An Experiment With Spectral Analysis Automation

Walt Truszkowski
NASA – Goddard Space Flight Center
Code 588
walt.truszkowski@nasa.gov

Mike Rilee
L-3 Communications
michael.l.rilee.1@gsfc.nasa.gov

Sidney Bailin
Knowledge Evolution, Inc.
sbailin@waves.kevol.com

ABSTRACT

This paper addresses an experiment with the concept of Spectral Analysis Automation (SAA). The long-term fully-realized SAA system will be a multi-agent system that is designed to provide automated support for two major functions: (1) the automatic remote filtering (on-board a spacecraft or robotic device) of spectral image data based on Principal Investigator (PI) guidance, goals and science agenda and (2) the packing and transmission of the selected spectral data to the PI for further processing. The focus of the research to-date has been to develop and evaluate a proof-of-concept SAA infrastructure. This paper will provide an overview of the multi-agent SAA infrastructure that has been prototyped. The near-term goal of the SAA project is to make operational that portion of the SAA system that can support the automated data mining of spectral data archives under PI guidance. The NEAR (Near Earth Asteroid Rendezvous) mission archive of spectral images will serve as an initial testbed for this aspect of the SAA activity. Both a neural net and a Bayesian filter have been developed for the near-term prototype of the SAA. These filters will be discussed. The longer-range goal is to have the SAA support real-time on-board spectral data filtering, selection, packing and transmission. The fact that the SAA is designed as a multi-agent system will provide the innovative flexibility that will be required to realize the progressive and adaptable autonomy needed for both the stated near-term and long-range goals. Additionally, the innovative multi-agent-based infrastructure for the SAA can be generalized in a way to enable it to support the type of progressive autonomy that will be needed to support an adaptive and growing autonomous behavior for other spacecraft or robotic subsystems (in addition to the subsystem dealing with onboard science data processing). This paper will address all these aspects of the SAA work.

1. OVERVIEW OF THE SAA ARCHITECTURE

We have developed a multi-agent-based architecture for filtering science data on-board a spacecraft prior to download, so as to maximize the efficient use of communications resources between the spacecraft and the ground. The architecture is depicted in Figure 1.

The flow of information in the filtering architecture is as follows. Data arrive from the spacecraft instrument and subsystems in the form of packets, which are assembled periodically. The period is called a Data Gathering Interval (DGI), and by an abuse of language we refer to the packet itself as a DGI too. A DGI contains spectral data from the instrument, as well as engineering data pertaining to both the instrument and the spacecraft, and tracking and ranging data to assist in the interpretation of the spectral data.

Each incoming DGI is placed in a database. The exact form of this database—e.g., whether it is stored in RAM or in a persistent storage device, whether it provides Database Management System (DBMS) functionality, etc.—is an open issue. The purpose of the database is to enable the filtering functions to consider DGIs in the context of other DGIs when deciding which of them should be downloaded. In addition, the database serves as a queuing area pending a downlink pass.

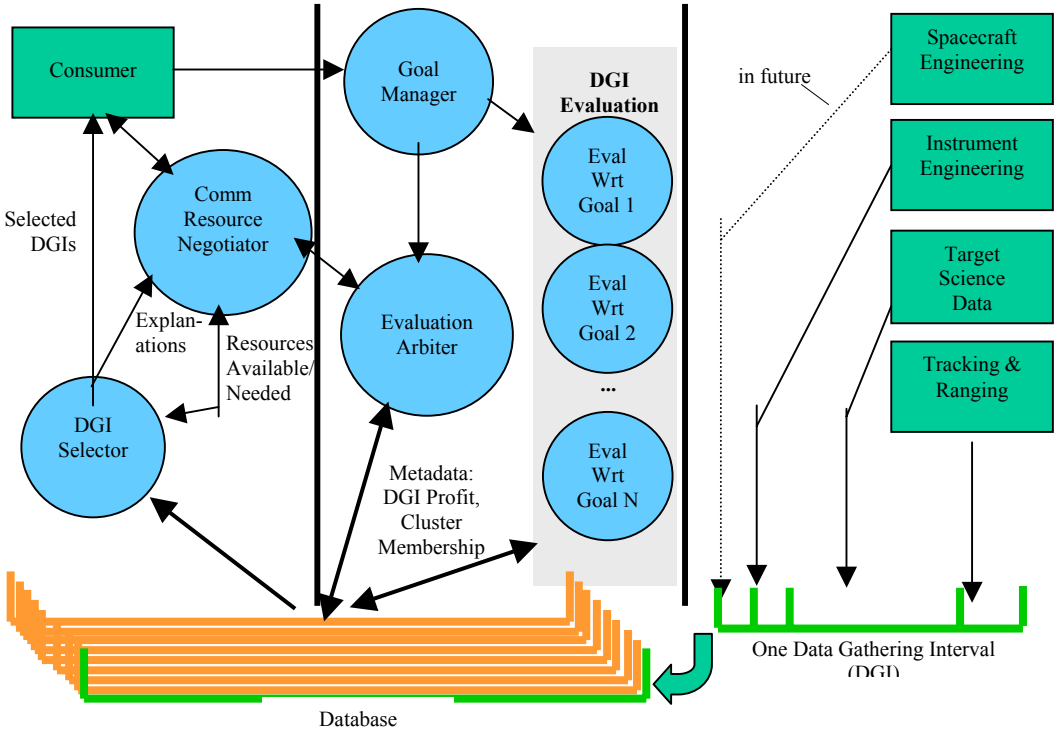

Figure 1. Spectral Analysis Automation Agent Filtering Architecture

When a DGI is placed in the database, several agents are notified about this event:

- Evaluation agents
- Evaluation arbiter

There is an Evaluation Agent for each mission goal, as defined by the Consumer. The Consumer may be the

science user on the ground; alternatively, it might be a supervisory or intermediate communications spacecraft. The Consumer conveys the goals, and their relative priorities, to the Goal Manager (GM). The GM is responsible for activating and deactivating the appropriate evaluation agents, and for communicating goal priorities to the Arbiter.

When a new DGI arrives in the database, each evaluation agent assigns a profit—that is, a measure of value—to the DGI. In the process, it may also revise its previous profit assignments to earlier DGIs. The Evaluator may also define clusters of DGIs and assign a profit to the entire cluster, meaning that the individual DGIs derive their value only in the context of the rest of the cluster. The Evaluators contribute their information by tagging the DGIs with metadata indicating profit (with respect to a particular goal), cluster membership, and potentially other forms of information. This approach provides a great deal of flexibility in the kinds of information that may be contributed by the Evaluators. (The current prototype uses only profit assignments to individual DGIs, although the mechanism is in place for recording cluster information.)

When the Evaluators have all finished evaluating the new DGI, the Arbiter derives an overall profit value for the DGI on the basis of the "votes" provided by the Evaluators. In the current prototype, several algorithms are available to the Arbiter to derive the overall profit value. The relative merit of the algorithms is a topic for further experimentation and analysis.

When a downlink pass occurs, the Selector agent uses the Arbiter's profit assignments to decide which DGIs should be downloaded to the Consumer. The Selector may simply download the DGIs in order of their profit values, until the capacity of the communications channel (and/or the time period of the pass) are exhausted; alternatively, the Selector may trade off the profit of a DGI against its size (also called the DGI's *weight*) in order to maximize the overall profit of the downloaded information. There are numerous issues concerning the utility of the science data that arise when trading off profit against size, and these are a topic of continued investigation.

One of the ways in which the tradeoff can be mitigated is by enlarging the capacity of the communications link. This may be appropriate, for example, if the recent DGIs indicate that large amounts of valuable science data are being collected. In such cases, the Communications Resource Negotiator may request additional bandwidth from the Consumer. The request is supported by information provided by the Arbiter, the Goal Manager, and the Selector concerning the value of the science data and the potential losses if the communications resources are not increased.

A summary of the roles of the various agents and other entities is provided in Table 1.

Table 1. Each Agent In The Filtering Architecture Has A Well-Defined Role

Agent	Role
Consumer	Entity for whom filtered spectral data are intended. Could be scientist on ground, or intermediate spacecraft in swarm
Information sources	Origin of data to be filtered and downloaded. Includes spacecraft engineering data, instrument engineering data, target science data, and tracking & ranging data

Data Gathering Interval (DGI)	One batch of source information. Collected over (and representing) a particular time interval
Database	On-board store of DGIs. Staging area prior to downlink of selected DGIs. Memory limited
DGI Evaluators	Agents responsible for assigning "profit" value to DGIs. Profit may be assigned to individual DGI or a cluster of DGIs determined by the evaluator. Evaluators may consider any or all of the current database contents, e.g., in light of most recently stowed DGI, or backing up to reconsider a previously stowed DGI. Evaluators output a profit for one or more DGIs, possibly in the context of other DGIs (i.e., requiring their presence too). In the simplest case (maybe sufficient) the assignment is to the latest DGI, by itself
Goal Manager	Agent responsible for creating, configuring and prioritizing the Evaluators on the basis of goals specified by Consumer
Evaluation Arbiter	Agent that arbitrates between conflicting profit assignments. Each evaluator represents a specific goal. Arbiter tries to balance the goals to derive an overall profit for each DGI and/or cluster. May query Communications Resource Negotiator about bandwidth possibilities. The resulting profit assessment summarizes the results of the Evaluation Arbiter. If context is used, this may be a complex data structure. Also, if context is used, it is an open issue what the DGI Selection algorithm should be (even Knapsack algorithm, which trades profit against weight, may not suffice)
DGI Selector	Agent responsible for choosing DGIs for downlink to consumer. Trades off profit against weight (= size of DGI) using one of several possible algorithms. Tries to produce downlink set of maximal usefulness given limited bandwidth. May request bandwidth change from Communications Resource Negotiator, or explain selection decisions in light of available bandwidth (as support info for negotiation). DGI selection uses the arbitrated profit assessment plus weights (sizes) of DGIs in database plus available communications resources to select DGIs to send to consumer. If weight is constant (i.e., constant-length DGIs), a simple "shop-til-you-drop" algorithm—highest profit DGIs first—suffices. If compression is used, Knapsack algorithm may be required to obtain maximal aggregate profit of the download. If individual DGIs are *not* assigned profit (i.e., complex context is used), this is an open issue
Communications resource negotiator	Agent that negotiates for downlink bandwidth. Interacts with Arbiter and Selector to stay informed of status and downlink needs. Resource negotiation may include negotiation of futures, e.g., "We're having a good day…" or "I'm especially interested in feature X…"

2.0 A CLOSER LOOK AT DGI EVALUATIONS - FILTERS

To date, two spectral data filters have been developed for DGI evaluation : a neural net filter and a Bayesian filter. These two filters were implemented and are being integrated into the SAA system. The following subsections provide brief overviews of these two filters.

2.1 NEURAL NET FILTER

This image (Figure 2) is a snapshot of taken from the NEAR X-ray spectrometer (XGRS) data analysis system. It contains two datasets. The bottom is a spectrogram of 350+ consecutive spectral integrations taken by the NEAR X-ray spectrometer. The x-axis represents increasing time, with a total time of about 6 hours of 1 minute integrations. The energy increases as you go up the y-axis. Brighter colors indicate higher levels of intensity in spectra relative to the rest of the image. The top plot is the output generated by a neural network that classifies each spectral integration (DGI) as to whether flourescence is detected in each spectrum. A "0" indicates that flourescence was detected and a "1" indicates background only. The x-axis of both the spectrogram and plot is directly correlated, so there is a vertical relationship between what one sees in the spectrogram and the neural net output directly above.

In this image NEAR was taking 1-minute integrations as it orbited between the 35 and 50 km from the 433Eros surface. During this time two solar flares occured, which hit the asteroid surface. Fluorescence was detected by the NEAR X-ray spectrometer as the two bright spots at the lower portions of the spectrogram. In the plot above a large sequence of "0"'s - fluorescence classifications were generated by the neural network that identified the high signal generated by each flare and tracked them until they subsided. Data

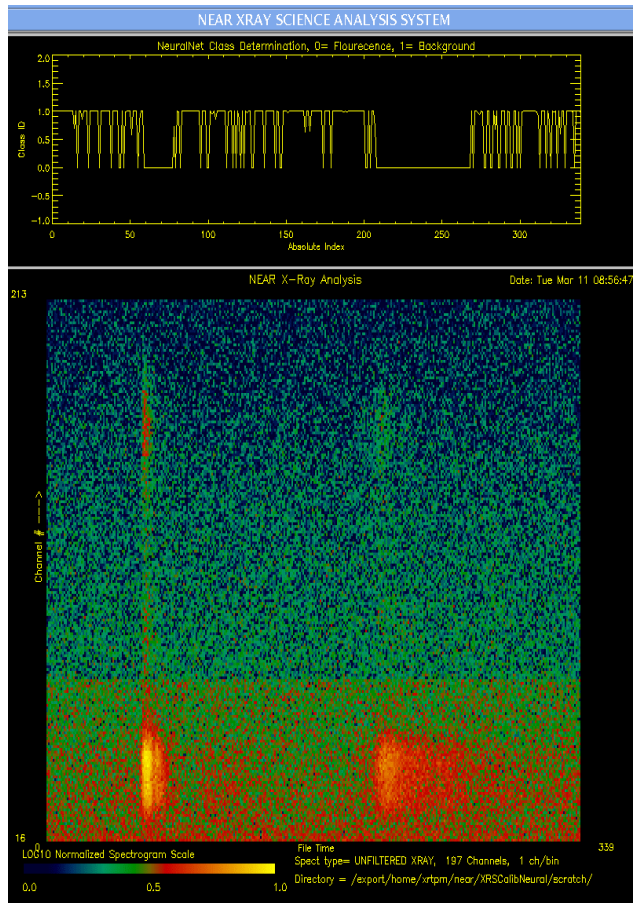

Figure 2. Example of Neural Net Filter

outside these two regions generally contains lower signal /noise implying lower scientific value to fluorescence studies of the asteroid surface. This automated classification is currently being done on the ground as a data mining activity, but could also be performed

in-situ to prioritize scientifically important information.

The intent of this system was to design an automated data mining mechanism to browse the 500,000 spectra that were collected by NEAR XGRS and attach a posteriori scientific value to each spectrum. The system is a 3 layer, backpropagation neural network, trained on 2 classes of data, Fluorescence "0" and Background "1". The Fluorescence class was obtained from several solar flares at different times of the mission. Background data was obtained during several times when NEAR XGRS was not pointed at the asteroid. The feature set is the contiguous set of channels in the unfiltered detector from Channel 20 to 51. We plan to use this system in ongoing NEAR data analysis activities

2.2 BAYESIAN FILTER

The modular construction of the SAA allows a great range of functionality to be added, particularly in the evaluation stage where a data's relevance to a mission goal is assessed. To demonstrate this range and flexibility a Bayesian Filter (BAF) has been developed that is based on an entirely different evaluation scheme than our Neural Net Filter (NNF). Even at this early stage of development, the SAA enables the comparison of different and important approaches to science data. The NNF contains information about system performance that it acquired during a training period. The BAF allows us to explore the use of *a priori* information and "experience" to answer quantitative questions about the data. For example, because of the tie between statistical inference and information theory, we should be able to quantify the rate at which knowledge about a particular object is being obtained. Thus our explorations with the BAF should lead to a quantitative method for determining when the further data acquisition will likely produce no further results. Going beyond ad hoc approaches and over general experiment design, the framework of Bayesian statistics provides a disciplined way to assess old and new information while adapting to changes in data or models [1].

The BAF is essentially a signal detection algorithm extending previous work on the detection of weak signals in radio spectra [2]. A priori knowledge about X-ray detectors, the NEAR/XGRS, and features of the background spectra are combined to construct a likelihood function that depends on the strength of spectral line and background emission. A model comparison is formed as an odds-ratio to find the most likely model given the data. Figure 3 shows a model that encapsulates a small amount of knowledge about the structure of X-ray spectra and is based on a model of X-ray background continuum and fluorescence line emission. For this example, a weak signal was deliberately chosen to show how the method works in the weak-signal regime (low signal-to-noise ratio). Figure 4 shows how the model in Figure 1 was chosen from a whole family of models indexed by the (hypothesized) strength of the spectral line. Figure 5 shows how the strength of X-ray fluorescence lines depends, among other factors, on the strength of solar X-ray flux. Preliminary results show how the filter assesses the odds for non-zero signals. Standard techniques for assessing the existence of a fluorescence

line in the data can handle neither weak signals, nor the transition from weak to nonexistent signals. The BAF provides a systematic way to plumb the weak-signal depths of data, opening up the disciplined study of marginal data in a way that can point towards reconfiguration of mission operations towards science opportunities that would otherwise be missed. Thus the impact of work on SAA and BAF extends beyond the field of X-ray, or more generally spectral, analysis, but across a wide range of observational sciences that deal with weak signals in noisy environments for which some a priori information is known.

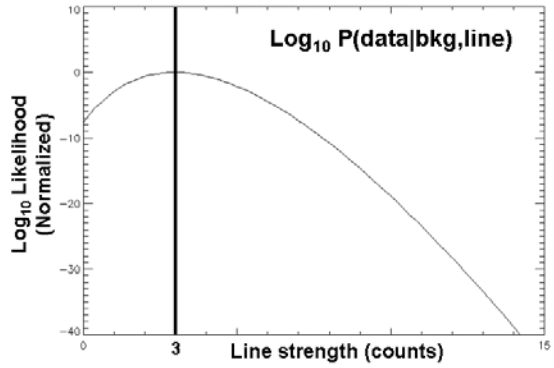

Figure 4. Likelihood of models given the data. Most likely is the model indexed by a strength of 3.

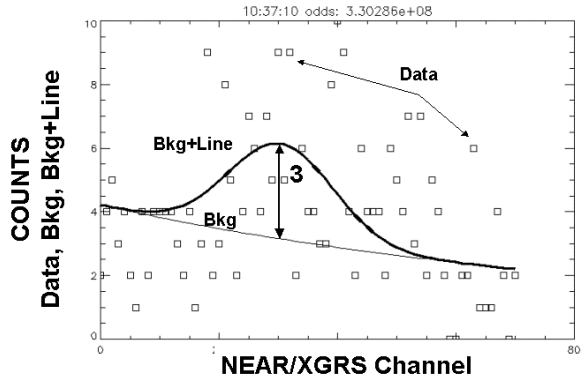

Figure 3. Measurements and model of x-ray line + background.

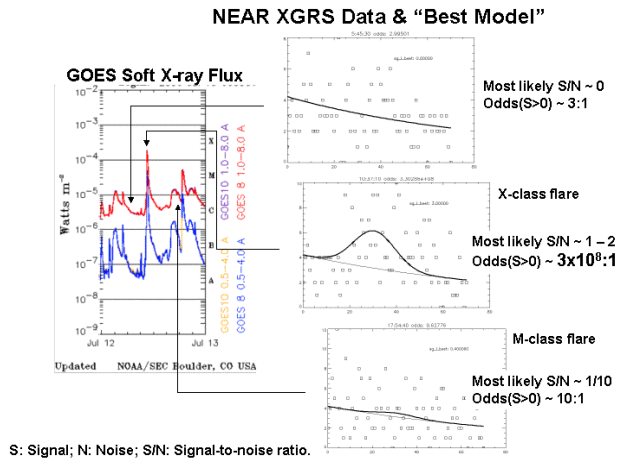

Figure 5. Models and maximum likelihood odds of signal-to-noise ratio.

3.0 GENERIC MULTI-AGENT FRAMEWORK

The power of the SAA concept lies in the innovative use of multi-agent communities. The PI has the capability of embedding his/her science data filtering goals into an agent which migrates to the community of agents engaged in spectral data filtering. As this community develops over time, the sophistication of the filtering process can grow. This allows the SAA to improve its performance over time and to increase the level of autonomy of its science data processing.

This same type of reasoning can be used to instrument the idea of progressive autonomy for other subsystems on the spacecraft or robotic device. This type of generic multi-agent framework for realizing progressive autonomy will be a future study.

4.0 CONCLUSIONS

A prototype of the SAA system architecture has been developed and demonstrated. Further work is planned on fleshing-out the SAA infrastructure over the coming year. It is planned that this SAA system will eventually be of use for both ground-based and space-based spectral data filtering.

ACKNOWLEDGEMENTS

The authors wish to acknowledge the contributions of Pam Clark, Tim McClanahan, Victoria Yoon, and Jay Karlin

REFERENCES

[1] Loredo, T. (1992). The promise of Bayesian inference for astrophysics. In Feigelson, E., & Babu, G. (Eds.), Statistical Challenges in Modern Astronomy. Springer-Verlag.

[2] Rilee, M.L. and J. L. Green, Automated Detection of the Magnetopause for Space Weather from the IMAGE Satellite, in Proceedings of SPIE's 14th Annual International AeroSense Symposium, April 2000.