Automated Data Analysis for Operator Modeling

JC LeMentec (le_mentec_jc@cat.com)
Caterpillar Inc.
Technical Center, Bldg. E
P.O. Box 1875
Peoria, Illinois 61656-1875 USA

Abstract

IDA is a new architecture that combines data analysis and operator model creation and simulation. It enables the development of operator models by finding causality patterns in recorded machine operation data. A new symbolic constraints based reasoning system is used to accurately and robustly segment data. These segments can then be used to discover an operator strategy expressed as hierarchical set of Tasks.

Keywords: Operator Modeling, Cognitive Model, Data Analysis, Data Segmentation, Autodig, Wheel Loader.

Introduction

To assess the design of a new machine design, Caterpillar relies on an interactive simulation of machine, soil and human operator. The operator model must reproduce the levers, steering and pedals commands in response to environmental inputs, as a human operator would do when performing particular tasks.

The expertise for conventional cognitive models is usually obtained through an interview process with domain experts. While still useful for operator models, interviews alone are not sufficient as a large part of the behavior is unconscious. Fortunately, operator activities can be scrutinized in great details through recorded data of machine operation. Machines are routinely instrumented with various sensors, and a large amount of data is available.

Data Collection

To develop an operator model, we need to collect all data related to the operator-machine interactions. Operator commands are usually directly measured and many perception stimuli can be inferred from various sensor data. The bucket position of a wheel loader, for example, can be determined by the various cylinders extension of the linkage. Other visual information, such as the position of the bucket in relation to a pile of dirt, can be deduced from the force exerted by the soil on the bucket (calculated from cylinder pressure). When analyzing the data with the appropriate tool some patterns start to emerge. These patterns, in turn can be used to identify some general principles of machine operation that can be used to create an operator model.

Looking at gigabytes worth of data can be an intimidating task. Common practice is to isolate one occurrence of the particular action intended to be modeled (for example a dig operation) and analyze it in detail. This approach however does not provide the distinction between random actions and regular patterns. To really understand the trends, a large number of digs must be analyzed and compared side by side.

Data Segmentation

The first step is to identify the part of the data that relate to the operation to be studied. A skilled engineer can do this by recognizing some patterns of cylinders extension and a few other sensors in time series data. Yet, automating this segmentation has been an elusive goal, suffering in time accuracy and having false positive or false negative detections. Recognizing the importance of such automated data segmentation, we designed our operator-modeling tool, named IDA, to integrate a new data visualization and segmentation tool.

Constraints Based Segmentation

IDA uses a concept of constraints based reasoning to perform segmentation. A segment is defined by a set of constraints chosen from eight possible types of constraints. IDA uses a special display to show numerical data channels and symbolic data channels simultaneously. Each Symbolic channel is displayed as a thick line where each segment appears as a colored rectangle corresponding to the color associated with this type of segment. Figure 1 shows an example of the data display and constraints definition for a dig segments during truck loading operation with a wheel loader.

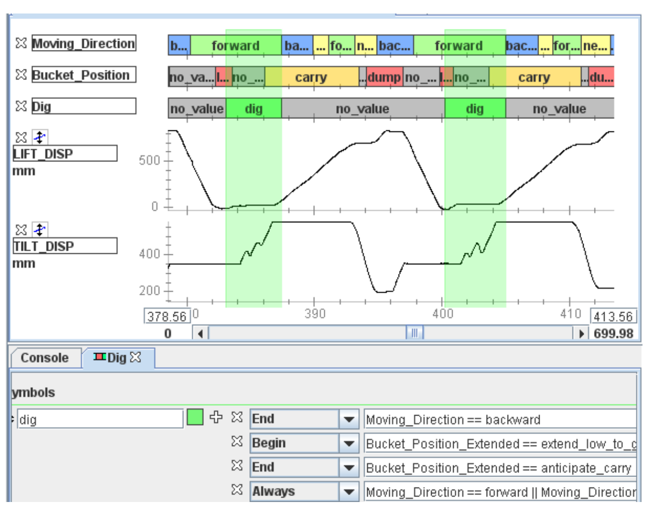

Figure 1: symbolic channels for data segmentation.

Numerical channels can also be calculated with user provided or predefined functions, such as filters or derivatives. Complex segmentations involve dozens of symbolic channels, each defined by its own set of constraints that refer to other symbolic channels or numerical channels. By combining numerical and symbolic processing it becomes possible to create segmentation that are both accurate and robust.

Finding Patterns In Data

Once particular segments of interest have been identified, IDA provides another view to show all the segments of a given type side by side. Figure 2 show a segmented view of dig segments for an operator. In this view, a regular pattern is visible for both tilt and lift command.

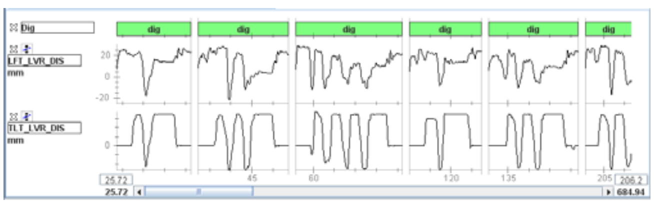

Figure 2: Tilt and Lift commands for an operator during dig

The view has been reduced to fit the available space in this paper. A large display would show more segments at a time and thus reinforce the side-by-side comparison. Finally, the horizontal scroll bar, at the bottom of the display, provides a quick way to browse hundreds of segments. Any irregular segment would stand-out and could be analyzed further. We believe that showing the data with an appropriate view enables the operator model developer to discover new patterns that would not be apparent with a more conventional data display.

Automated Pattern Discovery

A special type of function, called an aggregate function, is used to calculate a single value for each segment. This kind of data reduction can be used, for example, to calculate the energy spent on each dig by using an integral function (an aggregate function) applied to the power delivered by the cylinders over time. Another example of an aggregate function is the "begin_value" function that simply reports the value of a numerical channel at the beginning of a segment. When applied to the gear channel for dig segments, it shows the gear at the beginning of each dig. Figure 3 shows the begin values for gear, speed and throttle. Data channel values at the beginning of a segment present a particular interest when searching for causality of the initiation of a type of task (like digging in our example). If the operator starting the dig action once the machine had reached a specific speed during the deceleration caused by contacting the pile, we would expect to see a consistent value at the beginning of each dig.

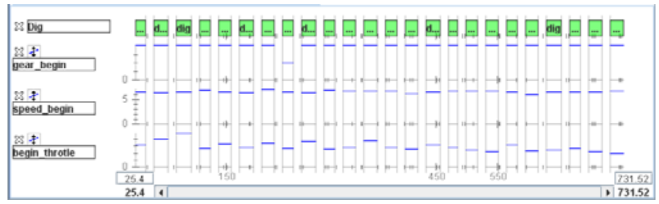

Figure 3: gear, speed and throttle values at the beginning of each dig segment

By calculating automatically the begin values for all the numerical channels available and sorting them by increasing order of their coefficient of variation, the most promising candidate for causality would start to emerge. By applying this calculation on a very large number of data, it is possible to automatically discover some general trends.

Conclusion

IDA, which stands for Integrated Development and Analysis, is currently used as an internal tool at Caterpillar for data analysis and operator model creation and simulation. The combination of both analysis and operator model building in the same architecture opens new possibilities for the creation of operator models. The models are, in a sense, created from the data and compared to actual data at different stages of their development.

The operator modeling approach in IDA relies on serial process definitions related to COGNET (Zachary, LeMentec & Loiederman 1999) and GOMS (Card, Moran & Newell 1983). The concept of data driven modeling could also be applicable to other cognitive modeling systems that rely on production rules such as ACT-R (Anderson & Lebiere 1998) or SOAR (Newel 1990) as long as observable data are available.

In the future, the IDA architecture could be extended to implement knowledge representation for different purposes.

Acknowledgments

The author thanks Caterpillar and its Virtual Product Development division for supporting this work.

References

Anderson, J. R. & Lebiere, C. (1998). *The atomic components of thought.* Mahwah, NJ: Erlbaum.

Newell, A. (1990). *Unified Theories of Cognition.* Harvard University Press. Cambridge, Massachusetts.

Card, Stuart; Thomas P. Moran and Allen Newell (1983). *The Psychology of Human Computer Interaction.* Lawrence Erlbaum Associates. ISBN 0-89859-859-1.

Glenn, F., Zachary, W., Le Mentec, J. C., & Loiederman, E. (1999). Modeling perceptual and motor behavior in a cognitive context. *Proceedings of SAE International Conference on Digital Human Modeling for Design and Engineering*. Warrendale, PA: Society of Automotive Engineers.