logi.DIAG
High-Volumn Real-time Data

Thomas Baier[1,*] **,Erich Neuwirth**[2]

1. logi.cals
2. University of Vienna

* Contact author: thomas.baier@logicals.com

Keywords: automation, industry, data mining, real-time

Industrial Automation in general and in particular PLCs (Programmable Logic Controllers) and embedded devices are a rapidly growing market. Embedded devices are found in small devices, like, e.g., watches or mobile phones, are used in everyday life as for example ABS systems or engine-monitoring systems in cars. In larger applications these system are typically called PLCs and used to control assembly lines, rolling mills or power plants.

Depending on the requirements on availability of automation systems on the one hand or safety considerations on the other hand, more and more effort is put into monitoring the system during its whole life time.

Typical aproaches for monitoring are either rule-based systems or open-loop control scnearios. In rule-based systems data is collected and processed according to statically defined rules (e.g., issuing an emergency shutdown if some safety-related devices fails). Open-loop are designed to collect data and present the results to an operator. The operator then has to decide on further actions (or if the operator fails to acknowledge an alarm message, an automatic procedure brings the whole automation system into a fail-stop or fail-safe operation mode).

As automation systems are getting more and more complex, the size of data sets is increasing far more than the size of the applications. New approaches requiring statistical methods to handle the large amounts of data will have to be established on the market in just a few years. Partners from both industry and academia are working on the funded research project "logi.DIAG" to find solutions for systems with such increasing complexity. See logi.DIAG (2008) for more information.

In contrast to typical applications of data-mining (or even data-warehouse systems), PLCs are very limited in both computing speed and memory (typically a few hunded kilobytes of RAM and no persistent storage at all). Therefore one of the issues is finding ways for data compression and storage which takes these kinds of resources into account. The approach described by Chambers et al. (2006) has been considered especially useful for this case and its adoption for the specific problem domain is one of the first analysis steps in this project.

R is been used for prototyping implementations and during analysis and adoption of algorithms and if possible for further analysis.

References

logi.DIAG (2008). logi.DIAG. Test-Driven Automation. `http://www.logidiag.at/`.

Chambers et al. (2006). Monitoring Networked Applications With Incremental Quantile Estimation. *Statistical Science, 2006, vol. 21, p. 463*,