Data Analytics and Visualization for Manufacturing Systems

Xinran Shi[1], Pengfei Chen[1], Yuchen Wen[1], Xiaowei Yue[1], and Yuqi Cao[2]
[1]School of Industrial and Systems Engineering, Georgia Tech
[2]School of Music, Georgia Tech

Introduction

Big data analytics and visualization hold enormous potential for enabling smart manufacturing.

The objective of this research is to develop a Big Data analytics and visualization framework for complex manufacturing systems through systematic and deep integration of data analytics, visualization and manufacturing domain knowledge.

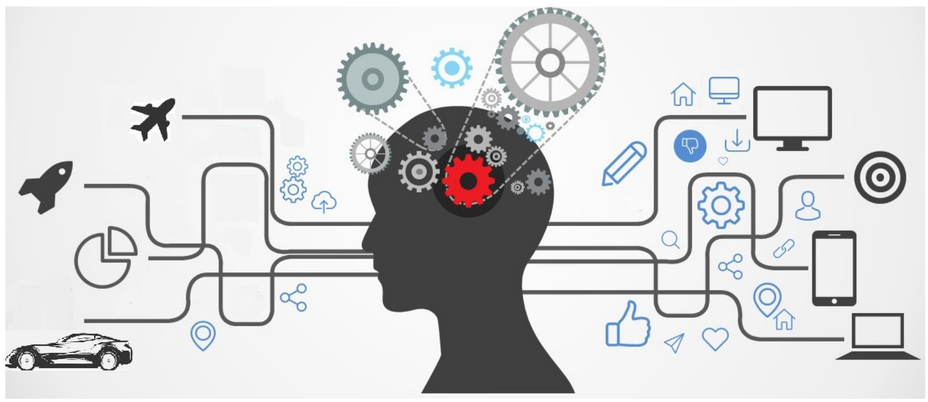
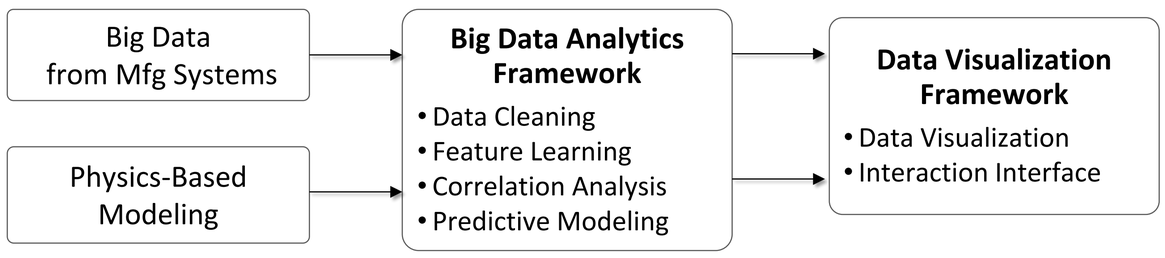


Fig. 1 Manufacturing Systems ⇒ Data Analytics ⇒ Data Visualization



Data from Bosch

Bosch records data at every step along its assembly lines, they have the potential to apply advanced big data analytics/visualization to improve these manufacturing processes. However, the intricacies of the data and complexities of the production line pose big challenges.

(more than 1 million parts samples, 4 production lines, 52 sections, more than a thousand features)

Data File	Description	Size
train_date.csv	Training set data features	2.7 GB
train_numeric.csv	Training set numeric features	2.0 GB
train_categorical.csv	Training set categorical features	2.5 GB

Approaches

Data cleaning for path identification

- OpenRefine failed to clean the dataset
- We used R to extract path (shown in Fig. 2)

Correlation Analysis

- Within-section correlation
- Between-section correlation
- Lay a foundation for causality assessment

Modeling

- Ward hierarchical clustering for failures
- Random forest
- Logistic regression

Visualization

- D3 user interface

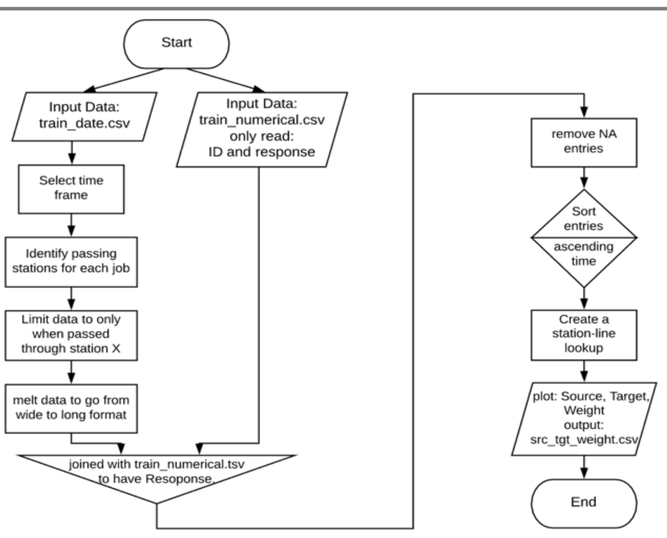


Fig. 2 Flowchart for path identification

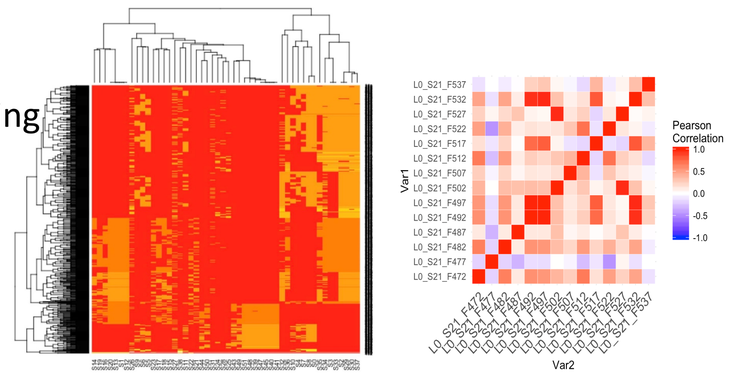


Fig. 3 (a) whole cluster correlation map (b) local correlation map

Experiments and Results

The proposed Big Data analytics and visualization methodology consists of following experimental results.

- Identify the global and local **correlation map** (shown in Fig. 3)
- Identify the **key sections** (**40, 43, 49, 50**) that result in large failure rate (shown in Fig. 4)
- Develop **an interactive interface** for visualization (shown in Fig. 5)

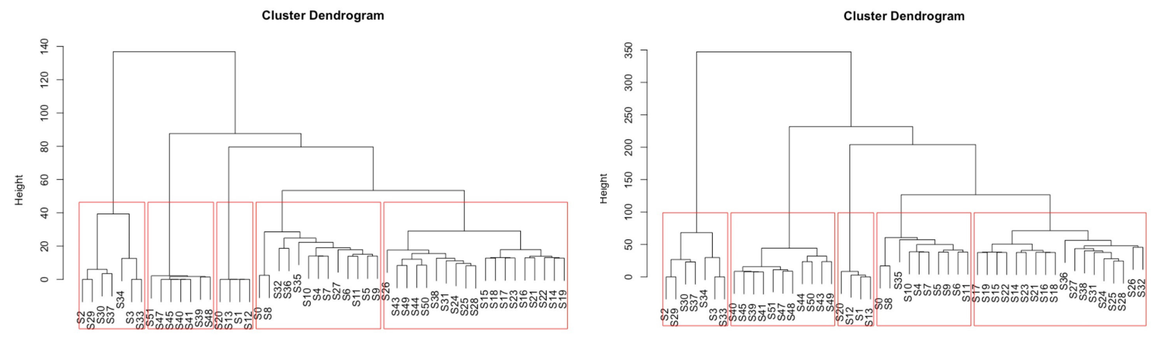


Fig. 4 Hierarchical clustering for (a) failure (b) success samples

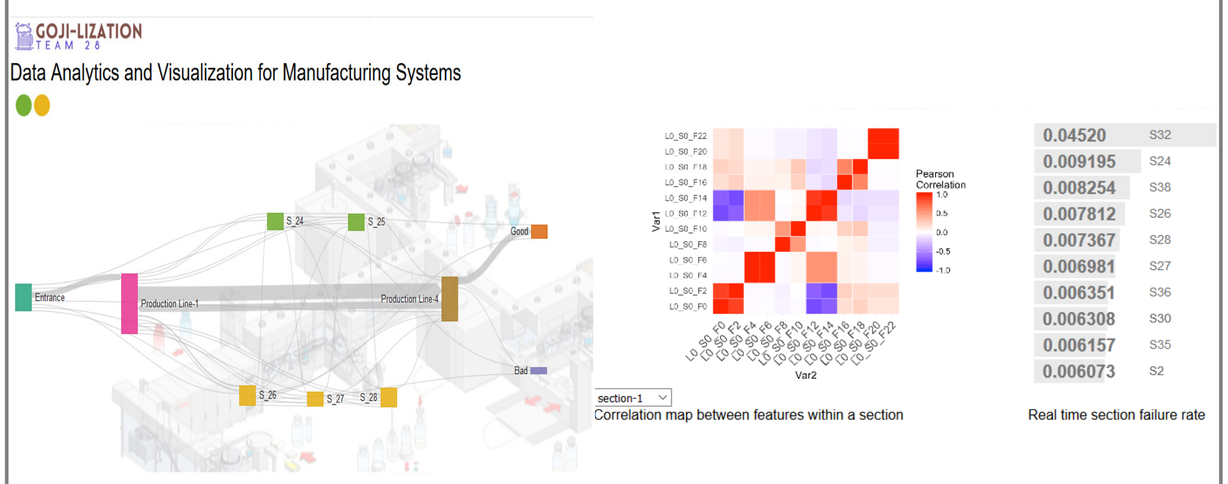


Fig. 5 Interactive interface via D3

Summary and Deliverables

- A systematic framework for big data analytics and visualization has been developed for smart manufacturing systems
- The global/local correlation map, hierarchical clustering, and modeling are explored
- We provide an interactive D3 interface for manufacturing engineers

Future Work: Cloud-Based Smart Manufacturing

- A cloud-based data analytics and visualization architecture will be utilized for smart manufacturing

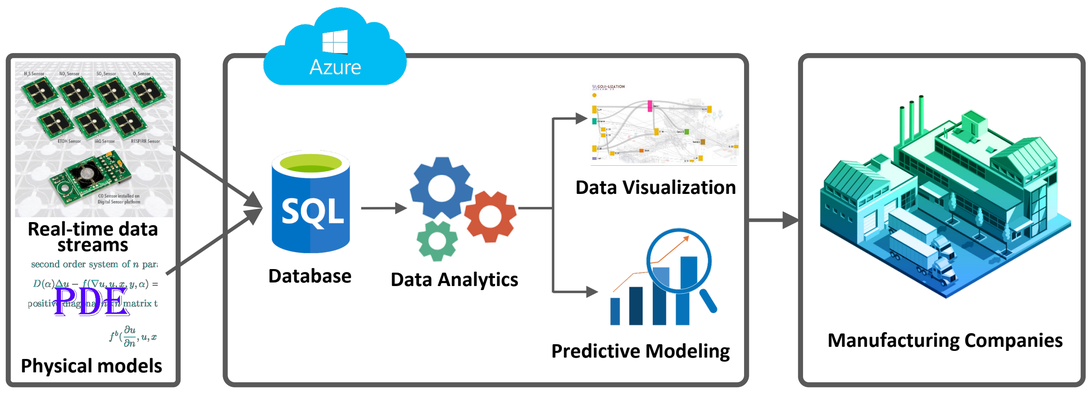


Fig. 6 Cloud-based Smart Manufacturing Architecture

References

1) Kaggle, "Bosch Production Line Performance," 2016. [Online]. Available: https://www.kaggle.com/c/bosch-production-line-performance.
2) Maclean, M., 2018. D3 Tips and Tricks v3.x. Interactive Data Visualization in a Web Browser.
3) Murray, S., 2017. Interactive Data Visualization for the Web: An Introduction to Designing with D3.
4) Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique. Journal of artificial intelligence research, 16, 321-357.