

Using Lexical Link Analysis as a Tool to Improve Sustainment

Edwin Stevens, Ying Zhao *
Naval Postgraduate School, Monterey, CA, USA

Abstract

A major challenge in the the complex enterprise of the US Navy global materiel distribution is that when a new operation condition occurs, the probability of fail or demand model of a Naval ship part or item needs to modify to adapt to the new condition. Meanwhile, historical supply databases include demand patterns and associations that are critical when the new condition enters the system as a perturbation or disruption which can propagate through the item association network. In this paper, we first show how the two types of item demand changes can be interacted and integrated to calculate the total demand change (TDC). We show a use case on how to apply the lexical link analysis (LLA) to discover the item association network that propagates the TDC.

Introduction

There are many challenges in the complex enterprise of the US Navy global materiel distribution. Forward deployed US Navy ships, particularly in the high operating tempo (OPTEMPO) areas such as the Seventh and Fifth Fleet, have challenges that arise in receiving logistical support when part failures occur. These failures manifest as either a demand on the supply system, a casualty report (CASREP), or a request for technical assistance. The toughest challenges arise when a high impact part fails, and is not immediately available. This can cause a “redline,” or a failure that stops the unit from being able to complete its’ mission until the problem can be resolved. The goal of any operational commander is 100% operation availability (AO), meaning their unit is always ready to be tasked for any situation that arises. Failures in contentious environments will stop the mission, and could have great effects on the international and political situation. The goal of a Navy logistician is to “not let the

logistical tail wag the operation dog,” in other words, a good logistician does not want to be the reason that the mission can’t go on. Limited manpower, funding, storage space, and resources for repair are all in high demand. A good system needs to be in place to determine the most efficient and effective method of stocking, forward staging, or contracting for the materials that have the highest likelihood of demand, balanced with the potential impact of failure. Since even one ship has hundreds of thousands of failed parts, many of which could cause a “redline.” It is of critical importance to consider the activities of all the parts as a complex system and predict the demand as a whole so that the supply system is as intelligently designed as possible in order to quickly handle part failures.

Uncertainty, Perturbation and Association

The probability of fail of a part can be affected by many factors. We need to consider the uncertainty, disruption and perturbation that can impact the logistics plans as a whole. For example, uncertainty factors related to environment and events in wide geographic areas, such as, weather change or mission change from a peace time to a conflict time, or a sudden event can cause a perturbation and disruption for previous logistics and supply plans. Previously high impact but low fail parts may suddenly become in high demand.

The probability of fail is also embedded in the historical supply and maintenance data. A failed part is considered to be fixed before a new one is ordered. A part order frequency in the historical supply data reflects its demand if the part can not be repaired within a certain period of time. The demand data in the supply data reflects partial probability of fail.

The complexity of predicting total probability of fail for a large list of the items calls for the integration of methods in data fusion, data mining, causal learning, and optimization for all the elements in a logistics when facing particular uncertainty and perturbation. The goal of this paper is to demonstrate the techniques such as data mining and lexical link analysis (LLA) to recalculate the probability of fail for the previously high impact and low failure parts or items when the whole system facing a perturbation, uncertainty, disruption, or a “redline” failure.

*This will certify that all author(s) of the above article/paper are employees of the U.S. Government and performed this work as part of their employment, and that the article/paper is therefore not subject to U.S. copyright protection. No copyright. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0). In: Proceedings of AAAI Symposium on the 2nd Workshop on Deep Models and Artificial Intelligence for Defense Applications: Potentials, Theories, Practices, Tools, and Risks, November 11-12, 2020, Virtual, published at <http://ceur-ws.org>

Lexical Link Analysis (LLA)

A data mining tool used for this research is Lexical Link Analysis (Zhao, MacKinnon, and Gallup 2015). LLA is an unsupervised machine learning method and describes the characteristics of a complex system using a list of attributes or features, or specific vocabularies or lexical terms. Because the potentially vast number of lexical terms from big data, the model can be viewed as a deep model for big data. For example, we can describe a system using word pairs or bi-grams as lexical terms extracted from text data. LLA automatically discovers word pairs, and displays them as word pair networks. Figure 1 shows an example of such a word network discovered from data. “Clean energy” and “renewable energy” are two bi-gram word pairs. For a text document, words are represented as nodes and word pairs as the links between nodes. A word center (e.g., “energy” in Figure 1) is formed around a word node connected with a list of other words to form more word pairs with the center word “energy.”

Discovering Item Associations Using LLA

Bi-grams allow LLA to be extended to numerical or categorical data. For example, using structured data, such as attributes from supply chain databases, we discretize numeric attributes and categorize their values to word-like features. The word pair model can further be extended to a context-concept-cluster model (Zhao and Zhou 2014). A context can represent a location, a time point, or an object shared across data sources. For example, the quarters in a year can be one of the contexts for item supply data. Items (parts) are the concepts.

In this paper, we use LLA for the structured data of supply databases. We want to show that the bi-gram generated by LLA can also be a form discovery of association among items demand for a Navy supply database.

The common consensus is that data-driven analysis or data mining can discover initial statistical correlations and associations from big data.

Figure 2 shows conceptually how the associations and correlations are discovered by LLA. We anticipate the demand change (DC) an item i might come from two types of sources: Type 1): A collection of outside perturbations such as the change of missions or new operational conditions; and

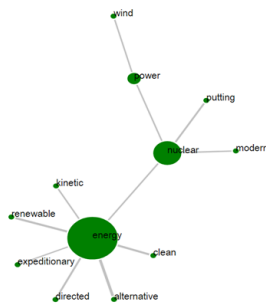


Figure 1: An example of lexical link analysis

Type 2): Item associations with other items where the associations could be due to physical linkages or linked demand based on past business practices. If an item i is ordered, item j is also likely to be ordered based on the historical data. Type 2) DCs can be mined from historical potentially big data, Type 1) DCs may come from expert and engineering knowledge and simulations.

In Figure 2, $Assoc_{ij}$ measures how strong item i and j are demanded together. Probability and lift are the two measures defined in Equation (1) and Equation (2) in LLA to measure the strength of an association.

$$prob_{ij} = \frac{\text{demand of item } i, j \text{ together out of demand of item } j}{\text{demand of item } i \text{ out of all demands}} \quad (1)$$

$$lift_{ij} = \frac{\text{demand of item } i, j \text{ together out of demand of item } j}{\text{demand of item } i \text{ out of all demands}} \quad (2)$$

In LLA, we first use $lift_{ij}$ to filter out the associations that are not strong enough, then apply $prob_{ij}$ to compute the total demand change (TDC) for item i as in Equation (3)

$$TDC_i = \sum_{m=1}^M DC_i | C_m + \sum_{j=1}^N prob_{ij} * TDC_j \quad (3)$$

In this paper, we show LLA can be used to compute the association network, $prob_{ij}$, and $lift_{ij}$ from historical demand data. When there is a perturbation such as a new operation condition C_m occurs that generates a $DC_j | C_m$ for item j , it causes a TDC_j for item j ; meanwhile, TDC_j propagates through the discovered association network from LLA to affect the whole demand system and forward predictions as shown in Equation (3).

Data Description and Initial Analysis Results

Currently, a part is reviewed to be stocked if it has more than two reorders in one year. This simple system is effective overall, but does not consider the reasons for failure, the reason it is being reordered, or the effect that the failure has on the ship. There are a small amount of parts, called “maintenance assist modules” that are carried onboard every ship due to engineering specifications calling for immediate availability if needed, but that is not enough to prevent “redline” failures. To show the feasibility of our methodology, we compiled a large selection of demand data over the last nine years, containing over 1,000,000 individual demands. This data was then compiled by Item Mission Essentiality Code (IMEC - impact code), quarter in which the demand occurred, and number of demands logged. Next, LLA

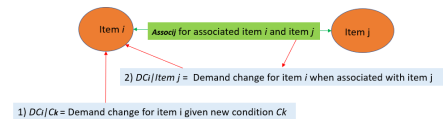


Figure 2: Total demand change (TDC) caused by new conditions and associations

was applied the data to help discover historical associations among the failures. The associations reflect the items that are ordered in the same contexts (e.g., the same quarter or same ship) historically. Associated parts might be stockpiled in the same manner should one fail suddenly in a new and disrupted condition. On a sample run, there were 50 connections found across 65,000 demands as illustrated in Figure 3, we only considered the associations among the high impact items (4) with quarterly demand > 51 (high) or low (= 1). For example, item “lwm048749” and “lwm048745” both have high impact 4, while “lwm048749” had high demand in some quarters when “lwm048745” had low demand. When drilling down using LLA as shown in Figure 4, “lwm048749” had high demand in two quarters (10 and 18) when “lwm048745” had low demand. “lwm048749” had high demand in two quarters out of the total 20 quarters. The probability for the association of the two items is 100% and lift is 10. Should “lwm048745” demand more in a new operation condition, associated parts such as “lwm048749” may demand even more in the new condition. LLA calculates the lift measure that is similar to the counterfactual reasoning in causal learning (Mackenzie and Pearl 2018; Pearl 2018; Zhao, MacKinnon, and Jones 2019), i.e., that there is indeed causal relationship between two demands.

Conclusion and Future Work

In this paper, we showed the feasibility on how to apply LLA to improve demand change predictions for a complex Navy supply database.

In the future research, we will consider the association



Figure 3: Total demand change (TDC) caused by new conditions and associations

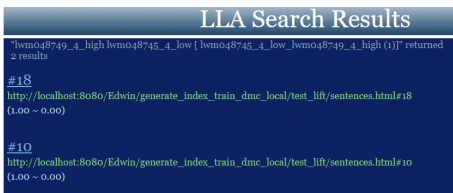


Figure 4: LLA allows a drill-down to see how many times (quarters) the two items are associated

contexts set to be ship type, unit identification code, IMEC, or shorter time period than the quarters, and then apply LLA to search for causal associations at higher or lower resolutions, or by stricter or looser requirements. In comparison, there is a current tool in place called Predictive Risk Sparing Matrix (PRiSM), which has been able to identify parts in various C4I systems that have had real world demands, which would not have been identified under the standard system. PRiSM uses mathematical algorithms from inventory sparing models to determine potential failures, and these algorithms could possibly be used in coordination with simulation and LLA to better determine future needs. We will also leverage the liaisons from NAVSUP and DLA at the Fifth and Seventh fleet naval bases, whose job is to track demand, and then to work with the DoD logistics organizations to improve operational availability. The LLA tool could be tested and then given to these liaisons to help them and to improve the overall area of operation (AO) for forward deployed ships and improve sustainment.

ACKNOWLEDGMENTS

Authors would like to thank the Office of Naval Research (ONR)’s Naval Enterprise Partnership Teaming with Universities for National Excellence (NEPTUNE 2.0) program. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied of the U.S. Government.

References

Book with Multiple Authors
Mackenzie, D. and Pearl, J. 2018. The Book of Why: The New Science of Cause and Effect. Penguin.

Journal Article
Zhao, Y. and MacKinnon, D.J. and Gallup, S.P., 2015. Big data and deep learning for understanding DoD data. *Journal of Defense Software Engineering, Special Issue: Data Mining and Metrics*, July/August 2015, Page 4-10. Lumin Publishing ISSN 2160-1577.

Proceedings Paper Published by a Society
Pearl, J. 2018. The Seven Pillars of Causal Reasoning with Reflections on Machine Learning. Retrieved from http://ftp.cs.ucla.edu/pub/sta_tser/r481.pdf

Zhao, Y. and Zhou, C. 2014. System and method for knowledge pattern search from networked agents. US Patent 8,903,756.

Proceedings Paper Published by a Press or Publisher
Zhao Y., MacKinnon, D.; and Jones, J. 2019. Causal Learning Using Pair-wise Associations to Discover Supply Chain Vulnerability. *Proceedings of the 11th International Conference on Knowledge Discovery and Information Retrieval (KDIR 2019)*, September 17-19, 2019, Vienna, Austria.