ON-BOARD DOWNLINK PRIORITIZATION BALANCING SCIENCE UTILITY AND DATA DIVERSITY

G. Doran, M. Wronkiewicz, and S. Mauceri. Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, CA, 91109-8099, USA (gary.doran@jpl.nasa.gov).

Introduction: Science instruments for planetary missions are becoming increasingly complex and able to collect ever larger data volumes. This poses a challenge for missions targeting the outer planets because communications with ground teams on Earth are highly constrained. Transmitting information over vast distances requires large amounts of energy and, therefore, has to be limited to what is most essential. Additionally, there is a significant time delay between send and receive operations [1]. These constraints mean that science teams do not have access to the full raw data sets and can not operate the instruments interactively. Instead, data must be carefully selected on-board for return to Earth in order to maximize science yield.

In this study, we focus on data from the Ocean Worlds Life Surveyor (OWLS) platform, which comprises several different instruments to search for life in water at both the molecular and cellular level. Because signs of life are expected to be rare within the data OWLS will collect, we have developed systems that will analyze data on-board the spacecraft to produce smaller summary products called Autonomous Science Data Products (ASDPs), which are highly compressed versions of the data that include only the most scientifically relevant information. For example, ASDPs include the tracks of moving particles in raw microscope images or the location of peaks within the spectra from a mass-spectrometer. By combining advanced instruments with on-board autonomy, OWLS is aimed at advancing our life detection capabilities on ocean worlds like Europa and Enceladus. Given a set of generated ASDPs available for downlink, the goal of this work is to sort and down-select these products while balancing two objectives: (1) maximize the estimated per-byte *science utility* of the downlinked products to achieve mission objectives and (2) maximize the *diversity* of the products downlinked to enable the discovery of rare science content. Below, we describe our prioritization approach and compare its performance to baselines.

Background: OWLS is designed to search for life in water samples on ocean worlds using several instruments including a Digital Holographic Microscopy (DHM) and a Capillary Electrophoresis-Electrospray Ionization (CE-ESI) mass spectrometer. If life does exists on these bodies, it might be rare and only appear sporadically within acquired samples [2]. OWLS instruments can produce hundreds of gigabytes of data across observations, but not all of this data will be scientifically valuable if the signs of life are rare. Limitations in the bandwidth of data transmission and mission lifespan will permit only about one hundred megabytes of returned data. This represents approximately 0.01% to 0.1% of the original data volume. Therefore, we must select only a small portion of the most relevant data for the purpose of life detection.

To downlink the most scientifically relevant data under such extreme bandwidth constraints, we have developed two systems to automatically analyze raw data on-board the spacecraft and produce summary products called ASDPs. The first system, Holographic Examination for Life-like Motility (HELM), analyzes DHM videos to track moving objects, then classifies these tracks as either exhibiting life-like “motility” behavior or randomly drifting non-motile behavior. Small cutouts surrounding the detected motile particles are extracted for downlink that make up a small fraction of the raw video data volume. The second system, Autonomous CE-ESI Mass-spectra Examination (ACME), searches for peaks within data from a CE-ESI mass spectrometer, and creates ASDPs that summarize detected peaks and their context within the raw observations. After extracting the ASDPs using either HELM or ACME, the autonomous software must decide how to prioritize these products for downlink. This key problem is addressed by the Joint Examination for Water-based Extant Life (JEWEL) component of the OWLS autonomy system. JEWEL will trade-off between the competing objectives of (1) maximizing the utility or scientific relevance of selected data while also (2) selecting a set of diverse observations to avoid redundancy and maximize the return of rare and novel features in the data.

Approach: In order to incorporate both science utility and diversity, JEWEL relies on two pieces of information provided by HELM, ACME. First is the Science Utility Estimate (SUE), which is an quantitative estimate of the scientific value of an observation. In the case of HELM, the SUE is derived from the number of tracked particles in an observation classified as “motile,” and for ACME, the SUE is based on the number of peaks detected in an observation, the number of these peaks corresponding to known organic compounds, and size of these peaks relative to the background noise levels. The second piece of information is a Diversity Descriptor (DD), a vector that encodes properties of the observation’s content such that a set of observations with large Euclidean distances between DDs are considered to be diverse. For HELM, the DD is comprised of the numbers of detected motile and non-motile particles, and for ACME, the DD is comprised of 13 factors describing the numbers of peaks, their sizes, shapes, and background information.

To incorporate both SUEs and DDs when prioritizing ASDPs for downlink, JEWEL uses an algorithm based on the Maximum Marginal Relevance (MMR) criterion for ranking [3]. MMR works by iteratively selecting the data product with the highest utility, but discounted by a factor proportional to the Gaussian similarity of each candidate ASDP to the most similar product already selected for downlink. A user-specified parameter controls the balance between the initial SUE value and the similarity-discounted SUE value based on DDs during the iterative selection process. This computationally simple approach is well-suited for JEWEL, which will need to run on-board OWLS under the computational constraints of a radiation-hardened flight processor. In additional to the MMR algorithm that balances utility and diversity, JEWEL also allows scientists to use standard bin-based prioritization if certain observations are known *a priori* to have higher downlink significance, and enables scientists to specify data volume constraints on individual instruments so that observations from a single instrument do not block data from another instrument, even in cases where utility might be higher. Thus, the data volume constraints ensures that JEWEL selects products that are diverse not only within instruments, but also across instruments.

Evaluation: To evaluate JEWEL, we compare its prioritization to that produced by other baseline approaches for ranking data. We use a dataset comprised of 12 DHM observations processed by HELM from both laboratory samples and samples acquired from Newport Beach, CA, and 7 samples processed by ACME that were prepared in a laboratory. We compare to two baseline strategies: random ordering and first-in first-out (FIFO) ordering, in which the ASDPs are downlinked in the order they were produced (which is more-or-less arbitrary in this case). For the MMR-based JEWEL algorithm, the trade-off parameter is set to 0.0 and 1.0, which ignores diversity, or completely relies on the diversity-discounted values based on DD similarities, respectively. For the random strategy, we compute the performance across 1000 random trials. Our evaluation metric is cumulative utility, the total utility as a function of data volume.

Results: Figure 1 shows the cumulative science utility as a function of the cumulative amount of data downlinked, prioritized by each of the three approaches compared. For the random downlink approach, the average across 1000 trials is plotted along with the standard deviation to show the variability across trials. A square and circle marker indicate where in the downlink the first ACME and HELM ASDPs appear for the JEWEL and FIFO prioritization approaches (one of these always occurs at the far left-hand side of the plot).

These results show that as expected, JEWEL selects the products with the largest per-byte SUE values first, so it achieves a larger cumulative utility faster than the other approaches. The other algorithms do not incorporate SUE values, so the largest increase in utility either comes later in the downlink for FIFO, or is spread out evenly for random downlink ordering. Furthermore, comparing the multiple runs of JEWEL that either ignore the diversity-based prioritization or rely entirely on the diversity-discounted SUE values (“max. diversity”), there is only a small difference in the ordering produced in this case. However, the first HELM product is downlinked sooner when taking diversity into account, likely because the *marginal* utility of the final ACME product is much less than its initial SUE. This example illustrates the benefit of accounting for diversity during prioritization.

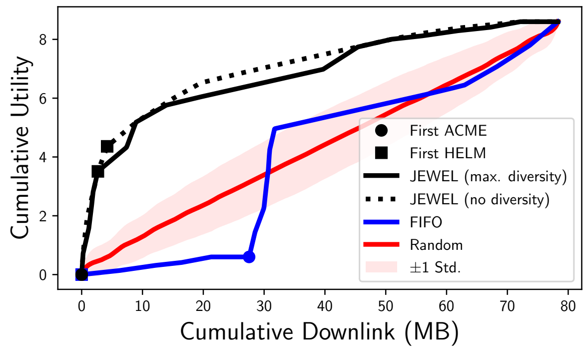

Figure 1: A comparison of the cumulative utility of selected data products using JEWEL, random, and FIFO prioritization algorithms. The square and circle indicate where the first of each type of data product appear in the downlink.

Conclusions: These experiments demonstrate the importance of content-based prioritization in missions to the outer solar system. Accounting for both diversity and utility is important to ensure that the most scientifically relevant data is returned to Earth as quickly as possible. Both factors are essential for platforms like OWLS that have the ability to collect much more data than is possible to transmit. In the future, we will consider alternative prioritization strategies to MMR, explore the incorporation of additional instruments and cross-instrument dependencies, and perform evaluations with scientists to measure their agreement with the prioritization of selected data products.

Acknowledgments: This research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. ©2021 California Institute of Technology. U.S. Government sponsorship acknowledged. Special thanks to Jake Lee, Jack Lightholder, Lukas Mandrake, and the rest of the OWLS team for providing data and software used by these experiments.

References: [1] Deutsch, L., Townes, S. (2016) *JPL White paper*. [2] Nadeau, J., et al. (2017) *Advances in Physics: X*. [3] Carbonell, J., Goldstein, J. (1998) *ACM SIGIR*.