Neural Visualization for the Analysis of Energy and Water Consumptions in the Automotive Industry

Raquel Redondo✉[1 [0000-0002-2444-5384]], Álvaro Herrero[1[0000-0002-9267-0620]], Emilio Corchado[2[0000-0001-8560-3991]], Javier Sedano[3[0000-0002-4191-8438]]

[1] Grupo de Inteligencia Computacional Aplicada (GICAP), Departamento de Ingeniería Civil, Escuela Politécnica Superior, Universidad de Burgos, Av. Cantabria s/n, 09006, Burgos, Spain
{rredondo, ahcosio}@ubu.es
[2]Departamento de Informática y Automática, Universidad de Salamanca, Spain
escorchado@usal.es
[3]Instituto Tecnológico de Castilla y León, Pol. Ind. Villalonquejar, C/ López Bravo 70, 09001, Burgos, Spain
javier.sedano@itcl.es

Abstract. This study presents the application of neural models to a real-life problem in order to study the energy and water consumptions of an automotive multinational company for resources saving and environment protection. The aim is to visually and naturally analyse different consumptions data for a whole year, month by month, from factories and locations worldwide where different kinds of products are produced. The data are studied in order to see whether the geographical location, the month of the year or the technology used in each factory are relevant in terms of consumptions and then take actions for a greener production. The consumptions dataset is analysed using different neural projection models: Principal Component Analysis and Cooperative Maximum-Likelihood Hebbian Learning. This unsupervised dimensionality reduction techniques have been applied, and subsequent interesting conclusions are obtained.

Keywords: soft computing, artificial neural networks, exploratory projection pursuit, industrial applications, energy consumption.

1 Introduction

In recent years, more and more companies and governments are concerned about the environment. Being respectful with the environment and consuming the minimum fossil energy and water is widely accepted as a strategy for greener industrial production as it greatly contributes to reduce atmospheric pollution [1].

Because of this commitment, companies carry out studies on their environmental impact. In order to do so, it is necessary to have relevant information and appropriate analysis tools to obtain indicators for taking decisions. In keeping with this idea, present study proposes soft computing techniques (more precisely, artificial neural networks) for the analysis of energy and water consumption of several factories from a multinational company in the automotive industry sector.

The consumption dataset is analysed using two projection methods: Principal Component Analysis (PCA) [2, 3] and Cooperative Maximum-Likelihood Hebbian Learning (CMLHL) [4], to identify the dataset structure and clearly identify the status of each one of the factories regarding environmental issues. If the initial collected dataset, once analysed shows a certain degree of clustering, it can be seen as a sign of a representative data set (there is not a single problem related when collecting the information and the process is well defined by such data set).

The data are provided by Grupo Antolin [5]. In order to characterize the different situations in which each one of its factories is, from an environmental point of view, it is necessary to analyse if it is a representative and informative enough dataset.

The remaining sections of this study are structured as follows: Section 2 presents the techniques and methods that are applied to analyse the data. Section 3 details the real-life case study, while Section 4 describes the experiments and results. Finally, Section 5 sets out the conclusions and future work.

2 Data structure analysis using connectionist techniques

Soft Computing is a set of several technologies whose goal is to resolve inexact and complex problems [6]. It investigates, simulates, and analyses very complex issues and phenomena in order to solve real-world problems [7]. Soft Computing has been successfully applied for data analysis, and plenty of algorithms are reported in the literature [8, 9].

The unsupervised learning is useful to explore a dataset when there is no a specific goal or it is not clear what information the data contains. It is also a way to reduce the dimensionality of a given dataset.

Following subsections describe the techniques applied in present study.

2.1 Principal component analysis

Principal Component Analysis (PCA) [2] is a well-known method that provides the best linear data compression in terms of least mean square error by addressing the data variance.

PCA generates new variables by linear combinations of the original variables so that small numbers of new variables captures most of the information. When the datasets have many variables, often they are redundant. PCA takes advantage of the redundancy of information on them.

With PCA, it is possible to discover a smaller group of underlying variables that define the data. PCA has been the most frequently reported linear operation involving unsupervised learning for data compression and feature selection.

Though it was proposed as a statistical method, it has been proven that it can be implemented by several ANNs [10, 11].

2.2 Cooperative maximum likelihood hebbian learning

Exploratory Projection Pursuit (EPP) [12, 13, 14] is a more recent statistical method designed to resolve the difficult problem of identifying structure in high dimensional data. It projects the data onto a low dimensional subspace in which the data structure is searched by eye. However, not all projections will reveal this structure equally well. It therefore defines an index that measures how "interesting" a given projection is, and then represents the data in terms of projections that maximize the index.

Cooperative Maximum Likelihood Hebbian Learning (CMLHL) [4] is an extended version of Maximum Likelihood Hebbian Learning (MLHL) [15], an EPP connectionist model. CMLHL incorporates lateral connections [4, 16] derived from the Rectified Gaussian Distribution (RGD) [17]. The RGD is a modification of the standard Gaussian distribution in which the variables are constrained to be non-negative, enabling the use of non-convex energy functions. The CMLHL architecture is represented in Fig. 1, where lateral connections are shown.

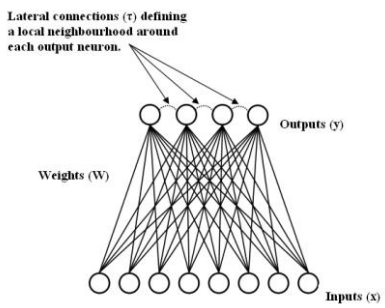

Fig. 1. CMLHL: lateral connections between neighbouring output neurons

The resultant net can find the independent factors of a data set but does so in a way that captures some type of global ordering in the data set.

Considering an N-dimensional input vector (x), an M-dimensional output vector (y) and with W_{ij} being the weight (linking input j to output i), CMLHL can be expressed as:

Feed-forward step:

$$y_i = \sum_{j=1}^{N} W_{ij} x_j, \forall i \qquad (1)$$

Lateral activation passing:

$$y_i(t+1) = \left[y_i(t) + \tau(b - Ay) \right]^+ \qquad (2)$$

Feedback step:

$$e_j = x_j - \sum_{i=1}^{M} W_{ij} y_i, \forall j \qquad (3)$$

Weight change:

$$\Delta W_{ij} = \eta . y_i . sign(e_j) | e_j |^p \qquad (4)$$

Where: η is the learning rate, τ is the "strength" of the lateral connections, b the bias parameter and p is a parameter related to the energy function [4, 15].

A is a symmetric matrix used to modify the response to the data, the effect of which is based on the relation between the distances between the output neurons. It is based on Cooperative Distribution, but to speed up the learning process, it can be simplified to:

$$A(i, j) = \delta_{ij} - \cos(2\pi(i - j)/M) \qquad (5)$$

where, δ_{ij} is the Kronecker delta.

3 A real case study: energy and water consumptions analysis of an automotive multinational company

As previously stated, in present research neural techniques are applied to analyse the energy and water consumptions of an automotive multinational company. To do this analysis, some data have been selected and gathered.

The data used in this study are taken from year 2016 and it comprises energy and water consumptions of all the worldwide factories and locations of Grupo Antolin [5]. This multinational company from the automotive industrial sector, is one of the largest players in the car interiors international market and the number 1 worldwide supplier of headliner substrates. It has presence in 26 countries with 167 production plants & Just in Time centers and 29 technical-commercial offices. Grupo Antolin manufactures products that are technologically sustainable, based on two premises: light and green, thereby contributing to lower CO2 emissions. Grupo Antolin is strongly committed to the environment so this study is a first step in order to analyse with soft computing techniques the different consumptions in the company.

The dataset under analysis comprises a total of 1090 samples. The data are from year 2016, and there are a sample for each month of the year from 91 different factories. The following consumption parameters (features) were gathered for each one of the factories:

- Electricity (KWh)
- Natural gas (m^3)
- Propane (m^3)
- Fuel (l)
- Gasoil (l)
- Water (m^3)

As explained before, this dataset gathers data from different factories or centers, in different countries worldwide, where several kinds of technologies and products are manufactured. Each center could has several kinds of consumes, for example, not all of

the locations consume propane or fuel but all the factories have electricity consumptions.

4 Results

The dimensionality reduction techniques have been applied in order to analyse and find the characteristics that best describe the collected data. The principal aim is to detect if a clear internal structure can be identified, this means that the selected data is informative enough. Otherwise, further data must be properly collected [18].

The above mentioned neural projection models have been applied to the data described in section 3, whose results are compiled in present section. PCA is the first technique that has been applied in order to find a possible structure in the data. In next figures, data samples are depicted in the PCA projection (2 first principal components), according to different criteria: continent were the factory is located, production technology, business unit and the month of the year.

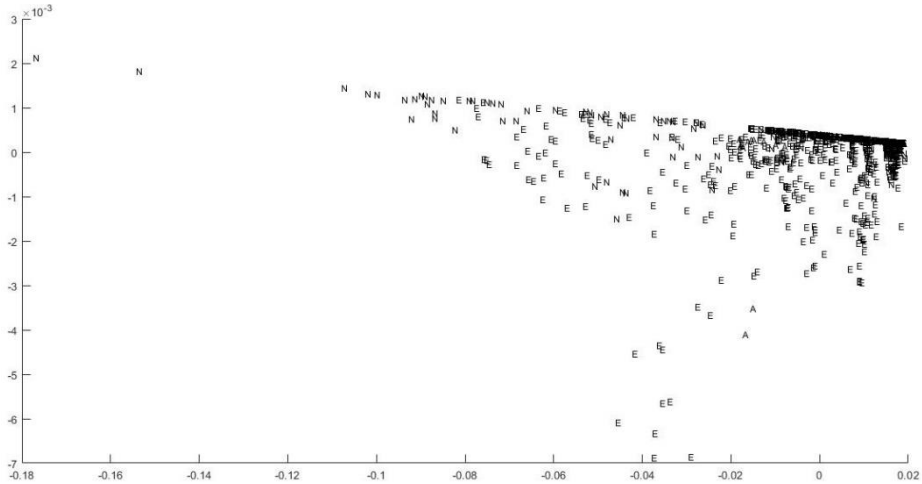

Fig. 2. PCA Projection–Continent: E: Europe, N: North America, A: Asia, F: Africa, S: South America.

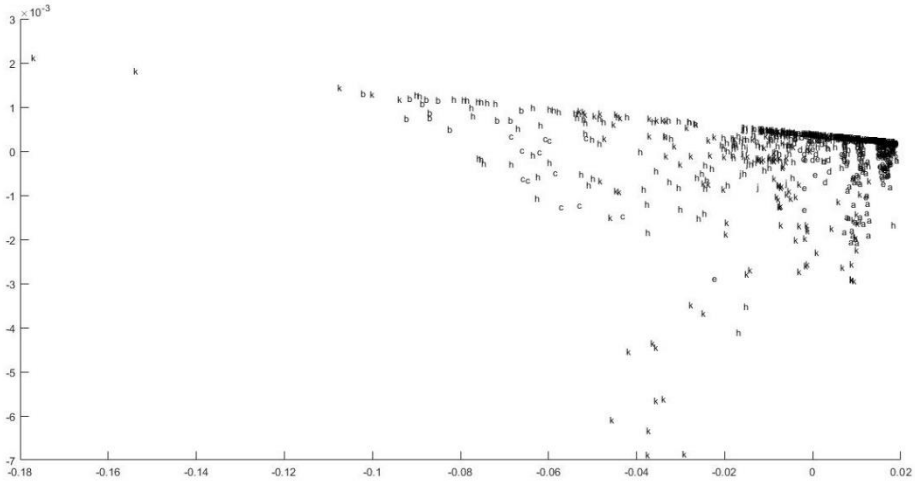

Fig. 3. PCA Projection–Technology.

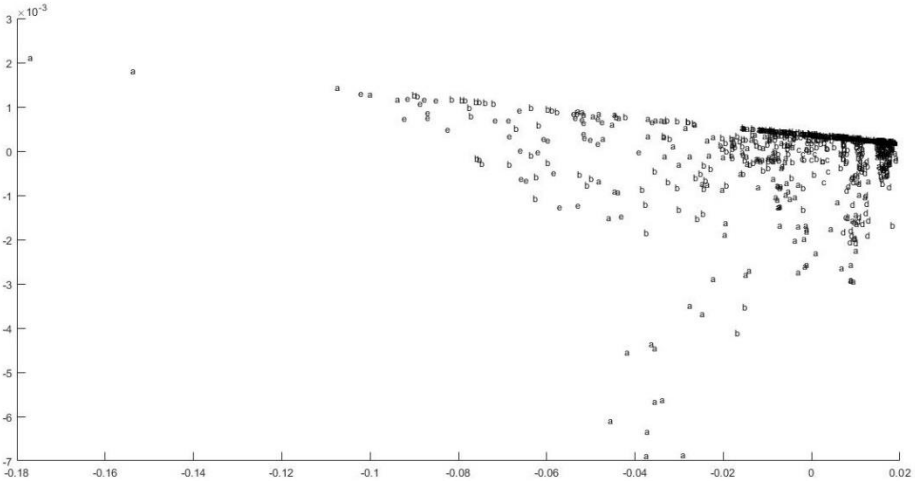

Fig. 4. PCA Projection–Business Unit.

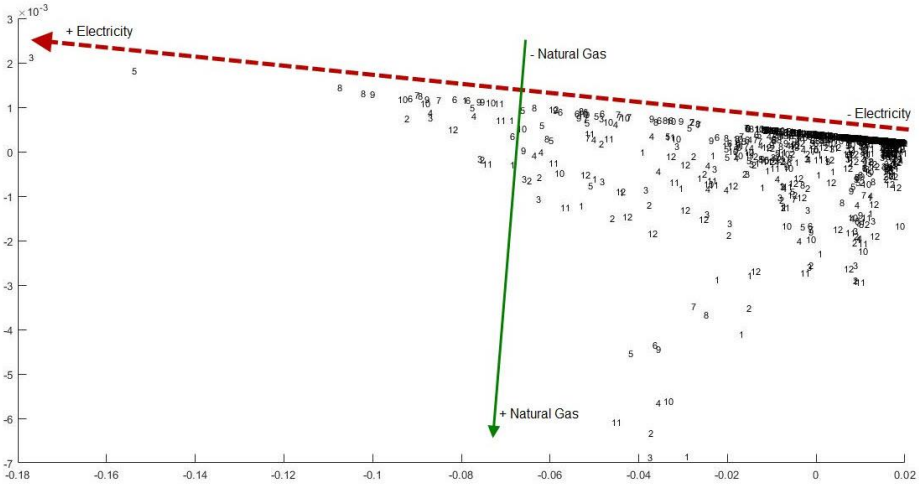

Fig. 5. PCA Projection–Month: one number per each month in the year (1-January, 2- February…)

Analysing the projection obtained with PCA (from Fig. 2 to Fig. 5) it can be concluded that PCA shows a clear internal structure in the dataset by identifying the consumption of electricity and natural gas as the most relevant features according to the data ordering, as stated in Fig. 5. On the other hand, different data groups can be identified in the projections, the samples located on the left of the projections are from big factories that consume more electricity, and the samples at the bottom are from big factories that consume more natural gas. These factories, as can be seen in Fig. 2, are from different continents; as a result, they use different types of energy.

CMLHL is the second technique that has been applied. In the following figures CMLHL results are shown, where data are depicted according to four criteria as in the case of PCA (continent, technology, business unit and month of the year).

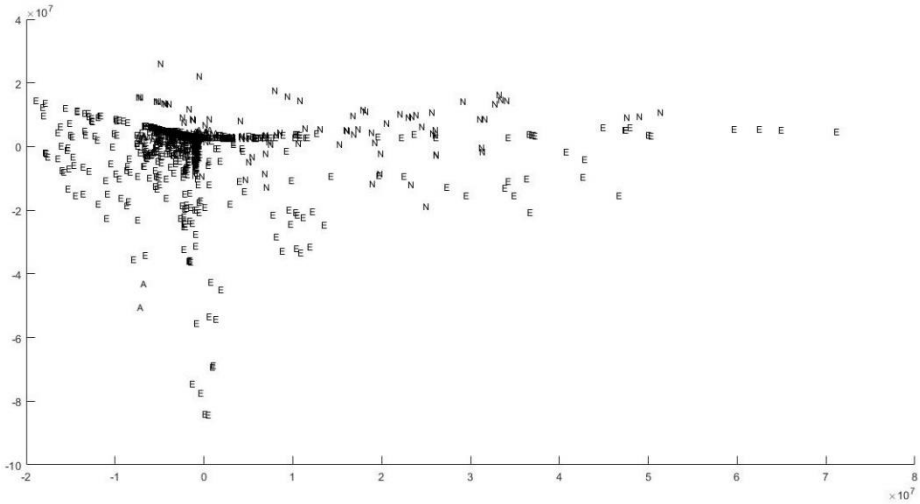

Fig. 6. CMLHL Projection–Continent: E: Europe, N: North America, A: Asia, F: Africa, S: South America

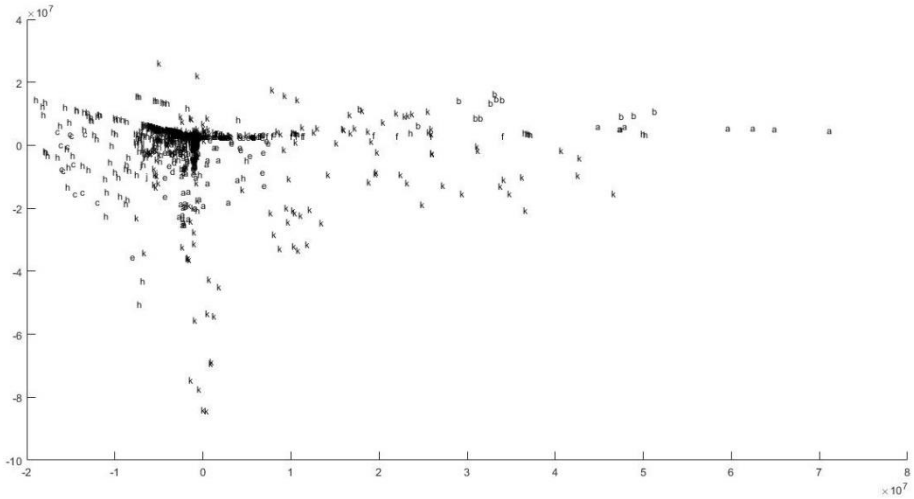

Fig. 7. CMLHL Projection–Technology.

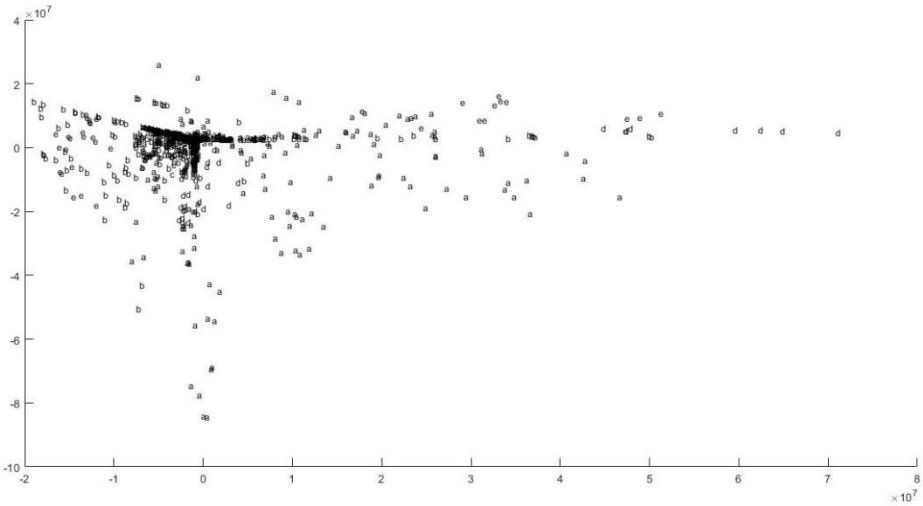

Fig. 8. CMLHL Projection–Business Unit.

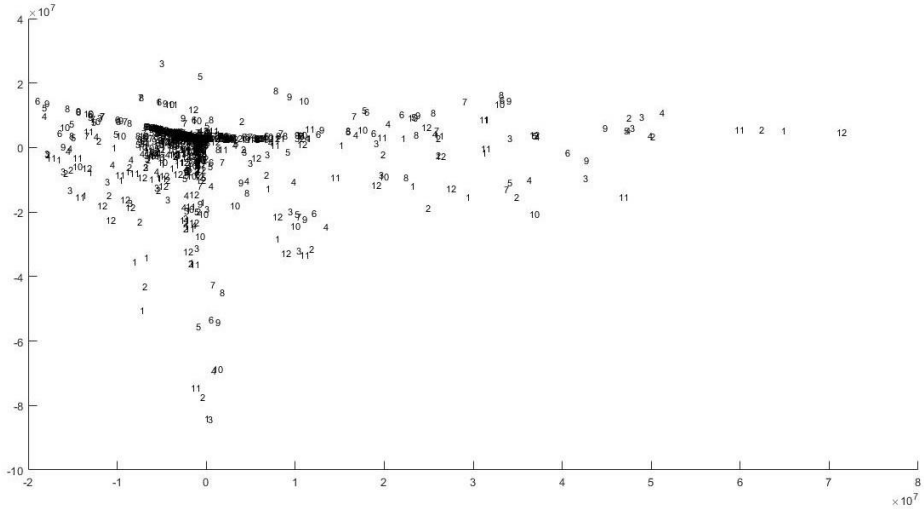

Fig. 9. CMLHL Projection–Month: one number per each month in the year

An analysis of the results obtained with the CMLHL model (from Fig. 6 to Fig. 9) leads to the conclusion that this method reveals the dataset structure in a clearer way than PCA.

In Fig. 6, it is clearly identified the grouping according to the continent where the factory is located. There are two main groups in this projection; one gathering factories from Europe, Asia and South America and another one those from North America. In Fig. 7, different groups of data can be identified, there are several clusters were the technology '*k*' is clearly identified. In Fig. 8, groups by business unit are visibly represented, the different business units usually group similar technologies, and because of

that the projection for technologies are very similar to the business unit one; and in Fig. 9, it can be concluded that there are small clusters grouped by season, points from winter are closer between them and in the same way summer points. One reason for the difference between summer and winter is that in winter it is necessary to use heating. Also some factories in summer have vacations period.

In this study PCA and CMLHL are applied in order to bear out that the dataset is representative enough. By comparing two different neural projection techniques, more evidences supporting conclusions are provided.

An analysis of the results obtained with the CMLHL model leads to the conclusion that this technique shows several different clusters in a clearer and sparser way than PCA; the points from the same factory or locations are grouped together.

5 Conclusions and futures lines of work

Once the dataset has been analysed, it can be concluded that a certain structure has been revealed and thus, it can be seen as a representative and informative dataset. PCA projections show samples ordered by electricity and natural gas consumption. In addition, with CMLHL it is obtained more information because it is observed that the data are grouped by continent, technology, business unit and there are even clusters by factory.

These results are relevant for the Grupo Antolin company. The obtained projections show certain degree of clustering, either by type of manufacturing technology, business unit, etc., which will allow the company to make decisions, taking into account, for example, the number of pieces produced in each factory. The company could study which kind of energy used in certain factories is the most economic or the most environment friendly or even if one factory has anomalous consumptions when compared with another similar factories.

This is a first study to analyse if the data have a structure or information that could be relevant. In the future, this study will go on with data from additional years and also it would be possible to work with more variables such as CO2 emissions, hazardous waste, number of pieces produced, etc. With a future analysis the company would be able to detect what topics could be developed in order to improve its commitment to be environmentally-friendly.

Acknowledgments

The authors would like to thank the vehicle interior manufacturer, Grupo Antolin, for its collaboration in this research.

References

1. R. San José, J. L. Pérez and R. M. González: An operational realtime air quality modelling system for industrial plants, Environmental Modelling & Software, vol. 22, pp. 297-307, (2007)
2. I.T. Jolliffe. Principal Component Analysis, (second ed.), Springer, New York, NY (2002)

3. R. Platon, V. R. Dehkordi, J. Martel. Hourly prediction of a building's electricity consumption using case-based reasoning, artificial neural networks and principal component analysis, Energy and Buildings, Volume 92, pp. 10-18, (2015)
4. E. Corchado, Á, Herrero. Neural visualization of network traffic data for intrusion detection. Appl. Soft Comput., vol. 11(2), pp. 2042-2056, (2011)
5. Grupo Antolin : http://www.grupoantolin.com last accessed 2018/04/03
6. E. Corchado, B. Baruque. Wevos-visom: An ensemble summarization algorithm for enhanced data visualization. Neurocomputing, vol. 75(1), pp. 171-184, (2012)
7. D.B. Araya, K. Grolinger, H. F. ElYamany, M. A.M. Capretz, G. Bitsuamlak. An ensemble learning framework for anomaly detection in building energy consumption, Energy and Buildings, Volume 144, pp. 191-206, (2017)
8. H. Garg, M. Rani, S.P. Sharma. An approach for analyzing the reliability of industrial systems using soft-computing based technique, Expert Systems with Applications, Volume 41, Issue 2, pp. 489-501, (2014)
9. H. Gitinavard, S. M. Mousavi, B. Vahdani, Soft computing based on hierarchical evaluation approach and criteria interdependencies for energy decision-making problems: A case study, Energy, Volume 118, pp. 556-577, (2017)
10. E. Oja: "Principal components, minor components, and linear neural networks", Neural Networks, vol. 5, pp. 927-935 (1992)
11. E. Oja: "Neural networks, principal components, and subspaces", International Journal of Neural Systems, vol. 1, pp. 61-68 (1989)
12. J. Friedman, Exploratory projection pursuit. Journal of the American Statistical Association, vol 82(397), pp. 249-266, (1987)
13. E. Corchado, D. MacDonald, C. Fyfe: Maximum and Minimum Likelihood Hebbian Learning for Exploratory Projection Pursuit, Data Mining and Knowledge Discovery, 8(3), pp. 203–225 (2004)
14. Á. Herrero, E. Corchado, L. Sáiz Bárcena, A, Abraham.. DIPKIP: A Connectionist Knowledge Management System to Identify Knowledge Deficits in Practical Cases. Computational Intelligence, vol. 26(1), pp. 26-56, (2010)
15. P. Krömer, E. Corchado, V. Snásel, J. Platos, L. García-Hernandez. Neural PCA and Maximum Likelihood Hebbian Learning on the GPU. ICANN (2), pp. 132-139, (2012)
16. E. Corchado, Y. Han, C. Fyfe: Structuring global responses of local filters using lateral connections, Journal of Experimental & Theoretical Artificial Intelligence 15(4), pp. 473–487 (2003)
17. H. Seung, N. Socci, D. Lee: The Rectified Gaussian Distribution, Advances in Neural Information Processing Systems 10, pp. 350–356 (1998)
18. V. Vera, J. Sedano, E. Corchado, R. Redondo, B. Hernando, M. Camara, A.Laham, A. E. Garcia: A Hybrid System for Dental Milling parameters optimisation. 6th International Conference on Hybrid Artificial Intelligence Systems. Wroclaw, Poland HAIS 2011, Part II, LNAI 6679, 2011, pp. 437–446 (2011)