A Flexible Architecture for Data Mining from Heterogeneous Data Sources in Automated Production Systems

Abstract:
Data heterogeneity and proprietary interfaces present a major challenge for big data analytics. The data generated from a multitude of sources has to be aggregated and integrated first before being evaluated. To overcome this, an automated integration of this data and its provisioning via defined interfaces in a generic data format could greatly reduce the effort for an efficient collection and preparation of data for data analysis in automated production systems. Besides, the sharing of specific data with customers and suppliers, as well as near real-time processing of data can boost the information gain from analysis. Existing approaches for automatic data integration lack the fulfillment of all these requirements. On this basis, a flexible architecture is proposed, which simplifies data integration, handling and sharing of data over organizational borders. Special focus is put on the ability to process near real-time data which is common in the field of automated production systems. An evaluation with technical experts from the field of automation was carried out by adapting the generic concept for specific use cases. The evaluation showed that the proposed architecture could overcome the disadvantages of current systems and reduce the effort spent on data integration. Therefore, the proposed architecture can be an enabler for automated data analysis of distributed data from sources with heterogeneous data formats in automated production systems.