Project title:	Map Reduce for Optimizing the Large-Scale Industrial IoT towards Digital Manufacturing
Research team:	Soundar Kumara, Hui Yang, and Daniel Finke
Industry collaborators:	N.A.
Thrust area:	Intelligence
Current TRL:	TRL - 2
Final TRL:	TRL - 6
Project type:	Proposed project
Start date:	10/1/2016
Completion date:	9/30/2017
Percent complete:	20%
Budget:	$50,000
IAB funding:	$0
Other funding:	$0

Industrial Relevance

The large footprint of manufacture sector in the US economy and the rapid advancement of industrial internet of things (IIOT) are driving the third industrial revolution, which provides an unprecedented opportunity for digital manufacturing and mass customization. However, a large-scale IIOT connects to numerous machines and brings the proliferation of big data. Realizing the full potential of big data for manufacturing decision making depends on the information-processing algorithms to extract pertinent knowledge about manufacturing operations. This project aims to provide parallel-computing algorithms that will enable efficient information processing and real-time decision making in large-scale IIOT.

Problem Statement

The ***research objective of this project*** is to develop parallel and distributed algorithms to leverage big data in the large-scale IIOT for data-driven modeling and optimization of the network based manufacturing systems. Through a DMDII project, we have fabricated a cost-effective hardware kit (namely, ThingBox) that includes multi-channel sensors for real-time monitoring of machine conditions (acoustic, vibration, and cutting force), ambient environment, and energy consumption. Together, these data provide real-time signature of machine states. This provides a great opportunity for real-time manufacturing decision making such as online machine scheduling, condition monitoring, predictive maintenance, and energy consuming optimization. However, large amounts of machines and big data in the IIOT pose significant challenges on analytical computing. A single computer is limited in memory capacity and processing power, thereby hampering real-time decision making. There is an urgent need to develop a novel parallel-computing infrastructure that will enable and assist (i) the handling of massive data communicated by the IIOT; (ii) the extraction of pertinent knowledge about operational and environmental dynamics of manufacturing systems; (iii) the exploitation of the acquired knowledge for data-driven modeling and optimization of manufacturing systems.

Approach and Method

Our research team plans to accomplish the overall objective by completing the following research tasks:
(1) ***Characterization and modeling of multi-channel signals for the extraction of machine signatures***: We will represent signal information in alternative domains, such as time, frequency, or time-frequency domains, or state space so that mathematical description of machine patterns contained in the signal is much simpler and efficient in the transformed space. Further, we will characterize machine signatures and quantify salient features that are sensitive to operational, eviromental, sustainable states of manufacturing machines, instead of extraneous noises.
(2) ***Develop parallel-computing algorithms for modeling the machine network in IIOT***:
We will develop a new parallel-computing scheme on a multitude of processors for modeling the large-scale machine network. First, we will develop quantitative measures of dissimilarity of machine signatures (or features). Second, a new stochastic gradient descent approach will be designed to

optimize the embedding of each machine in the network model. Finally, we will develop new map-reduce algorithms for the detection of machine communities in the network. Our preliminary work shows that the parallel-computing algorithms achieve 80-fold faster than conventional algorithms for network embedding and modeling of 20,000 machines.

(3) ***Simulation studies and real-world case studies for implementation, evaluation and validation***

We will perform simulation studies, as well as collect real-world data using the ThingBox developed in the DMDII project for algorithmic implementation, evaluation and validation.

Deliverables and Benefits

Accomplishment of the proposed project will establish a new parallel-computing scheme and software algorithms for modeling and optimization of machine networks towards digital manufacturing. For instance, new models will be availbale to extract machine signatures and forecast each machine's health condition. Such information will be used for machine scheduling and predictive maintenance. This, in turn, will prevent machine failures, reduce the machine downtime, and increase the revenue.

Potential Application Areas

The proposed parallel-computing scheme is developed for large-scale IIOT in manufacturing; however, the methodology and the software package is generally applicable to many other IoT systems that experience computing challenges due to the increasing number of entities (e.g., health IOT).

Project Plan and Progress

The project plan consists of 1) Reviewing current literature, IIOT data collection and computer simulation; 2) Multi-channel condition monitoring, signal modeling and forecasting; 3) Developing parallel-computing algorithms for data-driven prediction of machine conditions and service decision making; 4) Real-world case studies for implementation, evaluation and validation. The Gantt Chart is shown below.

Month	1	2	3	4	5	6	7	8	9	10	11	12
Literature review	■	■	■	■								
Machine signature extraction	■	■	■	■								
data collection	■	■	■									
modeling and analysis		■	■	■								
Development of large-scale machine network			■	■	■	■	■	■	■	■		
Characterization of machine similarity			■	■								
network embedding algorithm					■	■	■					
map reduce algorithm								■	■	■		
Evaluation and validation				■			■			■	■	■
Document and presentation												■

Current State of Practice and Research

The hardware infrastructure of Industrial IoT (IIoT) system has not become available until very recently. As a result, heterogeneous data (e.g. energy consumption, ambient temperature, and condition monitoring data) are readily available at the level of individual machine as well as the population level. Therefore, complex manufacturing structure and big data pose significant challenges in the design and development of large-scale IIOT. Very little work has been done to develop parallel-computing algorithms for efficient information processing and real-time decision making in IIOT. Most existing techniques in the literature focus on conventional manufacturing and cannot take full advantages of the availability of big data in the IIOT context. There is a pressing need to establish a parallel-computing infrastructure for data-driven modeling, manufacturing decision making that are tailored for both the opportunities and the needs of next-generation IIOT systems.

How Ours is Different

The proposed map reduce algorithm is the first of its kind, which will enable efficient information processing in large-scale IIOT system and help real-time decision making in manufacturing processes. Specifically, our contributions are as follows:

1) We propose ***a parallel computing algorithm*** (*rather than traditional serial computing*), which enables real-time optimization of large-scale machine network.
2) We propose a ***stochastic gradient decent algorithm*** (*rather than traditional deterministic algorithm*) to efficiently optimize the embedding of each machine in the network model,
3) The proposed algorithm characterizes and models ***multi-channel signals*** (*rather than traiditonal single-channel signals*) for the extraction of machine signatures.
4) The proposed research establishes a dynamic machine network (rather than static network), which is suitable to characterize and monitor the degradation of condition of each machine.