AIIM

Association for Information and Image Management

1100 Wayne Avenue, Suite 1100
Silver Spring, Maryland 20910

301/587-8202

Centimeter

Inches

MANUFACTURED TO AIIM STANDARDS
BY APPLIED IMAGE, INC.

V.

PNL-SA-22026

APPLICATIONS OF NEURAL NETWORKS TO REAL-TIME DATA PROCESSING AT THE ENVIRONMENTAL AND MOLECULAR SCIENCES LABORATORY (EMSL)[1]

P. E. Keller
R. T. Kouzes
L. J. Kangas

June 1993

Presented at the
8th Conference on Real-Time
Computer Applications in Nuclear
Particle & Plasma Physics
June 8-11, 1993
Vancouver, British Columbia

Prepared for
the U.S. Department of Energy
Contract DE-AC06-76RLO 1830

Pacific Northwest Laboratory
Richland, Washington 99352

APPLICATIONS OF NEURAL NETWORKS TO REAL-TIME DATA PROCESSING AT THE ENVIRONMENTAL AND MOLECULAR SCIENCES LABORATORY (EMSL)[1]

Paul E. Keller, Richard T. Kouzes, Lars J. Kangas
Pacific Northwest Laboratory, K1-87, P.O. Box 999
Richland, WA 99352, U.S.A.
pe_keller@pnl.gov, rt_kouzes@pnl.gov, lj_kangas@pnl.gov

Abstract

Detailed design of the Environmental and Molecular Sciences Laboratory (EMSL) at the Pacific Northwest Laboratory (PNL)[2] is nearing completion and construction is scheduled to begin later this year. This facility will assist in the environmental restoration and waste management mission at the Hanford Site. This paper identifies several real-time data processing applications within the EMSL where neural networks can potentially be beneficial. These applications include real-time sensor data acquisition and analysis, spectral analysis, process control, theoretical modeling, and data compression.

I. INTRODUCTION TO EMSL

Enormous amounts of hazardous waste were generated by more than 40 years of plutonium production at the Hanford Site. There are an estimated 1700 waste sites distributed around the 560 square miles of southeastern Washington that comprise the Hanford Site[1]. This waste includes nuclear waste (e.g., fission products), toxic chemical waste (e.g., carbon tetrachloride, ferrocyanide, nitrates, etc.), and mixed waste (combined radioactive and chemical waste). The current mission at the Hanford Site is environmental restoration and waste management. To assist in this effort, the U.S. Department of Energy is funding the construction of the EMSL at PNL. The role of the EMSL will be to provide the scientific and technical breakthroughs that will be necessary to accomplish environmental cleanup and restoration in a timely and cost effective manner.

Several topics will be investigated in the EMSL including analysis of large biological molecules, modeling and monitoring of groundwater transport, chemical sensing methods, chemistry in solution, and development of a molecular understanding of contaminant behavior. Exploration of these topics will require the development of advanced spectroscopic techniques, chemical sensor devices, sophisticated theoretical models of molecular structures and dynamics, computational methods, and systems capable of processing and analyzing large quantities of data. Part of the current EMSL planning process involves the development of analytical instruments including mass spectrometers, nuclear magnetic resonance (NMR) spectrometers, and chemical sensors. Artificial neural networks (ANNs) are used in a wide variety of data processing applications where real-time data analysis and information extraction are required. We have investigated various applications for ANNs within the EMSL facility. These applications are identified and discussed in this paper.

II. APPLICATIONS OF NEURAL NETWORKS IN EMSL

Several potential applications of ANNs have been identified in the EMSL including sensor data acquisition and analysis, classification and analysis of patterns in spectral data, process control, data compression, and theoretical modeling.

A. Sensor Data Acquisition and Analysis

There are many real-time and remote sensing applications on the Hanford Site including *insitu* monitoring of contaminants, and chemical and isotope identification. Many of these applications require an inexpensive, compact, and automated system for identifying and monitoring the object of interest (e.g., chemical, isotope). Such a system could be constructed with a sensor array and an automated pattern recognition system (such as a neural network). In hazardous environments, these systems have a distinct advantage over traditional sampling and laboratory analysis methods since an environment can be monitored without risk to human operators.

We are currently prototyping an ANN system that analyzes data from 18 thermal sensors on a high-level nuclear waste tank. We are also exploring the general application of neural networks to sensor data acquisition and analysis.

Part of the EMSL planning process involves the development of new chemical sensor arrays and exploration of new methods in sensor data collection and analysis. A sensor array is composed of several sensing elements, where each element measures a different chemical property of the sensed chemical sample. Data generated by the array is analyzed to determine the constituent components of the sample. When the sensor array is combined with an automated data analysis system, it is often referred to as an artificial nose [2]. Artificial noses will be used to examine and identify chemical waste samples and contamination on the Hanford Site.

The complexity of the data collected by large sensor arrays makes analysis with conventional chemometric methods difficult. ANNs, which are relatively easy to train for analyzing complex data, are likely to be a better choice for sensor data analysis. Several researchers have developed artificial noses that incorporate ANNs for use in applications including monitoring food and beverage odors [3], automated flavor control [2], analyzing fuel mixtures [4], and quantifying individual components in gas mixtures [5,6]. Several ANN configurations have been used in artificial noses including backpropagation-trained, feed-forward networks; Kohonen's

[1]This research was supported by the Northwest College and University Association for Science (Washington State University) under Grant DE-FG06-89ER-75522 with the U.S. Department of Energy.

[2]Pacific Northwest Laboratory is operated for the U.S. Department of Energy by Battelle Memorial Institute under contract DE-AC06-76RLO 1830.

self-organizing networks; Hamming networks; Boltzmann machines; and Hopfield networks [7-9].

Another project on the Hanford Site involves the deployment of a multi-instrument array into hot cells. This array consists of fiber optic sensors, radiation sensors (gamma and beta detectors), and ultrasonic devices. An ANN system would be useful in the analysis and identification of the data generated by the various sensors.

The development of new sensors requires that a methodology for sensor calibration and validation be established. This involves the identification of tested methods and the development of new methods. ANNs have been used in spectral peak verification [10] and will be considered for both the calibration and validation of new sensors, particularly for new complex sensors that may perform better than established calibration and validation methods.

B. Spectral Analysis

The EMSL will be equipped with advanced mass spectrometers, ion cyclotron resonance mass spectrometers (ICR-MS), and several high-field and ultrahigh-field NMR spectrometers. These instruments will be used in the analysis of large macromolecules, such as enzymes, to be used in environmental remediation. In mass spectrometry, ions and ionic fragments are produced from sample molecules and are sorted by their mass-to-charge ratio. This distribution or mass spectrum is used to identify the elemental composition of the sample molecules.

Automated identification of spectral data is an identified problem with applicability to ANNs. In this application, the chemical composition of a sample is determined from its spectral signature. ANNs have been successfully used to classify spectra from various modalities including infrared spectroscopy [11], mass spectrometry [12], and NMR spectroscopy [13]. In general, the backpropagation algorithm is used to train a feed-forward ANN for this application. A training set of labeled spectra are generated and presented to the training algorithm, which iteratively fixes the synaptic weights in the ANN. An important advantage that an ANN has over the traditional database search is the drastic reduction in classification time realized by the ANN. The ANN stores pattern data in a more compact form than the database that results in a more efficient search. Also, when it is implemented in a true parallel distributed processing system, the inherent parallelism of the ANN provides for a very rapid search since the comparison operation is distributed over the entire neural network.

Another potential application of ANNs is in the interpretation of important features in spectral data. A specific problem in this area is locating the spectral peaks of the lowest molecular-weight monoisotope in a mass spectrum of a large organic molecule. Potentially, an ANN could be trained to look at the distribution around the various peaks in the mass spectrum and infer the location of the lowest molecular-weight monoisotope.

C. Process Control

A more complex problem, is process control. The cleanup of Hanford will require that many controls be maintained over complex chemical processes. It would be difficult for human operators to closely monitor all key process parameters for a sophisticated chemical process in real-time. It is more effective to use automated systems in the process control and use human operators in a supervisory capacity. ANNs allow continuous, high-level monitoring of all process sensors and can function as adaptive controllers. In many systems, performance degrades over time due to deterioration of the system components. To compensate, operational parameters are dynamically adjusted to optimize system performance. An ANN can be used to monitor the process, make decisions about system operation, and adjust the appropriate controls to keep the process operating with optimal efficiency and safety. An advantage ANNs have over more traditional adaptive controllers is that the ANN can be continuously updated with new information by using a dynamic learning approach. ANNs have been successfully used to control complex processes and could be beneficial in the waste processing plants being designed for the cleanup of the Hanford Site. The backpropagation algorithm is commonly used to train ANNs in process control with the training set composed of historical data about the process. ANNs have been used in various process control applications including process fault diagnosis [14,15] and temperature control [16]. Several additional applications are discussed in two recent special issues of the **IEEE Control Systems** magazine devoted to neural networks [17,18].

D. Data Compression

The amount of data generated in the planned EMSL is likely to be overwhelming; therefore, construction of novel systems capable of compressing large quantities of data is necessary. For example, in an analytical system currently in use, photon counting of fluorescent molecules is performed. This procedure produces a two-dimensional histogram or image of the fluorescing surface. The size of the generated image is 1024 by 1024 pixels at 16 bits per pixel, which is equivalent to 2 million bytes of data. Larger images will be generated by systems in the EMSL. An examination of the structures in the image shows that only a small amount of information would be lost if a large amount of data compression was performed. A recognized approach in data compression of this form is to tile the image into subimages and then compress each individual subimage by using Principal Component Analysis (PCA). ANNs have been trained to perform efficient PCA data compression in real-time[19,20].

E. Theoretical Modeling

A fundamental understanding of the processes that occur in the complex environment of hazardous waste is necessary in the development of efficient and cost effective systems for environmental remediation. By simulating molecular structures and dynamics, one can gain insight into these processes. ANNs have been used in various theoretical applications including pattern recognition of molecular structures [21], modeling chemical systems, prediction of spectral data, [10] and prediction of energy levels [22].

Several theoretical models in molecular science involve search and optimization. For example, molecular structures can be determined by optimizing a set of structural parameters for a set of physical constraints. Unfortunately, as the number of components in the problem is increased, the computational complexity increases at a rate faster than can be described by a polynomial expression of the number of components. While generally producing suboptimal results, ANNs have been used to generate approximate solutions in relatively short computation times when compared with more rigorous optimization techniques (e.g., Boltzmann machine).

A sample application of ANNs is the modeling of various systematics of nuclear data[23,24]. We have implemented an ANN based nuclear mass model in a feed-forward network. It was trained with mass excess data and generates a mass excess prediction for a given isotope. Currently, our average prediction error is about 1 MeV, but should drop with refinements to our model.

III. CONCLUSIONS

This paper has identified several real-time data processing applications in the planned EMSL that can potentially benefit from ANNs. These applications include sensor data acquisition and analysis, spectral analysis, process control, data compression, and theoretical modeling.

We are currently working on prototype evaluation to determine if ANNs are appropriate for the aforementioned applications. This involves development of software ANN simulators and exploration of the capabilities and limitations of ANNs to these applications. If ANN solutions are judged appropriate after the evaluation, then a dedicated ANN hardware system will be considered.

IV. REFERENCES

[1] Levi, B.G., "Hanford seeks short- and long-term solutions to its legacy of waste," Physics Today **45(3)**, 17-21 (March 1992).

[2] Moriizumi, T., Nakamoto, T., and Sakuraba, Y., "Pattern Recognition in Electronic Noses by Artificial Neural Network Models," in **Sensors and Sensory Systems for an Electronic Nose,** Gardner and Bartlett, ed. (Kluweer Academic Publishers, Amsterdam, 1992) pp. 217-236.

[3] Gardner, J.W., Hines, E.L., and Wilkinson, M., "Application of Artificial Neural Networks to an Electronic Olfactory System," Meas. Sci. Tech. **1**, 446-451, (1990).

[4] Lauf, R.J. and Hoffheins, B.S., "Analysis of Liquid Fuels Using a Gas Sensor Array," Fuel **70**, 935-940 (1991).

[5] Hoffheins, B.S., Using Sensor Arrays and Pattern Recognition to Identify Organic Compounds, (MS-Thesis, University of Tennessee, Knoxville, TN, 1989).

[6] Sundgren, H., Winquist, F., Lukkari, I., and Lundström, I., "Artificial Neural Networks and Gas Sensor Arrays: Quantification of Individual Components in a Gas Mixture," Meas. Sci. Tech. **2**, 464-69, (1991).

[7] Rumelhart, D.E. and Hinton, G.E., Williams, R.J., "Learning Internal Representations by Error Propagation," in **Parallel Distributed Processing**, vol. 1, Rumelhart and McClelland, eds. (MIT Press, Cambridge, MA, 1986), ch. 8.

[8] Hinton, G.E. and Sejnowski, T.J., "Learning and relearning in Boltzmann machines," in **Parallel Distributed Processing**, vol. 1, Rumelhart and McClelland, eds. (MIT Press, Cambridge, MA, 1986), ch. 7.

[9] Kohonen, T., "Self-organized formation of topologically correct feature maps," Biol. Cybern. **43**, 62 (1982).

[10] Wythoff, B.J., Levine, S.P. and Tomellini, S.A., "Spectral Peak Verification and Recognition Using a Multilayered Neural Network," Anal. Chem. **62**, 2702-2709 (1990).

[11] Tanabe, K., Tamura, T. and Uesaka, H., "Neural Network System for the Identification of Infrared Spectra," Appl. Spectroscopy **46**, 807-810 (1992).

[12] Shadmehr, R., Angell, D., Chou, P.B., Oerlein, G.S., and Jaffe, R.S., "Principal Component Analysis of Optical Emission Spectroscopy and Mass Spectrometry: Application to Reactive Ion Etch Process Parameter Estimation Using Neural Networks," J. Electrochem. Soc. **139**, 907-914 (1992).

[13] Meyer, B., Hansen, T, Nute, D., Albersheim, P., Darvill, A., York, W. and Sellers, J., "Identification of the H-NMR Spectra of Complex Oligosaccharides with Artificial Neural Networks," Science **251,** 542-544 (1991).

[14] Venkatasubramanian, V. and Chan, K., "A Neural Networks Methodology for Process Fault Diagnosis," Am. Inst. Chem. Eng. J. **35**, 1993-2002 (1989).

[15] Unger, L.H., Powell, B.A., and Kamens, S.N., "Adaptive Networks for Fault Diagnosis and Process Control," Computers in Chem. Eng. 24, 561-572 (1990).

[16] Khalid, M. and Omatu, S., "A Neural Network Controller for a Temperature Control System," IEEE Control Systems **12(4)**, 58-64 (June 1992).

[17] IEEE Control Systems, **10(2)**, (April 1990).

[18] IEEE Control Systems, **12(2)**, (April 1992).

[19] E. Oja., "A Simplified Neuron Model as a Principal Component Analyzer," **Biology**, (Berlin, Springer Verlag, 1982), pp. 267-273.

[20] Anthony, D., Hines, E., Barham, J., and Taylor, D., "A Comparison of Image Compression by Neural Networks and Principal Component Analysis," in **Proceeding of the 1990 International Joint Conference on Neural Networks,** vol. 1, 339-344 (1990).

[21] Otto, M., George, T., Schierle, C., and Wegscheider, W., "Fuzzy logic and Neural Networks - Applications to Analytical Chemistry," Pure and Appl. Chem. **64**, 497-502 (1992).

[22] Peterson, K.L., "Classification of Cm I Energy Levels Using Counterpropagation Neural Networks," Phys. Rev. A **A41**, 2457-2461 (1990).

[23] Clark, J.W., Gazula, S., Bohr, H., "Nuclear Phenomenology with Neural Networks," in **Complex Dynamics in Neural Networks**, Caianiello, Clark, Cotterill, and Taylor, eds, (Springer-Verlag, Berlin, 1992), pp. 305-322.

[24] Gernoth, K.A., Clark, J.W., and Prater, J.S., "Neural network models of nuclear systematics," Phys. Lett. B **300**, 1-7 (1993).

DATE

FILMED

9/21/93

END