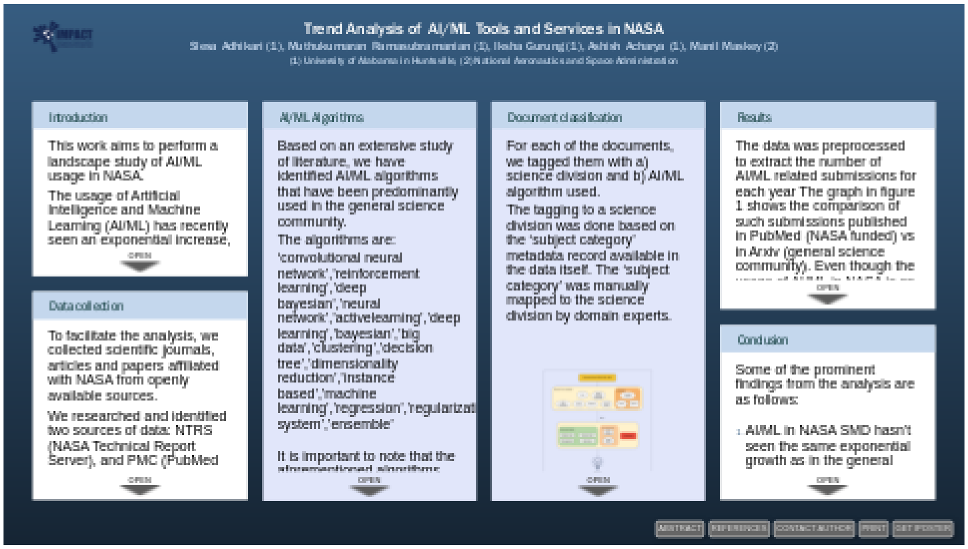
Trend Analysis of AI/ML Tools and Services in NASA



Slesa Adhikari (1), Muthukumaran Ramasubramanian (1), Iksha Gurung (1), Ashish Acharya (1), Manil Maskey (2)

(1) University of Alabama in Huntsville, (2) National Aeronautics and Space Administration

PRESENTED AT:

INTRODUCTION

This work aims to perform a landscape study of AI/ML usage in NASA.

The usage of Artificial Intelligence and Machine Learning (AI/ML) has recently seen an exponential increase, owing to high performance frameworks and rapid adoption of the technology. However, as the Strategy for Data Management and Computing for Groundbreaking Science report [1] finds, the advance of AI/ML has yet to be fully appreciated and understood by the SMD and its science disciplines. As such, it recommends making investments to incentivize the use of AI/ML for achieving breakthroughs in science. It also finds that collaborations on such analysis could be facilitated by sharing results and lessons learned among the community. To that end, an analysis of the current landscape is crucial.

DATA COLLECTION

To facilitate the analysis, we collected scientific journals, articles and papers affiliated with NASA from openly available sources.

We researched and identified two sources of data: NTRS (NASA Technical Report Server), and PMC (PubMed Central). NTRS provides access to NASA metadata records of full-text online documents, images, and videos. It includes conference papers, journal articles, meeting papers, patents, research reports, images, movies, and technical videos – scientific and technical information (STI) created or funded by NASA[2]. Similarly, PMC provides public access to the peer-reviewed papers resulting from NASA-funded research[3]. We consolidated the list of documents, from the two sources, into a database. The database includes documents from the year 2000 AD to present.

AI/ML ALGORITHMS

Based on an extensive study of literature, we have identified AI/ML algorithms that have been predominantly used in the general science community.

The algorithms are:

‘convolutional neural network’,’reinforcement learning’,’deep bayesian’,’neural network’,’activelearning’,’deep learning’,’bayesian’,’big data’,’clustering’,’decision tree’,’dimensionality reduction’,’instance based’,’machine learning’,’regression’,’regularization’,’rule system’,’ensemble’

It is important to note that the aforementioned algorithms aren’t an exhaustive list of AI/ML algorithms; however, these are sufficient to cover the AI/ML usage in NASA.

DOCUMENT CLASSIFICATION

For each of the documents, we tagged them with a) science division and b) AI/ML algorithm used.

The tagging to a science division was done based on the 'subject category' metadata record available in the data itself. The 'subject category' was manually mapped to the science division by domain experts.

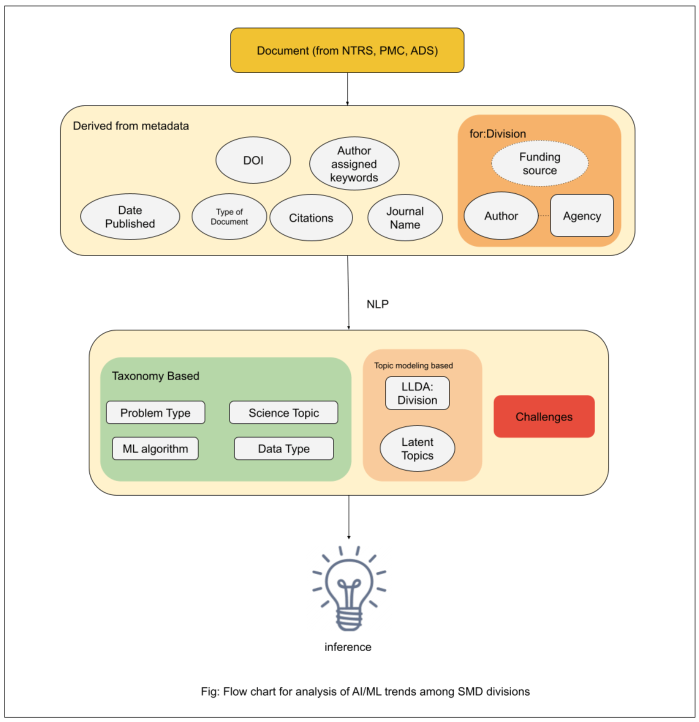


Figure 0

The tagging to an AI/ML algorithm was done by employing natural language processing to the text from the document title and abstract.

The workflow for analysis the trends of AI/ML in the NASA SMD is depicted in Figure 0.

RESULTS

The data was preprocessed to extract the number of AI/ML related submissions for each year The graph in figure 1 shows the comparison of such submissions published in PubMed (NASA funded) vs in Arxiv (general science community). Even though the usage of AI/ML in NASA is on the rise, it is evident that the growth rate in AI/ML in NASA isn't comparable with that of the general science community.

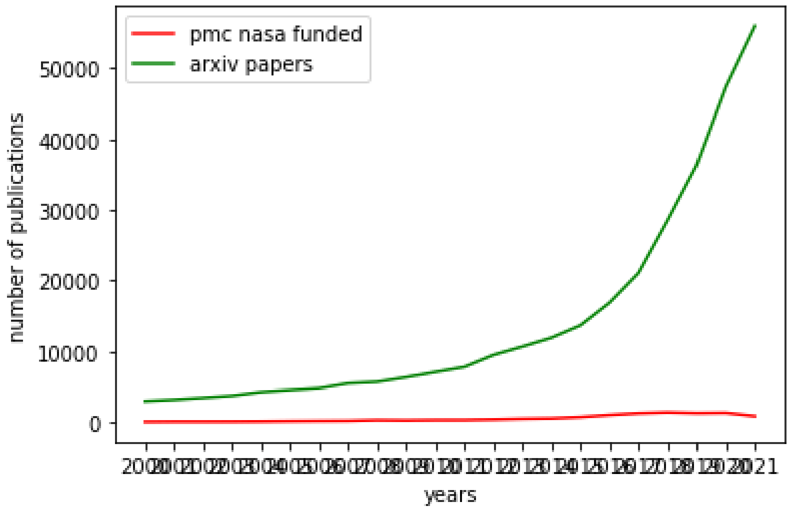


Figure 1

Furthermore, we looked at the distribution of those AI/ML related publications against the five SMD divisions, namely, Earth Science, Astrophysics, Planetary, Biophysical, Heliophysics, and the High End Computing for Science program. In figure 2, we see a huge discrepancy. The Earth Science division has the highest number of AI/ML submissions and the difference with the second highest contributing category 'High End Computing for Science' is staggering. However, it should be noted that these numbers have not been normalized based on the total number of submissions in each category.

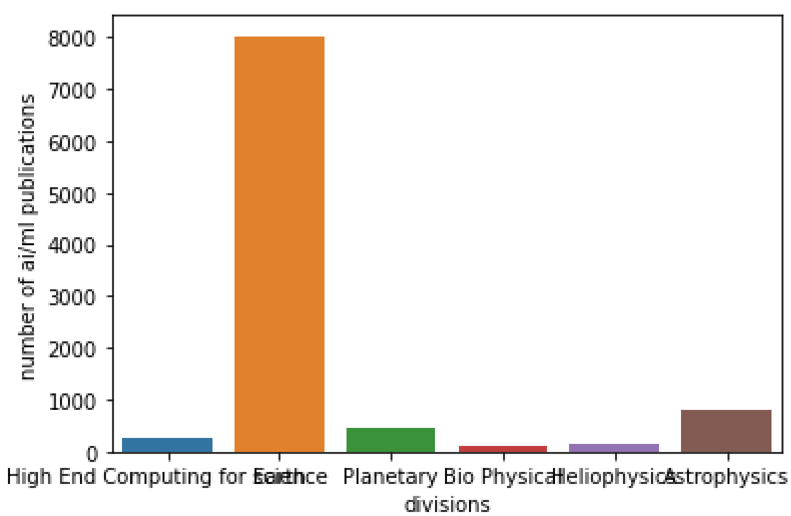


Figure 2

Further analysis on the taxonomy of algorithms shows that although deep learning - a class of neural networks that use multiple layers to capture highly non-linear dependencies in data - has been introduced and has been growing in the NASA SMD, the rate of increase is not on par with what is seen in the general science community. Figure 3 from [4] and figure 4 from NASA trends show this analysis.

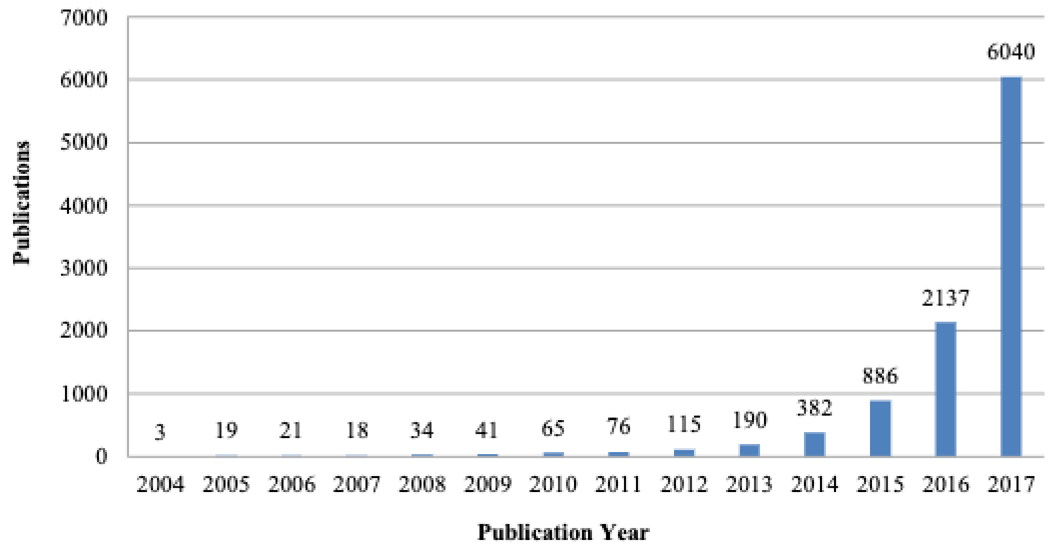


Figure 3[4]

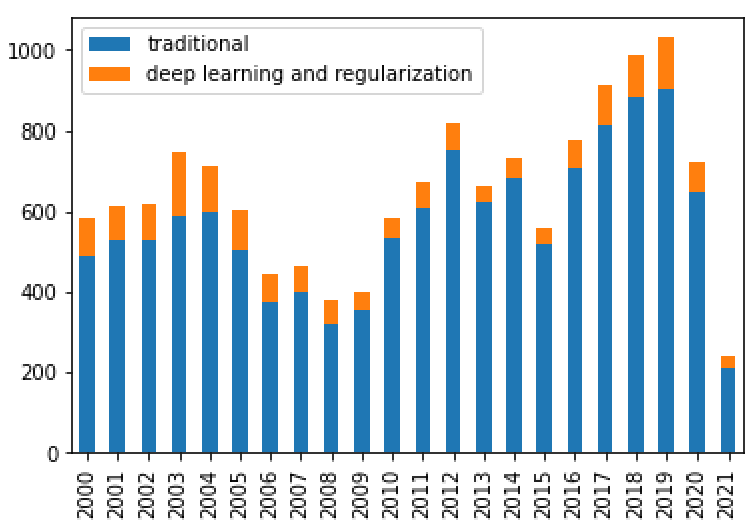


Figure 4

CONCLUSION

Some of the prominent findings from the analysis are as follows:

AI/ML in NASA SMD hasn’t seen the same exponential growth as in the general science community.

Among the NASA SMD Divisions, the Earth Science division is the forerunner of AI/ML usage.

The penetration of the newer, more advanced machine learning techniques like deep learning, neural networks, etc., is lacking in the NASA SMD.

ABSTRACT

Usage of Machine Learning (ML) algorithms within NASA's Science Mission Directorates have been increasing over the years. This can be quantitatively observed in the upward trends of ML usage found by analyzing the publications and presentations (in affiliation with NASA) available through NASA Technical Reports Server (NTRS) and PubMed Central (PMC). Identifying the problem types and class of ML algorithms used to tackle them across the divisions can present opportunities for collaborations, interdisciplinary projects and knowledge transfer for sustainable partnerships. In this presentation, we will present the trend analysis of ML algorithms used in different SMD divisions based on the publications and presentations publicly available. We identify these trends by leveraging ML algorithms which are able to search through the publication texts semantically; which are also highly scalable. We will also present an analysis on the available open source tools and services in NASA leveraging AI/ML algorithms. This work will provide ample avenues for collaborative efforts across different disciplines based on the surfaced trends.

REFERENCES

[1] Strategy for data management and computing for groundbreaking science. https://science.nasa.gov/science-pink/s3fs-public/atoms/files/SDMWG%20Strategy_Final.pdf (https://science.nasa.gov/science-pink/s3fs-public/atoms/files/SDMWG%20Strategy_Final.pdf)

[2] NASA Technical Reports Server. https://ntrs.nasa.gov/ (https://ntrs.nasa.gov/)

[3] NASA in PubMed Central. https://www.ncbi.nlm.nih.gov/pmc/funder/nasa/ (https://www.ncbi.nlm.nih.gov/pmc/funder/nasa/)

[4] Deep Learning Research: Scientometric Assessment of Global PublicationsOutput during 2004 -17. 10.28991/esj-2019-01165 (https://doi.org/10.28991/esj-2019-01165)

Upload new