Break through with
BIG DATA

IEs can play a key role in mining numbers for dramatic improvements

By Andrew Kusiak

The growth of data streams and data analysis will generate significant developments in the 21st century. Big data will play a key role in shaping the design of products, systems and services by making them smarter, better connected, more efficient and widely accessible. Data will become a powerful economic engine.

Data analysis has been enlarging its footprint in business and engineering applications for the last two decades. In particular, the two labels of data science and big data have emerged recently, attracting attention and receiving massive press coverage. Expectations from data science are enormous.

In the research community, the interest in data is not new, as it has accompanied developments in science and practice for centuries. The novelty is in the scale at which the data is available, its growth and the acknowledgement that data research has been equated to data science. The industrial engineering profession can be at the forefront of applying big data analysis and research in a number of business sectors.

What is big data?

The concept of big data can be illustrated by likening it to an Excel-like file. Such a file could include countless columns and rows of data. The rows of the Excel file would be generated at high frequency (data streaming). However, it needs to be realized that a big data scientist does not expect the data to appear in an Excel format. Big data sets seldom exist in a ready-to-use form. Instead, they need to be harvested.

Another important aspect of big data is its acquisition, which the author calls "data farming." To date, most data analysis projects have been based on the data generated for purposes other than data science research and applications. This limited scope of data constraints is what data science has to offer. To meet the expectations of data science, data farming needs to become a reality. The industrial engineering connection to data farming comes naturally through the design of experiments, which may be considered as a data farming tool. Other tools for data farming, e.g., sensor design and deployment, will emerge.

Despite prolific writings about big data, there are only a handful of places where big data are readily available. Many organizations store data in volumes measured in large numbers that we are unaccustomed to. The fact that the data stores are large does not make them truly big, however. For example, deep-sequenced genetic data may include tens of thousands of parameters (columns of an Excel file); however, in a healthcare environment, such data may be available for only a few patients. On the other hand, industrial process data may have been collected for years, although the number of parameters for which the data has been collected could be limited.

The question of who owns the data often is overlooked. It usually is assumed that an industry or an individual that has purchased equipment owns the data the equipment generates. Many individuals and companies may be surprised that the equipment manufacturer owns the data generated by the equipment, especially when the equipment is under warranty. Realizing the flow of data is important to the ownership of the problems and solutions that are developed through data analysis.

Areas ripe for big data mining

Embracing big data adds another dimension to industrial engineering research and practice. It translates into an area of intellectual and economic activity, thus creating new jobs. The fact that the term process is prominent in the definition of industrial engineering establishes a natural bridge to the phenomenon of big data.

After all, the very reason for collecting process data is to reflect process behavior. Such data can be used to model processes, which in turn can be optimized. Big data modeling offers a significant advantage, as the models created can be more accurate than models created with less data.

Such models may be considered digital replicas of the processes or phenomena of interest. Thus, it needs to be realized that models are created by machine learning algorithms rather than being developed by people. The models can be refreshed as frequently as needed to maintain the expected accuracy. Industrial experience in modeling stationary and nonstationary processes becomes useful in designing data-driven models.

As a profession, industrial engineering stands to gain tremendously by embracing and engaging in big data research and applications. Such opportunities either exist or will emerge in the near future across many industrial and service domains. Industrial engineering can apply big data research in areas that include manufacturing and remanufacturing; product design and development; the service, energy, transportation, healthcare and biology sectors; sustainability and innovation.

Manufacturing and remanufacturing. Traditional manufacturing and remanufacturing as well as emerging manufacturing technologies and systems involve data. Areas of industrial engineering activities such as continuous improvement, process modeling, production planning and scheduling, and quality improvement are naturally amenable to big data analysis.

Data science offers tools for fusing diverse data steams (parameters) that would not be possible with traditional approaches such as operations research or simulation. Models involving diverse parameters could be highly accurate, which could be a boon since forecasting in any industry usually includes the chance of significant error.

For example, predicting the demand for remanufactured parts and assemblies in a market where original and remanu-

Fix it with data

A data-driven model helped resolve manufacturing issues associated with rounded surfaces on wafers. Wafer surfaces with a roll-off greater than zero cannot be used for integrated circuits.
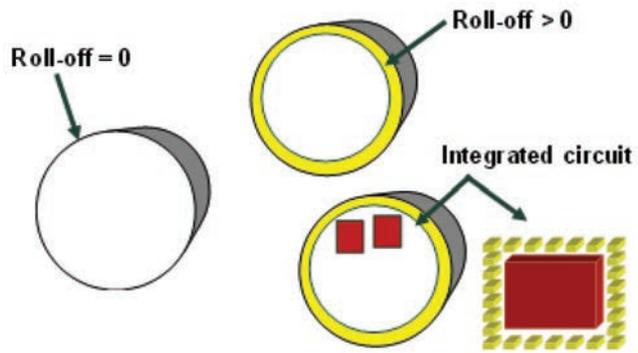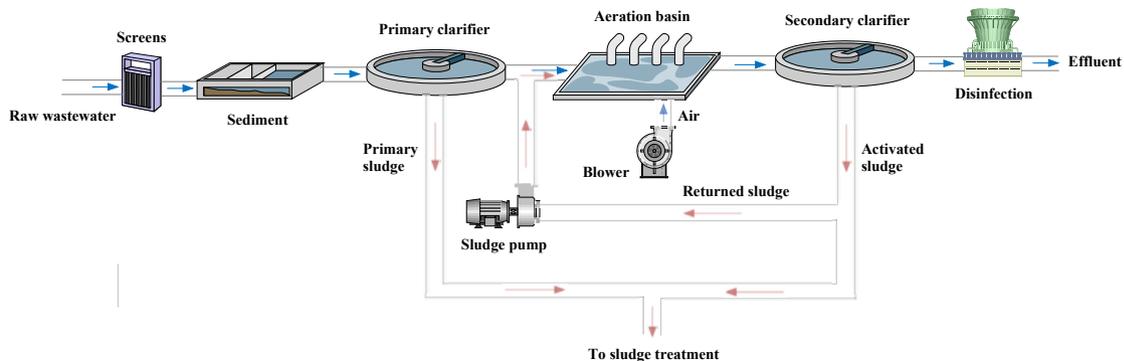

FIGURE 2

From dirty to clean

The basic processes in a wastewater processing plant are just waiting for data solutions provided by industrial engineers.

factured parts and assemblies are sold is difficult. Our studies have shown that data-driven models accurately predict production volumes using historical production volumes, stock prices, product reliability data and parameters as esoteric as average temperature in the region.

In one case, a data-driven model combining data about product, material and process parameters dramatically reduced the number of roll-off errors in wafer production. The roll-off error illustrated in Figure 1 involves a rounded surface of a wafer that cannot be used for integrated circuits. Such a rounded surface is created from shortcomings of the wafer manufacturing process.

Product design and development. Industrial engineering needs to be open to embracing all engineering areas where data offers value. Product design and development is a highly multidisciplinary process awaiting new solutions. It is widely recognized that design of an innovative product considers data streams originating at customers, experts, trails of data left by products throughout their lifetime, and cyberspace. It has been established that markets get conquered by products inspired by requirements that go beyond basic functions. Disruptive innovations often succeed because they meet this expanded set of requirements.

Service. Business and computing communities have en-gaged in research of service systems and service computing. The industrial engineering profession may learn from the frameworks developed by these sister communities and translate them into practical applications versed in data. Text (e.g., emails), images (e.g., facial features) and numerical data (process parameters) can be combined to support decision-making in industry. A model versed in data from sources that never have been used before could revolutionize decision-making. There is no doubt that data from heterogeneous data streams are truly big data.

Energy. The energy industry is undergoing significant changes, triggered largely by environmental concerns. Energy generation is becoming distributed rather than centralized as a shift to clean energy is taking place. A combustion engine or a nuclear power plant is replaced by dozens to thousands of wind turbines or solar panels, thus creating a large energy-supply network. Coordinating energy consumption by thousands or millions of households and industries with the energy supply makes it a complex optimization problem.

At some level, a wind farm is not that different from a manufacturing plant. A machine tool changes the material form; a wind turbine converts the energy of the wind into electrical energy. Besides controlling, monitoring and maintenance of turbines, the energy produced needs to be managed. The

emerging energy model could benefit from the logic and solutions embedded in enterprise resource planning systems widely known to industrial engineering practitioners.

Production system expertise is applicable to the energy industry, and this includes wind energy. Wind farms offer tremendous volumes of data that come at low cost. It is rare to see an industrial process generating as much data as wind turbines provide. This offers a new area for industrial engineering solutions.

Transportation. In the transportation domain, industrial engineering has been present for years. Now is the time to integrate the data modeling and human factors perspectives in a uniform research and development approach.

A car generates data from its engine and sensors installed in the cabin. It also receives data from external sources such as cameras, satellites and other vehicles. As the term autonomous vehicle gains attention, the industrial engineering community should act. Applying industrial engineering to these tools can yield better capacity utilization of highways and vehicles. How different is managing transportation assets and road capacity from that of production management?

The single most differentiating factor between the transportation and production domain is in the data. Vehicles move while production machinery remains stationary. Therefore, telematic systems limit the width and frequency of the volumes of data transmitted. Determining the type of data and the frequency of data transmission from a vehicle is, in fact, a valid industrial engineering problem. Making use of the data attributed to a transportation network is a great challenge.

Sustainability. Industrial engineers could not find a more interesting area of research and practice than sustainability. Partnerships could be formed with civil and geo-informatics communities to create data-driven models involving water, soil and climate applications. One of many applications awaiting attention is performance optimization of wastewater processing plants. The basic processes involved in these plants are shown in Figure 2.

A wastewater processing plant involves processes that consume energy, and it also produces energy (methane gas). It is not unusual to see at least 10,000 parameters monitored at different time scales. This is a large system that needs big data solutions due to the diversity of different processes that have to work seamlessly.

This industry is large in numbers, as there are municipal and industrial facilities. Facilities that process clean drinking water usually operate independently of wastewater plants and use different processes. In countries facing water shortages, such as Singapore, it is not unusual to see the waste- and clean-water processing plants combined into one, with wastewater flowing in and usable water flowing out of the plant.

Healthcare and biology. The industrial engineering profession has a long history of engagement in healthcare re-

Big data, minimal changes

A recent report in *Forbes* shows that the big data money train has yet to pull into the station.

According to a survey of 226 executives, only 13 percent have fully implemented their big data programs to where they are integrating predictive insights into business operations. While 79 percent have not fully integrated their data sources across the organization, 67 percent admit that they don't have well-defined criteria to measure the success of big data.

Even under such limited criteria, only 27 percent labeled their big data initiatives "successful," and only 8 percent called them "very successful."

According to *Forbes*, the problems are remarkably similar to those experienced when enterprises try to integrate any advanced business technology — client/server computing, ERP woes and service-oriented architecture.
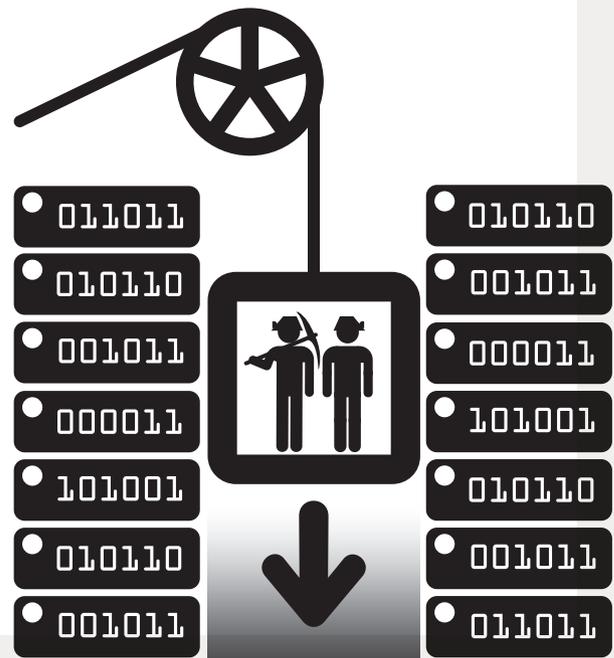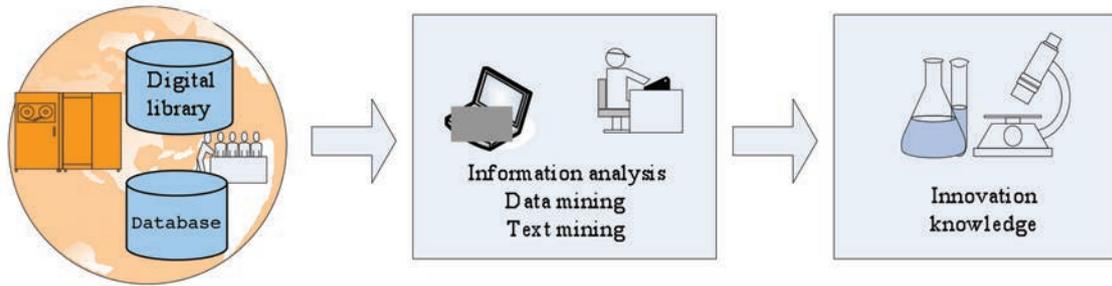
search and practice, particularly involving process improvement, safety and information technology. Embracing the data perspective offers the opportunity to enhance clinical research and applications.

Healthcare awaits solutions for disease diagnosis, treatment optimization and process improvement to benefit patients. Embracing clinical and genetic data, including data from deep-sequencing equipment, offers great opportunities for implementing process and modeling perspectives. Industrial engineering research in medical decision-making has been visible for decades, mostly in administrative domains. The key challenges ahead for medical research include discovering knowledge from all of the data gathered and making use

Innovation framework

Analyzing and mining data could yield knowledge that would improve the process of innovation.

In many ways, a wind farm is like a manufacturing plant, as turbines need to be controlled, monitored and maintained and the energy produced needs to be managed.

is usually backed up with science. For example, management has management science and manufacturing has manufacturing science. Though innovation appears to have the highest impact on economic output, its science is missing. One comes across methods, rules and tools supporting innovation; however, a unified theory of innovation is missing. The number of publications on innovation topics is growing rapidly, offering a hope for establishing innovation science.

The fact that meaningful innovations are rare can be interpreted to mean that our innovation processes carry a huge error. Depending on the assumptions and area of economic activity, this error is estimated to be as high as 95 percent and can even exceed 99 percent. Innovation needs a science that would reduce this error dramatically.

This large error suggests that massive experimentation is needed in the generation of high-impact solutions. Any innovative solution needs to be accepted by the market, which implies that a market model versed in data and allowing for massive simulations would lead to disruptive and continuous innovations. Simulation and industrial engineering go hand in hand.

Opportunities to imagine

Systematic studies and imagination aimed at improving processes and outcomes by better use of data are issues that are worth the attention of the industrial engineering profession. Big data has not been explored to its fullest, and IEs could be the first to power the data-driven economic engine.

Opportunities for discovery are plentiful. Even novice IEs introduced to a technology or research area could have a meaningful impact on products, processes and services by taking advantage of the value hidden in big data. ❖

Andrew Kusiak is professor and chair of the Department of Mechanical and Industrial Engineering at the University of Iowa. He previously served as associate professor at the University of Manitoba and assistant professor at the Technical University of Nova Scotia.

of this data in clinical practice. Such data may include patient data, medical tests, imaging and genetic data. This new information represents dozens to even thousands of parameters.

The new knowledge and models discovered from medical data could lead to dramatic productivity and quality improvements, thus making the healthcare system sustainable.

Innovation. Innovation has been studied for decades; however, long awaited breakthroughs are still to come. The four data sources discussed in the product design and development section are of great interest to innovation, particularly the data trail attributed to a product or a system over its lifetime and the data distributed in cyberspace, as shown in Figure 3. Analyzing and mining such data would enhance innovation.

Some industries use data in fostering innovation, while most do not use it at all. Any area of economic significance