Adaptive Envelopes for Engineering Data
Continuation Proposal for 632-07

Task Leader

Dennis DeCoste, JPL, decoste@aig.jpl.nasa.gov, 818-393-5366, PI

Product Description

The target product is algorithms for learning, adapting, and validating prediction models from historic engineering data. The primary deliverable is software (called ELMER) which takes multivariate time-series data (e.g. sensed and/or simulated) as inputs and automatically learns and outputs functions which define context-sensitive high and low limits on each quantity. ELMER discovers envelopes which are as tight as possible while still avoiding data falling outside of those bounds. It will also determine whether its training data is applicable to new data and adaptively adjust learned envelopes as required to track system drift and degradation.

In contrast to related time-series prediction work, this work emphasizes modeling uncertainty in terms of interval-valued upper and lower bounds (i.e. envelopes). Envelopes are well-suited for a wide variety of NASA onboard and ground operations contexts, including fault detection (e.g. FDIR for Deep Space Network), event detection for engineering data summarization (e.g. during Beacon Operations), resource best-case / worst-case profiling (e.g. to support rover resource planning), and general engineering data-mining (e.g. discovering unusual and poorly understood behaviors, such as thermal snapping of Origins Program interferometer engineering structures). For example, in the case of fault detection, envelopes provide context-sensitive functions which generalize the static red-line limits traditionally used currently in most NASA monitoring operations.

*This is a **continuing**, **push** task (with about 20% pull at this point).*

The following table gives some expected quantitative improvements and comparisons in capability that this work will provide over state-of-the-art (e.g. red-line fault detection):

metrics	state-of-art	**FY00**	**FY01**
average predictive precision (normalized mean-error of fit)	*0.5*	*0.1*	*0.05*
detection coverage (percentage of faults detected within 1 day of 1st symptom)	*25%*	*50%*	*90%*
average false alarm rate (per sensor)	*1 / week*	*1 / week*	*1 / month*
average detection delay (since early symptoms of fault, per sensor)	*1 day*	*1 hour*	*60 secs*
CPU time requirements (operations per second per sensor)	*10*	*1000*	*10,000*
RAM space requirements (bytes per sensor)	*16*	*128*	*512*

[This table illustrates typical good use of each technology; naturally some metrics can be traded-off (e.g. higher false alarm rates for higher detection coverage). Assumes a typical 1Hz sampling rate, 100's of sensors, and data similar to DS1 Mission. These metric improvements could reduction annual operations by at least 1 FTE per mission (> $150K savings).]

Benefits

Autonomy in NASA space systems requires automated reasoning about engineering data, such as detection of faults and prediction of resource availability over time. Robust reasoning requires adapting underlying models during a mission, to reflect both unexpected environmental conditions and system degradations. Key benefits of our work are:

- *increase science through-put, by enabling more optimal planning and FDIR decisions,*
- *reduce operations costs, by detecting faults and trends earlier and automatically,*
- *reduce knowledge-engineering costs, by automated learning of models from simulation, testbed, or mission data,*

- *automate model testing and validation, via the training process itself (e.g. cross-validation over simulation data).*

ELMER enables both ground and onboard operations which are lower-cost, more autonomous, and more reliable. For example, cost savings within an FDIR fault detection context would come from supplementing traditional red-line limits with functional limits learned from ELMER. This would reduce the need and costs of manually defining and maintaining red-line limits. Within NASA operations contexts, red-lines are typically wide static values that particular sensors are expected to never exceed, like a dangerously high pressure or temperature. In contrast, ELMER's limits are context-sensitive and tight, allowing automated detection of developing faults much earlier. By providing such earlier detection in a comprehensive and automated fashion across all sensors, ELMER would extend expected mission life with effectively no additional cost. Furthermore, such comprehensive fault detection enables even more aggressive and autonomous mission scenarios than NASA missions can currently afford using red-lines, leading to drastically higher science throughput.

Technical Approach

We view ELMER as a collection of techniques for performing the task of bounds estimation, *as opposed to traditional regression techniques which emphasis means estimation or general probability density estimation. We have formulated several methods for bounds estimation and have been exploring their various tradeoffs as well as comparing them to traditional techniques. Common to all our techniques is the notion that bounds estimation's key distinction is that it involves a special form of constrained optimization. In particular, a prediction from a learned high bounding function should not only be as close to the training target value as possible, but also strictly above that target.*

Methods we have coded and have been evaluating can be divided into two broad classes:

- ***memory-based methods*** *--- such as a variant of k-nearest neighbors, where the high (low) prediction is essentially the max (min) over the target values of the neighbors (as opposed to the mean, as in traditional nearest-neighbors regression).*
- ***model-based methods*** *--- such as a variant of traditional error bars, where an input-conditional mean estimation is used to divide the training set into "target is above the mean" and "target is below the mean" subsets. A prediction model of the error residual for each subset is learned, allowing error distributions which are asymmetric around the mean to be easily handled. The final learned high bounding function is H(X) = Hb + Hs * M(X), where M is the mean estimation, Hb is a bias shift value and Hs is a scaling factor playing a role similar to a standard deviation factor in a Gaussian model.*

Making such methods generalize well to future data requires further details beyond the basic sketches given above. For example, one way to determine reasonable values for the Hs and Hb parameters above is to use a simplex method of constrained optimization (over only those two variables, for fixed M(X)) and extensively use cross-validation to select the widest fit (i.e. highest Hb and Hs values) that any 90% subset of the training data requires to avoid alarms on 10% hold out sets. Similarly, for methods such as min/max k-nearest neighbors, the best k value and suitable shift offsets to bound all data can be determined via cross-validation as well.

The memory-based bounding methods have some key advantages, including often being more readily understood by humans during post-detection (e.g. diagnosis) analysis, since their bounds violations are grounded in terms of specific previous sensor behavior examples. We have focused more to date on model-based methods, mainly because onboard applications, including the DS1 Beacon experiment, have tight space constraints that preclude on-line access to vast historic databases. However, we are beginning to explore, within our bounding context, appropriate ways to combine both approaches. One such approach involves support vector machines, which identify subsets of examples which are most valuable for retaining in memory.

An important property of ELMER is that it is very scaleable with respect to the number of available sensor inputs, the available training data, the computational time available for learning and adaptation, and the real-time memory and CPU restrictions for representing and computing final bounding functions. It finds the best bounds it can with whatever it is given (even if that results in almost static red-line bounds at that point), and can incrementally improve bounds as more is given later.

Related Work

It is useful to view ELMER as a generalization of the static red-lines traditionally used in NASA fault monitoring operations. ELMER's bounds are intended to work just like red-lines in that data outside of those bounds should be suspect. A key problem with red-lines is that attempts to avoid "nuisance alarms", where red-lines are excessively tight, easily leads to red-lines that are much too wide to detect faults until very late (and often critical) stages. Indeed, our work on ELMER arose from attempting to better capture the context-sensitivity of the domains, for earlier detectability, while not making the strong error distribution assumptions that common statistical error bars approaches do. Note that data outside of error bars based on the mean plus or minus two standard deviations will still occur about 5% of the time. That is not acceptable for large-scale monitoring tasks, for which thousands of sensors are sampled every second.

The fundamental problem is that for complex engineering systems such as spacecraft, the error in achievable predictions based on available sensor data is not primarily Gaussian, nor any other kind of standard distribution. Even when the sensor data is sufficient to find a deterministic (plus small Gaussian white noise) model, from a practical point of view that does not help much if the step-wise regression technique being used has not yet selected all the right inputs, out of the thousands of (raw and transformed) candidates to consider. Furthermore, the mission's memory and CPU limitations might well require limiting each function to a handful of the relevant inputs.

A key distinction between ELMER and other machine learning technologies is that it learns and defines high and low bounds independently and without assuming a specific prediction error distribution. Other techniques, such as neural networks which learn "error bars" (e.g. estimates of the mean of the data as well as the variance of the data [Nix 1994]), assume specific types of distribution of error, often symmetric (e.g. Gaussian). ELMER handles well such asymmetric error distributions, which are common in spacecraft data (due to engineering set-points and other skewed behavior). The end result is that ELMER can produce tighter bounds which lead to better detections and trending predictions.

There does exist a class of techniques, called probability density estimation (PDE) (e.g. [Weigend 1995]), which, like ELMER, avoid the problems of assuming any specific class of error distributions. Conditional probability densities explicitly represent the probability of each possible output value, given the inputs. For example, instead of assuming that error is distributed as a single Gaussian, a PDE approach such as mixture density estimation might use hundreds of Gaussians of varying parameters (i.e. centers and widths). With sufficiently large mixtures, any distribution can eventually be modeled to any arbitrary precision using such PDE.

However, the generality of PDE is both its strength and its major weakness. To learn the parameters of the mixtures well typically requires orders of magnitude more data than the single regression that ELMER requires for each bound. Similarly, PDE's with hundreds of Gaussians are orders of magnitude more expensive to store and compute at execution time, making them much more expensive than ELMER to use for tasks such as real-time monitoring.

In short, PDE promises more than is necessary for tasks that only require bounds, and delivering on those promises requires excessive resources at both training and execution. Thus, we argue that ELMER's explicit focus on estimating bounds is more appropriate for many tasks, such as monitoring and resource profiling. For some tasks, most notably control, invertable models are critical. For such tasks, PDE of some precision is generally required. One planned extension for ELMER is to generalize it with PDE capability, so that in an anytime fashion it finds the best trade-off for a given task between high/low bounds and full precision PDE.

Future Work

Despite our initial promising results this year, a couple of key issues must still be addressed to mature this technology for practical applications.

First, ELMER needs to be extended to allow it to determine at runtime when the current (test) data is so dissimilar from the training data that the previously learned bounds are not applicable. For example, autonomously detecting such situations is required to avoid false alarms when the test context is radically different from the training

scenarios --- such as training during cruise phase of a mission and testing during orbit-insertion phase. In a general probability density estimation approach, such determination could be directly performed by evaluating some (previously learned) joint density estimate for the current values of all the input sensors. That is, a small likelihood in the conditional probability of some quantity should not itself be the cause for declaring a fault detection when the inputs for its estimation are in fact themselves very unlikely. Capturing this distinction between inputs joint probability and output conditional probability sufficiently for the goals of bounds estimation, without incurring the full cost of general density estimation, is the goal for this extension. An advantage of memory-based approaches to bounds estimation mentioned earlier is that their use of distance metrics between test and training data already provides some such distinctions (i.e. a nearest-neighbor which is still relatively far away could be an indication that the training data is insufficient to confidently bound the new data).

Second, ELMER needs to be extended to support robust on-line adaptation of bounding functions in light of new data during a mission. This capability is required to track non-stationarities due to system drift and degradation, as well when environmental conditions turn out to be different that initial expectations (e.g. ground testbed for Mars rover). To date, our work has focused on learning envelopes using batch training, due to its simplicity of implementation and evaluation. Also, addressing the issues of when to adapt and what portions of a model to retain requires first addressing the above issue of detecting significant differences between previous (training) and new (test) data.

Another goal for the ELMER work is to more formally incorporate an ability to learn probabilistic graphical models (e.g. Bayesian networks) from the data. ELMER currently uses basic concepts from the field of Bayes nets, such as partial correlations, to (heuristically) identify useful inputs for each bounding function. Useful extensions would include refining Bayes net algorithms for learning causal (directed) structures so that they work well within the bounds estimation framework of ELMER.

Status and Milestones

Status for FY 1999:

- *Trained envelopes from down-linked data and tested onboard DS1 mission during the DS1 Beacon Experiment (experiment declared "success" in June of 1999).*
- *Adapted our enveloping techniques to the task of resource profiling [DeCoste 1999].*
- *Developed a variety of new methods for bounds estimation and investigating tradeoffs (time, space, test errors).*
- *Evaluating (off-line, in lab) on archived data from multiple domains (DS1, IPEX / SIM, DSN, Rocky 7 rover).*

FY 2000 Milestones:

- *Develop effective methods for detecting the degree to which previous (training) data is relevant to current (test) data, without incurring the (often prohibitive) expensive of detailed probability density estimations. Test on relevant data (e.g. DS1 and IPEX).*

- *Extend current (mostly batch) algorithms to be more adaptive (i.e. on-line), suitable for onboard adaptation. Test on relevant historic data sets (e.g. DS1).*

FY 2001 Milestones:

- *Deliver robust envelope learning software package, suitable for adaptation by multiple NASA missions and tasks.*

Customer Relevance

This work is relevant to Level 1 Technology Requirements for all three Enterprises, such as:

- ***HEDS*** ---
- *H1 (intelligent FDIR for autonomous operations),*

- ***Earth Science*** *---*
- *E4 (lower operations cost, through automated fault detection),*
- ***Space Science*** *---*
- *S1 (e.g. data-mining testbed and mission data, to understand structure disturbances, such as thermal snapping, which impact interferometer science, i.e. SIM missions),*
- *S3 (e.g. resource profiling, e.g. for use in autonomous rover resource planning).*

Letters of support from the following two people are included with this proposal:

- ***Dr. Marie B. Levine****, Program Element Manager, Microdynamics; Space Interferometer Mission (SIM), JPL.*

- ***Dr. Daniel Dvorak****, Technical Lead for MDS State Determination, Mission Data Systems (MDS) Program, JPL.*

PI Qualifications

Dennis DeCoste is Senior Member of Technical Staff / Technical Group Leader in the Machine Learning Systems Group at the Jet Propulsion Laboratory. He received his Ph.D. in Artificial Intelligence / Computer Science from the University of Illinois at Urbana-Champaign in 1994. He has been technical lead of several projects in fault detection and time-series data mining at JPL for the last five years. He has served on the AAAI program committee and as reviewer for AIJ, IEEE PAMI, AAAI, and IJCAI. He has published several papers on detection and data-mining (e.g. AAAI, AI Journal, AI Magazine, KDD).

Technical References

[Bishop 1995] *Chris Bishop.* Neural Networks and Pattern Recognition, *1995.*

[DeCoste 1999] *Dennis DeCoste*. Adaptive Resource Profiling. *In Proceedings of Fourth International Symposium on Artificial Intelligence, Robotics, and Automation for Space (i-SAIRAS '99), June 1999.*

[DeCoste 1997] *Dennis DeCoste.* Mining Multivariate Time-Series Sensor Data to Discover Behavior Envelopes. *Proceedings of the Third Conference on Knowledge Discovery and Data Mining (KDD-97). August 1997.*

[Nix 1994] *David Nix and Andreas Weigend.* Learning Local Error Bars for Nonlinear Regression. *NIPS 7, 1994.*

[Weigend 1995] Andreas Weigend and Ashok Srivastava. *Predicting Conditional Probability Distributions: A Connectionist Approach..* International Journal of Neural Systems, Volume 6, 1995.

June 24, 1999

NASA Cross-Enterprise Technology Development (UPN 632-07) Thinking Space Systems Thrust

Subject: **Adaptive Envelopes for Engineering Data Task, Automated Fault Detection and Data-Driven Modeling Technology**

Reference: Letter of Support

The purpose of this letter is to indicate my strong support for the proposal on *Adaptive Envelopes for Engineering Data* by Dr. Dennis DeCoste. The proposed work will provide an important capability for use by customers of the Mission Data System (MDS).

There are many approaches to fault detection, but this one is particularly attractive for two reasons. First, the concept of envelope functions can be readily understood by flight project engineers as an improvement over static alarm limits, and the way in which the functions are derived is straightforward. That's important because flight projects *don't* fly technology that they don't understand and trust. Second, the fact that the envelope functions are learned from actual measurements—rather than engineered through laborious design work—makes this technology relatively easy to deploy. That's important because it helps projects be faster-better-cheaper in an era of tight schedules and tight budgets. In addition, this technology looks promising as a way of detecting unusual behavior during system testing.

Thus, I strongly support funding this task since it is directly relevant to the task of state determination in MDS and can contribute to the success of MDS-based flight projects in the Space Sciences Enterprise.

Dr. Daniel Dvorak
Technical Lead for MDS State Determination
Mission Data Systems (MDS) Program, JPL

June 25, 1999

NASA Cross-Enterprise Technology Development (UPN 632-07) Thinking Space Systems Thrust

Subject: **Adaptive Envelopes for Engineering Data Task, Automated Engineering Data-Mining and Data-Driven Modeling Technology**

Reference: Letter of Support

The purpose of this letter is to indicate my support for the work in the development and deployment of Automated Engineering Data-Mining and Data-Driven Modeling Technology being funded by the Adaptive Envelopes for Engineering Data Task, led by Dr. Dennis DeCoste at JPL and funded by the NASA 632 Program.

A central task in the Interferometer Program is shifting through enormous volumes of engineering flight and ground test data, to gain better understanding of the small-scale stability characteristics of space structures. Solid understanding of this relatively understudied field of microdynamics is necessary for better design, operations, and science scheduling for important future interferometer Space Science missions (e.g. the Origins program).

Automated data-mining and data-driven modeling technology represents a major component for a semi-automated approach to analyzing such large volumes of data. For example, in a single 50-hour Space Shuttle payload experiment (IPEX-II), we gathered over 10 gigabytes of sensor data, of which only about 5 minutes of data has been extensively (manually) analyzed to date. The understanding and better predictive models that can be obtained from more comprehensive (automated) analysis will likely lead to greatly improved science throughput and better (e.g. preventative) operations.

Thus, this work is directly relevant to the success of the Origins Program of the Space Sciences Enterprise and I strongly support funding this task.

Dr. Marie Levine
Space Interferometer Mission
Program Element Manager, Microdynamics,
JPL