Interactive Visual Data Exploration Solutions for Industrial Systems

Andy Wu[1], Chih-Pin Hsiao[2], Na Cheng[3], Waqas Javed[4], James J.W. Lin[5]

Connected Experience Labs, GE Global Research & Software Center of Excellence[1,3,4,5], Georgia Institute of Technology[2]

Abstract

The Industrial Internet provides the capability of acquiring, sharing, and utilizing data from industrial assets across global networks. To exploit the data collected from industrial systems analysts use software tools to understand the data and explore insights for decision support. In this paper we present our research on the practices and challenges of visual analysts at General Electric (GE) and the experimental interactive visualization tool built for visual data-exploration of industrial systems.

Keywords: Industrial Internet, visual data exploration, big data, predictive asset health management.

Index Terms: H.5.2 [Information Systems]: Information Interfaces and Presentation—User Interfaces.

1 Introduction

General Electric (GE) is an industrial leader across diverse fields including energy, healthcare, finance, aviation, transportation and appliances. To efficiently and automatically monitor complex systems in these fields a large number of sensors are installed inside machines provided by GE. Integrating the data from heterogeneous machines networked sensors and providing innovative services on the result forms the basis of the *Industrial Internet*. The volume of data collected from GE's Industrial Internet is considered *big data* since the amount of data collected from these machines is so large (e.g. a commercial jet engine can generate 20 terabytes of data per hour) and complex (e.g. data are stored in different formats and encrypted with different methods) that it is difficult to handle with traditional database management system and data processing tools [1].

In this work, we present preliminary research results through an example project called *'A'* here for confidentiality reasons. Project A focuses on *improving the efficiency of big data exploration for predictive management of industrial assets*. This research has three major emphases: *visual data exploration*, *integration of current analytic practices and visualizations*, and *intelligent feedback for visual analytics*. We will also highlight our design solutions and the development of the system that we are building.

2 Background

Asset failure is a critical and costly problem in the day-to-day operation of industrial systems. It impacts not only the industrial system itself, but usually, more severely, the business eco-system around it. For example, in the commercial aviation industry a jet engine failure could ripple to flight delays, requiring rearranging of passenger accommodations and disruption of maintenance plans. Similarly, a turbine failure in a power plant could trigger a power outage that causes significant financial and societal loss and could even lead to public safety concerns. A typical approach to reduce such asset faults is to exercise preventive maintenance of the assets in a planned and regular basis. The Industrial Internet allows analysts to receive data close to real-time from sensors inside the assets. By developing analytical intelligence data scientists further diagnose and predict failures of the assets, which significantly helps reduce the cost of unexpected failures and improves customer satisfaction.

The business objectives of our example project A includes reducing cost, providing better work flow for asset maintenance than current practices, increasing the asset fault detection rate, reducing the false alarm rate, and increasing the lead-time between when a fault is detected and when it will cause a delay in service. To achieve these goals we interviewed professionals involving in building analytic models, investigated the needs and gaps in their process, and developed an interactive prototype emphasizing visual data exploration for predictive asset health management.

2.1 User Roles

For project A we first targeted data scientists / analytic researchers, and analytic application developers. Our data scientists / analytic researchers are people who create and improve analytic algorithms to gain insights into problems that they can then share with the community. They need to review existing algorithms, create new algorithms, test the algorithms and publish them. Application developers then identify the best algorithms and build applications for business analyst's needs, e.g. early fault detection, diagnostics, and prognostics. Both roles need to quickly understand and discover useful patterns or insights from the data. One common pain point for the users is the lack of effective tools supporting their data exploration process. The tool focuses on developing the data exploration capabilities needed to support these users.

2.2 Challenges in Data Exploration

Each of the analytic roles uses data exploration to analyze large, unorganized sets of data to uncover useful information for further insight discovery. The complexity of the data collected from the asset sensors makes it difficult to analyze with just computational tools. Visual data exploration has shown the potential to be more effective than traditional data mining [2]. Analytic application developers often have to collaborate with subject matter experts (SMEs) to understand the data. Their current tools, however, are disconnected and require manual management of the underlying data. The whole process involves several steps and there is no efficient technology to help them keep track of the progress of analysis. Often, users utilize visualizations to help them understand the data, but these general-purpose tools do not always generate the insight the users expect.

2.3 The Data

Figure 1 shows a visualization of the underlying multivariate temporal data chosen to demonstrate the prototype. The data presented in Figure 1 contains 330 variables collected from 42 different assets over 5 years. Inspecting the temporal data set in Figure 1(f), missing data is noticeable, as indicated by the gaps in

[1]wu@ge.com,
[2]chsiao9@gatech.edu
[3]chengna@ge.com
[4]javed@ge.com
[5]jlin@ge.com

the chart. Large gaps could be because of asset failures or maintenance, while narrow ones could be due to the routine work schedule of the asset. In order to predict future asset faults analysts examine the data collected prior to the failure and try to identify prominent patterns associated with the failure. Metadata such as an asset's manufacture information or factory settings are also helpful to identify problems. Also, contextual data like weather conditions and geographic information from external sources could also be valuable for analysts to correlate the faults with the prevailing surrounding conditions.

Figure 1: The system has six views, asset, variable, intelligent search, canvas, detail chart, and timeline views (from a to f).

3 Design Approaches

Our user research revealed users highly desired capabilities that allow them to effectively discover patterns and relationships between multi-dimensional variables and different assets. In our vision, users could research and formulate potential patterns-of-interest progressively using various cross-referencing interactive visualization tools for pattern discovery. Therefore, we designed an integrated environment that enables users to save patterns on the server and share them with others.

User scenario: A user starts with loading a set of sensor data, pre-processed by analytics algorithm servers (Figure 2 left). The system then groups and colors the variables according to their hiearchy. The user next identifies variables or variable-groups of interest and starts the exploration process. Once the user selects a pattern-of-interest, potential similar patterns would then be suggested by the system. The system also captures the user's discovery paths to intelligently suggest patterns-of-interest for future explorations (Figure 2 right).

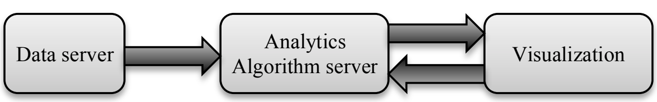

Figure 2: Intelligent feedback and learning in Project A. The analytics algorithm server learns users' feedback from the visualization, analyzes it, and sends refined results to them.

4 System

The proposed workflow integrates the strength of current computational analytic methods (proprietary clustering algorithms) and a new visualization interface. The visualizations presented in this interface are based on analyses done beforehand, e.g. the variable view (Figure 1(b)) is generated by a precomputed classification analytic. In addition, the intelligent feedback allows the system to adjust the output based on users' interaction with the visualizations (Figure 2). The system captures the search patterns the users enter and compares them with previous selections performed by other users. The recommended patterns are shown in the intelligent search view. This intelligent process is already adopted in GE's product [3] to strength the workflow.

We implemented a web-based application prototype using D3.js. The user to performs data exploration using six different visualization views.

Asset View: Figure 1(a) allows users to dynamically group different assets or asset models by certain variables and analytics. In the example the assets are grouped by fleet, which is a group of assets, and categorized by asset models and types of machines. When the cursor hovers over an asset the classification is shown in the variable view, Figure 1(b).

Variable View: Figure 1(b) shows a hierarchical structure of all variables based on analytical classifications. The size of a cell reveals the deviation of the corresponding variable. High deviation often implies occurrences of abnormal conditions.

Intelligent Search View: Users can drag an interesting pattern to this view (Figure 1(c)) where the system will search for similar patterns in the dataset. The system learns the users' selection patterns and provides better search results cumulatively as more patterns are entered.

Canvas View: Figure 1(d) presents each data point as a resizable block. The x-axis in this view represents time and the y-axis an asset's maintenance cycles. The time-span of the x-axis is adjustable. When adjusting the time-span of the x-axis to a shorter time period, say one week, all data is cut in a series of weeklong pieces and stacked along the x-axis in parallel (the two charts in Figure 1(d) right). This interaction is particularly useful when a user aligns the time-span to a specific event, e.g. an asset's operating cycle that ends with an asset fault. Clicking on a highlighted pattern realigns the entire data to this pattern for users to compare (the two vertically aligned graphs in Figure 1(d) left).

Detail Chart View: This view (Figure 1(e)) enables the user to examine the most likely patterns that can cause asset failures. These patterns can be suggested by the system through prior knowledge from the database. Users can also dynamically modify or specify search patterns on this view by dragging a time-window on this view. Once the user has selected a particular time-window, the system then finds the most similar patterns (Figure 1(d)) of the same variable/variable-sets across different assets. The outcome of this is highlighted by level of similarity in the comparison view in Figure 1(d). Furthermore, the corresponding assets will be highlighted in the asset view.

Timeline View: The timeline view in Figure 1(f) shows a temporal overview of datasets related to multiple selected assets. The user can select a time-window to analyze the details of multiple assets in the detail chart view in Figure 1(e).

5 Discussion and Future Work

In this paper, we proposed an interactive and intelligent visual data exploration prototype that is designed to discover patterns-of-interest from the data of industrial assets. It was designed based on the needs and major pain points collected from the user research in an industrial setting to be effective, efficient, and collaborative. We consider this new, proposed workflow an enhancement to the users' current practices. While current development shows promising improvement for the process, we will be continuously enhancing the capabilities of this tool. As this work is integrated into GE's ecosystem of analytical development tools we will conduct further user validation.

6 References

[1] P. Evans and M. Annunziata, "Industrial Internet: Pushing the Boundaries of Minds and Machines," Nov. 2012.

[2] D. A. Keim, "Information visualization and visual data mining," *IEEE Transactions on Visualization and Computer Graphics*, vol. 8, no. 1, pp. 1–8, 2002.

[3] A. Sherman, S. Dekel, and S. R. Ranjan, "Smart 3d pacs workflow by learning," 2013.