Chapter 7
Data Integration Technology for Industrial and Environmental Research via Air Quality Monitoring Network

Tianxing Cai
Lamar University, USA

ABSTRACT

Industrial and environmental research will always involve the study of the cause-effect relationship between emissions and the surrounding environment. Qualitative and mixed methods researchers have employed a variety of Information and Communication Technology (ICT) tools, simulated or virtual environments, information systems, information devices, and data analysis tools in this field. With the collection and representation of information in a range of ways, software tools have been created to manage and store this data. This data management enables more efficient searching ability of various types of electronic and digitized information. Various technologies have made the work of research more efficient. The results of the qualitative or mixed methods research may be integrated to reach the research target. Right now, a lot of software tools are available for analysis to identify patterns and represent new meanings. The programs extend the capabilities of the researcher in terms of information coding and meaning-making. Machine-enhanced analytics has enabled the identification of aspects of interest such as correlations and anomalies from large datasets. Chemical facilities, where large amounts of chemicals and fuels are processed, manufactured, and housed, have high risks to originate air emission events, such as intensive flaring and toxic gas release caused by various uncertainties like equipment failure, false operation, nature disaster, or terrorist attack. Based on an available air-quality monitoring network, the data integration technologies are applied to identify the scenarios of the possible emission source and the dynamic pollutant monitor result, so as to timely and effectively support diagnostic and

DOI: 10.4018/978-1-4666-6493-7.ch007

prognostic decisions. In this chapter, several systematic methodologies and preliminary data integration system designs for such applications are developed according to the real application purpose. It includes two stages of modeling and optimization work: 1) the determination of background normal emission rates from multiple emission sources and 2) single-objective or multi-objective optimization for impact scenario identification and quantification. They have the capability of identifying the potential emission profile and spatial-temporal characterization of pollutant dispersion for a specific region, including reverse estimation of air quality issues. The chapter provides valuable information for accidental investigations and root cause analysis for an emission event, and it helps evaluate the regional air quality impact caused by such an emission event as well. Case studies are employed to demonstrate the efficacy of the developed methodology.

BACKGROUND

Data integration techniques or information and communication technologies have been intensively used in different data mining applications such as data clustering, classification, association rules mining, sequential pattern mining, outlier detection, feature selection, and information extraction in the industrial and environmental research via air quality monitoring network. A huge increase in the number of papers and citations in the area has been observed in the previous decade, which is clear evidence of the popularity of these techniques. These have included the adoption of such kind of methodologies in the research field of polarization-difference imaging for observation through scattering media (Rowe, Pugh, Tyo, & Engheta, 1995), biologically inspired self-adaptive multi-path routing in overlay networks (Leibnitz, Wakamiya, & Murata, 2006), a biologically inspired system for action recognition (Jhuang, Serre, Wolf, & Poggio, 2007), programmable self-assembly using biologically-inspired multiagent control (Nagpal, 2002), biologically inspired growth of hydroxyapatite nanocrystals inside self-assembled collagen fibers (Roveri, Falini, Sidoti, Tampieri, Landi, Sandri, & Parma, 2003), biologically inspired cognitive radio engine model utilizing distributed genetic algorithms for secure and robust wireless communications and networking (Rieser, 2004), biomimetics of biologically inspired technologies (Bar-Cohen, 2005), biologically inspired computing (De Castro & von Zuben, 2005), and biologically inspired algorithms for financial modeling (Brabazon & O'Neill, 2006). Before we start to give the introduction of these techniques in the research field of industrial operation and environment sustainability, the brief introduction will be given for these techniques.

1. **Artificial Neural Networks:** In computer science and related fields, artificial neural networks are models are derived from animal central nervous systems (Wang & Fu, 2008). The biologically neural networks are capable of machine learning and pattern recognition. They can be regarded as systems of internally connected neurons. They can compute values from inputs by feeding information through the network (Stevens & Casillas, 2006). For example, in a neural network for image recognition, a set of input neurons may be activated by the pixels of an input image representing a shape or color. The activations of these neurons are then passed on, weighted and transformed by some function determined by the network's designer, to other neurons, etc., until finally an output neuron is activated that determines which image was recognized. Similar with other methods of machine learning, neural networks have been applied to solve a wide range of jobs which are difficult to solve using ordinary rule-