Hybridizing Deep Learning and Neuroevolution: Application to the Spanish Short-Term Electric Energy Consumption Forecasting

PINV18-846: Análisis de la eficiencia energética en edificios no residenciales mediante técnicas metaheurísticas y de inteligencia artificial

11/Octubre/2021

Miguel García Torres
Data Science & Big Data Research Lab
Lenguajes y Sistemas Informáticos, Universidad Pablo de Olavide

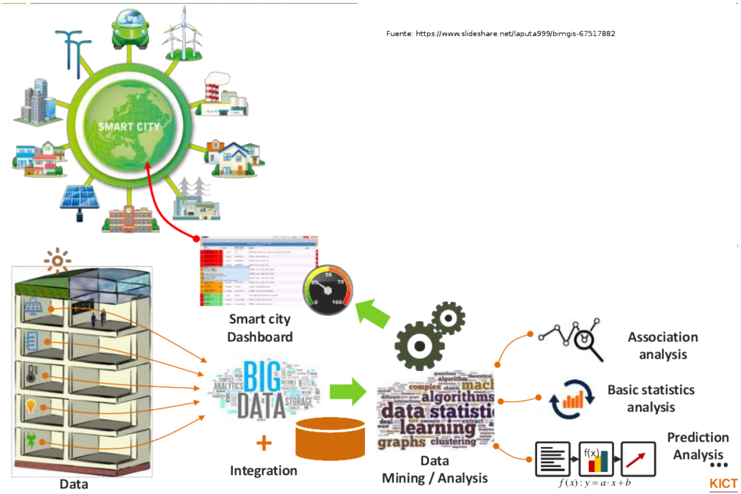
What is a smart city?

Impact

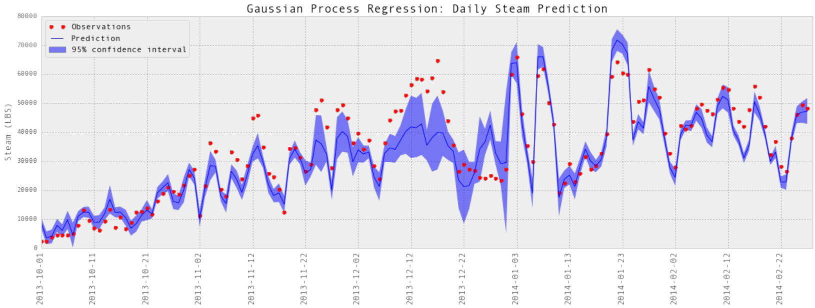
What is a smart building?

- Automated and integrated management system.
  - HCAV (heating, ventilation, and air conditioning).
  - Lighting, access control, security, etc.
- Remote monitoring with sensors.
- Decision-making support system.
- Benefits:
  - Energy efficiency.
  - Security.
  - Usability.
  - Accessibility.

Workflow


Fuente: https://www.slideshare.net/laputa999/bimgis-67517882
SMART CITY
Smart city
Dashboard
Data
BIG
DATA
ANALYTICS
STORAGE
Integration
Data
Mining / Analysis
Association
analysis
Basic statistics
analysis
Prediction
Analysis
...
$f(x): y = a \cdot x + b$
KICT

Time series data



Data

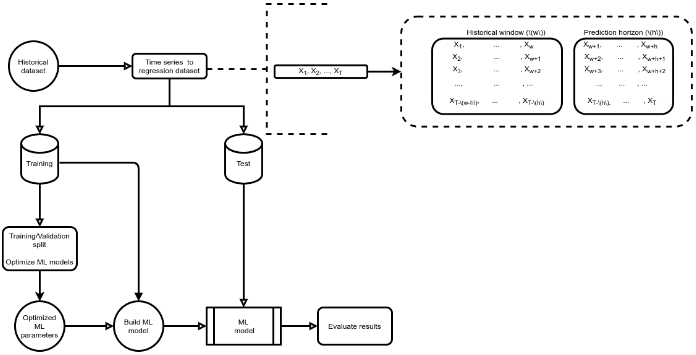
Training data

Date	T (C)	Humidity	Energy demand (mw)
23 Feb	9.47	0.89	20820
24 Feb	8.28	0.83	19950
25 Feb	8.75	0.83	19825
26 Feb	16.02	0.67	15437
...	...		
16 Apr	10.70	0.95	12375

Test data

Date	T (C)	Humidity	Energy demand (mw)
17 Jun	9.87	0.75	?
18 Jun	12.04	0.50	?

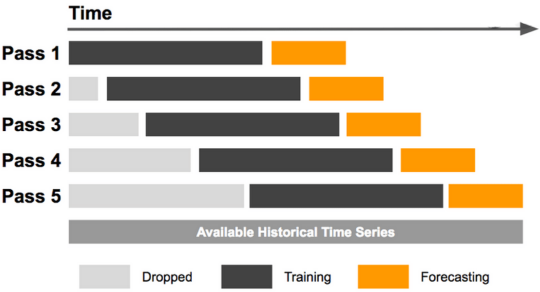
Data analysis workflow



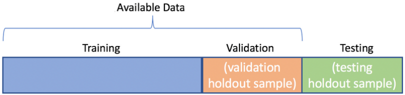
Model evaluation

Train-Test split: Walk Forward Validation

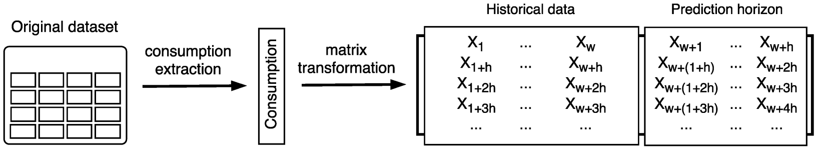
Walk Forward Validation



Validation for hyperparameter tuning



From time series to a regression problem



Training data

Date	**Cons. (mw)**
23 Feb	20820
24 Feb	19950
25 Feb	19825
26 Feb	15437
27 Feb	19825
28 Feb	15437
...	...

Historical: 3, Horizon: 2

Historical data (w)			Prediction horizon (h)	
23 Feb	24 Feb	25 Feb	26 Feb	27 Feb
24 Feb	25 Feb	26 Feb	27 Feb	28 Feb
25 Feb	26 Feb	27 Feb	28 Feb	1 mar
...	...			

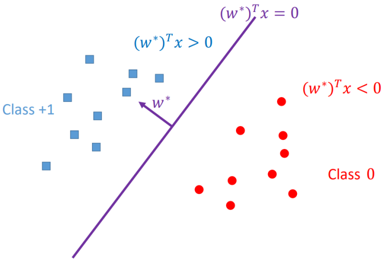
The Perceptron (Minsky-Papert, 1969)

The linear classifier

Inputs: feature values
Parameters: weights
Hypothesis: $f(x) = w^T x$

$$y = \begin{cases} 1 & \text{if } w^T x > 0 \\ 0 & \text{if } w^T x \leq 0 \end{cases}$$

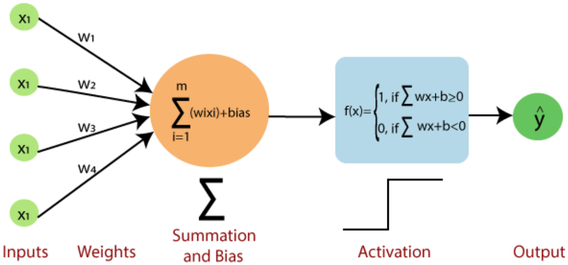
Prediction: $y = sign(f(x)) = sign(w^T x)$



The Perceptron (Minsky-Papert, 1969)

The linear classifier

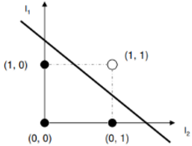
Inputs: feature values



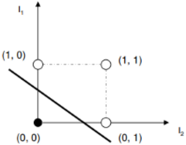
The Perceptron

The linear classifier

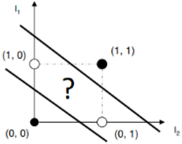
AND		
I_1	I_2	out
0	0	0
0	1	0
1	0	0
1	1	1



OR		
I_1	I_2	out
0	0	0
0	1	1
1	0	1
1	1	1

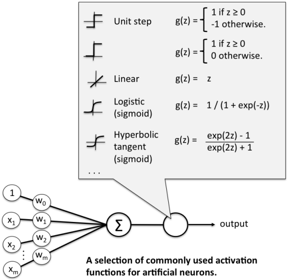


XOR		
I_1	I_2	out
0	0	0
0	1	1
1	0	1
1	1	0



The Perceptron

The nonlinear classifier



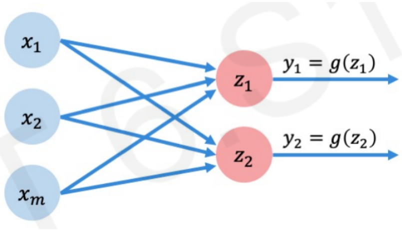
A selection of commonly used activation functions for artificial neurons.

$$y = g(z)$$

$$z = w_0 + \sum_{j=1}^{m} x_j w_j$$

The Perceptron

Multi output



$$y_i = g(z_i)$$

$$z_i = w_{0,i} + \sum_{j=1}^{m} x_j w_{j,i}$$

Single Layer Neural network

Multi output

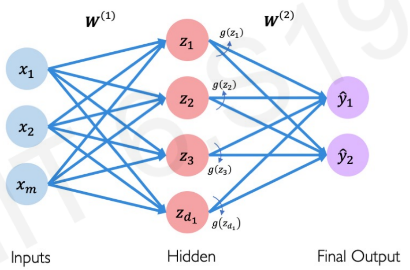
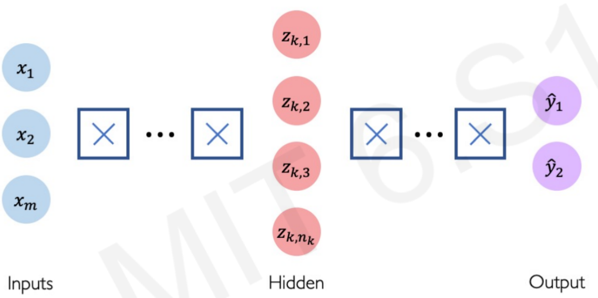


Figure adapted from A. Amini, Introduction to Deep learning.

$$z_i = w_{0,i}^{(1)} + \sum_{j=1}^{m} x_j w_{j,i}^{(1)}$$

$$\hat{y}_i = g(w_{0,i}^{(2)} + \sum_{j=1}^{d_1} g(z_j) w_{j,i}^{(2)})$$

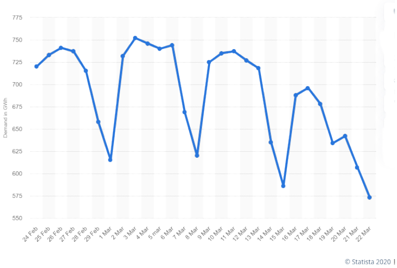
Deep Neural Network

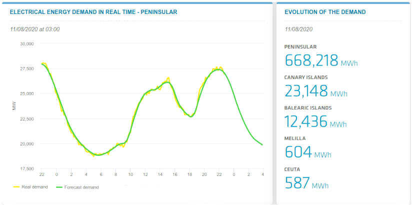


$$x_{k,i} = w_{0,i}^{(k)} + \sum_{j=1}^{n_{k-1}} g(z_{k-1,j}) w_{j,i}^{(k)}$$

Characteristics

- Data provided by the Spanish Nominated Electricity Market Operator (NEMO)
- Spanish electricity consumption from January 1, 2007 to June 21, 2016.
- 497832 Mesurements recorded every 10 minutes without neither missing values nor outliers.





Experiment setup

Magnitud of Relative Error

$$MRE = \frac{1}{n}\sum_{i=1}^{n}\frac{|Y_i - \hat{Y}_i|}{Y_i}$$

Machine Learning Algorithms: Random Forest (RF), Artificial Neural Networks (ANN), Evolutionary Decision Trees (EV), the Auto-Regressive Integrated Moving Average (ARIMA), the Gradient Boost Method (GBM), Decision Tree (DT) and an hybrid approach (ENSEMBLE).
Deep Learning Algorithms: Feed-Forward Neural Network (FFNN), Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM).

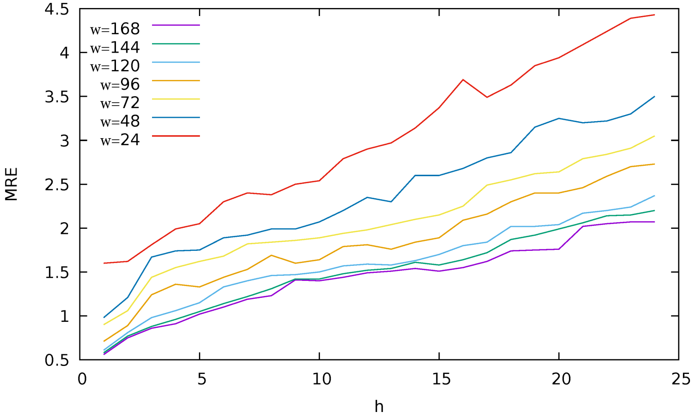
Experiment setup

Experiments

- Exp#1: Analysis of the historical window (w) impact on models quality.
- Exp#2: Hyperparameter tuning using a Genetic Algorithm (GA).
- Exp#3: Models comparison.

Results

Exp#1: Analysis of the historical window (w)



The proposed strategy obtains similar results for $w = 168, 144, 120$ on all the considered values of the prediction horizon h.
$W = 168$ was selected!

Results

Exp#2: Hyperparameter tuning

Parameters selected by the GA for each h using a w of 168

h	Layers	Neurons	λ	ρ	ϵ	Activation	Distribution
1	52	942	4.09×10^{-10}	1.00	6.43×10^{-12}	Tanh	Gaussian
2	68	921	0	1.00	0	Maxout	Huber
3	75	880	0	1.00	0	Maxout	Huber
4	68	921	0	1.00	0	Maxout	Huber
5	88	504	0	1.00	0	Maxout	Huber
6	80	789	0	1.00	0	Maxout	Huber
7	74	892	0	1.00	0	Maxout	Huber
8	46	300	0	1.00	0	Maxout	Huber
9	75	889	5.57×10^{-10}	0.99	6.74×10^{-10}	Tanh	Gaussian
10	25	852	0	1.00	0	Maxout	Huber
11	58	843	3.69×10^{-10}	1.00	2.45×10^{-10}	Tanh	Gaussian
12	41	491	0	1.00	0	Maxout	Huber
13	17	552	0	0.99	0	Maxout	Huber
14	26	661	0	0.99	0	Maxout	Huber
15	89	811	5.61×10^{-10}	0.99	4.23×10^{-10}	Tanh	Gaussian
16	98	697	0	1.00	0	Maxout	Huber
17	74	478	1.46×10^{-10}	1.00	3.58×10^{-10}	Tanh	Gaussian
18	62	705	2.74×10^{-10}	0.99	6.64×10^{-10}	Tanh	Gaussian
19	65	879	0	0.99	0	Maxout	Huber
20	81	780	7.62×10^{-10}	0.99	5.21×10^{-10}	Tanh	Gaussian
21	27	931	0	1.00	0	Maxout	Huber
22	95	745	0	1.00	0	Maxout	Huber
23	41	923	0	1.00	0	Maxout	Huber
24	80	754	0	1.00	0	Maxout	Huber

Results

Exp#3: Models comparison

Average results obtained by different methods for different historical window values. Standard deviation between brackets.

	w						
	24	48	72	96	120	144	168
NDL	3.01 (0.90)	2.38 (0.69)	2.08 (0.57)	1.85 (0.55)	1.60 (0.46)	1.51 (0.46)	1.44 (0.42)
CNN	4.08 (0.04)	3.16 (0.03)	2.69 (0.02)	2.51 (0.02)	2.30 (0.02)	1.71 (0.02)	1.79 (0.02)
LSTM	2.43 (0.03)	2.05 (0.02)	1.82 (0.02)	2.08 (0.02)	1.74 (0.02)	1.78 (0.02)	1.97 (0.02)
FFNN	4.51 (0.52)	3.46 (0.33)	3.39 (0.30)	3.12 (0.42)	2.98 (0.28)	2.32 (0.29)	2.46 (0.29)
ARIMA	8.82 (5.31)	8.26 (4.73)	11.37 (10.43)	14.03 (13.00)	6.79 (2.53)	7.63 (2.54)	6.92 (2.97)
DT	9.52 (1.55)	9.45 (1.48)	9.33 (1.39)	9.40 (1.45)	9.08 (1.12)	8.86 (1.01)	8.79 (0.96)
GBM	8.07 (3.82)	6.59 (2.71)	5.73 (2.23)	5.33 (2.08)	5.02 (1.81)	4.49 (1.54)	4.45 (1.56)
RF	4.39 (2.13)	3.69 (1.71)	2.93 (1.16)	2.78 (1.04)	2.45 (0.79)	2.22 (0.71)	2.15 (0.69)
EV	4.49 (1.91)	3.98 (1.52)	3.48 (1.18)	3.42 (1.15)	3.19 (0.95)	3.15 (0.90)	3.09 (0.84)
NN	4.39 (2.23)	4.27 (2.16)	4.13 (2.05)	3.55 (1.56)	3.15 (1.41)	2.16 (0.78)	2.08 (0.74)
ENSEMBLE	3.58 (1.65)	2.95 (1.19)	2.64 (0.99)	2.57 (0.97)	2.38 (0.81)	1.94 (0.69)	1.88 (0.67)

Conclusions and future works

- The proposed methodology is efficient for short-term electric energy forecasting (it achieved the best performance).
- The best models performance is achieved for large values of w.
- Apply the framework proposed to Paraguay data and other datasets .
- Improve the hyperparameter tuning stage (due to technical limitations).