Available online at www.sciencedirect.com

ScienceDirect

Procedia Manufacturing 35 (2019) 629–634

Procedia MANUFACTURING

www.elsevier.com/locate/procedia

2nd International Conference on Sustainable Materials Processing and Manufacturing

(SMPM 2019)

Big data analysis for gas sensor using convolutional neural network and ensemble of evolutionary algorithms

Ima Essiet[a], Yanxia Sun[a]*, Zenghui Wang[b]

[a]*Dept. of Electrical and Electronic Engineering Science, University of Johannesburg, Johannesburg, 2006, South Africa*
[b]*Dept. of Electrical and Mining Engineering, University of South Africa, Florida, 1710, South Africa*

Abstract

Big data analysis has gained popularity over the years as a result of developments in computing and electronics. Several methods have been proposed in literature for efficiently mining data from dedicated databases and a wide range of electronic sensors. However, as the volume of data grows, diversity and velocity of the data also grows (sometimes exponentially). Neural networks have been proposed in literature for optimal big data mining; however, they suffer from problems of over-fitting and under-fitting. In this paper, an ensemble of evolutionary algorithms is proposed, namely: improved non-dominated sorting genetic algorithm (NSGA), differential evolution (DE) and multi-objective evolutionary algorithm based on dominance and decomposition (MOEAD/D). These algorithms are each combined with a convolutional neural network (CNN); performance is evaluated using root mean square error (RMSE), and mean absolute percentage error (MAPE). The test data consists of gas sensor readings obtained from an array of 16 metal oxide semiconductor sensors. The gases being detected are Carbon Monoxide/Ethylene in air, and Methane/Ethylene in air. 4,178,504 data points were collected over an uninterrupted 12-hour period. Preliminary results show improved RMSE and MAPE values over 50 learning cycles compared to a case where the CNN learned on its own.

Peer-review under responsibility of the organizing committee of SMPM 2019.

Keywords: convolutional neural network (CNN), deep learning, classification accuracy, evolutionary algorithms, big data

1. Introduction

Big data processing involves techniques which reliably and efficiently interpret or make inferences based on the velocity, volume, variety, value and veracity of the data [1-3]. Data velocity refers to the speed at which data moves

* Corresponding author..
E-mail address: sunyanxia@gmail.com

2351-9789
Peer-review under responsibility of the organizing committee of SMPM 2019.
10.1016/j.promfg.2019.06.005

(or is transferred) from one source to another. Volume is concerned with the size of data collected from different digital sources and devices. The variety of big data describes the different forms of data collected from various digital devices i.e. audio, video, images, data logs and so on [4]. Data value is concerned with techniques employed to extract meaningful or usable information from the data being analyzed. Data veracity refers to the integrity or reliability of the results or inferences made from the data that has been analyzed. All these aspects are what constitute the idea of big data analysis. Therefore, big data can be discussed from the viewpoint of one or more of these aspects.

Big data analytics are particularly concerned with veracity and value of the data being analysed. This is because the results of these analytical tools and methods are used to make decisions which can impact positively or negatively on decision making for business owners and policy makers. Another definition of big data highlights the fact that only specific data analysis tools can be used to successfully mine big data: 'the exponential growth and wide availability of digital data that are difficult or even impossible to be managed and analyzed using conventional software tools and technologies' [5]. Convolutional neural networks (CNNs) are an instance of one such tool that has been used to successfully analyze big data. A CNN is composed of successive layers of convoluted and sub-sampled data based on the data originally inputted into the network. Feature maps are formed by convoluting the input data with a number of predetermined filters. The input data can either be from the outside world, or from a previously sub-sampled layer within the network [5]. In addition to convolution, operations like non-linear activation and spatial pooling are used to create feature maps [6]. A typical CNN is shown in Fig. 1.

In this work, we propose the use of evolutionary algorithms (EAs) to optimize an unsupervised learning method called predictive sparse decomposition (PSD) [7]. PSD involves finding an optimal setting for the basis function sets and filter bank such that both reconstruction error and code prediction error are minimized simultaneously. This approach is proposed as a result of the prevalent problem of handling high-volume, high-dimensionality data using CNNs. Although CNNs have the capability to quickly learn patterns in data, it is more challenging to make accurate inferences when the training method is unsupervised. Evolutionary algorithms are capable of using heuristics to find optimal solutions to complex, multi-modal problems [8].

Contribution

The approach proposed in this paper aims to improve classification accuracy of convolutional neural network (CNN) compared to a case where the CNN learns with back-propagation. The rest of the paper is organized as follows: Section 2 discusses the use of CNN for image data classification and recognition. Section 3 discusses the proposed approach of optimizing CNN classification of sensor data using EAs for error reduction with PSD. Section 4 discusses results obtained with the proposed approach, and Section 5 concludes the paper.

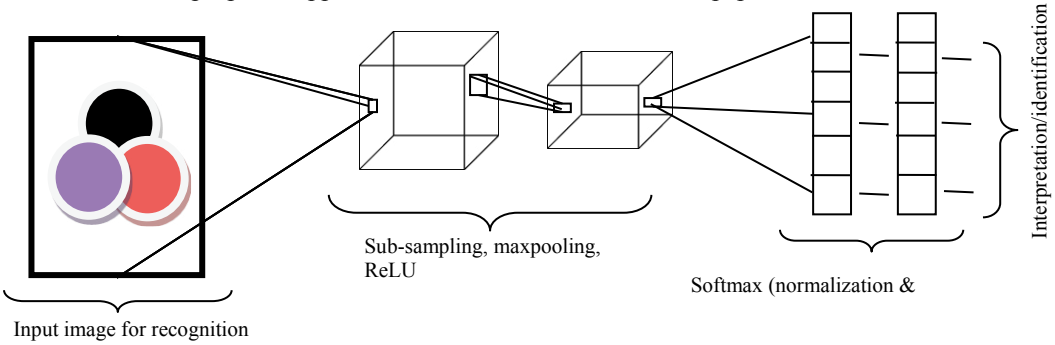


Fig. 1. Convolutional Neural Network for Image Classification and Recognition

2. Convolutional Neural Networks in Image Data Classification and Recognition

CNNs are particularly useful for data mining because they are capable of adapting to good feature hierarchies automatically, while ensuring that translational and distortional invariances are maintained [5]. In particular, a deep CNN was proposed in [6] to implement scene details learning from multiple patches of a cropped image. The first convolution layer filters consisted of cropped images from 5 regions of the original image (sourced from GoogLeNet). These cropped regions were flipped horizontally to create 10 sub-crops which were inputted into the first CNN layer. Sub-sampling was done using maxpooling. Training was done using back-propagation by stochastic gradient descent.

A batch size of 40 samples was used. From results obtained, the proposed deep CNN outperformed 6 other state-of-the art methods in terms of classification accuracy.

In [9], a random forest strategy based on classification and regression trees was proposed for data classification of an electronic tongue. This method was earlier proposed in [10] to overcome the problem of high dimensionality of data in the presence of many variables. Principal component analysis (PCA) was used to distribute samples of 4 different data sets. Classification accuracy of the random forest strategy was compared to that of back-propagation neural network and support vector machine (SVM). Average correct classification rates for the three methods on the selected data sets were 86.68%, 66.45% and 99.07% respectively.

A restricted Boltzmann machine (RBM) was used for 2-lead electrocardiography (ECG) classification to detect heart defects in [11]. The widely used ANSI/AAMI EC57 classification database was not capable of handling large, high-dimensional data sets because classification was done using supervised learning. Therefore, an unsupervised RBM was used to train data from the MIT-BIH arrhythmia database [12]. Results showed that the proposed unsupervised RBM achieved classification accuracy of 98.829%. Other applications of CNNs to improve classification accuracy of high-dimensional data can be found in [13-17].

From the research highlighted, there is need for data classification algorithms that can accurately classify data characterized by high volume and high dimensionality (particularly with unsupervised learning). In this research, we propose the use of CNN with an ensemble of evolutionary algorithms for optimal tuning of basis function sets and filter bank. The aim is to establish the capability of EAs to optimally tune network parameters, thus enhancing learning accuracy of CNNs.

3. Proposed Methodology

. Data classification using CNNs involves three basic steps: Data pre-processing, dimensionality and error reduction and classification. The first step involves noise removal from the raw sensor data which was done using the moving average filtering approach [18]. The filtered data samples were then encoded into three-channel images using the approach in [19]. The Cartesian coordinates of the data points were converted to polar coordinates according to:

$$\sigma_t = \cos^{-1}(S_n^t),\ 0 \le S_n^t \le 1 \quad (1)$$

where S_n^t is the sensor response for the *n*-th sensor at time *t*. In this paper, *n*= [1,….16].

Pixel normalization was performed on transformed sensor images using the approach in [20]. The min-max normalization was used with interval [0, 255] according to:

$$N = \alpha + \frac{(N-N_{min})(\beta-\alpha)}{N_{max}-N_{min}} \quad (2)$$

where α and β are minimum and maximum values of the color space, N is the normalized pixel, N_{min} and N_{max} are minimum and maximum values of data set *n* respectively.

The method of unsupervised learning considered in this paper is the predictive sparse decomposition (PSD) approach [5]. The main contribution of the paper is to simultaneously minimize reconstruction error and code prediction error using EAs. The results of optimal error reduction are compared with the performance of CNN without EAs. The PSD approach is described according to:

$$\Psi = |||I - FS|||_2^2 + \gamma|S|_1 + \delta\|S - G \cdot \tanh(PI)\|_2^2 \quad (3)$$

Ψ is the matrix of predicted sparse coefficients, I is the input matrix to be approximated, F is a linear basis matrix, S is the matrix of sparse coefficients, G is diagonal gain matrix, P is a matrix of filter coefficients. For the sake of clarity, we refer to γ and δ as sparse representation coefficient and code prediction coefficient respectively.

The EAs are used to optimally minimize the reconstruction error represented by the first term of Equation 3, and the code prediction error which is the third term. This is done according to the steps shown in Algorithm 1. The objective minimization function is given by Equation (4).

$$\min f(\mathbb{A} + \mathbb{B})\ s.t. 0 < \gamma, \delta \le 1 \quad (4)$$

where $\mathbb{A} = |||I - FS|||_2^2$ and $\mathbb{B} = \delta\|S - G \cdot \tanh(PI)\|_2^2$

The specification of parameters for the selected EAs (NSGA, DE and MOEAD/D) is given in Table 1. The CNN structure being proposed consists of 5 stages (6 layers) of convolutional filtering with maxpooling at each stage for downsampling to ensure dimensionality reduction. The structure of the CNN is shown in Fig. 2. The classification stage consists of a 2-dimensional vector for the classification of both Carbon Monoxide and Methane respectively. Softmax was used to provide the final classification. Matlab neural network toobox was used to simulate the CNN.

60% of the sensor data was used to train the CNN while 40% was used to test the network. Hidden layer weights were adjusted using rectified linear unit (ReLU) approach. The number of training and testing cycles (epochs) was selected as 50. Gas sensor data were obtained from [21].

Algorithm 1

Start
1 Input I, F, S, P, G, maxgen
2 Initialize γ, δ, Ψ
3 Use EA to search for optimal F, S, P, G based on specified objective function
4 Compute $\mathbb{A}$ *and* $\mathbb{B}$
5 Compute new Ψ *and compare approximate inputs of I with initial* Ψ
6 If computed Ψ *is optimal then*
7 Output Ψ, $\mathbb{A}$ *and* $\mathbb{B}$
8 Else
9 Repeat steps 2 to 5
10 End if
End

Table 1. Parameter Settings for MOEAD/D, NSGA and DE.

Parameter	Setting
Mutation factor (M)	0.8
Number of simulation runs per case	10
Mutation rate	1/*N*
Crossover probability (ρ_c)	0.7
Population size	*N*=50
Number of generations	*maxgen*=500
Mutation probability (ρ_m)	0.8
Vector size for EA encoding	2×16

4. Results and Discussion

Classification results of the CNN for Methane and Carbon Monoxide are shown in Table 2. From the results obtained, it can be seen that using EAs for selecting the sparse coefficient matrix yielded higher classification accuracy. In particular, MOEAD/D was the best-performing EA and performed better than CNN (trained using back-propagation) by 7.12% and 7.76% for classification of Methane and Carbon Monoxide respectively. These results demonstrate that EAs are capable of balancing bias and variance for the neural network, which results in better classification accuracy. In terms of the root mean square error (RMSE) and mean absolute percentage error (MAPE), the average error rates of EAs combined with CNN were lower than those of the CNN alone. Results are given in Table 3.

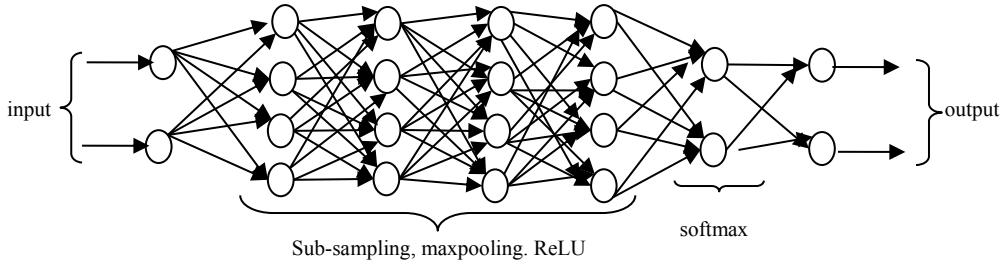


Fig.2. Structure of Fully-connected CNN used for Gas data Classification

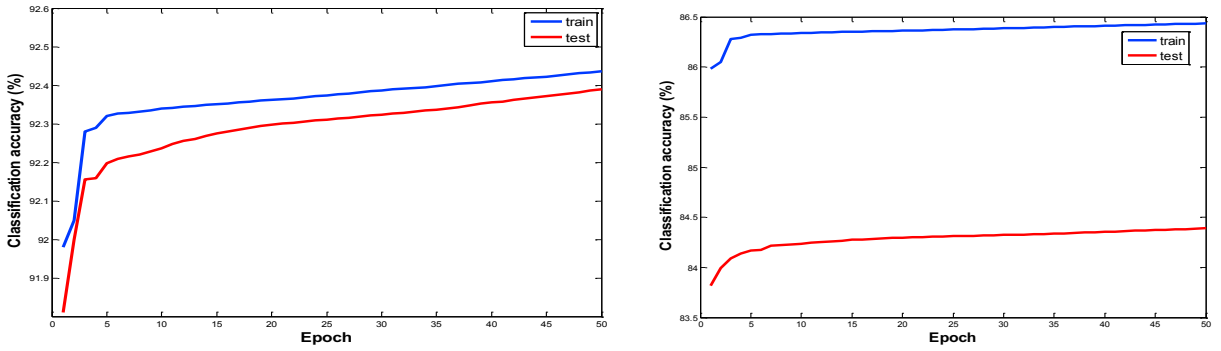
Table 2. Average Percentage Accuracy of Target Gas Identification based on CNN Training Method over 50 Epochs

	Training Method			
Target Gas	CNN+DE	CNN+MOEAD/D	CNN+NSGA	CNN+BP
Methane	90.68	92.86	88.71	85.74
Carbon Monoxide	89.37	91.45	90.05	83.69

Fig. 3 shows the variation between the training and test data sets for MOEAD/D and CNN, as well as CNN alone. From the variation over 50 epochs, it is seen that the training data matches the test data more closely for the EA and CNN combination. This demonstrates that EAs are capable of making optimal tradeoff between variance and bias of training and testing data with respect to data classification. The RMSE variation over 50 epochs for MOEAD/D+CNN, NSGA+CNN, DE+CNN and CNN alone is shown in Fig. 4.

Table 3. Average RMSE and MAPE of Target Gas Identification based on CNN Training Method over 50 Epochs

	Training Method			
Target Gas	CNN+DE	CNN+MOEAD/D	CNN+NSGA	CNN+BP
Methane	RMSE: 5.5E-02	RMSE: 2.5E-02	RMSE: 3.52E-02	RMSE: 6.5E-01
	MAPE: 4.08E-02	MAPE: 3.08E-02	MAPE: 3.64E-02	MAPE: 6.71E-01
Carbon Monoxide	RMSE: 6.02E-02	RMSE: 3.1E-02	RMSE: 4.03E-02	RMSE: 6.91E-01
	MAPE: 6.22E-02	MAPE: 3.17E-02	MAPE: 4.16E-02	MAPE: 7.02E-01



(a) (b)

Fig. 3, Classifcation accuracy for (a) CNN+MOEAD/D and (b) CNN+BP for training and test data sets for Methane gas

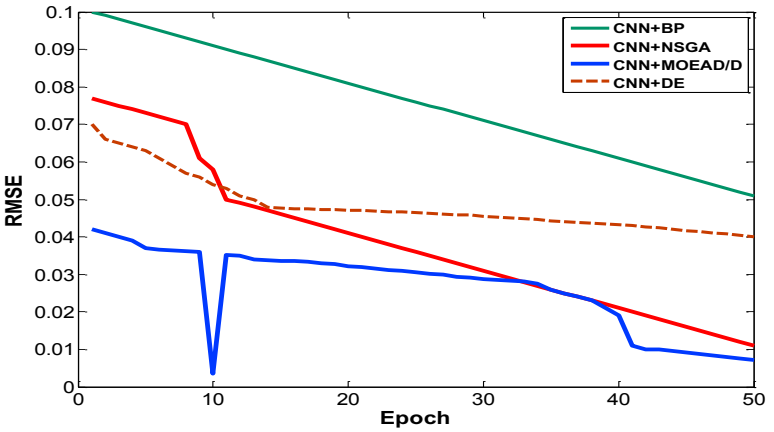


Fig. 4. RMSE Variation for DE, NSGA, MOEAD/D and Back-propagation with CNN over 50 Epochs for Methane gas

5. Conclusion and Future Work

This paper has presented preliminary work on the use of evolutionary algorithms in parameter selection for predictive sparse decomposition for CNN. The RMSE and MAPE were used to evaluate the performance of EA-based unsupervised learning. From results obtained, it can be seen that using EAs to minimize error rate for PSD improves performance of CNN. In particular, the MOEAD/D algorithm performed very well compared to the other EAs. This

is likely due to the capability of MOEAD/D to balance dominance with decomposition for a given optimization task. Therefore, EAs are able to simultaneously minimize reconstruction error and code prediction error for PSD. By this approach, the problems of over-fitting and under-fitting are minimized.

Future work will focus on the effect of varying the number of convolution layers on the performance of CNN for data classification tasks. Alternative methods of adapting network weights would also be investigated in order to improve classification accuracy. In particular, research would be carried out on the capability of EAs to optimize bias and variance in hidden layers based on their unique search strategy.

Acknowledgements

This research is supported partially by South African National Research Foundation Grants (No. 112108 and 112142), and South African National Research Foundation Incentive Grant (No. 95687 and 114911), Eskom Tertiary Education Support Programme Grants (Y. Sun, Z. Wang), Research grant from URC of University of Johannesburg.

References

[1] Laney, D. (2001). *3D Data Management: Controlling Data Volume, Velocity and Variety.*
[2] Beya, M., & Laney, D. (2012). *The Importance of 'Big Data': A Definition.* Stamford: Gartner.
[3] Hashema, I., Yaqooba, I., Anuara, N., Mokhtara, S., Gania, A., & Khaub, S. (2015). The Use of Big Data on Cloud Computing: Review and Open Research Issues . *Information Systems*, 98-115.
[4] Orgaz, G.-.., Jung, J., & Camacho, D. (2016). Social Big Data: Recent Achievements and New Challenges. *Information Fusion*, 45-59.
[5] Chen, X.-W., & Lin, X. (2014). Big Data Deep Learning: Challenges and Perspectives. *IEEE Access*, 514-525.
[6] Tan, Y., Tang, P., Zhou, Y., Luo, W., Kang, Y., & Li, G. (2017). Photograph Aesthitical Evaluation and Classification with Deep Convolutional Neural Networks. *Neurocomputing*, 165-175.
[7] Kavukcuoglu, K., Ranzato, M., Fergus, R., & LeCun, Y. (2009). Learning Invariant Features through Topographic Feature Maps. *Proc. Int. Conf. CVPR*, (pp. 1605-1612).
[8] Maltese, J., Ombuki-Berman, B., & Engelbrecht, A. (2016). Scalability Study of Many-objective Optimization Algorithms. *IEEE Transactions on Evolutionary Computation*, 1-18.
[9] Liu, M., Wang, M., Wang, J., & Li, D. (2013). Comparison of Random Forest, Support Vector Machine and Back Propagation Neural Network for Electronic Tongue Data Classification. *Sensors and Actuators B: Chemical*, 970-980.
[10] Breiman, L. (2001). Random Forests. *Machine Learning*, 5-32.
[11] Yan, Y., Qin, X., Wu, Y., & Zhang, N. (2015). A Restricted Boltzmann Machine based Two-lead Electrocardiography Classification. *2015 IEEE 12th Int. Conf. on Wearable and Implantable Body Sensor Networks* (pp. 1-9). IEEE.
[12] Goldberger, A., Amaral, L., Glass, L., Hausdorff, J., Ivanov, R., Mark, R., . . . Stanley, H. (2000). PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Search Resource for Complex Physiologic Signals. *Circulation*, 215-220.
[13] Lee, H., Park, M., & Kim, J. (2016). Plankton Classification on Imbalanced Large Scale Database via Convolutional Neural Networks with Transfer Learning. *2016 IEEE Int. Conf. on Image Processing* (pp. 3713-3717). IEEE.
[14] Zhu, X., Zhou, J. H., & Han, Y. (2016). Chinese Article Classification Oriented to Social Network based on Convolutional Neural Network. *2016 IEEE First Int. Conf. on Data Science in Cyberspace* (pp. 33-36). IEEE.
[15] Hu, J., Mou, L., Schmitt, A., & Zhu, X. (2017). FusioNet: A Two-Stream Convolutional Neural Network for Urban Scene Classification using PolSARand Hyperspectral Data. *2017 Joint Urban Remote Sensing Event (JURSE)*, (pp. 1-4).
[16] Guo, T., Dong, J., Li, H., & Gao, Y. (2017). Simple Convolutional Neural Network on Image Classification. *2017 IEEE 2nd Int. Conf. on Big Data Analysis* (pp. 721-724). IEEE.
[17] Abroyan, N. (2017). Convolutional and Recurrent Neural Networks for Realtime Data Classification. *7th Int. Conf. on Innovative Computing Technology* , (pp. 42-45).
[18] Alexander, B., Ivan, T., & Denis, B. (2016). Analysis of Noisy Signal Restoration Quality with Exponential Moving Average Filter. *2016 Int. Siberian Conference on Control and Communications (SIBCON)*, (pp. 1-4).
[19] Liu, Y.-J., Meng, Q., & Zhang, Z. (2018). Data Processing for Multiple Electronic Noses using Sensor Response Visualization. *IEEE Sensors Journal*, 1-10.
[20] Sare, P., & R, A. (2017). Pixel Normalization from Numeric Data as Input to Neural Networks for Machine Learning and Image Processing. *IEEE 2017 Int. Conf. on Wireless Communications, Signal Processing and Networking (WISPNET 2017)* (pp. 2221-2225). IEEE.
[21] UCI Machine Learning Repository, Center for Machine Learning and Intelligent Systems, University of California, Irvine. Available at: https://archive.ics.uci.edu/ml/datasets.html?sort=nameUp&view=list