio monitor social media metrics and can show trends in sentiment across relevant brand mentions from social media sites like Facebook or Twitter. Companies like Scrapehero ingest not only text, but video and images, and then apply NLP and computer vision techniques to analyze content. Companies like Auris use news streams, social media, blogs, and review boards to collect data and then analyze it with NLP and sentiment analysis. All these processes are automated, and the data examined in real-time and processed much faster than is possible manually. Although these companies often conduct market research or follow a particular brand, the same techniques could be used to follow a particular acquisition program and uncover potential problems or identify successes.

In a public-sector example, Kate et al. (2014) implement a text analytics approach to track the World Wide Web as a source of news feeds for food safety articles for the National Environment Agency of Singapore. Their approach uses machine-learning techniques to identify and rank relevant content about emerging food safety risks.

3. Examples

We used Twitter as a data source to create a prototype that demonstrates the recent public conversation about acquisition programs. We performed exploratory analysis by searching on tweets containing the name of an MDAP. In these examples, many of the tweets are not relevant to acquisition. To obtain better search results, OUSD(A&S) staff suggested reviewing the twitter accounts from members of Congress. As an example, they suggested the account for U.S. Representative Mac Thornberry (@MacTexPress). First, we performed a network analysis to identify a larger set of relevant accounts.[9] We then identified 57 accounts corresponding to defense news organizations and national security correspondents followed by @MacTexPress. Next, we identified the set of accounts each of the 57 accounts followed and selected a subset of about 200 accounts that were each followed by at least 30 of the 57 accounts. We then performed a similar analysis by examining the accounts followed by those of the Armed Services committees for both parties for both the House and Senate to identify 185 accounts related to members of Congress. Next, we collected up to 3,000 tweets for each the 57 accounts from news organizations and reporters, about 200 accounts identified in the news network analysis and 185 accounts from Congress. Through the network analysis, we constructed a database of tweets from more relevant and reliable sources, resulting in a set of almost 1 million tweets.

[9] The free version of the Twitter application programming interface (API) allows users to download up to 3,000 recent tweets for a specific account and tweets related to search terms within the last 7 days. The Twitter API allows users to purchase access to historical data, thus relaxing those constraints.

We filtered the 1 million tweets to include only those accounts that consistently tweeted from November 2018 through November 2019. The resulting subset of tweets totaled about a quarter of a million. Within those tweets, we searched for the names of several acquisition programs and calculated the sentiment of those tweets; Figure 4 shows these results. The size of the circle is correlated with the number of tweets in that month. The y-axes of the graphs represent the sentiment for that program's tweets. Circles with low values on the y-axis represent more negative sentiment for tweets that month on average relative to circles with higher values on the y-axis.

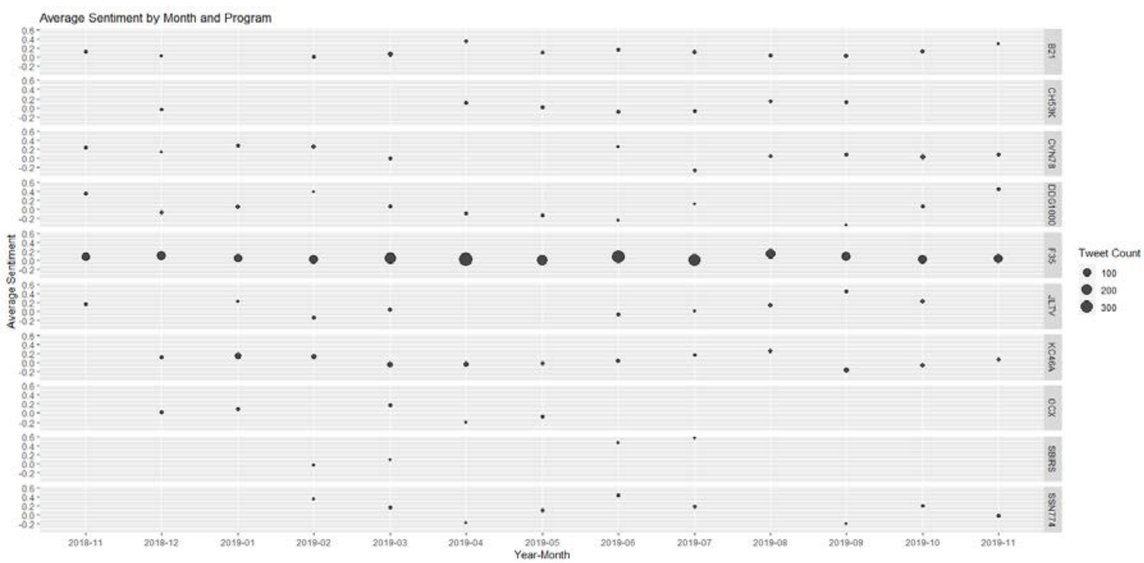

Figure 4. Tweet Count and Average Sentiment by Month and Program

Figure 4 provides an overall view of the conversation of a set of programs, but we can explore the individual tweets corresponding to these data to better understand the issues. For example, Figure 5 shows tweets about Boeing's KC-46A aircraft in September 2019 with the most negative sentiment. Each tweet discusses issues with the program.

Figure 5. KC-46A Tweets

4. Next Steps

Although news alerts can provide information on programs, text analytics can help prioritize this information. The next step would involve staff who prepare background material for meetings with stakeholders use the output from a prototype. A simple prototype would be to expand the list from Figure 4 to a larger set of acquisition programs, which could be sorted by average sentiment. The tweets for a particular program could then be sorted by sentiment within a program. Acquisition staff could give feedback on how this process could be improved. A list of relevant search terms could be refined as part of this process. The sentiment algorithms could also be customized with dictionaries that are relevant to acquisition. In addition, the network analysis could be refined to ensure that the most relevant accounts were included. Further, the network could be expanded to include more relevant accounts. This analysis would need to consider the tradeoff between adding more accounts to reduce the risk of missing information with the potential increase in the amount of irrelevant information.

E. Examine Program Office Emails

1. Description

Program office emails are another potential source of data on acquisition programs. Industry has used employee emails to gain insights about their companies. Similarly, this approach could be used to gain information about acquisition programs. Emails from personnel in the program offices could be anonymized to identify general issues in programs without identifying specific individuals. These issues could be negative sentiment about a program or even key topics of potential concern. This analysis does not have to be restricted to emails. Other anonymous text such as surveys or whistleblower data could be analyzed in a similar manner.

2. Previous Research

We found KeenCorp to be the most relevant example of a company using email analysis. KeenCorp is a Netherland-based data analytics company that measures employee sentiment in emails. Their tool reports only aggregated trends and does not report individual information. A popular data source for email is the Enron Corpus,[10] which is a database of more than 600,000 emails from almost 150 employees of Enron. At the conclusion of the investigation by the Federal Energy Regulatory Commission, the Commodity Futures Trading Commission, and Department of Justice, the information was made public and KeenCorp analyzed the dataset. Although their algorithms indicated the

---

[10] B. Klimt and Y. Yang, "The Enron Corpus: A New Dataset for Email Classification Research," *Proceedings of the 15th European Conference on Machine Learning*, 2004, 217–226.

lowest sentiment score for emails sent at the end of 2001, when Enron filed for bankruptcy, their analysis also indicated that June 28, 1999, was a significant date. By consulting with Enron's former chief financial officer, KeenCorp learned that a proposal was discussed at a board meeting that day, leaving many employees uncomfortable. This proposal eventually resulted in Enron's corporate fraud problems and bankruptcy.[11]

Other companies use a variety of data sources, such as surveys or internal social networking platforms, to analyze employee satisfaction.[12] A similar approach to detect issues within a program could be used if appropriate data were available from the personnel in acquisition program offices.

Still other companies use organizational network analysis to improve their efficiency. Organizational network analysis is a technical approach to organizing communication within an organization with data such as surveys or emails.[13] Genpact, a professional services company, studied the communication patterns of their top 650 leaders. Their study used only metadata on emails and not their actual content. Genpact found best practices among high-performing leaders, such as using simpler words to communicate, responding faster, and communicating more often.[14] A similar approach could be used to identify best practices in high-performing acquisition programs.

3. Next Steps

As demonstrated by the previous research, tools exist to measure the general sentiment of an organization using text data. To consider this option further, the appropriate dataset in defense acquisition would need to be identified. This dataset should include programs experiencing problems and those that are performing well. Another consideration in choosing the data source is how using the data to assess programs may eventually influence the data source itself. For example, if people knew their emails were being analyzed, they might change how they communicate. After a data source is identified, the next step is to conduct a prototype analysis to determine whether problems could be detected within that data source using sentiment analysis and/or machine learning. Our

---

[11] Frank Partnoy, "What Your Boss Could Learn by Reading the Whole Company's Emails," *The Atlantic*, September 2018, https://www.theatlantic.com/magazine/archive/2018/09/the-secrets-in-your-inbox/565745/, accessed June 24, 2020.

[12] Kaveh Waddell, "The Algorithms That Tell Bosses How Employees Are Feeling," *The Atlantic*, September 29, 2016, https://www.theatlantic.com/technology/archive/2016/09/the-algorithms-that-tell-bosses-how-employees-feel/502064/?utm_source=feed, accessed June 24, 2020.

[13] Agron Fazliu, "How to Use Corporate E-mail Analysis to Reveal Hidden Stars and Ensure Equal Opportunities!" December 30, 2018, https://towardsdatascience.com/how-to-use-corporate-e-mail-analysis-to-reveal-hidden-stars-and-ensure-equal-opportunities-90bb77d61a7f, accessed June 24, 2020.

[14] J. Bersin, "What Emails Reveal About Your Performance at Work," October 16, 2018, https://joshbersin.com/2018/10/what-emails-reveal-about-your-performance-at-work/, accessed June 24, 2020.

study considered only what is possible from an analytical standpoint and has not explored any potential legal issues or privacy concerns associated with analyzing these data.

F. Facilitate the Extraction of Metadata

1. Description

Another option for using text analytics to review programs is to extract metadata from unstructured data in documents. This approach would reduce the effort required for a person to read a document and extract the data manually. It may also increase data accuracy by reducing the risk of typos. After metadata is extracted from many documents, it could be analyzed to gain insights about particular programs or examine trends across programs.

2. Previous Research

Industry uses these metadata techniques in a variety of applications. One firm, iManage, has a tool for legal firms that reads and extracts business information from large volumes of documents and unstructured data. The tool extracts content and organizes it into document sets or clusters to expedite processing. This approach allows law firms to harness data from disparate locations, save time on previously manual processes, and reduce errors, which can tarnish their reputations. Another firm, Uptake, uses artificial intelligence in the Industrial, Energy, and Utilities sector to complete missing data, suggest asset labels, and create asset category schema for messy operational and maintenance data.

In the Government, the United States Air Force (USAF) has demonstrated a project where metadata has been extracted from contract announcements.[15] The USAF shows an example where such announcements, in the form of single-paragraph, semi-structured text, were scraped from public websites. Next, key data items such as company name, company city/state, award amount, fiscal year, and so on were automatically extracted from the text of the contract announcements. The data was then stored in a structured database for future reference or for additional analyses. The study showed limited technical accuracy, but the area holds promise. We note that databases such as the Federal Procurement Data System contain some metadata on contracts, but the concept has wider application, particularly in the acquisition realm, such as pulling key data from other documents such as Acquisition Decision Memoranda.

3. Next Steps

It is possible for similar techniques to be used on defense acquisition contracts or some of the other numerous required documents within defense acquisition. If information

---

[15] Spectrum, SAF-AQX Analytics Phase II Outbrief, April 19, 2017.

across contracts could be gathered and collated in a structured format, additional analyses could be performed on that dataset. An analyst could potentially gain insights into issues in acquisition and then seek mitigation strategies.

G. Identify Differences among Acquisition Documents

1. Description

In addition to extracting data from a particular document, an analyst may want to identify differences among documents. One example of this approach is tracking changes in versions of the same document, while another is determining whether two different documents relay the same or different information. For the acquisition community, it may be advantageous to track the same document over time to determine how requirements, key performance parameters, or costs have changed. This approach could also be useful in isolating text that is unique within a particular document from the corpus of documents required for a single program. Isolating unique text would help the analyst focus on the distinct information in each piece of documentation, rather than needing to read the same background, summary, or descriptive text that might be very similar across documents. For example, if periodic assessments of the program are developed, this approach could identify what is new in the latest assessment relative to the previous period.

In addition, we learned in discussions with AAP that it might be worthwhile to know whether DoD and Congress received the same information from the program offices. Artificial intelligence could streamline this comparison.

2. Previous Research

Industry uses several variations of similar software to accomplish these goals. A company called Kira has a tool that allows the user to import contracts or other documents, which are then processed and analyzed so that the user can quickly see what changes have been made between versions, find trends across documents, identify hidden risks, and export summaries. This software can help with the contract review process in the areas of compliance, lease abstraction, or due diligence in mergers and acquisitions. Similarly, a firm called Luminance provides a service that can identify similarities, differences, and anomalies in legal review documentation, which provides insights across contracts and allows the user to quickly identify and manage document revisions.

IDA has an in-house tool, IDATA, which performs a similar function and has been used for government sponsors for various purposes. For example, for the Office of Legislative Affairs, the tool was used to track changes between the House and Senate versions of a National Defense Authorization Act, historically a labor-intensive and repetitive process.[16]

3. Examples

As a proof-of-concept for the comparing information, we used Congressional testimony from October 2011 and the DAES report from January-February 2012 for the KC-46A aircraft. We partitioned each document into smaller sections roughly the size of a page. The result was 33 pages for Congressional testimony and 11 for DAES. Next, we fit a topic model with 6 topics across the 44 combined pages without considering differences between source documents. We then matched each page to the topic it was most closely associated with. In this matching, we found that some topics were related only to Congressional testimony pages, some topics were related only to DAES report pages, and one topic spanned both. See Appendix B for more details on the prototype analysis performed in this area, marked "For Official Use Only" and issued separately.

4. Next Steps

The next step in using this method is to identify a use case for comparing documents. Further analysis could involve tracking how particular documents change over time to allow analysts to focus their efforts analyzing new information. For instance, automatically comparing the performance assessment reports over time could illuminate the way the program has changed. Another potential use case could be to compare documents to identify discrepancies between two sources of information.

H. Estimate Whether a Document is Ready for Approval

1. Description

Artificial intelligence can be used to estimate whether a document is ready for approval or needs further work. For example, AI could be used to assist in the contract review process to identify documents that may have issues. This method could be used within the acquisition community to process documents as numerous acquisition contracts and documents undergo an approval process.

2. Previous Research

This AI application is particularly useful for legal firms, who need to route many documents through approval processes. A company called LawGeex has a tool that

---

[16] Laura Odell, Katherine Burton, and Miranda Seitz-McLeese, "Comparing the House and Senate Versions of the National Defense Authorization Act," IDA Research Notes – IDA Text Analytics (Alexandria, VA: Institute for Defense Analyses, October 2018).

automatically reviews contracts and, if no problems are found, automatically approves and routes the contract. If issues are found, pre-approved language is offered so that lawyers can edit documents expeditiously and with very little effort. This approach increases the amount of time users can spend on more critical work.

3. Next Steps

The next step to use this approach is to identify a use case. Potential documents for a use case could be contracts or other program office documents that require approval, such as a Test and Evaluation Master Plan. After identifying the use case, the next step is to develop a set of training documents with documents ready for approval and others with issues that need to be addressed. Given these training data, the techniques described to predict outcomes could be used to predict whether a document is ready for approval. Modelers who predict whether a document is ready for approval also need to work closely with subject matter experts to develop a model that allows the acquisition analysts to find issues efficiently in documents that are not ready for approval.

4. Conclusions

A. Observations

We engaged with several acquisition experts and demonstrated examples to give a better understanding of the potential use of text analytics to improve acquisition program assessment. Through this process, we gathered observations on the key elements of applying these tools to acquisition programs.

Although some decisions can be fully automated through AI, acquisition experts are still required in the process of assessing acquisition programs. Assessing an acquisition program is a complex process with numerous factors, and each program has its own unique characteristics. Our study suggests that human-machine teaming would be the best next steps to use AI in acquisition program assessment.

Another key element of applying AI to acquisition program assessment is the data. Analytics is only as good as the quality of the data. Data required by a contract, from an audit, or another mechanism that encourages high data quality could be good input for text analytics. Another challenge with data for DoD acquisition, in particular, is that there are a limited number of overall outcomes, given the limited number of total historical acquisition programs. DoD acquisition does have a large volume of unstructured data about each of these programs that can be analyzed through text analytics. Success in acquisition can be subjective and is assessed across multiple criteria. This issue could be overcome if we simplify and use some quantifiable metrics about a program or expert judgement to label a specific element of these data.

Another challenge in the implementation phase is preparing the data for the tools. Many of these data are not readily accessible in a format to conduct analysis. Data access for acquisition documents is often requires that analysts download individual documents to review as part of their manual review process. For example, an analyst might download a particular earned value report for a program that they are analyzing, but data is likely not configured for API access. Sharing sensitive data and the using software on Government machines is another challenge in implementing an AI tool. Before a prototype could be shifted to a production model, a significant effort would be required to rectify these data issues. Until a production model could be implemented, continuing to integrate personnel with analytics backgrounds into the acquisition assessment process could be a way to incrementally introduce emerging techniques.

Finally, for AI to be useful for decision-makers, the results need to be explainable. Acquisition decisions are large in magnitude and the processes are complex. A black box tool indicating a problem has limited utility. A tool would be much more valuable if text analytics could help an analyst identify an issue that they can describe. This approach would allow decision-makers to debate and further explore the issue before deciding on an appropriate action.

B. Path Forward

Our study suggests that the most promising options are using tools like sentiment analysis, topic models, or predicting outcomes to prioritize useful information from a corpus for acquisition staff to review. The key next step is to demonstrate improved efficiency for a particular use case. Regardless of the use case, the text analytics algorithms could be further improved by customizing these algorithms for acquisition-specific information.

Several possible use cases with different questions and data sources could be candidates for more detailed analysis, but one example use case is identifying challenges in a program that need management attention. These challenges could include the supply of labor, such as welders; weight growth in an aircraft; or issues like those with the KC-46A boom. We could analyze the narrative assessments in PARs to demonstrate the following concepts:

- Whether analysts who use text analytics to prioritize documents can review information more quickly
- Whether analysts who use text analytics to prioritize documents get improved results compared to analysts who perform reviews manually
- Whether analysts tracking a program over time can save time by using text analytics to focus reviews on documents that have changed since the last month
- Whether issues that ultimately required management attention could be identified sooner

Our feasibility study identifies several options for how AI techniques might be applied to acquisition and illustrates how these techniques might be used. The next step before deploying a tool is to estimate how text analytics would improve existing processes. In some cases, the benefit from a tool would not justify the resources required. One benefit could be analyst time saved using text analytics. A better outcome would be to demonstrate that an analyst using text analytics would not only save time using these tools, but would also be able to develop better insights about a program.

Appendix A.
Example Analysis: Identifying Correlations across Data in Different Functional Areas

This appendix is marked FOUO and issued as a separate document: “A Feasibility Study on the Use of Artificial Intelligence for Defense Acquisition Program Review, Volume II: Example Analyses,” IDA Paper P-13239, Log number: H 20-000223/1.

Appendix B. Example Analysis: Identifying Differences in Acquisition Documents

This appendix is marked FOUO and issued as a separate document: "A Feasibility Study on the Use of Artificial Intelligence for Defense Acquisition Program Review, Volume II: Example Analyses," IDA Paper P-13239, Log number: H 20-000223/1.

Illustrations

Figures

References

Bersin, J. “What Emails Reveal About Your Performance At Work.” October 16, 2018. https://joshbersin.com/2018/10/what-emails-reveal-about-your-performance-at-work/, accessed June 24, 2020.

Fazliu, Agron. “How to Use Corporate E-mail Analysis to Reveal Hidden Stars and Ensure Equal Opportunities!” December 30, 2018. https://towardsdatascience.com/how-to-use-corporate-e-mail-analysis-to-reveal-hidden-stars-and-ensure-equal-opportunities-90bb77d61a7f, accessed June 24, 2020.

Freeman, Charlton E. “Multivariate and Naïve Bayes Text Classification Approach to Cost Growth Risk in Department of Defense Acquisition Programs.” 2013.

Gwozdz, Lawrence T., and Paul M. Kodzwa. “The DAES Process.” *Defense Acquisition Magazine Blog*. January 30, 2019. https://www.dau.edu/library/defense-atl/blog/The-DAES-Process-, accessed May 22, 2020.

Hassler, M. and T. Clark. “Application of Natural Language Processing to Defense Acquisition Executive Summary Reports.” Naval Postgraduate School: Acquisition Research Symposium. May 2020.

House Hearing, 114th Congress - [H.A.S.C. No. 114-35]. “Update on the F-35 Joint Strike Fighter Program and the Fiscal Year 2016 Budget Request.” October 14, 2015.

Kate, K., S. Chaudhari, A. Prapanca, and J. Kalagnanam. “FoodSIS: A Text Mining System to Improve the State of Food Safety in Singapore.” *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2014. 1709–1718.

Klimt, B. and Y. Yang. “The Enron Corpus: A New Dataset for Email Classification Research.” *Proceedings of the 15th European conference on Machine Learning*. 2004. 217–226.

Kloke, J., B. Murphy, E. Murray, and V. Lourdes. “Program Performance Intelligence.” Acquisition Analytics Forum, Briefing. December 13, 2019.

Maiya, A., J. Thompson, F. Loaiza-Lemos, and R. Rolfe. “Exploratory Analysis of Highly Heterogeneous Document Collections.” *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2013. 1375–1383.

Miller, Trevor. “Acquisition Program Problem Detection Using Text Mining Methods.” 2012. https://scholar.afit.edu/etd/1022/.

Odell, Laura, Katherine Burton, and Miranda Seitz-McLeese. “Comparing the House and Senate Versions of the National Defense Authorization Act.” IDA Research Notes – IDA Text Analytics (Alexandria, VA: Institute for Defense Analyses, October 2018).

Partnoy, Frank. “What Your Boss Could Learn by Reading the Whole Company’s Emails.” *The Atlantic*. September 2018. https://www.theatlantic.com/magazine/archive/2018/09/the-secrets-in-your-inbox/565745/, accessed June 24, 2020.

Senate Hearing, 114th Congress - [S. Hrg. 114-658] Part 1. “Department of Defense Authorization for Appropriations for Fiscal Year 2017 and the Future Years Defense Program.” February–April 2016.

Shah, D., H. Isah, and F. Zulkernine. “Stock Market Analysis: A Review and Taxonomy of Prediction Techniques.” *International Journal of Financial Studies* 7 (2) (2019): 1–22.

Thomas, M. “How AI Trading Technology is Making Stock Market Investors Smarter.” April 10, 2020. https://builtin.com/artificial-intelligence/ai-trading-stock-market-tech, accessed June 24, 2020.

Waddell, Kaveh. “The Algorithms That Tell Bosses How Employees Are Feeling.” *The Atlantic*. September, 29, 2016. https://www.theatlantic.com/technology/archive/2016/09/the-algorithms-that-tell-bosses-how-employees-feel/502064/?utm_source=feed, accessed June 24, 2020.

Abbreviations

AAM	Acquisition Approaches and Management
AAP	Acquisition Analytics and Policy
AI	Artificial Intelligence
ALIS	Autonomic Logistics Information System
API	Application Programming Interface
ASD(A)	Assistant Secretary of Defense for Acquisition
AUC	Area under the Curve
DAES	Defense Acquisition Executive Summary
DAMIR	Defense Acquisition Management Information Retrieval
DCMA	Defense Contract Management Agency
DoD	Department of Defense
EV	Earned Value
EVM-CR	Earned Value Management Central Repository
GAO	Government Accountability Office
IDA	Institute for Defense Analyses
LCO	Legislative and Congressional Oversight
LDA	Latent Dirichlet Allocation
MDAP	Major Defense Acquisition Program
MTA	Middle Tier of Acquisition
NLP	Natural Language Processing
OUSD(A&S)	Office of the Under Secretary of Defense for Acquisition and Sustainment
OUSD(R&E)	Office of the Under Secretary of Defense for Research and Engineering
PAR	Performance Assessment Report
SAR	Selected Acquisition Report
USAF	United States Air Force
XGBOOST	Extreme Gradient Boosting

REPORT DOCUMENTATION PAGE	*Form Approved* *OMB No. 0704-0188*

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
xx-07-2020	Final	May 2019 -May 2020

4. TITLE AND SUBTITLE
A Feasibility Study on the Use of Artificial Intelligence for Defense Acquisition Program Review, Volume I: Main Report

5a. CONTRACT NUMBER
HQ0034-14-D-0001

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Rieksts, Brian, Q.
Guerrera, Kristen, M.

5d. PROJECT NUMBER
AB-7-4624

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Institute for Defense Analyses
4850 Mark Center Drive
Alexandria, VA 22311-1882

8. PERFORMING ORGANIZATION REPORT NUMBER
IDA Paper P-13239
H 20-000222

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Office of the Undersecretary of Defense (Acquisition and Sustainment)/Acquisition Analytics and Policy
The Pentagon, Room 3B948A
Washington, DC 20301

10. SPONSOR/MONITOR'S ACRONYM(S)
OUSD(A&S) / ASD(A) / DASD(AE) / AAP

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES
This document contains a Volume II, Example Analyses (Appendices A and B) that is marked FOUO is issued under a separate cover.

14. ABSTRACT
The Office of the Assistant Secretary of Defense for Acquisition (ASD(A)) tasked the Institute for Defense Analyses (IDA) to explore the feasibility of applying artificial intelligence (AI) and related tools to help assess defense acquisition programs. The objective of this project is to identify options to streamline acquisition processes, enhance predictive power, and better and more efficiently inform senior decision-makers about trends and possible acquisition outcomes. As part of its analysis, IDA engaged with acquisition staff who assess programs, identified AI activities in industry where unstructured data has been used to measure performance, and analyzed options to use AI to help assess acquisition programs. We identified several options to apply text analytics to unstructured data on major defense acquisition programs. Our study suggests that the most promising option is to use AI tools such as sentiment analysis and topic models to prioritize useful information from large datasets for acquisition staff to analyze.

15. SUBJECT TERMS
Text Analytics; Acquisition; Topic Models; Sentiment Analysis

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
Unclassified	Unclassified	Unclassified	Same as Report	41	Joseph, Brian
					19b. TELEPHONE NUMBER *(Include area code)* (703) 697-8076

Reset